AI in Medical Questionnaires: Scoping Review

Xuexing Luo[1*]; Yiyuan Li[1*]; Jing Xu[1*]; Zhong Zheng[1]; Fangtian Ying[2,3]; Guanghui Huang[1,4]

[1]Faculty of Humanities and Arts, Macau University of Science and Technology, Macau, China

[2]Industrial and Manufacturing Engineering, European Academy of Engineering, Gothenburg, Sweden

[3]College of Computer Science and Technology, Zhejiang University, Hangzhou, China

[4]Zhuhai M.U.S.T. Science and Technology Research Institute, Zhuhai, China

[*]these authors contributed equally

Corresponding Author:
Guanghui Huang
Faculty of Humanities and Arts
Macau University of Science and Technology
Avenida Wai Long
Taipa
Macau, 999078
China
Phone: 86 88972767
Email: ghhuang1@must.edu.mo

Related Article:
This is a corrected version. See correction statement in: http://www.jmir.org/2025/1/e80644/

Abstract

Background: The World Health Organization reports that >1 billion people worldwide experience mental disorders, with the prevalence of depression and anxiety among children and adolescents at 2.6% and 6.5%, respectively. However, commonly used clinical questionnaires such as the Hamilton Depression Rating Scale and the Beck Depression Inventory suffer from several problems, including the high degree of overlap of symptoms of depression with those of other psychiatric disorders and a lack of professional supervision during administration of the questionnaires, which often lead to inaccurate diagnoses. In the wake of the COVID-19 pandemic, the health care system is facing the dual challenges of a surge in patient numbers and the complexity of mental health issues. Artificial Intelligence (AI) technology has now been shown to have great promise in improving diagnostic accuracy, assisting clinical decision-making, and simplifying questionnaire development and data analysis.

Objective: This review aimed to explore the current applications, potential benefits, and issues of AI in medical questionnaires, focusing on its role in 3 main functions: assessment, development, and prediction. The global mental health burden remains severe.

Methods: The review included peer-reviewed studies that applied AI technologies to medical, psychological, or physiological questionnaires and reported measurable outcomes; non–peer-reviewed, non-English/Chinese, ethically noncompliant, or AI-unrelated studies were excluded. Five databases (PubMed, Embase, Cochrane Library, Web of Science, and CNKI) were searched from inception through September 2024. Three independent reviewers conducted data extraction, quality appraisal using the Joanna Briggs Institute tools, and narrative synthesis of AI applications across questionnaire assessment, development, and prediction tasks.

Results: Of 49,091 publications, a total of 14 (0.03%) studies met the inclusion criteria. AI technologies showed advantages in assessment, such as distinguishing myalgic encephalomyelitis or chronic fatigue syndrome from long COVID-19 with 92.18% accuracy. In questionnaire development, natural language processing using generative models such as ChatGPT was used to construct culturally competent scales. In terms of disease prediction, one study had an area under the curve of 0.790 for cataract surgery risk prediction. Overall, 24 AI technologies were identified, covering traditional algorithms such as random forest, support vector machine, and k-nearest neighbor, as well as deep learning models such as convolutional neural networks, Bidirectional Encoder Representations From Transformers, and ChatGPT. Despite the positive findings, only 21% (3/14) of the studies had entered the clinical validation phase, whereas the remaining 79% (11/14) were still in the exploratory phase of research. Most of the studies (10/14, 71%) were rated as being of moderate methodological quality, with major limitations including lack of a control group, incomplete follow-up data, and inadequate validation systems.

XSL•FO

RenderX

Conclusions: In summary, the integrated application of AI in medical questionnaires has significant potential to improve diagnostic efficiency, accelerate scale development, and promote early intervention. Future research should pay more attention to model interpretability, system compatibility, validation standardization, and ethical governance to effectively address key challenges such as data privacy, clinical integration, and transparency.

KEYWORDS

artificial intelligence; AI; medical questionnaires; questionnaire-based prediction; questionnaire development; diagnostic accuracy

Introduction

According to World Health Organization (WHO) data from 2022, approximately 1 billion individuals worldwide experience mental disorders [1,2]. Worldwide, the prevalence of depression and anxiety disorders among children and adolescents is estimated at approximately 2.6% and 6.5%, respectively [3,4]. This high disease burden necessitates reliable screening tools, yet current gold-standard questionnaires face two critical challenges. First, although these authoritative psychological assessment tools are used during screening and diagnosis, depressive symptoms often overlap with those of other psychiatric disorders, such as bipolar affective disorder and obsessive-compulsive disorder [5-9], making accurate diagnosis difficult when relying on a single questionnaire. Second, in many health care settings, patients complete questionnaires without proper guidance or oversight due to inadequate clinical supervision, resulting in distorted outcomes when various psychological and physiological self-assessment tools are applied in practice [10-12]. These limitations continually underscore the inefficiencies in the diagnostic utility and administration of traditional medical questionnaires. To obtain accurate health data and assist physicians during patient consultations, there is an urgent need to explore more efficient and precise assessment techniques. This study aimed to investigate how artificial intelligence (AI) technologies can address these fundamental limitations of traditional assessment tools while enhancing clinical decision-making processes.

Following the outbreak of COVID-19 in 2019, hospitals have faced a dual challenge of surging patient volumes and increasingly complex mental health care needs [13-15]. These unprecedented circumstances have exposed the inadequacies of traditional assessment approaches, creating an opportunity for technological innovation in mental health care delivery. Advancements in machine learning (ML) and large language models (LLMs) have demonstrated significant potential to reduce analytical biases, support clinical decision-making, and improve data processing efficiency [16-20], drawing medical experts' attention to how intelligent technologies can compensate for the inefficiency and subjectivity of traditional scales. Since 2013, the integration of ML with LLMs has resulted in breakthroughs in natural language processing (NLP) [21], complex reasoning, multilingual support, and multidimensional data analysis [22-26]. By 2024, these developments evolved into specialized mental health–oriented LLMs capable of identifying stress, depression, and suicidal ideation, thereby facilitating disease screening and early intervention [27-31]. This technological evolution directly responds to the postpandemic challenges by offering more accurate, efficient, and accessible screening tools that can identify psychological conditions early and create critical windows for timely intervention [32-34].

Multiple research studies across different application domains have substantiated the technological advantages of AI in mental health assessment. On the basis of the developments described previously, AI systems demonstrate several key capabilities that directly address the limitations of traditional questionnaires. First, ML algorithms excel at detecting subtle patterns across multiple variables that human clinicians might miss, as demonstrated by the high accuracy of the study by McGarrigle et al [35] in distinguishing among myalgic encephalomyelitis or chronic fatigue syndrome, post–COVID-19 condition, and healthy controls using random forest (RF) algorithms. LLMs' NLP capabilities capture contextual meanings and emotional undertones in patient narratives that structured questionnaires cannot. Unlike traditional categorical approaches, ML enables dimensional symptom representation, aligning with contemporary understanding of spectrum disorders, as shown in the approach by De Luca et al [36] to modeling suicide risk. In addition, AI-based systems implement adaptive questioning pathways that enhance user satisfaction while maintaining diagnostic validity, as Nam et al [37] demonstrated with their conversational AI for spinal pain assessment. Finally, AI can integrate multimodal data, as evidenced by the improvement in pathological voice assessment by Kojima et al [38] by combining acoustic features with questionnaire responses. These capabilities directly address traditional questionnaires' limitations in handling symptom overlap and contextual interpretation, providing a foundation for the practical applications of AI in clinical settings that will be discussed in the following section.

The comprehensive technical advantages are obvious. Rapid developments in ML, data science, and neural networks have allowed AI to be involved in the evaluation, development, and predictive modeling phases of medical questionnaires within clinical practice [39-41]. With its efficiency, accuracy, and capacity to handle large-scale data, AI enables clinicians and health care professionals to swiftly access and evaluate patient data derived from medical questionnaires, thereby improving primary health care efficiency [42-45]. For instance, advances in NLP have supported rapid screening for depression and anxiety, reducing initial consultation times [21,46-48], whereas deep learning (DL) has contributed to a range of medical tasks, including large-scale health data screening [49-59], pathological segmentation [53-56], and disease monitoring [57-59]. Integrating LLMs with traditional psychological assessments

can reduce human error and ensure consistency in professional diagnoses. By 2024, core NLP technologies integrated with ML methods were being used to classify depression and its severity [60].

This paper identifies key limitations in traditional medical scales, including substantial diagnostic bias, low screening efficiency, and data distortion. It explores the potential applications of AI in the health care domain, focusing on how AI-driven approaches can enhance evaluation, development, and prediction—3 critical diagnostic stages—while examining recent algorithmic research and clinical findings. In addition, it summarizes the impacts of AI-augmented traditional questionnaires on patient outcomes. Finally, it addresses broader societal and ethical considerations, such as privacy, fairness, transparency, and ethical challenges. This review concludes by discussing future developmental trajectories and scientific hypotheses concerning the integration of AI into medical questionnaires.

Methods

Data Sources and Search Strategies

This study was designed according to the latest version of the PRISMA-ScR (Preferred Reporting Items for Reviews and Meta-Analyses extension for Scoping Review) checklist [61] (Multimedia Appendix 1). This study was conducted through a systematic literature search in several databases (PubMed, Embase, Cochrane Library, Web of Science, and China National Knowledge Infrastructure on the China Knowledge Network) with the aim of exploring the innovation, diagnosis, and impact of AI in medical scales. The search keywords and other related terms in this study are shown in Table 1. In addition, the search was conducted from the inception of each database to September 2024 to ensure the inclusion of the latest relevant studies and was limited to English- and Chinese-language literature.

Table 1. Search strategies for English- and Chinese-language databases.

	Search term
1	"Artificial intelligence" (MeSH[a])
2	"Machine learning" (MeSH)
3	"Deep learning" (MeSH)
4	1 OR 2 OR 3
5	"Questionnaire"
6	"Scale"
7	"Medical questionnaire"
8	"Medical scale"
9	"Psychological questionnaire"
10	"Psychological scale"
11	"Physiological questionnaire"
12	"Physiological scale"
13	"Mental questionnaire"
14	"Mental scale"
15	5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14
16	4 AND 15
17	"Rengongzhineng" (artificial intelligence)
18	"Jiqixuexi" (machine learning)
19	"Shenduxuexi" (deep learning)
20	17 OR 18 OR 19
21	"Diaochawenjuan" (questionnaire)
22	"Liangbiao" (scale)
23	"Yixue diaochawenjuan" (medical questionnaire)
24	"Yixue liangbiao" (medical scale)
25	"Xinli diaochawenjuan" (physiological questionnaire)
26	"Xinli liangbiao" (psychological scale)
27	"Shengli diaochawenjuan" (physiological questionnaire)
28	"Shengli liangbiao" (physiological scale)
29	"Jingshen diaochawenjuan" (mental questionnaire)
30	"Jingshen liangbiao" (mental scale)
31	21 OR 22 OR 23 OR 24 OR 25 OR 26 OR 27 OR 28 OR 29 OR 30
36	20 AND 31

[a]MeSH: Medical Subject Headings.

Inclusion and Exclusion Criteria

The inclusion criteria for the studies can be specifically divided into three categories: (1) studies related to the application of AI technology to disease management, psychological, or physiological questionnaires; (2) articles that provided relevant data to support and validate the effectiveness of the application of AI technology to questionnaires; and (3) peer-reviewed literature, which may include but was not limited to randomized controlled trials, cohort studies, reviews, meta-analyses, and cross-sectional studies. On the basis of the inclusion criteria, the exclusion criteria for this review were as follows: (1) studies that did not use AI technologies and did not involve medical questionnaires; (2) literature in the category of *gray literature*, such as non–peer-reviewed literature, unpublished manuscripts, or conference abstracts; (3) literature not in English or Chinese (the authors' native language) and literature in Chinese that did not meet the inclusion criteria; (4) experimental literature that did not have ethics approval or obtain informed consent; (5) articles on the uses of AI not being ethical; and (6) articles on conflicts of interest related to AI technology.

Data Charting and Synthesis

In this study, a comprehensive literature screening and evaluation process was implemented to ensure the objectivity and accuracy of the research. This process was executed by 3 independent reviewers: YL, JX, and XL. Initially, YL was responsible for downloading and conducting a preliminary review of the screened literature to exclude documents unrelated to the study's topic. Subsequently, the relevant literature from the preliminary screening was passed to JX for a more detailed eligibility assessment. The included literature was then double-checked by YL and JX. In cases of disagreement between YL and JX, XL made the final decision. Only literature agreed upon by all 3 reviewers was included in the final analysis. Furthermore, throughout the data extraction process, the 3 reviewers (XL, YL, and JX) created 5 distinct tables and figures, each with a specific function, to systematically organize and analyze the collected data. The primary purpose of creating these tables and figures was to enhance the transparency, systematic approach, and scientific rigor of this review by organizing and analyzing data in a standardized manner. Table 1 presents the search strategies and keywords used in both English- and Chinese-language databases.

Quality Evaluation Methods

To ensure the quality of the selected literature, this review used the Joanna Briggs institute (JBI) critical appraisal tools [62]. The final quality assessment of the included literature was conducted using the scoring system provided by the JBI guidelines. This system assigns 1 point for each criterion fully met, with a score of 1 for *yes* and 0 for *no* or *unclear* responses. This scoring method facilitates a horizontal comparison of study quality, allowing for ranking based on the total score. For qualitative research, studies with <5 points were considered low quality, studies with 5 to 7 points were considered moderate quality, and studies with ≥8 points were considered high quality. For review studies, scores of <6 points indicated low quality, scores of 6 to 8 points indicated moderate quality, and scores of ≥9 points indicated high quality. Given the specific design and implementation of quasi-experimental studies, the scoring criteria were adjusted accordingly—scores of <5 points indicated low quality, scores of 5 to 7 points indicated moderate quality, and scores of ≥8 points indicated high quality. This systematic assessment approach ensured the reliability and scientific rigor of the review findings.

Quality Assessment

JBI Checklist for Systematic Reviews and Research Syntheses

The items in this checklist were as follows: (1) is the review question clearly and explicitly stated? (2) Were the inclusion criteria appropriate for the review question? (3) Was the search strategy appropriate? (4) Were the sources and resources used to search for studies adequate? (5) Were the criteria for appraising the studies appropriate? (6) Was critical appraisal conducted by ≥2 reviewers independently? (7) Were there methods to minimize errors in data extraction? (8) Were the methods used to combine studies appropriate? (9) Was the likelihood of publication bias assessed? (10) Were recommendations for policy or practice supported by the reported data? (11) Were the specific directives for new research appropriate?

JBI Checklist for Quasi-Experimental Studies

The items in this checklist were as follows: (1) is it clear in the study what is the *cause* and what is the *effect* (ie, there is no confusion about which variable comes first)? (2) Was there a control group? (3) Were participants included in any comparisons similar? (4) Were the participants included in any comparisons receiving similar treatment or care other than the exposure or intervention of interest? (5) Were there multiple measurements of the outcome both before and after the intervention or exposure? (6) Were the outcomes of participants included in any comparisons measured in the same way? (7) Were outcomes measured in a reliable way? (8) Was follow-up complete and, if not, were differences between groups in terms of their follow-up adequately described and analyzed? (9) Was appropriate statistical analysis used?

JBI Checklist for Qualitative Research

The items in this checklist were as follows: (1) is there congruity between the stated philosophical perspective and the research methodology? (2) Is there congruity between the research methodology and the research question or objectives? (3) Is there congruity between the research methodology and the methods used to collect data? (4) Is there congruity between the research methodology and the representation and analysis of the data? (5) Is there congruity between the research methodology and the interpretation of the results? (6) Is there a statement locating the researcher culturally or theoretically? (7) Is the influence of the researcher on the research and vice versa addressed? (8) Are participants, and their voices, adequately represented? (9) Is the research ethical according to current criteria or, for recent studies, is there evidence of ethics approval by an appropriate body? (10) Do the conclusions drawn in the research report flow from the analysis or interpretation of the data?

Results

Overview

The screening process for inclusion of studies is shown in Figure 1. An initial 49,091 records were identified through a systematic database search. After removing duplicates and excluding irrelevant articles, of the 49,091 initial articles, 3651 (7.44%) remained for title and abstract screening. After this step, of the 3651 articles, 3625 (99.29%) were excluded. The remaining 0.71% (26/3651) of the articles were assessed in full text and included various types of reviews, clinical studies, and case reports. Of these 26 articles, 12 (46%) were excluded because they did not meet the inclusion criteria. These reasons included a lack of peer review, failure to use questionnaire methodology, and insufficient relevance to the focus of the study. Ultimately, 14 articles were included in the final review.

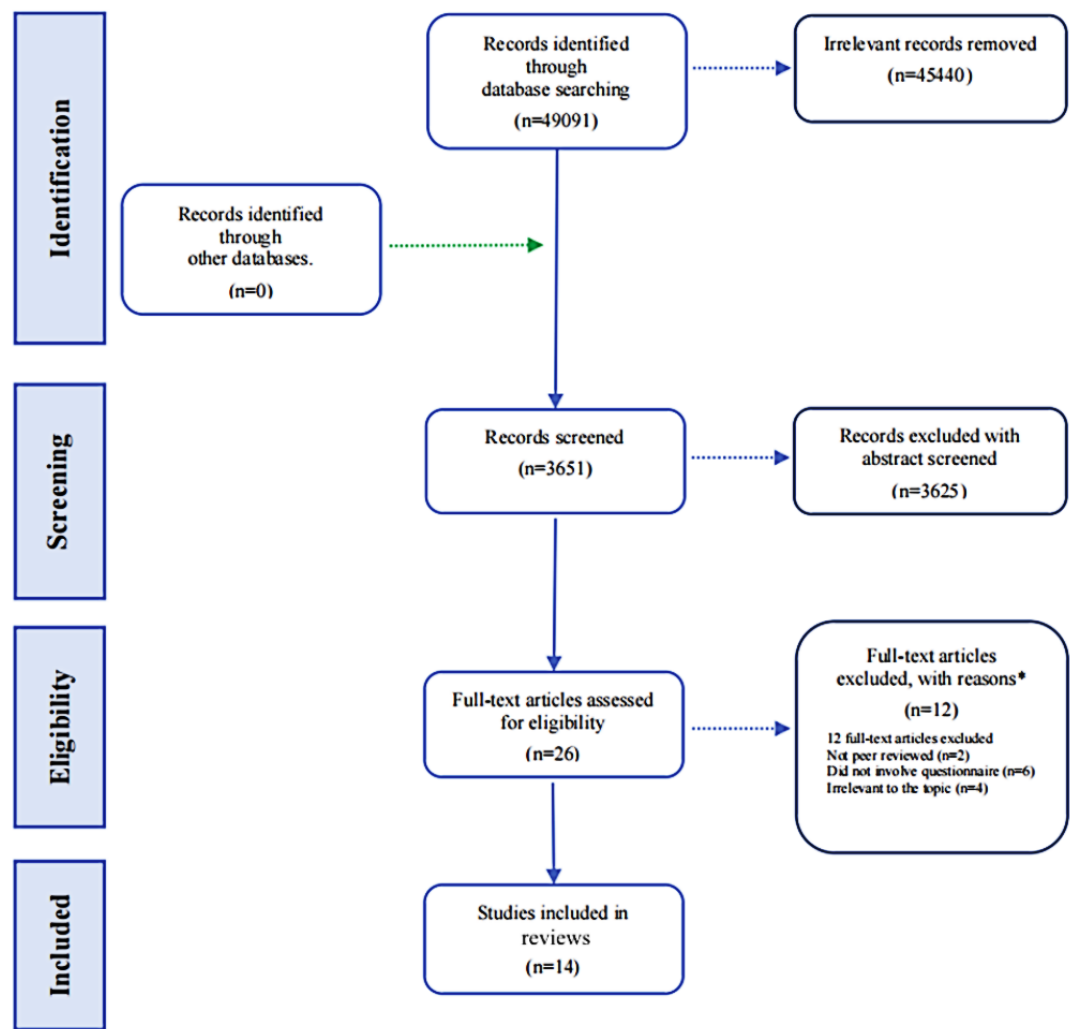**Figure 1.** Flow diagram for the included and excluded articles.

Table 2 assesses the quality of the studies using standardized JBI tools, categorizing the studies into low, medium, and high quality. Figure 2 shows the distribution of 24 different AI technologies (ML and DL) in the diagnosis of physiological and psychological conditions based on questionnaires. Multimedia Appendix 2 details the application of intelligent technologies in clinical and research environments (N=14), emphasizing how different AI methods facilitate the development, prediction, and evaluation of questionnaires. Finally, Figure 3 illustrates the distribution patterns of ML and DL technologies across the various studies. This framework ensured comprehensive data extraction, quality control, and systematic synthesis of AI applications in the implementation of medical questionnaires.

Table 2. Summary of the quality evidence in the included 14 reports.

Study	Study design	Assessment of the quality of the study													Overall score
		1	2	3	4	5	6	7	8	9	10	11	12	13	
Siddiqua et al [63], 2023	Quasi-experimental study	Yes	No	Yes	Unclear	No	Yes	Yes	No	Yes	__[a]	—	—	—	Medium quality (5/9)
van Buchem et al [64], 2022	Qualitative research	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	—	—	—	Medium quality (8/10)
Coraci et al [65], 2023	Quasi-experimental study	Yes	Yes	Yes	Yes	No	Yes	Yes	Unclear	Yes	—	—	—	—	Medium quality (7/9)
Nam et al [37], 2022	Quasi-experimental study	Yes	No	Yes	Yes	No	Yes	Yes	Unclear	Yes	—	—	—	—	High quality (6/9)
De Luca et al [36], 2024	Systematic review and research synthesis	No	No	No	No	No	No	Yes	No	No	Yes	Yes	—	—	Low quality (3/11)
McGarrigle et al [35], 2024	Quasi-experimental study	Yes	Yes	Unclear	Yes	No	Yes	Yes	No	Yes	—	—	—	—	Medium quality (6/9)
Kojima et al [38], 2024	Quasi-experimental study	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	—	—	—	—	Medium quality (6/9)
Wang et al [66], 2021	Systematic review and research synthesis	No	No	No	No	No	No	Yes	No	No	Yes	Yes	—	—	Low quality (3/11)
Ha et al [67], 2023	Quasi-experimental study	Yes	Yes	Unclear	Yes	No	Yes	Yes	No	Yes	—	—	—	—	Medium quality (6/9)
Shetty et al [68], 2024	Quasi-experimental study	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	—	—	—	—	Medium quality (6/9)
McCartney et al [69], 2014	Quasi-experimental study	Yes	No	Unclear	Yes	No	Yes	Yes	No	Yes	—	—	—	—	Medium quality (5/9)
Sali et al [70], 2013	Quasi-experimental study	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	—	—	—	—	Medium quality (6/9)
Ferreira Freitas et al [71], 2021	Quasi-experimental study	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	—	—	—	—	Medium quality (6/9)
Li et al [72], 2022	Quasi-experimental study	Yes	No	Unclear	No	No	Yes	Yes	No	Yes	—	—	—	—	Low quality (4/9)

[a]Not applicable.

Figure 2. 2013-2023 distribution of 24 different artificial intelligence technologies (machine learning and deep learning) in the diagnosis of physiological and psychological conditions based on medical questionnaires. ACML:Apple's Create ML; ANN: artificial neural network; BERT: bidirectional encoder representations from transformers; BN: Bayesian network; CNN: convolutional neural network; DL: multilayer feedforward deep learning; DT: decision tree; GA: genetic algorithm; GB: gradient boosting; GTF: Google's TensorFlow; KNN: k-nearest neighbor; LR: logistic regression; NB-G: naive Bayes–Gaussian; NB-M: naive Bayes–multinomial; NLP: natural language processing; RF: random forest; SVC: support vector classifier; SVM: support vector machine; TRM: traditional regression model; VA: voting algorithm; XGBoost: extreme gradient boosting; ZR-C: ZeroR classifier.
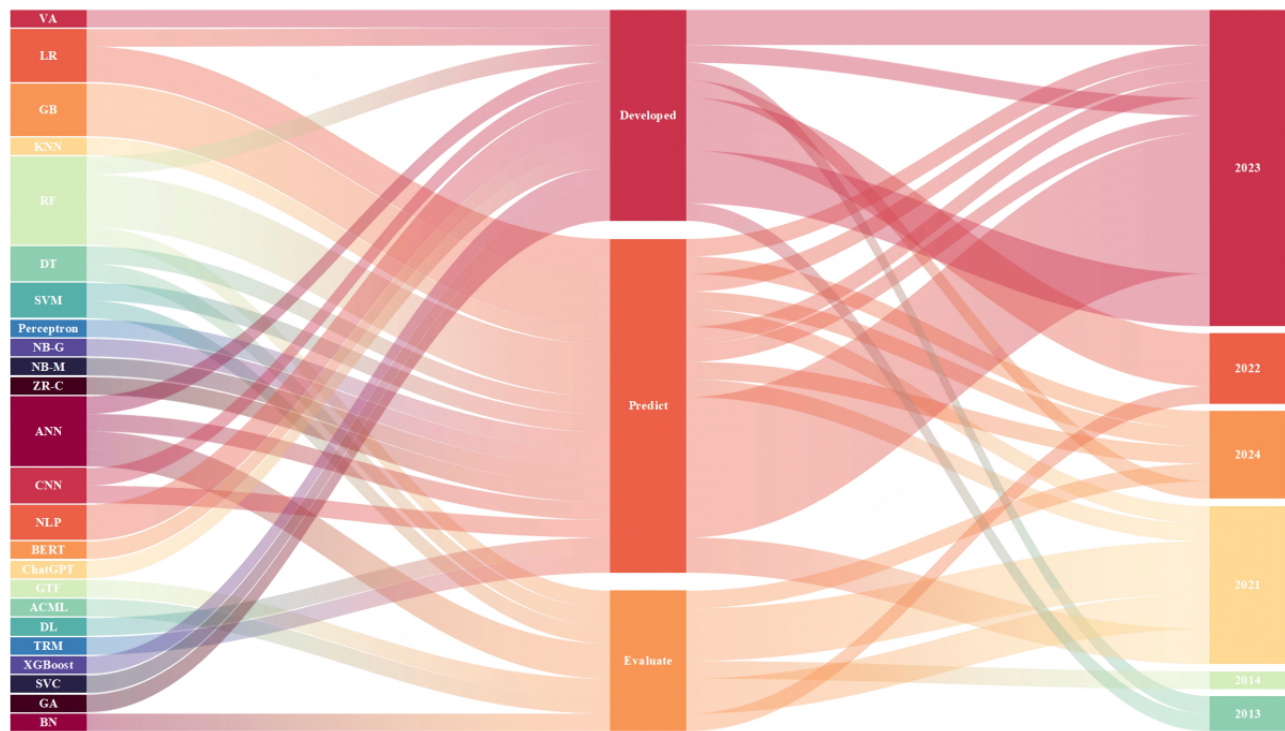
Figure 3. Distribution of machine learning and deep learning technologies across the studies. ACML: Apple's Create ML; ANN: artificial neural network; BERT: bidirectional encoder representations from transformers; BN: Bayesian network; CNN: convolutional neural network; DL: multilayer feedforward deep learning; DT: decision tree; GA: genetic algorithm; GB: gradient boosting; GTF: Google's TensorFlow; KNN: k-nearest neighbor; LR: logistic regression; NB-G: naive Bayes–Gaussian; NB-M: naive Bayes–multinomial; NLP: natural language processing; RF: random forest; SVC: support vector classifier; SVM: support vector machine; TRM: traditional regression model; VA: voting algorithm; XGBoost: extreme gradient boosting; ZR-C: ZeroR classifier.
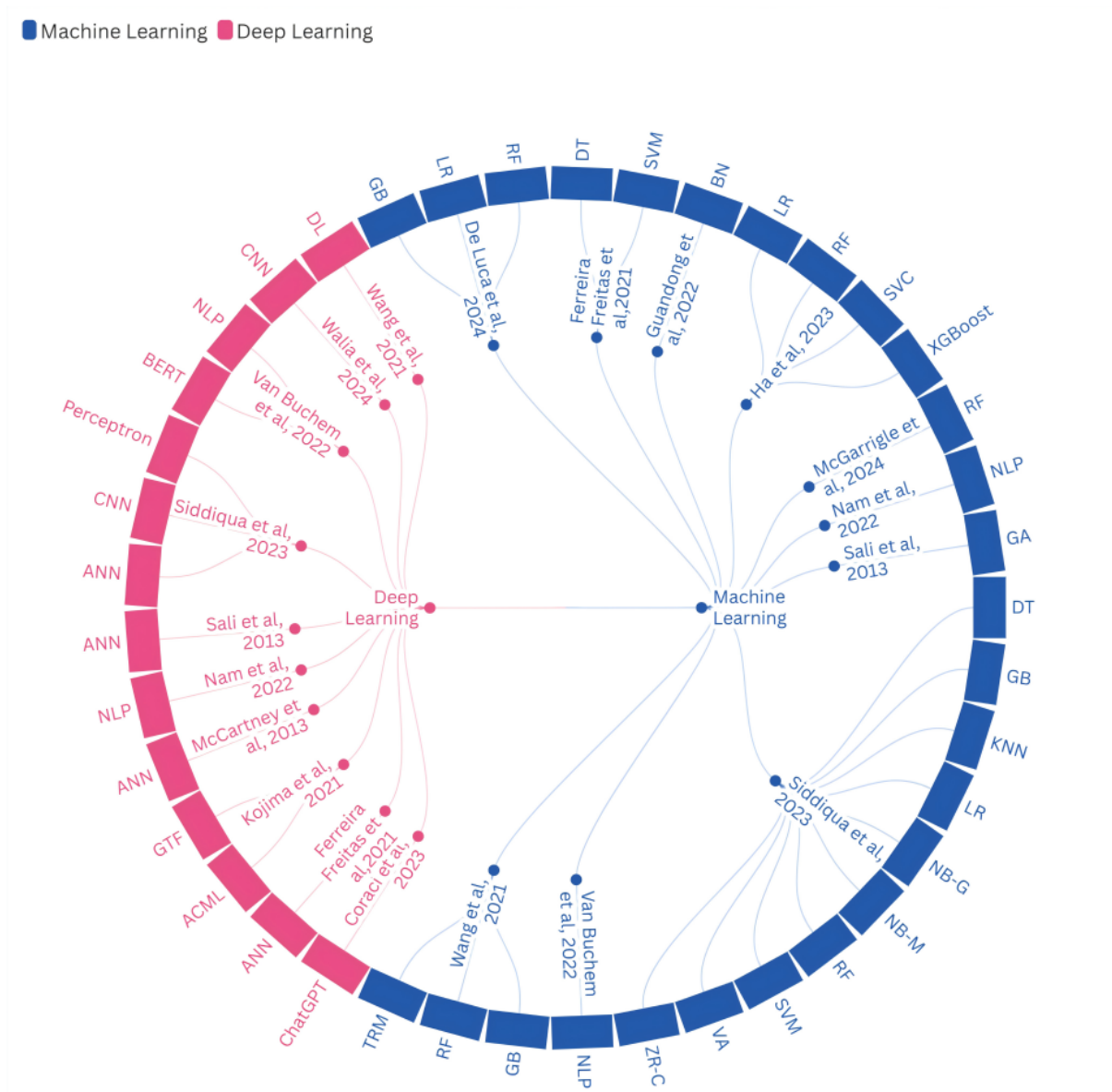
Quality Assessment

The quality of the 14 included studies was assessed using the JBI critical appraisal tools (Table 2). Each study was evaluated using the appropriate checklist based on its design (Table 2)—the JBI critical appraisal checklist for experimental studies (11 items), the JBI qualitative research checklist (1 item), or the JBI systematic review and research synthesis checklist (2 items). Most of the included studies (10/14, 71%) were of moderate methodological quality, with only 7% (1/14) of the studies rated as high quality [66] and 21% (3/14) of the studies classified as low quality [36,66,72]. Methodological strengths commonly observed in quasi-experimental studies included clear causal relationships, reliable outcome measurements, and appropriate statistical analyses. However, the absence of control groups and incomplete descriptions of follow-up were common

limitations. A total of 14% (2/14) of the studies, which were systematic reviews, identified flaws in the search strategies and the critical appraisal methodology. This quality assessment provided essential context for interpreting the findings related to the application of AI in medical questionnaires and highlighted the need for more rigorous methodological validation in this rapidly evolving field. Overall, the predominance of moderate-quality studies indicates significant room for improvement in study design and reporting.

AI and Medical Questionnaires

Overview

The use of AI for medical questionnaires harnesses DL and ML to enhance disease evaluation, facilitate the development of novel questionnaires, and improve data predictive capacities. These 3 facets are essential diagnostic stages for assessing both

physiological conditions and psychological issues. By reviewing recent algorithmic advances and clinical practice, this study uncovered the potential of incorporating AI into traditional medical questionnaires, offering new perspectives for broader clinical applications and informed medical diagnoses (Figure 4).

Figure 4. Artificial intelligence in medical questionnaires.

Artificial Intelligence in Medical Questionnaires: Innovations, Diagnosis, and Implications
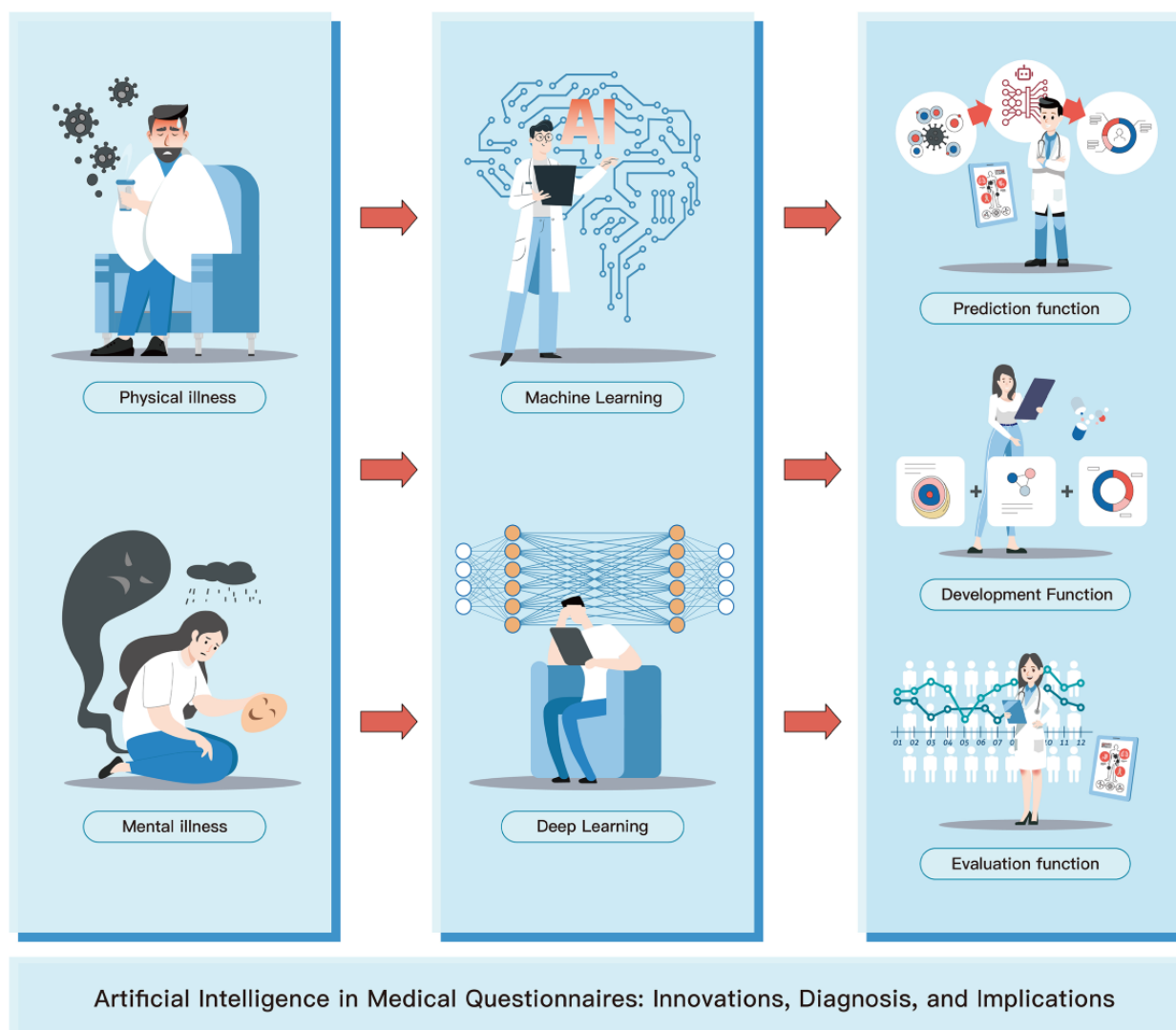
Enhancing the Efficiency of Medical Questionnaire Assessments Through AI

According to WHO data, since the onset of the COVID-19 pandemic, the global population experiencing anxiety and depression has markedly increased, accompanied by a 40% surge in the use of standardized questionnaires [73-76]. Traditional questionnaires offer convenient dissemination and, in some cases, self-administration [77-79]. However, as definitive diagnoses still require physician consultations, the assessment process remains labor intensive. Due to its speed and accuracy in data processing, AI has progressively been integrated into clinical support [43,44]. Notably, DL and NLP technologies—such as the pretrained bidirectional encoder representations from transformers (BERT) model and conversation-based models such as ChatGPT—observed rising adoption rates from 2023 to 2024 [77-80]. Furthermore, traditional algorithms, including support vector machine and k-nearest neighbor, continue to effectively evaluate patient data for complex disorders. Recent developments in the technical frameworks for applying these methods to medical questionnaire assessments are illustrated in Figure 2.

Analysis of AI applications in medical questionnaires showed that only 21% (3/14) of the studies involved AI-assisted questionnaires that had entered the clinical validation stage, whereas the remaining 79% (11/14) of the articles described the technology as being in the research stage (Multimedia Appendix 2). Clinical validation was been achieved in patient experience assessment using NLP-based sentiment analysis, in spinal pain evaluation through conversational AI systems [37], and in surgical risk prediction using ML for cataract surgery [66]. These applications share the characteristics of established clinical workflows, model interpretability, and structured validation frameworks. Most AI applications in mental health assessment, pain evaluation, disease differentiation, and specialized assessments remain in the research phases despite promising performance. This pattern stems from psychological assessment complexity [81], data standardization challenges, stringent clinical implementation requirements [82], and

emerging technologies' nascent nature in medical contexts. The findings indicate that AI-assisted medical questionnaires are technically feasible but still transitioning from research innovation to clinical implementation.

Enhancing the Efficiency of Dynamic Assessments

By leveraging DL on patient data, AI effectively evaluates how medical questionnaires adaptively distinguish between patient symptom variations, thereby accurately capturing changes in health status and demonstrating advantages in dynamic assessment. During prediagnostic evaluations, many pathological conditions have overlapping manifestations. For instance, post–COVID-19 sequelae resemble acute COVID-19 symptoms, and depression in bipolar disorders overlaps with that observed in unipolar depression [83-86]. To enhance diagnostic accuracy, one study used a series of RF algorithms to assess the psychometric properties of the DePaul Symptom Questionnaire–Short Form in classifying individuals with post–COVID-19 sequelae, those with myalgic encephalomyelitis or chronic fatigue syndrome (unrelated to COVID-19), and healthy controls. The results indicated that the DePaul Symptom Questionnaire–Short Form successfully distinguished patients with post–COVID-19 sequelae from healthy controls, achieving an accuracy of 92.18% [35].

Moreover, AI-based evaluations validated the diagnostic sensitivity of traditional questionnaires for subtle disorders such as women's physical and mental health issues influenced by hormonal fluctuations. Researchers used support vector machine, artificial neural networks (ANNs), and decision trees to confirm the validity and accuracy of the International Physical Activity Questionnaire for menopausal women [71]. Currently, integrating multiple algorithms helps improve both model interpretability and patient classification precision. One team, for example, used a hybrid model (genetic algorithms combined with ANNs) to evaluate the effects of different weighting schemes in a traditional scale applied to daily stress events; after weight adjustments, they reported that the traditional questionnaire achieved a sensitivity of 83% and a specificity of 81% for stress detection, positioning the modified instrument as a high-performance screening tool [70].

Collectively, these findings demonstrate that AI enhances the precision with which traditional questionnaires differentiate between disease categories and unique patient populations while simultaneously refining traditional questionnaire metrics.

Construction of Intelligent Data Systems

One of the key advantages of AI-driven data training is its capacity to construct intelligent data systems that assist clinicians in making more accurate diagnoses. Although clinical medical questionnaires undergo lengthy validation to ensure efficacy, traditional instruments often fail to comprehensively capture authentic patient behavior. For example, patients may complete questionnaires too hastily without supervision, leading to distorted results, and even after traditional assessments, diagnostic discrepancies of up to 17% may arise among different therapists [87]. Consequently, multidimensional approaches are needed in psychological and psychiatric diagnostics to reduce subjective influences and create intelligent data systems. A study on pathological voice data used AI to enhance the objectivity of the grade, roughness, breathiness, asthenia, and strain scale comparing 2 convolutional neural network (CNN)–based models—one built on Google's TensorFlow and the other on Apple's Core ML—both trained on identical pathological voice datasets [67]. By comparing these 2 models in classifying severity levels in pathological speech, it became possible to validate grade, roughness, breathiness, asthenia, and strain questionnaire results in real time without specialized equipment, thereby increasing the clinical objectivity of traditional questionnaires (Figure 5 [71]). In another effort, researchers designed a secure neural network–based application, randomly training and testing an ANN for diagnosing facial pain syndromes [67]. The system achieved a sensitivity of 92.4% and a specificity of 87.8%. Ultimately, AI-based data training promises to integrate patient information worldwide into intelligent online databases, thus expanding the accessibility and dissemination of intelligent data in the medical field.

Figure 5. Evaluation of pathological voice data by TensorFlow and Apple's Create ML (modified from the work by Ferreira Freitas et al [71]). AUROC: area under the receiver operating characteristic curve; COMISA: comorbid insomnia and sleep apnea; EUMCSH: Ewha Woman's University Medical Center Seoul Hospital; ISI: Insomnia Severity Index; OSA: obstructive sleep apnea; ROC: receiver operating characteristic; SHAP: Shapley Additive Explanations; SMC: Samsung Medical Center; SRQ: Sleep Regularity Questionnaire; XGBoost: Extreme Gradient Boosting.
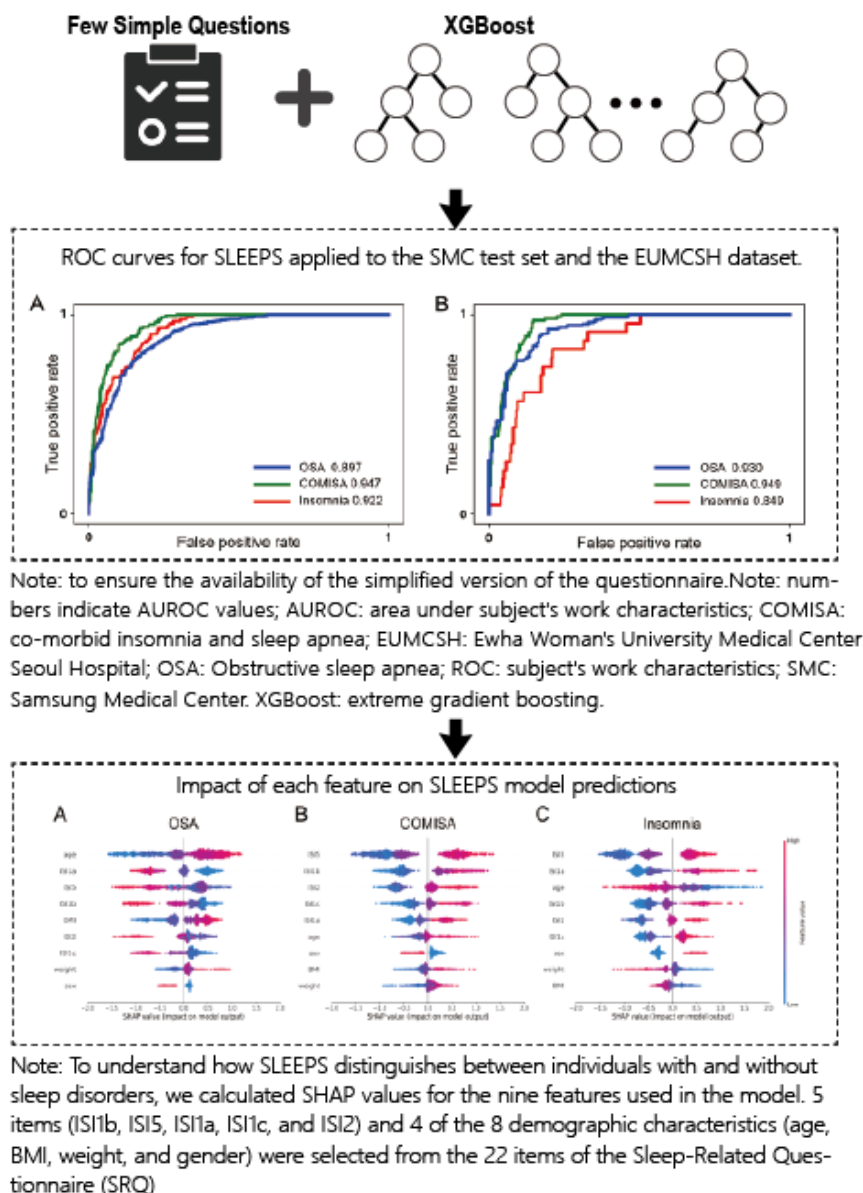
Optimizing the Development of Medical Questionnaires Through AI

Since 2013, ANNs and genetic algorithms have been used in the development of novel medical assessment models, enhancing parameter selection for psychological questionnaires and improving the accuracy of predicting patients' psychological states [88,89]. In 2021 and 2022, as the range of assessment methods broadened, integrating traditional ML algorithms (logistic regression [LR], RF, and decision tree) with ANN and DL tools further advanced the design of complex medical questionnaires and refined the evaluation of symptoms related to anxiety and depression (Figure 3). Nevertheless, to achieve more profound and long-term predictions of patients' psychological conditions, it remains essential to optimize both questionnaire item formulation and the assessment processes.

Enhancing the Patient Applicability of Questionnaires

In developing medical questionnaires suited to diverse patient populations and conditions, AI integration confers distinct advantages. Traditional assessment methods, particularly those used to diagnose late-life depression, face notable limitations. Late-life depression is often accompanied by various comorbidities, and subtle symptom distinctions can be difficult to capture using conventional questionnaires [90-93]. To address this complexity, especially in older adults, NLP techniques can analyze patients' linguistic patterns, thereby detecting latent symptoms that may be overlooked by traditional methods and, ultimately, expanding the questionnaire's applicability. Moreover, Bayesian network–based analyses have been used to generate highly matched items, effectively facilitating the development of a potential risk assessment system for geriatric conditions [94,95]. AI-driven approaches have also improved applicability for other vulnerable groups. For instance, the

Raghavendra Manjunath Shetty Digital Anxiety Scale leverages AI-generated facial expressions for children [68]. By enabling children to select expressions that best match their emotions, this interactive approach overcomes the limitations of traditional questionnaires, which often struggle to accurately convey or capture younger children's emotional states.

In addition, scholars have applied the extreme gradient boosting algorithm to streamline risk assessment questionnaires for insomnia and obstructive sleep apnea (OSA) [38] (Figure 6 [38,96]). By avoiding cumbersome overnight polysomnography measurements, researchers focused on feature importance to design a simplified questionnaire that accurately predicts the risks of 3 sleep disorders—OSA, comorbid insomnia and sleep

apnea, and insomnia—with an area under the receiver operating characteristic curve (AUROC) exceeding 0.897 for each. Figure 6 [38] illustrates the performance of a simplified questionnaire in predicting sleep disorder risk, demonstrating high accuracy in identifying OSA (AUROC=0.897), comorbid insomnia and sleep apnea (AUROC=0.947), and insomnia (AUROC=0.922). Figure 5 [86] also highlights the influence of various features on model predictions, ultimately improving the questionnaire's applicability across different severities of sleep disorders. By replacing complex, burdensome items with a high-confidence, streamlined set of questions, patients face a reduced response burden, and applicability to diverse patient populations is enhanced.

Figure 6. Artificial intelligence–assisted diagnostic data for system optimization and construction (modified from the work by Kojima et al [38], which is published under Creative Commons Attribution 4.0 International License [96]). CNN: convolutional neural network.
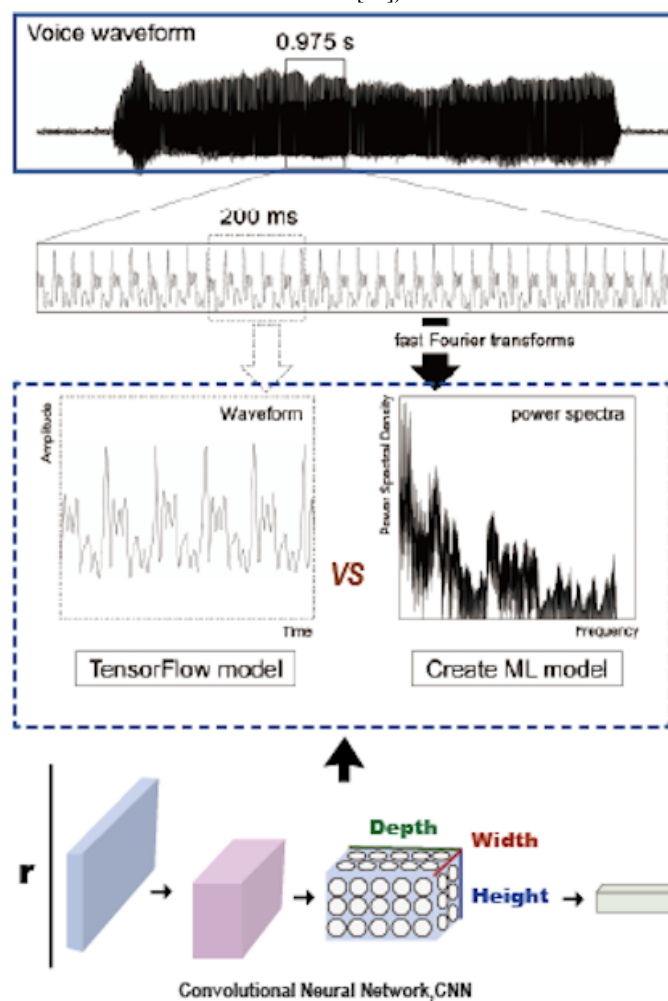
Enhancing the Cultural Adaptability of Questionnaires

In traditional clinical assessments, patients often interpret and respond differently due to varied cultural backgrounds, language habits, and life experiences. Questionnaire development must consider these cultural factors and provide timely feedback. Failure to account for the patient's experience can result in extreme responses that skew diagnostic outcomes. Thus, incorporating appropriate open-ended questions can flexibly capture differences in patients' understanding of medical questionnaires. One study used a multilingual and multicultural version of ChatGPT to generate a low back pain assessment

questionnaire [65]. Its findings demonstrated that language training technology could effectively overcome linguistic and cultural barriers, offering strong support for future cross-cultural medical evaluations. Another study leveraged a speech dialogue system trained to develop a new pain assessment tool for patients with spinal issues [37]. By integrating natural language understanding and speech recognition, the system quickly recorded patients' pain information and improved clinician-patient interactions. The error rates in speech recognition for physicians, nurses, and patients were 13.5%, 16.8%, and 34.7%, respectively. AI's real-time feedback

capabilities enable it to identify key factors from patient assessments across different backgrounds. One study showed that an AI-based patient-reported experience measure developed through NLP techniques could extract crucial information from patients' open-ended responses [64], promptly relaying insights to clinicians. This timely exchange not only saves analytic time but also strengthens trust and understanding for patients from diverse cultural contexts during the diagnostic process.

Enhancing the Predictive Accuracy of Medical Questionnaires Through AI

Currently, AI serves as a crucial and promising tool for supporting physiological treatment and enabling early psychological interventions [97]. In predictive tasks, DL models (such as ANNs and CNNs) excel at extracting features from complex datasets, whereas ensemble learning algorithms—gradient boosting, RF, and Extreme Gradient Boosting—specialize in classification and prediction. By combining these technical strengths, AI models have transcended traditional medical constraints, effectively increasing the precision of disease prediction and diagnosis at multiple critical time points [98]. From 2022 onward, DL (CNNs and DL) and NLP (BERT, NLP, and ChatGPT) have observed growing adoption rates. Notably, the application of CNNs surged in 2023 (Figure 3), and the use of BERT and ChatGPT continued to expand in health care throughout 2024 [99]. These trends collectively underscore the advantages of integrating a range of AI methodologies to enhance predictive tasks in medical questionnaires.

Early Prediction of Age-Related Diseases

According to 2024 WHO data, global aging is intensifying, and chronic diseases now account for >70% of all worldwide deaths [100]. In this context, AI-enabled analysis of global patient data can facilitate disease prediction. Ophthalmological conditions such as cataracts present an increasing diagnostic burden, with surgery remaining the only effective clinical intervention. Thus, the timely identification of risk factors for patients with cataracts is essential [101-103]. A 2022 study found that DL models surpass traditional statistical models in screening accuracy for age-related diseases, enabling more rapid identification of high-risk older populations. Further research has examined the use of AI for predicting cataract surgery risk. By leveraging questionnaires and medical records, the study compared its results to those of a traditional LR model. ML models achieved an area under the curve (AUC) between 0.781 and 0.790, outperforming the LR model's AUC of 0.767. The gradient boosting machine model demonstrated the best performance, attaining an AUC of 0.790 [101]. These findings indicate that ML can accurately forecast disease occurrence by rapidly assimilating diverse data inputs even in the absence of biological data.

Timely Attention to Mental Health

At present, the application of NLP techniques in mental health issue prediction is rapidly expanding [104]. ML methods (eg, LR and gradient boosting) and feature selection strategies have shown particular promise in early warning systems for mental health issues. One study on depression assessment used multiple ML and DL models to predict different levels of depressive states—normal, moderate, and severe. Among them, the RF model achieved an accuracy of 98.08%, outperforming the gradient boosting model (94.23%) and the CNN (92.31%) [103]. By applying feature selection and hyperparameter optimization, the study further enhanced model performance.

In addition, another research team used LR, RF, and gradient boosting models to evaluate suicidal intent and behaviors. Through Shapley Additive Explanations analysis, they identified the most effective features and reconstructed a simplified version of the Suicide Crisis Inventory–2 [36]. The LR model performed best, and the simplified questionnaire efficiently assessed and predicted suicidal crises in clinical settings, thereby reducing the supervision burden on health care providers [66]. The integration of AI not only improves data interpretability and predictive reliability but also contributes to timely health interventions in future health care systems.

Potential of AI

Overview

The global population is projected to grow continuously by approximately 1.3 billion people between 2020 and 2050, representing a 17% increase from the 2020 population, with a peak of approximately 9.73 billion by 2064 [105]. Health informatics is a fast-growing area in the health care field. These instruments serve as critical tools for assessing population health [106]. Currently, AI technologies have already demonstrated significant potential and advantages in the development, analysis, and prediction processes of medical questionnaires.

Despite the remarkable progress of AI in this domain, numerous challenges and research gaps remain. Therefore, this section presents several new research directions based on the potential of AI to enhance rapid assessment, precise prediction, and adaptable response methods in medical questionnaires, aiming to further promote AI's application and innovation in this field.

Potential for Rapid Assessment and Accurate Prediction

With the growing global population, the volume of questionnaires that health care institutions must collect and analyze is steadily increasing [107]. Traditional questionnaire evaluation methods are often time-consuming and require substantial human and economic resources [108]. Applying AI technologies to questionnaire assessments can significantly reduce labor and financial costs, a benefit that is especially pronounced in large-scale evaluations. Rapid questionnaire evaluation also enables researchers to quickly optimize questionnaires and establish related case identification systems, thereby allowing health care institutions to promptly allocate public health resources and respond more efficiently to public health events [109].

As global life expectancy continues to rise, the number and proportion of older individuals also increase. By 2030, a total of 1 in 6 people worldwide will be aged >60 years, and the population aged ≥60 years will grow from the current 1 billion to 1.4 billion, doubling to 2.1 billion by 2050 [110]. Common health issues among older adults include cognitive impairment, cataracts, and depression [111]. Integrating AI with medical

questionnaires to predict disease occurrence can help family members and health care providers implement early interventions or treatments [112]. However, predictions based solely on questionnaires may have limitations. AI shows high predictive accuracy when incorporating physiological data such as speech signals, heart rate, and medical imaging [113-115]. Future research could combine questionnaire-based predictions with additional physiological information to improve and confirm the accuracy of disease forecasting.

New Response Modalities for Medical Questionnaires

It is estimated that 1.3 billion people worldwide have severe disabilities, and this number is expected to rise [116]. Individuals with disabilities who have experienced stroke, limb loss, or spinal cord injuries often exhibit impaired hand function [117-119], significantly reducing their efficiency in completing traditional paper-based questionnaires. Integrating traditional questionnaires with AI-driven LLMs to create an *AI questionnaire* is one solution. By leveraging LLM technologies, users can respond verbally to questionnaire items, circumventing traditional written formats [120,121] (Figure 7).

Figure 7. New response modalities for medical questionnaires. AI: artificial intelligence.
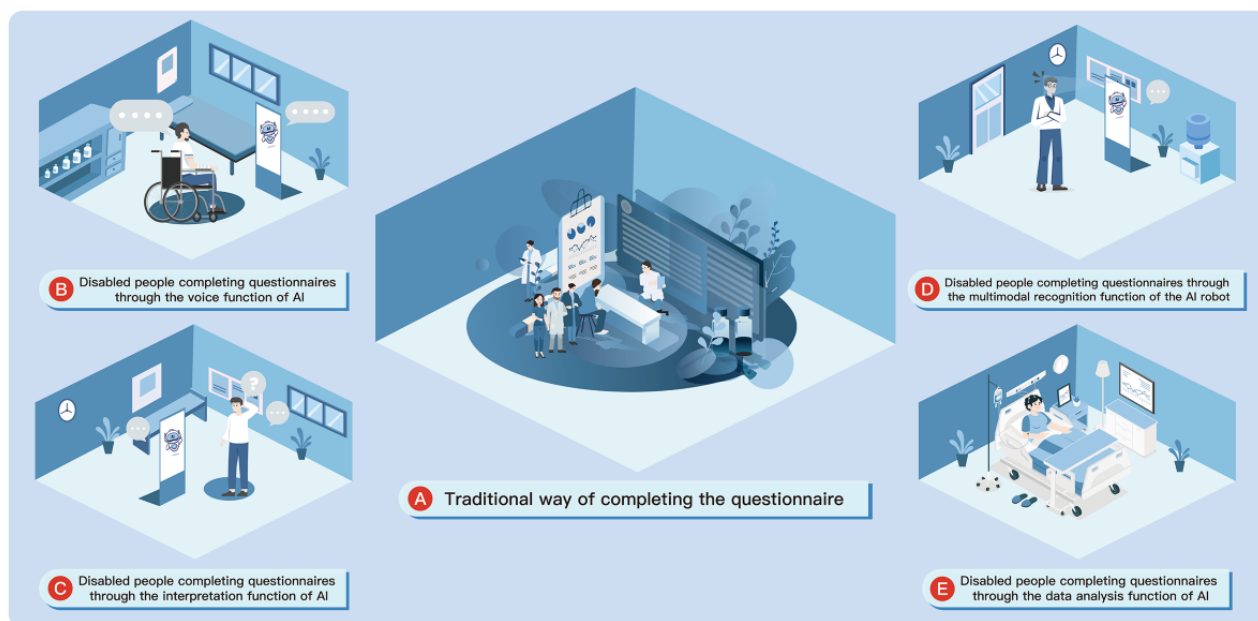
Persons with disabilities may face discrimination in various aspects of life, resulting in generally lower educational attainment than that of other populations [122]. For these groups, especially individuals with mental disabilities such as autism or dementia, challenging vocabulary, complex sentence structures, and ambiguous answer choices pose major obstacles to questionnaire completion [123]. One approach involves using NLP-enhanced questionnaires; AI models such as ChatGPT or Bard can provide immediate explanations of terminology, thereby reducing the difficulty and improving the efficiency of questionnaire completion for individuals with lower educational levels [124,125]. Another solution integrates AI image generators such as DALL-E to add graphical support, thereby increasing the clarity and comprehensibility of questionnaire options [126]. In addition, combining virtual reality paradigms with AI-driven visual questionnaires may offer more adaptive and inclusive assessment options for individuals with cognitive limitations. This strategy could further expand the range of response modalities available in digital health evaluation [127,128].

Some individuals with mental disabilities may resist answering questionnaires through writing or speech. This reluctance can compromise the completion rate of traditional medical questionnaires, disrupt data collection, and undermine data validity. In fact, beyond textual and linguistic information, patients' facial expressions, vocal emotions, and behaviors also carry clinically valuable data. Embodied AI robots with multimodal data collection capabilities may offer a solution [129]. By using cameras, microphones, and other devices, these robots can capture facial expressions, vocal emotional cues, and behavioral indicators during questionnaire administration [130,131]. AI-driven evaluation and analysis of these data can, in turn, guide the refinement of patient health questionnaires.

Individuals with severe disabilities—such as those with cerebral palsy, amyotrophic lateral sclerosis, or locked-in syndrome—may be unable to respond to certain medical questionnaires using normal speech or gestures for physiological or psychological reasons [132,133]. For these populations, clinicians might consider using AI technologies, brain-computer interfaces, or brainwave data to interpret and analyze patients' eye movements, enabling more personalized completion of medical questionnaires. However, in doing so, health care professionals must ensure the accuracy of these physiological measurements [134]. Preventing the misinterpretation of patient information is essential to maintaining the validity and reliability of medical questionnaire data.

Challenges of AI

Overview

AI technologies have already demonstrated substantial potential in the realm of medical questionnaires. However, the widespread implementation of AI still faces a range of complex challenges, including data privacy and security, data quality, system integration, and social equity. This section will systematically examine these issues and provide an in-depth analysis of possible strategies to overcome them. The goal is to furnish a scientific foundation and direction for effectively applying AI in medical questionnaires, thereby driving its comprehensive implementation and optimization within the health care sector.

Limitations of AI Technologies in Medical Questionnaires

Despite the potential that AI brings to the development, prediction, and evaluation of medical questionnaires, it also exposes multidimensional structural limitations in the context of actual clinical requirements.

ML models such as RF and LR are frequently adopted not because they represent the cutting edge of technology but because they are easier for health care professionals to understand, are less costly to validate, and do not require a reconfiguration of the hospital's information system [135,136]. For example, Siddiqua et al [63] achieved high classification accuracy using an integrated ML model in their study of depression risk, but the choice of model was largely motivated by considerations of interpretability rather than absolute optimality of technical performance. This phenomenon suggests that the trust mechanisms required for clinical adoption are far more dominant than model complexity in many settings.

In contrast, DL models embody a different paradigm. These methods are able to mine complex patterns in high-dimensional data and even achieve learning without predefined feature structures [137]. The advantages are particularly significant when dealing with multimodal data or unstructured text [138]. However, in the specific scenario of medical questionnaires, DL models often struggle to generalize stably due to small sample sizes, high data annotation costs, and large model variance [137,139,140]. Even though ANNs performed well when used to classify facial pain in the study by McCartney et al [69], the model still faces greater challenges in clinical settings subject to legal liability due to the lack of a clear and traceable explanatory path for its diagnostic mechanism.

The emergence of natural language models, on the other hand, brings new possibilities for questionnaire development and interactive evaluation. Systems such as ChatGPT are capable of generating linguistically fluent question items and simulating preliminary physician-patient dialogues [141]. However, this class of models suffers from well-known problems such as factual errors, semantic drift, and contextual mismatch [142-144]. Coraci et al [65] validated the correlation of an AI-generated questionnaire for low back pain with a traditional scale but pointed out problems such as incomplete logic of its content and inconsistency in its scoring criteria. In the absence of strong constraints, the variability of such systems is highly likely to undermine the standardized properties on which medical questionnaires are based.

Therefore, although each of the current AI approaches has its own technical advantages, their actual value still depends on whether they can be adapted to specific clinical needs, system constraints, and resource conditions. When the task emphasizes interpretability, limited data samples, or low deployment costs, ML is still a safe choice. When the problem involves high-dimensional inputs or the identification of subtle diagnostic clues, DL is advantageous provided that the output has a verifiable path of interpretation. Natural language modeling contributes most when interactivity and content adaptation are core requirements, but it must operate under controlled mechanisms to avoid destabilizing the data. Therefore, future research should not stop at improving technical performance but should aim to build model selection strategies that respond to real clinical situations.

Data Privacy and Security

The application of AI in medical questionnaires necessitates handling large volumes of personal health data, making data privacy and security primary concerns [145]. The sensitivity of patient information demands stringent privacy safeguards throughout data collection, storage, and sharing processes. Federated learning has emerged as an innovative approach to protect data privacy, allowing AI models to learn from distributed datasets without centralized data sharing, thereby mitigating the risk of data breaches [146]. However, this approach still faces limitations, particularly when dealing with large-scale, heterogeneous data. In such complex scenarios, the privacy-enhancing capabilities of federated learning may be constrained [147]. Hence, future research must focus on developing more effective data privacy protection techniques. Concurrently, as AI systems become more powerful, data security requirements intensify. Related studies can further explore techniques such as homomorphic encryption and differential privacy [148], ensuring data security during transmission and processing.

In fact, safeguarding data privacy and security requires not only technical solutions but also robust legal support. Regulations such as the European Union's General Data Protection Regulation and the US Health Insurance Portability and Accountability Act set legal standards for the legitimate use of personal health data [149]. Nonetheless, striking a balance between efficient data use and privacy protection remains a complex challenge in practice. Future research should continue exploring the equilibrium between privacy safeguards and data availability, especially when sharing data across multiple institutions. Ensuring patient information security while enhancing data usability is a key issue warranting focused attention.

Data Quality and Bias

As global health care systems improve, medical questionnaire data are often collected from multiple institutions. During data sharing among health care facilities, inconsistencies can arise due to variations in language translation, data formats, recording methods, and collection standards. When integrating AI into

the evaluation and prediction processes for medical questionnaires, data discrepancies—such as bias, inaccuracies, or inconsistencies—can significantly undermine the reliability of AI assessments [150,151]. To ensure stable AI model performance across diverse questionnaire types, future research should consider standardizing data collection and processing methods.

Beyond data quality, data bias also critically influences the performance of AI systems in medical questionnaire development, evaluation, and prediction [152]. Bias in AI systems often stems from imbalanced training data [153]. For example, if the questionnaire data used for training underrepresent certain racial or marginalized groups or groups with certain disabilities, the model's accuracy and predictive capabilities will diminish for these populations, potentially hindering their access to timely and effective health care services [154]. To address this issue, it is critical to ensure the diversity and representativeness of training data during questionnaire collection and development. Future research could use standardized methods such as the integration of the WHO's Data Quality Review framework and the Metric framework designed specifically for health care AI to systematically validate the quality of training datasets [155,156]. In addition, future studies could use targeted data collection strategies such as NLP-based terminology simplification, visual questionnaire enhancement, and culturally adapted translation [157]. These might help reduce the risk of exclusion of specific populations, particularly older people, people with disabilities, and linguistically or culturally marginalized groups [158]. These combined efforts can contribute to the development of more inclusive and bias-resistant AI health assessment models.

Technical Limitations and Clinical Implementation Challenges

While the combination of AI with the health care clinical environment offers transformative potential for the evaluation, development, and prediction of health care questionnaires, AI continues to suffer from significant technical limitations that hinder its full application and effectiveness. The accuracy of AI diagnosis is highly dependent on the quality, quantity, and diversity of the training data [159]. If models are developed based on limited or homogeneous datasets, they tend to show a decrease in predictive performance when applied to more diverse or different patient populations, thus increasing the risk of misdiagnosis or missed diagnoses [160]. In addition, the diagnostic conclusions drawn by AI are themselves subject to a degree of uncertainty [161]. Particularly when dealing with microscopic lesions or early stages of a disease, this uncertainty may manifest itself in a high false-positive or false-negative rate, which can negatively affect clinical decision-making and may even delay patients receiving effective treatment [162].

Practical implementation challenges complicate the integration of AI-enhanced questionnaires into clinical practice, especially when considering novel response modalities (Figure 7). Integrating AI-based questionnaire tools into existing clinical workflows can be resource intensive, involving significant initial setup costs and ongoing training requirements for health care providers [163]. Clinicians are often skeptical due to the limited

transparency and interpretability of AI algorithms, which further hinders the widespread adoption of AI [164,165]. Therefore, the future development of transparent and explainable AI systems and the provision of targeted training and education for health care professionals are important strategies to overcome these barriers and facilitate the successful integration of AI into clinical practice.

Integration With Existing Health Care Systems

For AI-based questionnaire tools to be successful in health care environments, the ability to integrate with existing health care infrastructure is critical. However, the complexity of health care systems and the lack of data interoperability pose significant challenges to such integration. Current health care systems are often intricate, with established workflows and legacy technologies that may be incompatible with new AI tools [153]. Implementing AI-based questionnaires typically requires significant infrastructure upgrades, capital investments, and time-consuming workflow redesigns. In addition, different health care institutions often use disparate data standards and formats within their electronic health record systems [166]. This variation complicates the exchange and integration of data from AI-based questionnaire tools.

Related current studies, such as the Mayo Clinic's successful integration of an AI-based mental health screening questionnaire into its Epic electronic health record system using the Fast Healthcare Interoperability Resources standard, demonstrate advancements in health care technology [167,168]. This study significantly improved diagnostic efficiency and clinician satisfaction. Similarly, the UK National Health Service seamlessly integrated an AI-based dementia screening tool using the Health Level Seven protocol [169]. This technology streamlined clinical workflow and improved patient prognosis. These studies demonstrate that standardized interoperability frameworks can play a key role in bridging the technological divide between AI tools and clinical systems.

To facilitate scalable and sustainable AI integration, future research could prioritize standardized data formats such as the Health Level Seven or Fast Healthcare Interoperability Resources protocols to build flexible and scalable technology interfaces. Such solutions help lower the threshold of system integration and reduce the disruption to established processes, which will be beneficial in enhancing the acceptance of AI systems by medical staff and promoting the sustainable development of AI technologies in diverse health care scenarios.

Ethical and Regulatory Considerations

Deploying AI in health care settings raises key ethical and regulatory issues [170]. The transparency of algorithms is critical, and lack of interpretability may undermine clinician trust and potentially compromise patient safety through opaque clinical recommendations. Ensuring that AI decisions are clearly explainable and illustrative is essential to promote clinician acceptance and reduce risk [171,172].

Informed consent is also a key ethical consideration [173,174]. Patients need to be fully informed about how their data will be used by AI systems, including the potential risks and benefits, highlighting the need for transparent communication and robust

consent processes. In addition, the allocation of responsibilities in AI-assisted diagnostics is an ongoing regulatory challenge [175,176]. It is essential to clearly define responsibilities among health care providers, AI developers, and organizations, particularly in the event of errors or adverse events. This requires comprehensive regulatory guidelines and liability frameworks to address the issue of liability in a transparent and effective manner.

It is important to note that the use of AI must comply with strict regulatory standards, including the General Data Protection Regulation and Health Insurance Portability and Accountability Act [177,178]. This is crucial for the legal and ethical use of AI-powered medical questionnaires. Future research should meet these stringent data protection and privacy requirements to better ensure a foundation for the safe and ethical integration of AI technologies into medical practice.

Discussion

This study provided a comprehensive examination of the current applications, potential benefits, and challenges faced by AI in health care questionnaires. AI technologies show significant potential in all phases of questionnaire assessment, development, and prediction. In particular, AI has a positive effect on diagnostic and prognostic support in questionnaire applications for large-scale data processing, construction of personalized assessment tools, and integrated management of complex health information. These innovations highlight the transformative potential of AI in modern health care questionnaires. This is especially true in areas where medical manpower is strained, diagnostic resources are scarce, and data infrastructure is weak. Lightweight ML models and automated questionnaire processes can take on initial screening and triage functions, alleviating manpower pressure and improving early identification. AI questionnaire systems with adaptive capabilities can improve the relevance of questions and reduce the burden of answering in environments where professionals are lacking.

Although AI has demonstrated great advantages in its application to medical questionnaires, its interpretability and adoption mechanisms in clinical decision-making still lack systematic understanding. Questionnaire designs with dynamic logic capabilities still face challenges in maintaining measurement validity, especially in user interaction scenarios that emphasize semantic consistency. Developing AI models that are maintainable and adaptable to deployment in resource-constrained environments also remains a key challenge. In addition, data privacy, data quality bias, and ethical and regulatory barriers are still the core issues constraining the implementation of the technology.

This review has several limitations. First, the number of eligible studies was small, which may restrict the generalizability of the findings. Second, most included studies were of moderate methodological quality, with limited clinical validation and frequent lack of control groups or standardized evaluation metrics. Third, the review was restricted to English- and Chinese-language literature, potentially omitting relevant studies in other languages. Lastly, the heterogeneity of AI models, clinical contexts, and questionnaire types limited the ability to perform quantitative synthesis or direct comparison across studies.

In response to the aforementioned challenges, future research should further promote the evolution of AI systems toward interpretability, contextual adaptability, and cross-platform compatibility and promote the establishment of a more standardized, safe, and controllable AI architecture. The validation of the applicability of AI-based questionnaire tools under diverse health systems should be strengthened, and localized deployment strategies for low-resource environments should be developed. Through the focus on these key issues, AI is expected to continue to drive the overall improvement of health care questionnaires in terms of diagnostic efficiency, patient engagement, and health equity.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA-ScR Checklist.
[PDF File (Adobe PDF File), 1945 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Advances in smart technology in the clinical and research environments in the included studies (N=14).
[DOCX File , 23 KB-Multimedia Appendix 2]

References

1. Saraceno B. [Mental health worldwide: commentary on the WHO 2022 Report]. Assist Inferm Ric. 2024;43(1):48-50. [doi: 10.1702/4250.42272] [Medline: 38572709]

2. Saraceno B, Caldas de Almeida JM. An outstanding message of hope: the WHO World Mental Health Report 2022. Epidemiol Psychiatr Sci. Jul 14, 2022;31:e53. [FREE Full text] [doi: 10.1017/S2045796022000373] [Medline: 35833232]

3. de Castro F, Cappa C, Madans J. Anxiety and depression signs among adolescents in 26 low- and middle-income countries: prevalence and association with functional difficulties. J Adolesc Health. Jan 2023;72(1S):S79-S87. [FREE Full text] [doi: 10.1016/j.jadohealth.2022.03.022] [Medline: 36528385]

4. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry. Mar 2015;56(3):345-365. [doi: 10.1111/jcpp.12381] [Medline: 25649325]

5. Shafer AB. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. J Clin Psychol. Jan 2006;62(1):123-146. [doi: 10.1002/jclp.20213] [Medline: 16287149]

6. Bech P, Kessing LV, Bukh JD. The validity of dysthymia to predict clinical depressive symptoms as measured by the Hamilton Depression Scale at the 5-year follow-up of patients with first episode depression. Nord J Psychiatry. Nov 2016;70(8):563-566. [doi: 10.1080/08039488.2016.1180712] [Medline: 27181030]

7. Schneibel R, Brakemeier E, Wilbertz G, Dykierek P, Zobel I, Schramm E. Sensitivity to detect change and the correlation of clinical factors with the Hamilton Depression Rating Scale and the Beck Depression Inventory in depressed inpatients. Psychiatry Res. Jun 30, 2012;198(1):62-67. [doi: 10.1016/j.psychres.2011.11.014] [Medline: 22445070]

8. Aben I, Verhey F, Lousberg R, Lodder J, Honig A. Validity of the Beck depression inventory, hospital anxiety and depression scale, SCL-90, and Hamilton depression rating scale as screening instruments for depression in stroke patients. Psychosomatics. 2002;43(5):386-393. [doi: 10.1176/appi.psy.43.5.386] [Medline: 12297607]

9. Wu P. Longitudinal measurement invariance of Beck depression inventory-II in early adolescents. Assessment. Apr 2017;24(3):337-345. [doi: 10.1177/1073191115608941] [Medline: 26423351]

10. Hyrkäs K, Appelqvist-Schmidlechner K, Haataja R. Efficacy of clinical supervision: influence on job satisfaction, burnout and quality of care. J Adv Nurs. Aug 2006;55(4):521-535. [doi: 10.1111/j.1365-2648.2006.03936.x] [Medline: 16866847]

11. Edwards D, Burnard P, Hannigan B, Cooper L, Adams J, Juggessur T, et al. Clinical supervision and burnout: the influence of clinical supervision for community mental health nurses. J Clin Nurs. Aug 2006;15(8):1007-1015. [doi: 10.1111/j.1365-2702.2006.01370.x] [Medline: 16879545]

12. Koivu A, Hyrkäs K, Saarinen PI. Who attends clinical supervision? The uptake of clinical supervision by hospital nurses. J Nurs Manag. Jan 2011;19(1):69-79. [doi: 10.1111/j.1365-2834.2010.01185.x] [Medline: 21223407]

13. Yoshifuji A, Nakahara S, Oyama E, Kobayashi R, Shimizu M, Sakamoto A, et al. Managing interhospital referrals during a COVID-19 patient surge in Japan: creating available beds by exchanging patients. Health Secur. 2023;21(3):165-175. [doi: 10.1089/hs.2022.0094] [Medline: 37093031]

14. Meille G, Decker SL, Owens PL, Selden TM. COVID-19 admission rates and changes in US hospital inpatient and intensive care unit occupancy. JAMA Health Forum. Dec 01, 2023;4(12):e234206. [FREE Full text] [doi: 10.1001/jamahealthforum.2023.4206] [Medline: 38038986]

15. Chu K, Lan CE. The impact of the COVID-19 surge on phototherapy in Taiwan: focusing on the patient profile, adherence, and attitude before and after the surge. Skin Res Technol. Apr 2023;29(4):e13314. [FREE Full text] [doi: 10.1111/srt.13314] [Medline: 37113097]

16. Su Z, Tang G, Huang R, Qiao Y, Zhang Z, Dai X. Based on medicine, the now and future of large language models. Cell Mol Bioeng. Aug 2024;17(4):263-277. [doi: 10.1007/s12195-024-00820-3] [Medline: 39372551]

17. Sblendorio E, Dentamaro V, Lo Cascio A, Germini F, Piredda M, Cicolini G. Integrating human expertise and automated methods for a dynamic and multi-parametric evaluation of large language models' feasibility in clinical decision-making. Int J Med Inform. Aug 2024;188:105501. [FREE Full text] [doi: 10.1016/j.ijmedinf.2024.105501] [Medline: 38810498]

18. Park Y, Pillai A, Deng J, Guo E, Gupta M, Paget M, et al. Assessing the research landscape and clinical utility of large language models: a scoping review. BMC Med Inform Decis Mak. Mar 12, 2024;24(1):72. [FREE Full text] [doi: 10.1186/s12911-024-02459-6] [Medline: 38475802]

19. Tailor PD, D'Souza HS, Li H, Starr MR. Vision of the future: large language models in ophthalmology. Curr Opin Ophthalmol. Sep 01, 2024;35(5):391-402. [doi: 10.1097/ICU.0000000000001062] [Medline: 38814572]

20. Lawson McLean A, Wu Y, Lawson McLean AC, Hristidis V. Large language models as decision aids in neuro-oncology: a review of shared decision-making applications. J Cancer Res Clin Oncol. Mar 19, 2024;150(3):139. [FREE Full text] [doi: 10.1007/s00432-024-05673-x] [Medline: 38503921]

21. Teferra BG, Rueda A, Pang H, Valenzano R, Samavi R, Krishnan S, et al. Screening for depression using natural language processing: literature review. Interact J Med Res. Nov 04, 2024;13:e55067. [FREE Full text] [doi: 10.2196/55067] [Medline: 39496145]

22. de Arriba-Pérez F, García-Méndez S. Leveraging large language models through natural language processing to provide interpretable machine learning predictions of mental deterioration in real time. Arab J Sci Eng. Aug 27, 2024:33. [doi: 10.1007/S13369-024-09508-2]

23. Corda E, Massa S, Riboni D. Context-aware behavioral tips to improve sleep quality via machine learning and large language models. Future Internet. Jan 30, 2024;16(2):46. [FREE Full text] [doi: 10.3390/fi16020046]

24. Ma J, Zheng Z, Zhu P, Liu Z. A machine learning and large language model-integrated approach to research project evaluation. J Database Manag. 2024;35(1):1-14. [doi: 10.4018/JDM.345400]

25. Scott IA, Zuccon G. The new paradigm in machine learning - foundation models, large language models and beyond: a primer for physicians. Intern Med J. May 2024;54(5):705-715. [doi: 10.1111/imj.16393] [Medline: 38715436]

26. Cao J, Xu X, Zhang X. Application of machine learning and large language model module for analyzing gut microbiota data. In: Proceedings of the 20th International Conference on Advanced Intelligent Computing in Bioinformatics. 2024. Presented at: ICIC '24; August 5-8, 2024:37-48; Tianjin, China. URL: https://dl.acm.org/doi/10.1007/978-981-97-5689-6_4 [doi: 10.1007/978-981-97-5689-6_4]

27. Zhou J, Shao K, Yu C, Hao Y, Hu L, Chen M. MDEA: modeling depressive emotions aligned with knowledge graphs and large language models. In: Proceedings of the 2024 IEEE Canadian Conference on Electrical and Computer Engineering. 2024. Presented at: CCECE '24; August 6-9, 2024:84-85; Kingston, ON. URL: https://ieeexplore.ieee.org/document/10667092 [doi: 10.1109/ccece59415.2024.10667092]

28. Heston T. Safety of large language models in addressing depression. Cureus. Dec 2023;15(12):e50729. [FREE Full text] [doi: 10.7759/cureus.50729] [Medline: 38111813]

29. Omar M, Levkovich I. Exploring the efficacy and potential of large language models for depression: a systematic review. J Affect Disord. Feb 15, 2025;371:234-244. [doi: 10.1016/j.jad.2024.11.052] [Medline: 39581383]

30. Nowacki A, Sitek W, Rybi?ski H. LLMental: classification of mental disorders with large language models. In: Proceedings of the 27th International Symposium on Foundations of Intelligent Systems. 2024. Presented at: ISMIS '24; June 17-19, 2024:35-44; Poitiers, France. URL: https://link.springer.com/chapter/10.1007/978-3-031-62700-2_4 [doi: 10.1007/978-3-031-62700-2_4]

31. Omar M, Soffer S, Charney AW, Landi I, Nadkarni GN, Klang E. Applications of large language models in psychiatry: a systematic review. Front Psychiatry. 2024;15:1422807. [FREE Full text] [doi: 10.3389/fpsyt.2024.1422807] [Medline: 38979501]

32. Bernert RA, Hilberg AM, Melia R, Kim JP, Shah NH, Abnousi F. Artificial intelligence and suicide prevention: a systematic review of machine learning investigations. Int J Environ Res Public Health. Aug 15, 2020;17(16):5929. [doi: 10.3390/ijerph17165929]

33. Abdelmoteleb S, Ghallab M, IsHak WW. Evaluating the ability of artificial intelligence to predict suicide: a systematic review of reviews. J Affect Disord. Aug 01, 2025;382:525-539. [doi: 10.1016/j.jad.2025.04.078] [Medline: 40274119]

34. Fonseka TM, Bhat V, Kennedy SH. The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviors. Aust N Z J Psychiatry. Oct 2019;53(10):954-964. [doi: 10.1177/0004867419864428] [Medline: 31347389]

35. McGarrigle WJ, Furst J, Jason LA. Psychometric evaluation of the DePaul Symptom Questionnaire-Short Form (DSQ-SF) among adults with Long COVID, ME/CFS, and healthy controls: a machine learning approach. J Health Psychol. Sep 2024;29(11):1241-1252. [doi: 10.1177/13591053231223882] [Medline: 38282368]

36. De Luca GP, Parghi N, El Hayek R, Bloch-Elkouby S, Peterkin D, Wolfe A, et al. Machine learning approach for the development of a crucial tool in suicide prevention: the Suicide Crisis Inventory-2 (SCI-2) Short Form. PLoS One. May 10, 2024;19(5):e0299048. [FREE Full text] [doi: 10.1371/journal.pone.0299048] [Medline: 38728274]

37. Nam KH, Kim DY, Kim DH, Lee JH, Lee JI, Kim MJ, et al. Conversational artificial intelligence for spinal pain questionnaire: validation and user satisfaction. Neurospine. Jun 2022;19(2):348-356. [FREE Full text] [doi: 10.14245/ns.2143080.540] [Medline: 35577340]

38. Kojima T, Fujimura S, Hasebe K, Okanoue Y, Shuya O, Yuki R, et al. Objective assessment of pathological voice using artificial intelligence based on the GRBAS scale. J Voice. May 2024;38(3):561-566. [doi: 10.1016/j.jvoice.2021.11.021] [Medline: 34973892]

39. Rezazadeh H, Ahmadipour H, Salajegheh M. Psychometric evaluation of Persian version of medical artificial intelligence readiness scale for medical students. BMC Med Educ. Jul 24, 2023;23(1):527. [FREE Full text] [doi: 10.1186/s12909-023-04516-6] [Medline: 37488522]

40. Boillat T, Nawaz FA, Rivas H. Readiness to embrace artificial intelligence among medical doctors and students: questionnaire-based study. JMIR Med Educ. Apr 12, 2022;8(2):e34973. [FREE Full text] [doi: 10.2196/34973] [Medline: 35412463]

41. Hamad M, Qtaishat F, Mhairat E, Al-Qunbar A, Jaradat M, Mousa A, et al. Artificial intelligence readiness among Jordanian medical students: using medical artificial intelligence readiness scale for medical students (MAIRS-MS). J Med Educ Curric Dev. 2024;11:23821205241281648. [FREE Full text] [doi: 10.1177/23821205241281648] [Medline: 39346121]

42. Lee EE, Torous J, De Choudhury M, Depp CA, Graham SA, Kim H, et al. Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. Biol Psychiatry Cogn Neurosci Neuroimaging. Sep 2021;6(9):856-864. [FREE Full text] [doi: 10.1016/j.bpsc.2021.02.001] [Medline: 33571718]

43. Olawade D, Wada OZ, Odetayo A, David-Olawade A, Asaolu F, Eberhardt J. Enhancing mental health with artificial intelligence: current trends and future prospects. J Med Surg Public Health. Aug 2024;3:100099. [FREE Full text] [doi: 10.1016/j.glmedi.2024.100099]

44. Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. J Ambient Intell Humaniz Comput. 2023;14(7):8459-8486. [FREE Full text] [doi: 10.1007/s12652-021-03612-z] [Medline: 35039756]

45. Hirani R, Noruzi K, Khuram H, Hussaini AS, Aifuwa EI, Ely KE, et al. Artificial intelligence and healthcare: a journey through history, present innovations, and future possibilities. Life (Basel). Apr 26, 2024;14(5):557. [FREE Full text] [doi: 10.3390/life14050557] [Medline: 38792579]

46. Mesquita F, Mauricio J, Marques G. Depression detection using deep learning and natural language processing techniques: a comparative study. In: Proceedings of the 26th Iberoamerican Conference on Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. 2023. Presented at: CIARP '23; November 27-30, 2023:327-342; Coimbra, Portugal. URL: https://dl.acm.org/doi/10.1007/978-3-031-49018-7_24 [doi: 10.1007/978-3-031-49018-7_24]

47. DeSouza DD, Robin J, Gumus M, Yeung A. Natural language processing as an emerging tool to detect late-life depression. Front Psychiatry. Sep 6, 2021;12:719125. [FREE Full text] [doi: 10.3389/fpsyt.2021.719125] [Medline: 34552519]

48. Oyong I, Utami E, Luthfi E. Natural language processing and lexical approach for depression symptoms screening of Indonesian twitter user. In: Proceedings of the 10th International Conference on Information Technology and Electrical Engineering. 2018. Presented at: ICITEE '18; July 24-26, 2018:359-364; Bali, Indonesia. URL: https://ieeexplore.ieee.org/document/8534929 [doi: 10.1109/iciteed.2018.8534929]

49. Park JY, Seo EH, Yoon H, Won S, Lee KH. Automating Rey complex figure test scoring using a deep learning-based approach: a potential large-scale screening tool for cognitive decline. Alzheimers Res Ther. Aug 30, 2023;15(1):145. [FREE Full text] [doi: 10.1186/s13195-023-01283-w] [Medline: 37649070]

50. Bouzid K, Sharma H, Killcoyne S, Castro DC, Schwaighofer A, Ilse M, et al. Enabling large-scale screening of Barrett's esophagus using weakly supervised deep learning in histopathology. Nat Commun. Mar 11, 2024;15(1):2026. [FREE Full text] [doi: 10.1038/s41467-024-46174-2] [Medline: 38467600]

51. Cao K, Xia Y, Yao J, Han X, Lambert L, Zhang T, et al. Large-scale pancreatic cancer detection via non-contrast CT and deep learning. Nat Med. Dec 20, 2023;29(12):3033-3043. [FREE Full text] [doi: 10.1038/s41591-023-02640-w] [Medline: 37985692]

52. Zhou W, Cheng G, Zhang Z, Zhu L, Jaeger S, Lure FY, et al. Deep learning-based pulmonary tuberculosis automated detection on chest radiography: large-scale independent testing. Quant Imaging Med Surg. Apr 2022;12(4):2344-2355. [FREE Full text] [doi: 10.21037/qims-21-676] [Medline: 35371946]

53. Liu X, Song L, Liu S, Zhang Y. A review of deep-learning-based medical image segmentation methods. Sustainability. Jan 25, 2021;13(3):1224-1222. [FREE Full text] [doi: 10.3390/su13031224]

54. Conze P, Brochard S, Burdin V, Sheehan FT, Pons C. Healthy versus pathological learning transferability in shoulder muscle MRI segmentation using deep convolutional encoder-decoders. Comput Med Imaging Graph. Jul 2020;83:101733. [FREE Full text] [doi: 10.1016/j.compmedimag.2020.101733] [Medline: 32505943]

55. Sekuboyina A, Kukacka J, Kirschke J, Menze B, Valentinitsch A. Attention-driven deep learning for pathological spine segmentation. In: Proceedings of the 5th International Workshop, MSKI 2017, Held in Conjunction with MICCAI 2017 on Computational Methods and Clinical Applications in Musculoskeletal Imaging. 2018. Presented at: MSKI '17; September 10, 2017:108-119; Quebec City, QC. URL: https://doi.org/10.1007/978-3-319-74113-0_10 [doi: 10.1007/978-3-319-74113-0_10]

56. Devda J, Eswari R. Pathological myopia image analysis using deep learning. Procedia Comput Sci. 2019;165:239-244. [FREE Full text] [doi: 10.1016/j.procs.2020.01.084]

57. Brown JM, Campbell JP, Beers A, Chang K, Donohue K, Ostmo S. Fully automated disease severity assessmenttreatment monitoring in retinopathy of prematurity using deep learning. In: Proceedings of the 2018 International Conference on Imaging Informatics for Healthcare, Research, and Applications. 2018. Presented at: IIHRA '18; February 13-15, 2018:33; Houston, TX. URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10579/2295942/Fully-automated-disease-severity-assessment-and-treatment-monitoring-in-retinopathy/10.1117/12.2295942.short?tab=ArticleLinkCited [doi: 10.1117/12.2295942]

58. Ramkumar G, Seetha J, Priyadarshini R, Gopila M, Saranya G. IoT-based patient monitoring system for predicting heart disease using deep learning. Measurement. Aug 2023;218:113235. [doi: 10.1016/j.measurement.2023.113235]

59. Goschenhofer J, Pfister FM, Yuksel KA, Bischl B, Fietzek U, Thomas J. Wearable-based Parkinson's disease severity monitoring using deep learning. In: Proceedings of the 2019 Conference on Machine Learning and Knowledge Discovery in Databases. 2019. Presented at: ECML PKDD '19; September 16-20, 2019:400-415; Würzburg, Germany. URL: https://link.springer.com/chapter/10.1007/978-3-030-46133-1_24 [doi: 10.1007/978-3-030-46133-1_24]

60. Hutto A, Zikry TM, Bohac B, Rose T, Staebler J, Slay J, et al. Using a natural language processing toolkit to classify electronic health records by psychiatric diagnosis. Health Informatics J. 2024;30(4):14604582241296411. [doi: 10.1177/14604582241296411] [Medline: 39466373]

61. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. Oct 02, 2018;169(7):467-473. [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]

62.   Munn Z, Aromataris E, Tufanaru C, Stern C, Porritt K, Farrow J, et al. The development of software to support multiple systematic review types: the Joanna Briggs Institute System for the Unified Management, Assessment and Review of Information (JBI SUMARI). Int J Evid Based Healthc. Mar 2019;17(1):36-43. [doi: 10.1097/XEB.0000000000000152] [Medline: 30239357]

63.   Siddiqua R, Islam N, Bolaka J, Khan R, Momen S. AIDA: artificial intelligence based depression assessment applied to Bangladeshi students. Array. Jul 2023;18:100291. [FREE Full text] [doi: 10.1016/j.array.2023.100291]

64.   van Buchem MM, Neve OM, Kant IMJ, Steyerberg EW, Boosman H, Hensen EF. Analyzing patient experiences using natural language processing: development and validation of the artificial intelligence patient reported experience measure (AI-PREM). BMC Med Inform Decis Mak. Jul 15, 2022;22(1):183. [FREE Full text] [doi: 10.1186/s12911-022-01923-5] [Medline: 35840972]

65.   Coraci D, Maccarone MC, Regazzo G, Accordi G, Papathanasiou JV, Masiero S. ChatGPT in the development of medical questionnaires. The example of the low back pain. Eur J Transl Myol. Dec 15, 2023;33(4):12114. [FREE Full text] [doi: 10.4081/ejtm.2023.12114] [Medline: 38112605]

66.   Wang W, Han X, Zhang J, Shang X, Ha J, Liu Z, et al. Predicting the 10-year risk of cataract surgery using machine learning techniques on questionnaire data: findings from the 45 and Up Study. Br J Ophthalmol. Nov 2022;106(11):1503-1507. [FREE Full text] [doi: 10.1136/bjophthalmol-2020-318609] [Medline: 34039562]

67.   Ha S, Choi SJ, Lee S, Wijaya RH, Kim JH, Joo EY, et al. Predicting the risk of sleep disorders using a machine learning-based simple questionnaire: development and validation study. J Med Internet Res. Sep 21, 2023;25:e46520. [FREE Full text] [doi: 10.2196/46520] [Medline: 37733411]

68.   Shetty RM, Walia T, Osman OT. Reliability and validity of artificial intelligence-based innovative digital scale for the assessment of anxiety in children. Eur J Paediatr Dent. Jan 01, 2024:1. [FREE Full text] [doi: 10.23804/ejpd.2024.1937] [Medline: 38197569]

69.   McCartney S, Weltin M, Burchiel KJ. Use of an artificial neural network for diagnosis of facial pain syndromes: an update. Stereotact Funct Neurosurg. 2014;92(1):44-52. [doi: 10.1159/000353188] [Medline: 24217113]

70.   Sali R, Roohafza H, Sadeghi M, Andalib E, Shavandi H, Sarrafzadegan N. Validation of the revised stressful life event questionnaire using a hybrid model of genetic algorithm and artificial neural networks. Comput Math Methods Med. 2013;2013:601640. [FREE Full text] [doi: 10.1155/2013/601640] [Medline: 23476715]

71.   Ferreira Freitas R, Santos Brant Rocha J, Ives Santos L, de Carvalho Braule Pinto AL, Rodrigues Moreira MH, Piana Santos Lima de Oliveira F, et al. Validity and precision of the international physical activity questionnaire for climacteric women using computational intelligence techniques. PLoS One. 2021;16(1):e0245240. [FREE Full text] [doi: 10.1371/journal.pone.0245240] [Medline: 33444409]

72.   Li G, Zou X, Wang Y, Xu Y, Sun Y, Ma Z, et al. A simplified method for elderly comprehensive geriatric assessment process based on Bayesian network. Beijing Biomed Eng. 2024;41(6):589-596.

73.   Levis B, Benedetti A, Thombs BD. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ. Apr 12, 2019;365:l1781. [FREE Full text] [doi: 10.1136/bmj.l1781] [Medline: 30979729]

74.   Negeri ZF, Levis B, Sun Y, He C, Krishnan A, Wu Y, et al. Depression Screening Data (DEPRESSD) PHQ Group. Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: updated systematic review and individual participant data meta-analysis. BMJ. Oct 05, 2021;375:n2183. [FREE Full text] [doi: 10.1136/bmj.n2183] [Medline: 34610915]

75.   Chang J, Ji Y, Li Y, Pan H, Su P. Prevalence of anxiety symptom and depressive symptom among college students during COVID-19 pandemic: A meta-analysis. J Affect Disord. Sep 01, 2021;292:242-254. [FREE Full text] [doi: 10.1016/j.jad.2021.05.109] [Medline: 34134022]

76.   Li Y, Scherer N, Felix L, Kuper H. Prevalence of depression, anxiety and post-traumatic stress disorder in health care workers during the COVID-19 pandemic: a systematic review and meta-analysis. PLoS One. 2021;16(3):e0246454. [FREE Full text] [doi: 10.1371/journal.pone.0246454] [Medline: 33690641]

77.   Kjaergaard M, Arfwedson Wang CE, Waterloo K, Jorde R. A study of the psychometric properties of the Beck Depression Inventory-II, the Montgomery and Åsberg Depression Rating Scale, and the Hospital Anxiety and Depression Scale in a sample from a healthy population. Scand J Psychol. Feb 2014;55(1):83-89. [doi: 10.1111/sjop.12090] [Medline: 24256317]

78.   Westhoff-Bleck M, Winter L, Aguirre Davila L, Herrmann-Lingen C, Treptau J, Bauersachs J, et al. Diagnostic evaluation of the hospital depression scale (HADS) and the Beck depression inventory II (BDI-II) in adults with congenital heart disease using a structured clinical interview: impact of depression severity. Eur J Prev Cardiol. Mar 2020;27(4):381-390. [doi: 10.1177/2047487319865055] [Medline: 31349778]

79.   Steer RA, Rissmiller DJ, Beck AT. Use of the Beck Depression inventory-II with depressed geriatric inpatients. Behav Res Ther. Mar 2000;38(3):311-318. [doi: 10.1016/s0005-7967(99)00068-6] [Medline: 10665163]

80.   Mullis J, Chen C, Morkos B, Ferguson S. Deep neural networks in natural language processing for classifying requirements by origin and functionality: an application of BERT in system requirements. J Mech Des. 2024;146(4):23-29. [FREE Full text] [doi: 10.1115/1.4063764]

81.   De Luca L, Pastore M, Palladino BE, Reime B, Warth P, Menesini E. The development of Non-Suicidal Self-Injury (NSSI) during adolescence: a systematic review and Bayesian meta-analysis. J Affect Disord. Oct 15, 2023;339:648-659. [FREE Full text] [doi: 10.1016/j.jad.2023.07.091] [Medline: 37479039]

82.   Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell. 2023;6:1169595. [FREE Full text] [doi: 10.3389/frai.2023.1169595] [Medline: 37215063]

83.   Wu Z, Wang J, Zhang C, Peng D, Mellor D, Luo Y, et al. Clinical distinctions in symptomatology and psychiatric comorbidities between misdiagnosed bipolar I and bipolar II disorder versus major depressive disorder. BMC Psychiatry. May 10, 2024;24(1):352. [FREE Full text] [doi: 10.1186/s12888-024-05810-3] [Medline: 38730288]

84.   Akiskal HS, Benazzi F. Continuous distribution of atypical depressive symptoms between major depressive and bipolar II disorders: dose-response relationship with bipolar family history. Psychopathology. 2008;41(1):39-42. [doi: 10.1159/000109954] [Medline: 17952020]

85.   Hu C, Xiang Y, Ungvari GS, Dickerson FB, Kilbourne AM, Si T, et al. Undiagnosed bipolar disorder in patients treated for major depression in China. J Affect Disord. Oct 2012;140(2):181-186. [doi: 10.1016/j.jad.2012.02.014] [Medline: 22397888]

86.   Xiang Y, Zhang L, Wang G, Hu C, Ungvari GS, Dickerson FB, et al. Sociodemographic and clinical features of bipolar disorder patients misdiagnosed with major depressive disorder in China. Bipolar Disord. Mar 2013;15(2):199-205. [FREE Full text] [doi: 10.1111/bdi.12052] [Medline: 23437963]

87.   Lutz W, Leon S, Martinovich Z, Lyons J, Stiles W. Therapist effects in outpatient psychotherapy: a three-level growth curve approach. J Couns Psychol. 2007;54(1):32-39. [FREE Full text] [doi: 10.1037/0022-0167.54.1.32]

88.   Erguzel TT, Ozekes S, Tan O, Gultekin S. Feature selection and classification of electroencephalographic signals: an artificial neural network and genetic algorithm based approach. Clin EEG Neurosci. Oct 2015;46(4):321-326. [doi: 10.1177/1550059414523764] [Medline: 24733718]

89.   Adams LJ, Bello G, Dumancas GG. Development and application of a genetic algorithm for variable optimization and predictive modeling of five-year mortality using questionnaire data. Bioinform Biol Insights. 2015;9(Suppl 3):31-41. [FREE Full text] [doi: 10.4137/BBI.S29469] [Medline: 26604716]

90.   Bottino CM, Barcelos-Ferreira R, Ribeiz SR. Treatment of depression in older adults. Curr Psychiatry Rep. Aug 2012;14(4):289-297. [doi: 10.1007/s11920-012-0281-z] [Medline: 22627999]

91.   Marvanova M, McGrane IR. Treatment approach and modalities for management of depression in older people. Sr Care Pharm. Jan 01, 2021;36(1):11-21. [doi: 10.4140/TCP.n.2021.11] [Medline: 33384030]

92.   Kok RM, Reynolds CF. Management of depression in older adults: a review. JAMA. May 23, 2017;317(20):2114-2122. [doi: 10.1001/jama.2017.5706] [Medline: 28535241]

93.   Barkin RL, Schwer WA, Barkin SJ. Recognition and management of depression in primary care: a focus on the elderly. A pharmacotherapeutic overview of the selection process among the traditional and new antidepressants. Am J Ther. May 2000;7(3):205-226. [doi: 10.1097/00045391-200007030-00008] [Medline: 11317169]

94.   McLachlan S, Dube K, Hitman GA, Fenton NE, Kyrimi E. Bayesian networks in healthcare: distribution by medical condition. Artif Intell Med. Jul 2020;107:101912. [FREE Full text] [doi: 10.1016/j.artmed.2020.101912] [Medline: 32828451]

95.   Sihag G, Delcroix V, Grislin-Le Strugeon E, Siebert X, Piechowiak S, Puisieux F. Combining real data and expert knowledge to build a Bayesian network — application to assess multiple risk factors for fall among elderly people. Expert Systems with Applications. Oct 2024;252:124106. [FREE Full text] [doi: 10.1016/j.eswa.2024.124106]

96.   Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2025-06-20]

97.   Gillham JE, Brunwasser SM. Psychological interventions to prevent depression: a cause for hope. Lancet Psychiatry. Dec 2024;11(12):947-948. [doi: 10.1016/S2215-0366(24)00364-X] [Medline: 39572107]

98.   Islam MM, Hassan S, Akter S, Jibon FA, Sahidullah M. A comprehensive review of predictive analytics models for mental illness using machine learning algorithms. Healthc Anal. Dec 2024;6:100350. [FREE Full text] [doi: 10.1016/j.health.2024.100350]

99.   R S, Mujahid M, Rustam F, Shafique R, Chunduri V, Villar MG, et al. Analyzing sentiments regarding ChatGPT using novel BERT: a machine learning approach. Information. Aug 25, 2023;14(9):474. [doi: 10.3390/info14090474]

100. Chang Y, Liu M, Zhao S, Guo W, Zhang M, Zhang L. Impact of modifiable healthy lifestyles on mortality in Chinese older adults. Sci Rep. Nov 21, 2024;14(1):28869. [FREE Full text] [doi: 10.1038/s41598-024-79072-0] [Medline: 39572655]

101. Zhavoronkov A, Mamoshina P, Vanhaelen Q, Scheibye-Knudsen M, Moskalev A, Aliper A. Artificial intelligence for aging and longevity research: recent advances and perspectives. Ageing Res Rev. Jan 2019;49:49-66. [FREE Full text] [doi: 10.1016/j.arr.2018.11.003] [Medline: 30472217]

102. Dixon D, Sattar H, Moros N, Kesireddy SR, Ahsan H, Lakkimsetti M, et al. Unveiling the influence of AI predictive analytics on patient outcomes: a comprehensive narrative review. Cureus. May 2024;16(5):e59954. [FREE Full text] [doi: 10.7759/cureus.59954] [Medline: 38854327]

103. Wassan JT, Zheng H, Wang H. Role of deep learning in predicting aging-related diseases: a scoping review. Cells. Oct 28, 2021;10(11):2924. [FREE Full text] [doi: 10.3390/cells10112924] [Medline: 34831148]

104. Si D, Cheng SC, Xing R, Liu C, Wu HY. Scaling up prediction of psychosis by natural language processing. In: Proceedings of the 31st International Conference on Tools with Artificial Intelligence. 2019. Presented at: ICTAI '19; November 4-6, 2019:339-347; Portland, OR. URL: https://ieeexplore.ieee.org/document/8995325 [doi: 10.1109/ictai.2019.00055]

105. Haines A, Frumkin H. The role of health professionals in fostering planetary health. In: Haines A, Frumkin H, editors. Planetary Health: Safeguarding Human Health and the Environment in the Anthropocene. Cambridge, MA. Cambridge University Press; 2021:360-378.

106. Javaid M, Haleem A, Singh RP. Health informatics to enhance the healthcare industry's culture: an extensive analysis of its features, contributions, applications and limitations. Inform Health. Sep 2024;1(2):123-148. [doi: 10.1016/j.infoh.2024.05.001]

107. Batko K, Ślęzak A. The use of big data analytics in healthcare. J Big Data. 2022;9(1):3. [FREE Full text] [doi: 10.1186/s40537-021-00553-4] [Medline: 35013701]

108. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. J Big Data. Jun 19, 2019;6(1):54. [doi: 10.1186/S40537-019-0217-0]

109. Jia Q, Guo Y, Wang G, Barnes SJ. Big data analytics in the fight against major public health incidents (including COVID-19): a conceptual framework. Int J Environ Res Public Health. Aug 25, 2020;17(17):6161. [FREE Full text] [doi: 10.3390/ijerph17176161] [Medline: 32854265]

110. Dai Y, Teng D, Zhang C, Wang H, Lai Y, Ding S, et al. Priorities in tackling noncommunicable diseases among the population aged 60 years and older in China, 1990-2021: a population-based study. Ageing Res Rev. Dec 2024;102:102574. [doi: 10.1016/j.arr.2024.102574] [Medline: 39522888]

111. Xiong Z, Li X, Yang D, Xiong C, Xu Q, Zhou Q. The association between cataract and incidence of cognitive impairment in older adults: a systematic review and meta-analysis. Behav Brain Res. Jul 26, 2023;450:114455. [doi: 10.1016/j.bbr.2023.114455] [Medline: 37148915]

112. Hurvitz N, Azmanov H, Kesler A, Ilan Y. Establishing a second-generation artificial intelligence-based system for improving diagnosis, treatment, and monitoring of patients with rare diseases. Eur J Hum Genet. Oct 2021;29(10):1485-1490. [FREE Full text] [doi: 10.1038/s41431-021-00928-4] [Medline: 34276056]

113. Borna S, Haider CR, Maita KC, Torres RA, Avila FR, Garcia JP, et al. A review of voice-based pain detection in adults using artificial intelligence. Bioengineering (Basel). Apr 21, 2023;10(4):500. [FREE Full text] [doi: 10.3390/bioengineering10040500] [Medline: 37106687]

114. Nahavandi D, Alizadehsani R, Khosravi A, Acharya UR. Application of artificial intelligence in wearable devices: opportunities and challenges. Comput Methods Programs Biomed. Jan 2022;213:106541. [doi: 10.1016/j.cmpb.2021.106541] [Medline: 34837860]

115. Xu X, Li J, Zhu Z, Zhao L, Wang H, Song C, et al. A comprehensive review on synergy of multi-modal data and AI technologies in medical diagnosis. Bioengineering (Basel). Feb 25, 2024;11(3):219. [FREE Full text] [doi: 10.3390/bioengineering11030219] [Medline: 38534493]

116. Gréaux M, Moro MF, Kamenov K, Russell AM, Barrett D, Cieza A. Health equity for persons with disabilities: a global scoping review on barriers and interventions in healthcare services. Int J Equity Health. Nov 13, 2023;22(1):236. [FREE Full text] [doi: 10.1186/s12939-023-02035-w] [Medline: 37957602]

117. Barry AJ, Triandafilou KM, Stoykov ME, Bansal N, Roth EJ, Kamper DG. Survivors of chronic stroke experience continued impairment of dexterity but not strength in the nonparetic upper limb. Arch Phys Med Rehabil. Jul 2020;101(7):1170-1175. [FREE Full text] [doi: 10.1016/j.apmr.2020.01.018] [Medline: 32113974]

118. Cappello L, Meyer JT, Galloway KC, Peisner JD, Granberry R, Wagner DA, et al. Assisting hand function after spinal cord injury with a fabric-based soft robotic glove. J Neuroeng Rehabil. Jun 28, 2018;15(1):59. [FREE Full text] [doi: 10.1186/s12984-018-0391-x] [Medline: 29954401]

119. Moritz C, Field-Fote EC, Tefertiller C, van Nes I, Trumbower R, Kalsi-Ryan S, et al. Non-invasive spinal cord electrical stimulation for arm and hand function in chronic tetraplegia: a safety and efficacy trial. Nat Med. May 2024;30(5):1276-1283. [FREE Full text] [doi: 10.1038/s41591-024-02940-9] [Medline: 38769431]

120. Wang X, Zhang H, Zhao S, Chen H, Cheng B, Ding Z, et al. HiBERT: Detecting the illogical patterns with hierarchical BERT for multi-turn dialogue reasoning. Neurocomputing. Mar 2023;524:167-177. [doi: 10.1016/j.neucom.2022.12.038]

121. Ni J, Young T, Pandelea V, Xue F, Cambria E. Recent advances in deep learning based dialogue systems: a systematic survey. Artif Intell Rev. Aug 20, 2022;56(4):3055-3155. [doi: 10.1007/S10462-022-10248-8]

122. Acharya Y, Yang D. The effect of disability on educational, labor market, and marital outcomes in a low-income context. SSM Popul Health. Sep 2022;19:101155. [FREE Full text] [doi: 10.1016/j.ssmph.2022.101155] [Medline: 35813185]

123. Nicolaidis C, Raymaker DM, McDonald KE, Lund EM, Leotti S, Kapp SK, et al. Creating accessible survey instruments for use with autistic adults and people with intellectual disability: lessons learned and recommendations. Autism Adulthood. Mar 01, 2020;2(1):61-76. [FREE Full text] [doi: 10.1089/aut.2019.0074] [Medline: 32355908]

124. Steimetz E, Minkowitz J, Gabutan EC, Ngichabe J, Attia H, Hershkop M, et al. Use of artificial intelligence chatbots in interpretation of pathology reports. JAMA Netw Open. May 01, 2024;7(5):e2412767. [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.12767] [Medline: 38776080]

125. Chen X, Zhang W, Xu P, Zhao Z, Zheng Y, Shi D, et al. FFA-GPT: an automated pipeline for fundus fluorescein angiography interpretation and question-answer. NPJ Digit Med. May 03, 2024;7(1):111-118. [FREE Full text] [doi: 10.1038/s41746-024-01101-z] [Medline: 38702471]

126. Goparaju N. Picture this: text-to-image models transforming pediatric emergency medicine. Ann Emerg Med. Dec 2024;84(6):651-657. [doi: 10.1016/j.annemergmed.2024.07.019] [Medline: 39365207]

127. Makmee P, Wongupparaj P. VR cognitive-based intervention for enhancing cognitive functions and well-being in older adults with mild cognitive impairment: behavioral and EEG evidence. Psychosoc Interv. Jan 2025;34(1):37-51. [doi: 10.5093/pi2025a4] [Medline: 39781014]

128. Makmee P, Wongupparaj P. Virtual reality-based cognitive intervention for enhancing executive functions in community-dwelling older adults. Psychosoc Interv. Jul 2022;31(3):133-144. [FREE Full text] [doi: 10.5093/pi2022a10] [Medline: 37361011]

129. Shoumy NJ, Ang L, Seng KP, Rahaman D, Zia T. Multimodal big data affective analytics: a comprehensive survey using text, audio, visual and physiological signals. J Netw Comput Appl. Jan 2020;149:102447. [doi: 10.1016/j.jnca.2019.102447]

130. Yip M, Salcudean S, Goldberg K, Althoefer K, Menciassi A, Opfermann JD, et al. Artificial intelligence meets medical robotics. Science. Jul 14, 2023;381(6654):141-146. [doi: 10.1126/science.adj3312] [Medline: 37440630]

131. Bartolozzi C, Indiveri G, Donati E. Embodied neuromorphic intelligence. Nat Commun. Mar 23, 2022;13(1):1024. [FREE Full text] [doi: 10.1038/s41467-022-28487-2] [Medline: 35197450]

132. Harvey A, Smith N, Smith M, Ostojic K, Berryman C. Chronic pain in children and young people with cerebral palsy: a narrative review of challenges, advances, and future directions. BMC Med. Jun 11, 2024;22(1):238. [FREE Full text] [doi: 10.1186/s12916-024-03458-0] [Medline: 38862988]

133. Vidal F. Phenomenology of the locked-in syndrome: an overview and some suggestions. Neuroethics. Oct 31, 2018;13(2):119-143. [doi: 10.1007/S12152-018-9388-1]

134. Angrick M, Luo S, Rabbani Q, Candrea DN, Shah S, Milsap GW, et al. Online speech synthesis using a chronically implanted brain-computer interface in an individual with ALS. medRxiv. Preprint posted online July 1, 2023. [FREE Full text] [doi: 10.1101/2023.06.30.23291352] [Medline: 37425721]

135. Nasarian E, Alizadehsani R, Acharya U, Tsui K. Designing interpretable ML system to enhance trust in healthcare: a systematic review to proposed responsible clinician-AI-collaboration framework. Inf fusion. Aug 2024;108:102412. [doi: 10.1016/j.inffus.2024.102412]

136. Lu SC, Swisher CL, Chung C, Jaffray D, Sidey-Gibbons C. On the importance of interpretable machine learning predictions to inform clinical decision making in oncology. Front Oncol. 2023;13:1129380. [FREE Full text] [doi: 10.3389/fonc.2023.1129380] [Medline: 36925929]

137. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform. Nov 27, 2018;19(6):1236-1246. [FREE Full text] [doi: 10.1093/bib/bbx044] [Medline: 28481991]

138. Shamshirband S, Fathi M, Dehzangi A, Chronopoulos AT, Alinejad-Rokny H. A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. J Biomed Inform. Jan 2021;113:103627. [FREE Full text] [doi: 10.1016/j.jbi.2020.103627] [Medline: 33259944]

139. Chang K, Balachandar N, Lam C, Yi D, Brown J, Beers A, et al. Distributed deep learning networks among institutions for medical imaging. J Am Med Inform Assoc. Aug 01, 2018;25(8):945-954. [FREE Full text] [doi: 10.1093/jamia/ocy017] [Medline: 29617797]

140. Rahman A, Debnath T, Kundu D, Khan MS, Aishi AA, Sazzad S, et al. Machine learning and deep learning-based approach in smart healthcare: recent advances, applications, challenges and opportunities. AIMS Public Health. 2024;11(1):58-109. [FREE Full text] [doi: 10.3934/publichealth.2024004] [Medline: 38617415]

141. Harrison CJ, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction to natural language processing. BMC Med Res Methodol. Jul 31, 2021;21(1):158. [FREE Full text] [doi: 10.1186/s12874-021-01347-1] [Medline: 34332525]

142. Hager P, Jungmann F, Holland R, Bhagat K, Hubrecht I, Knauer M, et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nat Med. Sep 04, 2024;30(9):2613-2622. [doi: 10.1038/s41591-024-03097-1] [Medline: 38965432]

143. Meng X, Yan X, Zhang K, Liu D, Cui X, Yang Y, et al. The application of large language models in medicine: a scoping review. iScience. May 17, 2024;27(5):109713. [FREE Full text] [doi: 10.1016/j.isci.2024.109713] [Medline: 38746668]

144. Vrdoljak J, Boban Z, Vilović M, Kumrić M, Božić J. A review of large language models in medical education, clinical decision support, and healthcare administration. Healthcare (Basel). Mar 10, 2025;13(6):603. [FREE Full text] [doi: 10.3390/healthcare13060603] [Medline: 40150453]

145. Khalid N, Qayyum A, Bilal M, Al-Fuqaha A, Qadir J. Privacy-preserving artificial intelligence in healthcare: techniques and applications. Comput Biol Med. May 2023;158:106848. [FREE Full text] [doi: 10.1016/j.compbiomed.2023.106848] [Medline: 37044052]

146. Yazdinejad A, Dehghantanha A, Karimipour H, Srivastava G, Parizi RM. A robust privacy-preserving federated learning model against model poisoning attacks. IEEE Trans Inform Forensic Secur. 2024;19:6693-6708. [doi: 10.1109/tifs.2024.3420126]

147. Beitollahi M, Lu N. Federated learning over wireless networks: challenges and solutions. IEEE Internet Things J. Aug 15, 2023;10(16):14749-14763. [doi: 10.1109/jiot.2023.3285868]

148. Yang M, Guo T, Zhu T, Tjuawinata I, Zhao J, Lam K. Local differential privacy and its applications: a comprehensive survey. Comput Stand Inter. Apr 2024;89:103827. [doi: 10.1016/j.csi.2023.103827]

149. van Bekkum M, Zuiderveen Borgesius F. Using sensitive data to prevent discrimination by artificial intelligence: does the GDPR need a new exception? Comput Law Secur Rev. Apr 2023;48:105770. [doi: 10.1016/j.clsr.2022.105770]

150. Cross JL, Choma MA, Onofrey JA. Bias in medical AI: implications for clinical decision-making. PLOS Digit Health. Nov 2024;3(11):e0000651. [doi: 10.1371/journal.pdig.0000651] [Medline: 39509461]

151. Li YH, Li YL, Wei MY, Li GY. Innovation and challenges of artificial intelligence technology in personalized healthcare. Sci Rep. Aug 16, 2024;14(1):18994. [FREE Full text] [doi: 10.1038/s41598-024-70073-7] [Medline: 39152194]

152. Paulus JK, Kent DM. Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. NPJ Digit Med. 2020;3:99. [FREE Full text] [doi: 10.1038/s41746-020-0304-9] [Medline: 32821854]

153. Albahri A, Duhaim AM, Fadhel MA, Alnoor A, Baqer NS, Alzubaidi L, et al. A systematic review of trustworthy and explainable artificial intelligence in healthcare: assessment of quality, bias risk, and data fusion. Information Fusion. Aug 2023;96:156-191. [doi: 10.1016/j.inffus.2023.03.008]

154. Sun M, Oliwa T, Peek ME, Tung EL. Negative patient descriptors: documenting racial bias in the electronic health record. Health Aff (Millwood). Mar 2022;41(2):203-211. [FREE Full text] [doi: 10.1377/hlthaff.2021.01423] [Medline: 35044842]

155. Data quality assurance: module 3: site assessment of data quality: data verification and system assessment. World Health Organization. URL: https://www.who.int/publications/i/item/9789240049123 [accessed 2025-05-29]

156. Schwabe D, Becker K, Seyferth M, Klaß A, Schaeffter T. The METRIC-framework for assessing data quality for trustworthy AI in medicine: a systematic review. NPJ Digit Med. Aug 03, 2024;7(1):203-275. [FREE Full text] [doi: 10.1038/s41746-024-01196-4] [Medline: 39097662]

157. Alkhawaldeh M, Khasawneh M. Advancing natural language processing for adaptive assistive technologies in reading and writing disabilities. J Namib Stud Hist Polit Cult. 2023;35(1):841-860. [doi: 10.25365/jnshpc.2023.35.19]

158. Almufareh M, Kausar S, Humayun M, Tehsin S. A conceptual model for inclusive technology: advancing disability inclusion through artificial intelligence. J Disabil Res. 2024;3(1):5. [doi: 10.57197/jdr-2023-0060]

159. Jayakumar S, Sounderajah V, Normahani P, Harling L, Markar SR, Ashrafian H, et al. Quality assessment standards in artificial intelligence diagnostic accuracy systematic reviews: a meta-research study. NPJ Digit Med. Jan 27, 2022;5(1):11. [FREE Full text] [doi: 10.1038/s41746-021-00544-y] [Medline: 35087178]

160. Celi LA, Cellini J, Charpignon ML, Dee EC, Dernoncourt F, Eber R, et al. for MIT Critical Data. Sources of bias in artificial intelligence that perpetuate healthcare disparities-a global review. PLOS Digit Health. Mar 2022;1(3):e0000022. [FREE Full text] [doi: 10.1371/journal.pdig.0000022] [Medline: 36812532]

161. Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. NPJ Digit Med. Jan 05, 2021;4(1):4. [FREE Full text] [doi: 10.1038/s41746-020-00367-3] [Medline: 33402680]

162. Kim HE, Kim HH, Han BK, Kim KH, Han K, Nam H, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. Lancet Digit Health. Mar 2020;2(3):e138-e148. [FREE Full text] [doi: 10.1016/S2589-7500(20)30003-0] [Medline: 33334578]

163. Petersson L, Larsson I, Nygren JM, Nilsen P, Neher M, Reed JE, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. BMC Health Serv Res. Jul 01, 2022;22(1):850. [FREE Full text] [doi: 10.1186/s12913-022-08215-8] [Medline: 35778736]

164. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. Nov 30, 2020;20(1):310. [FREE Full text] [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]

165. Hassija V, Chamola V, Mahapatra A, Singal A, Goel D, Huang K, et al. Interpreting black-box models: a review on explainable artificial intelligence. Cogn Comput. Aug 24, 2023;16(1):45-74. [doi: 10.1007/S12559-023-10179-8]

166. Cerchione R, Centobelli P, Riccio E, Abbate S, Oropallo E. Blockchain's coming to hospital to digitalize healthcare services: designing a distributed electronic health record ecosystem. Technovation. Feb 2023;120:102480. [doi: 10.1016/j.technovation.2022.102480]

167. Hong N, Wen A, Shen F, Sohn S, Wang C, Liu H, et al. Developing a scalable FHIR-based clinical data normalization pipeline for standardizing and integrating unstructured and structured electronic health record data. JAMIA Open. Dec 2019;2(4):570-579. [FREE Full text] [doi: 10.1093/jamiaopen/ooz056] [Medline: 32025655]

168. Tabari P, Costagliola G, De Rosa M, Boeker M. State-of-the-art fast healthcare interoperability resources (FHIR)-based data model and structure implementations: systematic scoping review. JMIR Med Inform. Sep 24, 2024;12:e58445. [FREE Full text] [doi: 10.2196/58445] [Medline: 39316433]

169. Rajab RM, Abuhmida M, Wilson I, Ward RP. A Review of IoMT Security and Privacy related Frameworks. In: Proceedings of the 23rd European Conference on Cyber Warfare and Security. 2024. Presented at: ECCWS '24; June 27-June 28, 2024:733-743; Jyväskylä, Finland. [doi: 10.34190/eccws.23.1.2239]

170. Farhud DD, Zokaei S. Ethical issues of artificial intelligence in medicine and healthcare. Iran J Public Health. Nov 2021;50(11):i-v. [FREE Full text] [doi: 10.18502/ijph.v50i11.7600] [Medline: 35223619]

171. Mennella C, Maniscalco U, De Pietro G, Esposito M. Ethical and regulatory challenges of AI technologies in healthcare: a narrative review. Heliyon. Mar 29, 2024;10(4):e26297. [FREE Full text] [doi: 10.1016/j.heliyon.2024.e26297] [Medline: 38384518]

172. Li F, Ruijs N, Lu Y. Ethics and AI: a systematic review on ethical concerns and related strategies for designing with AI in healthcare. AI. Dec 31, 2022;4(1):28-53. [doi: 10.3390/ai4010003]

173. Cohen IG. Informed consent and medical artificial intelligence: what to tell the patient? SSRN Journal. Preprint posted online February 26, 2020. [FREE Full text] [doi: 10.2139/ssrn.3529576]

174. Schiff D, Borenstein J. How should clinicians communicate with patients about the roles of artificially intelligent team members? AMA J Ethics. Mar 01, 2019;21(2):E138-E145. [FREE Full text] [doi: 10.1001/amajethics.2019.138] [Medline: 30794123]

175. Naik N, Hameed BM, Shetty DK, Swain D, Shah M, Paul R, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility? Front Surg. 2022;9:862322. [FREE Full text] [doi: 10.3389/fsurg.2022.862322] [Medline: 35360424]

176. Bottomley D, Thaldar D. Liability for harm caused by AI in healthcare: an overview of the core legal concepts. Front Pharmacol. 2023;14:1297353. [FREE Full text] [doi: 10.3389/fphar.2023.1297353] [Medline: 38161692]

177. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. In: Gerke S, Minssen T, editors. Artificial Intelligence in Healthcare. the Netherlands. Academic Press; 2020:295-336.

178. Sivarajah U, Wang Y, Olya H, Mathew S. Responsible artificial intelligence (AI) for digital health and medical analytics. Inf Syst Front. Jun 05, 2023;16(3):1-6. [FREE Full text] [doi: 10.1007/s10796-023-10412-7] [Medline: 37361886]

Abbreviations

AI: artificial intelligence
ANN: artificial neural network
AUC: area under the curve
AUROC: area under the receiver operating characteristic curve
BERT: bidirectional encoder representations from transformers
CNN: convolutional neural network
DL: deep learning
JBI: Joanna Briggs Institute
LLM: large language model
LR: logistic regression
ML: machine learning
NLP: natural language processing
OSA: obstructive sleep apnea
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RF: random forest
WHO: World Health Organization

XSL•FO
RenderX