International Journal of Medical Informatics 205 (2026) 106103

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Review article

Scoping review on the economic aspects of machine learning applications in healthcare

Hanna von Gerich [a,b,*], Mikael Helenius [c,a], Iiris Hörhammer [d], Hans Moen [d], Laura-Maria Peltonen [e]

[a] Department of Nursing Science, University of Turku, Finland
[b] Department of Health and Social Management, University of Eastern Finland, Finland
[c] Turku University Hospital, Finland
[d] Department of Computer Science, Aalto University, Finland
[e] Department of Health and Social Management, University of Eastern Finland, and Kuopio University Hospital, Finland

ARTICLE INFO

Keywords:
Artificial intelligence
Machine learning
Costs and Cost Analysis
Healthcare

ABSTRACT

Background: The development and use of artificial intelligence and machine learning technologies in healthcare have increased, prompting a need for evidence on their safety and value. Economic evaluations support healthcare decision-making and resource allocation. This scoping review aimed to map and synthesize current approaches to evaluating the economic aspects of machine learning based technologies implemented in healthcare.

Methods: Following the updated JBI guidance for scoping reviews, six databases (PubMed, CINAHL, Cochrane Library, Embase, Scopus, and IEEE Xplore) were searched for studies evaluating the economic aspects of machine learning-based technologies within healthcare. No exclusions were applied to healthcare settings, healthcare professionals or used economic evaluation methods. The results of data extraction were analyzed using descriptive statistics and inductive coding. The reporting of the studies was compared against the CHEERS-AI statement.

Results: A total of 6332 references were retrieved, with 18 studies included in the review. The studies comprised economic evaluations (n = 9), impact evaluations (n = 5), and performance evaluations (n = 4), with cost-effectiveness analysis being the most frequently used economic evaluation method (n = 8). The comparison of the studies to the reporting guidelines revealed gaps in the reporting of details from economic evaluations and the artificial intelligence nature of the technologies. Overall, the study alignment with the CHEERS-AI items on average was 39.6 %, with 64.1 % alignment with economic evaluation details, and 21.3 % alignment with key details related to the artificial intelligence nature of the evaluated technologies.

Conclusions: The current literature evaluating the economic aspects of machine learning-based technologies implemented in healthcare reveals gaps in coherence and coverage. Frameworks guiding artificial intelligence development should be refined to incorporate components related to system evaluation and post-implementation considerations. Further, multidisciplinary collaboration should be enhanced and promoted.

1. Introduction

The constant expansion and storage of healthcare data have increased the significance of artificial intelligence (AI) and, specifically, machine learning (ML)-based technologies in healthcare [1]. ML serves as the most notable subset of AI used in healthcare, leveraging real-world data to create and refine predictions or classifications [2]. Within the scope of ML, a variety of deep learning models have emerged as powerful and effective methods for analyzing complex datasets and addressing multifaceted problems related to healthcare [3]. Research to support, streamline or automate complex tasks has exponentially increased, utilizing a vast variety of different ML methods with different health data types. These tasks include but are not limited to image-based liver organ and tissue segmentation [4,5] and retinal image processing [6], signal-based Electroencephalography analysis [7], and clinical data −based disease detection and diagnosis [8,9] and risk management

* Corresponding author.
E-mail address: hanna.m.vongerich@utu.fi (H. von Gerich).

https://doi.org/10.1016/j.ijmedinf.2025.106103
Received 25 July 2025; Received in revised form 28 August 2025; Accepted 29 August 2025
Available online 30 August 2025

[10,11]. Applying ML to multimodal data can contribute to a more comprehensive understanding of the patient health status, promote clinical decision making and personalized care, and automate and optimize administrative tasks [12,13].

The growing interest in applying and developing novel AI-driven solutions for healthcare has led to an increase in the number of frameworks guiding AI implementation and evaluation. Following the hype, most frameworks focus on the early technology development stage, whereas healthcare professionals report challenges related to system adoption in clinical settings [14]. Similarly, the evaluation of AI-based technologies is largely focused on model performance metrics such as precision or accuracy, while research on their clinical impact remains comparatively rare [15]. The sustainable use of AI requires careful preparation and examination of the quality, safety, and potential bias caused by the used data, as well as the explainability and unforeseen outcomes of the systems [16]. Before their widespread implementation, healthcare professionals and policy leaders require strong empirical evidence regarding the safe and effective use of these technologies [17]. The rapid growth in healthcare expenditures has pressured researchers and policymakers to pursue strategies to optimize healthcare resource allocation and management [18]. Economic evaluations support healthcare decision-making by enabling the assessment and comparison of interventions based on their scalability, sustainability, and impact on population health [19]. They are a crucial step in health technology assessment but need to be compared against one another to guide informed and transparent decision-making [20].

AI is widely expected to decrease healthcare costs while enhancing the outcomes, but the economic impacts are still relatively unexplored [21,22]. Recent reviews have explored the economic aspects of AI-based technologies in healthcare, with the current evidence predominantly limited to specific economic evaluation methods [23,24,25,26,27] or specific settings within healthcare, such as clinical care [24]. These reviews suggest that the evaluation methods used are limited [27], and the reporting of the evaluations as well as the technologies is insufficient [24,25,26], with many studies focusing on technologies not yet in clinical use [23].

In response to the current evidence, there is a need to enhance the quality and quantity of evaluating the economic aspects of AI in healthcare. Whilst previous reviews offer a broad overview, their primary focus lies on economic details, without addressing the distinctive features associated with AI or pinpointing the deficiencies in reporting them. The Consolidated Health Economic Evaluation Reporting Standards for Interventions that use AI (CHEERS-AI) statement is a 38-item set of standards developed for reporting economic evaluations of AI interventions [28]. The framework builds on the established CHEERS-2022 statement [29], adding more nuance and detail about reporting and capturing the AI nature of the intervention. Comparing the current body of knowledge with the framework can provide a refined understanding of how AI technologies are reported and assessed.

This study aimed to map and synthesize current approaches to evaluating the economic aspects of ML-based technologies implemented in healthcare. AI is a rather broad term that encompasses a wide range of techniques, including rule-based systems, expert systems, symbolic reasoning, and robotics, many of which do not involve data-driven learning. Narrowing the scope to ML models, the studies are more likely to employ algorithms that learn from relevant data and are tailored to a specific task. A gap exists in the literature regarding the comprehensive economic analysis of state-of-the-art ML-based technologies currently implemented in healthcare. The results serve as a basis for economic evaluation in the development and implementation of AI-based technologies in healthcare, as well as informing the refinement of frameworks and guidelines that steer future AI development.

2. Methods

The scoping review methodology was employed to encompass all available methods for evaluating, assessing, and analyzing the economic aspects of implementing ML-based technologies in healthcare. It was conducted following the updated JBI guidance for scoping review [30]. A protocol for this study was published in OSF registries (https://www.osf.io), DOI https://doi.org/10.17605/OSF.IO/7KCV8.

2.1. The review questions

To address the research aim, the following primary research question, with two sub-questions, was conducted:

1. Which ML-based technologies used by healthcare professionals and healthcare users have been evaluated for economic aspects in healthcare?
2. Which evaluation methods have been used?
3. How do the studies evaluate and present the economic aspects of implementing ML-based technologies in healthcare?

2.2. Inclusion criteria

The inclusion criteria for the reviewed study population were all healthcare providers working in various healthcare settings, such as clinical settings and healthcare management. Technologies were strictly limited to explicit descriptions of ML. All methods for evaluating the economic aspects of the technologies, including both direct and indirect costs, were considered. The study characteristics covered primary research studies, including quantitative, qualitative and mixed-methods research designs. We included original, published, peer-reviewed articles and proceedings in English.

The exclusion criteria were used to eliminate technologies that were commercial, private use without the intervention or oversight of a healthcare professional, or that were targeted for the education or training of healthcare professionals. The excluded study characteristics included literature reviews and *meta*-analyses, as well as modeling and simulation studies. Study protocols and abstracts were excluded.

2.3. Search strategy

The search strategy was developed and refined within the research team and tested against a set of relevant benchmark articles. The search was conducted on November 14, 2024, using the most appropriate databases to locate research literature from biomedical and life sciences, as well as information technology. These included PubMed (MEDLINE), CINAHL (EBSCO), Cochrane Library (Wiley), Embase (Elsevier), Scopus (Elsevier) and IEEE Xplore (IEEE/IEE Electronic Library databases).

The search phrases were formulated using the MeSH entry terms and categories. The categories for "artificial intelligence" and "machine learning" were explored for AI-based technologies, and the category "Costs and Cost Analysis" was used for economic aspects. The search phrase was complemented using applicable database-specific subject headings (ie. MeSH, CINAHL Subject Headings) and Boolean operators OR and AND, with NEAR used when appropriate. The date range was 10 years, and the title, abstract and keywords were searched:

("Artificial intelligence" OR (Comput AND (reasoning OR intelligence OR vision OR (knowledge AND (acquisition OR representation))) OR "neural networks") OR "Machine Intelligence" OR "machine learning" OR "Deep Learning" OR "Support Vector Machine" OR "natural language processing" OR "sentiment analysis") AND (cost* OR expen* OR value) AND (assess* OR evalua* OR feasib* OR analys* OR estim* OR calcul*) AND (healthcare*).*

2.4. Evidence screening and selection

The retrieved data was downloaded into Covidence systematic review tool (www. covidence.org) for duplicate detection and manual removal, title and abstract screen, full text screen and data extraction.

The title and abstract screening, as well as full-text screening, were conducted by two blinded individual reviewers, with a third reviewer resolving any conflicting decisions. The inclusion and exclusion criteria were discussed and refined by the search team during the early phases of the screening process.

Following the scoping review methodology recommendations, no formal quality assessment was conducted on the admitted studies. Instead, the reporting characteristics of the studies were examined using the CHEERS-AI statement [28] by one researcher.

2.5. Data extraction

The data extraction template was created in Covidence and pilot tested by three reviewers using a small sample of the included studies. Following the template refinement, the data were extracted by two independent researchers, with a third reviewer finalizing the consensus on the data extractions. Guided by the JBI standardized data extraction form [31], the following data elements were extracted from the included studies:

- General information: Citation details (e.g. author/s date, title, journal, volume, issue, pages), country, received funding
- Guidelines or criteria used for reporting the study.
- Study characteristics: study aim, design, setting and participants.
- Description of the evaluated technology: intended use and users, ML methods.
- Description of the economic evaluation: evaluation methodology, intervention, economic modeling, included cost elements, considered outcomes.

2.6. Data analysis

The data were analyzed using basic descriptive analysis methods, such as frequency counts, and qualitative data were coded inductively into categories.

The full economic evaluation methods were grouped following the methodological categorization outlined by Turner et al [32] and Pandit [33]. *Cost-effectiveness analysis* compares the costs of two or more interventions with measurable, one-dimensional outcomes, such as averted cases or increased lifespan. *Cost-utility analysis* compares the costs against multidimensional outcomes, such as quality-adjusted life years. *Cost-benefit analysis* evaluates two or more alternatives, where both costs and outcomes are valued in financial terms. In contrast, *cost-minimization analysis* assumes equal outcomes and focuses on comparing the costs of the interventions. Finally, *cost-consequence analysis* lists the costs and multiple outcomes of interventions separately.

Partial economic evaluations, i.e., analyses that were not comparative, did not contain a comparator or were one-dimensional, were categorized as *cost analyses* [33]. In cases where the used economic evaluation methods were inconsistently reported or included a combination of approaches, the categorization provided by the authors was used.

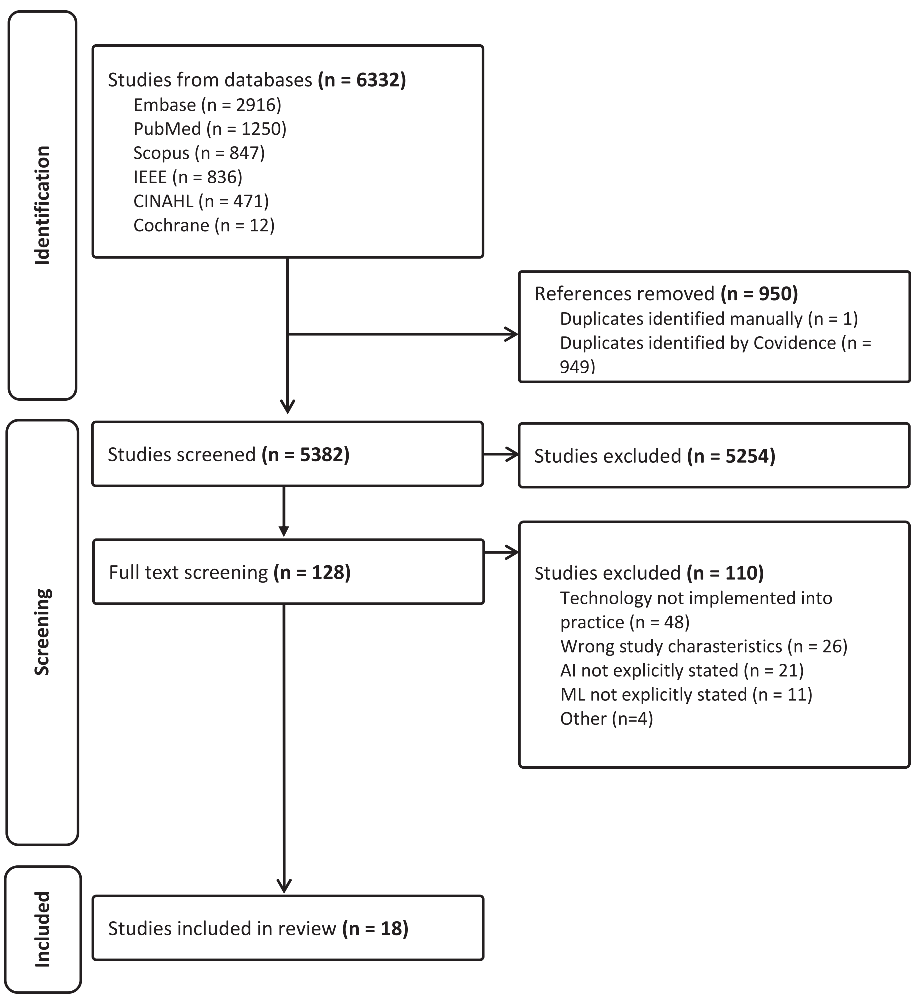

Fig. 1. PRISMA flow diagram presenting the literature review results.

2.7. Presentation of the results

The reporting of this study follows the Prisma for Scoping Reviews (PRISMA-ScR) checklist [34]. The study selection is presented using the PRISMA flow diagram for systematic reviews [35], and extracted results are presented in a descriptive format using tables and graphs.

3. Results

3.1. Description of the included studies

The literature search across six databases yielded a total of 6332 references, as presented in Fig. 1. After duplicates were removed, 5,382 studies were screened for titles and abstracts, resulting in a total of 128 articles being retrieved and screened for full-text evaluation. Finally, 18 studies were included in the review.

The studies included in the review were published between 2018 and 2024, with the majority published in 2023 and 2024 (n = 4, 22.2 %, and n = 7, 38.9 %, respectively). The studies were conducted predominantly in Asia (n = 9, 50.0 %) and Europe (n = 7, 38.9 %) and were funded by governmental or other public agencies (n = 10, 55.6 %). The full summary of the review findings is presented in Appendix 1.

Although all the included studies contained an economic evaluation, only half of the main aims focused on economic evaluations (n = 9, 50.0 %), as illustrated in Fig. 2. The studies were predominantly prospective (n = 12, 66.7 %), including cohort studies (n = 8, 44.4 %) and randomized controlled trials (n = 6, 33.3 %), conducted in outpatient care settings (n = 13, 72.2 %). Most of the study participants were healthcare users (n = 16, 88.9 %). The majority of the studies did not report using any reporting guidelines or standards (n = 11, 61.1 %).

3.2. Overview of the technologies

The technologies were most often developed for diagnostic support and detection (n = 10, 55.6 %) and predominantly intended for clinician use (n = 15, 83.3 %), as described in Fig. 3. The models utilized predominantly deep learning methods, where convolutional neural networks were mentioned in eight studies (44.4 %). Nearly half of the technologies utilized unspecified ML methods (n = 8, 44.4 %). Most (n = 13, 72.2 %) of the studies provided references for further information regarding the models used. Medical images were the primary source (n = 10, 55.6 %) for training and implementing the technologies, followed by clinical data (e.g., electronic health record data and other input data, n = 4, 22.2 %), and sensor data (n = 4, 22.2 %).

3.3. Economic evaluation of the implemented technologies

The most used economic evaluation method was cost-effectiveness analysis (n = 8, 44.4 %), as presented in Table 1. Half of the studies used descriptive approaches without formal modeling (n = 9, 50.0 %), seven (38.9 %) studies modeling based economic evaluation methods, such as Markov Modeling and decision trees, and two (11.1 %) studies using inferential statistics. The cost elements included in the evaluations were divided between medical costs (n = 12, 66.7 %), technology costs (n = 12, 66.7 %) and personnel costs (n = 7, 38.9 %), with one study not providing any cost information. The medical cost elements included direct and indirect facility and material costs related to patient hospitalization, medical treatments and procedures, excluding personnel

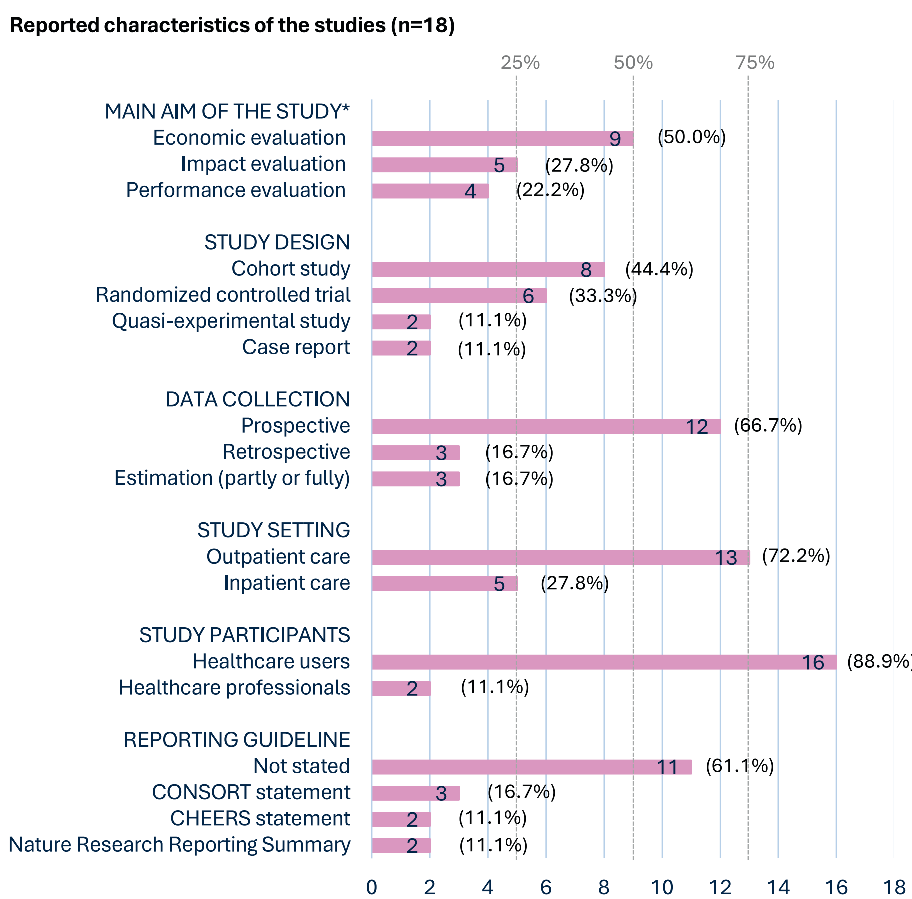

Fig. 2. Overview of the studies included. In Main aim of the stydy*, Economic evaluation refers to the main aim of the study being the cost and outcome comparison of two or more interventions, whilst in Impact evaluation the main aim is to investigate the causal effects of the technology and in Performance evaluation the outputs, efficiency or quality of the technology.

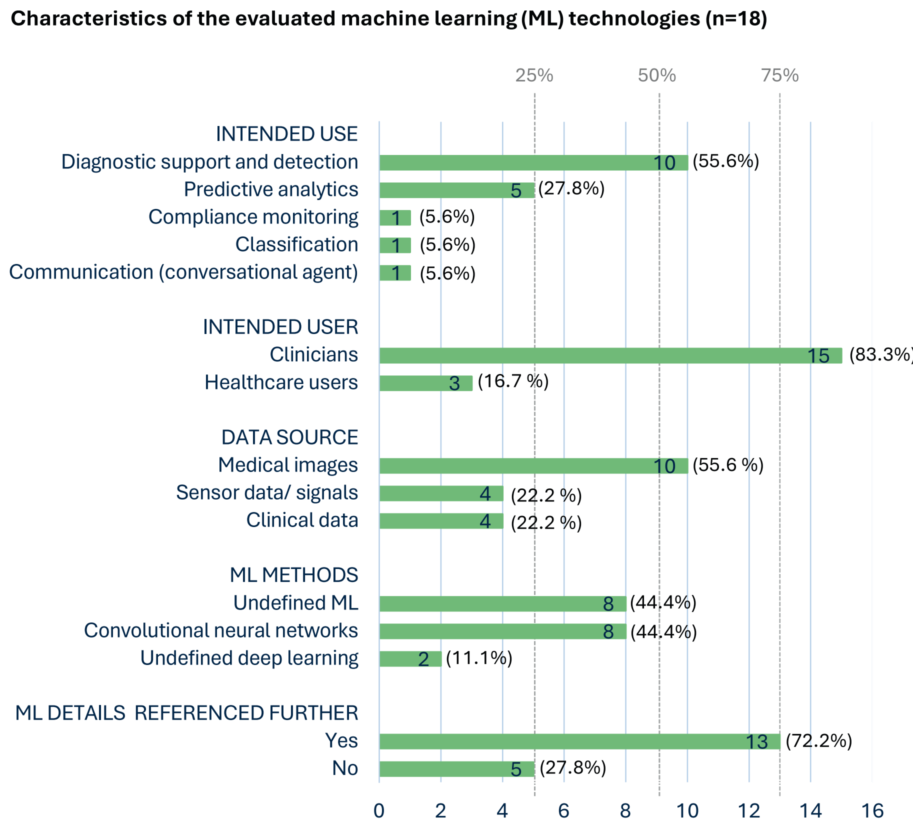

Fig. 3. Overview of the machine learning technologies (n = 18) in the included studies.

costs. Technology costs included implementation and maintenance costs of both the implemented ML-based technology, but also other technological equipment or device component costs related to either intervention or the comparator. Lastly, personnel costs included all direct and indirect costs related to healthcare workforce employment.

The outcomes considered in the economic evaluations were categorized into process outcomes (n = 8, 44.4 %), health economic ratios (n = 8, 44.4 %), clinical outcomes (n = 8, 44.4 %), costs (n = 9, 50.0 %) and quality adjusted life years (n = 5, 27.8 %). Process outcomes included measures related to treatment or procedure rates and effectiveness. Health economic ratios included summary measures representing the cost effectiveness (Incremental cost effectiveness ratio, ICER) or cost utility (Incremental cost utility ratio, ICUR) of the implemented technologies. Clinical outcomes could be operationalized as identified illnesses or hospitalization, whereas costs were predominantly increases or reductions of costs related to the implemented technology.

3.4. Reported characteristics of the studies

Comparing the results to the CHEERS-AI [28], all the admitted studies provided detailed descriptions of the methods used for data analytics (n = 18, 100.0 %) and stated conflicts of interest (n = 18, 100.0 %), as presented in Fig. 4. Most of the studies summarized the main results related to costs and outcomes (n = 17, 94.4 %) and reported all analytic inputs and study parameters (n = 16, 88.9 %), and more than half (n = 11, 61.1 %) discussed the effect of uncertainty in the results section. The studies also provided context and explained the practical relevance of the study in the background and objectives (n = 16, 88.9 %) and disclosed their sources of funding and the funders' role in the study (n = 16, 88.9 %). More than half of the studies described and justified the timeline of the study (n = 12, 66.7 %), provided sufficient details regarding the study population (n = 11, 61.1 %), described the demographics and techniques to support the valuation of outcomes (n = 10, 55.6 %) and stated the adopted perspective of the analysis (n = 10, 55.6 %) in the methods section. A sufficient summary, including relevant details of the used AI methods, was provided in 10 (55.6 %) study abstracts. On average, 39.6 % of the CHEERS-AI criteria were met on the admitted studies.

When examining the economic evaluation details, none of the studies reported an approach to (n = 18, 100.0 %) or the effect of (n = 18, 100.0 %) engaging with patients or other stakeholders in the study. Most (n = 16, 88.9 %) did not indicate the development of a health economic analysis plan for the study. The discount rate (n = 11, 61.1 %) or the currency, price date, and conversion (n = 12, 66.7 %) were not widely reported, nor was their rationale provided in the methods section. Additionally, 10 (55.6 %) studies did not identify the study as an economic evaluation in the title. On average, 64.1 % of the CHEERS-AI statement criteria for reporting the economic evaluation details were met.

The potential effect of the AI component was not clearly specified regarding their selection (n = 17, 94.4 %) or measurement (n = 16, 88.9 %). Likewise, there were shortcomings in the description of the comparatives: key details regarding the AI intervention, such as technological details, intended users, or user requirements, were absent primarily (n = 15, 83.3 %). The valuation of costs in the economic evaluation predominantly did not include relevant purchase, implementation, or maintenance costs of the used AI system (n = 16, 88.9 %), and the discussions did not address issues related to potential AI bias and its impact on economic equity (n = 16, 88.9 %). On average, 21.3 % of the CHEERS-AI statement criteria for reporting the key details related to the AI nature of the evaluated technology were met.

Looking at the AI-specific characteristics, none of the included studies accounted for the possibility of the AI intervention learning over time (n = 18, 100.0 %) or included it in the economic modeling (n = 18, 100.0 %). The potential impacts of the data used for model training (n = 18, 100.0 %) or the validated findings influencing the outcome analysis

Table 1
Overview of the economic evaluation and modeling methods per evaluated technology, with the included cost components and the considered outcomes. Cost-effectiveness analysis compares the costs of two or more interventions, and cost-utility analysis the costs against multidimensional outcomes. Cost-benefit analysis evaluates two or more alternatives, where both costs and outcomes are valued in financial terms, whereas cost-minimization analysis assumes equal outcomes and focuses on comparing the costs of the interventions. Partial economic evaluations are categorized as cost analyses. In cases where the used economic evaluation methods are inconsistently reported or include a combination of approaches, the categorization provided by the authors is used.

Evaluated technologies	Economic evaluation methods	Economic modeling methods	Included cost elements	Considered economic evaluation outcomes
AI-assisted follow-up conversational agent for post-surgery medical consultation [36]	Cost analysis	Descriptive approaches	Not stated	Process
Deep Resolve Boost (DRB) from Siemens Healthcare for radiology [37]	cost-minimization analysis	Descriptive approaches	Technology, Personnel	Costs Process,
GI Genius™ Intelligent Endoscopy Module, US-DG-2000309 © 2021 Medtronic for polyp detection [38]	Cost-effectiveness analysis	Descriptive approaches	Technology	Cost Process
Hypotension Prediction Index (Edwards Lifesciences) [39]	cost-benefit analysis	Descriptive approaches	Medical, Technology, Personnel	Clinical Costs
Hand hygiene notification system [40]	Cost-effectiveness analysis	Descriptive approaches	Technology, Personnel	Clinical, Costs, Process
Risk prediction algorithm for undiagnosed atrial fibrillation [41]	Cost-utility analysis	Markov Modeling, Decision tree	Medical, Technology, Personnel	Clinical, Health economic ratios QALY (quality adjusted life years)
ECG Real-Time Diagnostic Platform for anomaly detection [42]	Cost-effectiveness analysis	Descriptive approaches	Personnel	Cost Process
EyeWisdom (Visionary Intelligence Ltd., Beijing, China) for retinal image analysis [43]	Cost-utility analysis	Markov Modeling	Medical, Technology, Personnel	Health economic ratios, QALY
Alarm system for ECG anomaly prediction [44]	Cost-utility analysis	Markov Modelling	Medical, Technology	Health economic ratios, QALY
EndoBRAIN (Cybernet System Corp and Olympus Corp, Tokyo, Japan) for polyp classification [45]	cost-minimization analysis	Descriptive approaches	Technology	Costs
DentalXrai Pro 1.0.4 (dentalXrai Ltd, Berlin, Germany) for teeth classification [46]	Cost-effectiveness analysis	Markov Modeling	Medical, Technology	Health economic ratios, Process
SkinVision −app for skin lesion evaluation [47]	Cost-effectiveness analysis	Inferential statistics	Medical, Technology	Health economic ratios
Rapid nutritional diagnostic system for nutritional assessment [48]	Cost-effectiveness analysis	Decision tree	Medical	Clinical, Costs Health economic ratios
Computer-aided detection (CADe) of adenomas [49]	Cost-effectiveness analysis	Markov Modeling	Medical	Clinical, Health economic ratios, QALY
MiSAOS monitoring system for CPAP compliance [50]	Cost-effectiveness analysis	Inferential statistics	Medical, Technology	Clinical Costs
Assisted workflow for breast cancer metastase detection [51]	cost-minimization analysis	Descriptive approaches	Medical	Clinical, Process
Semi-automated approach to grade fundus photographs [52]	cost-minimization analysis	Descriptive approaches	Medical, Personnel	Clinical, Costs, Process
Deep Learning Assistant for Retinoblastoma (DLA-RB) [53]	Cost-utility analysis	Markov Modeling	Medical, Technology	Health economic ratios, QALY

of the AI system (n = 17, 94.4 %) were not identified or described. The uncertainties and their implications related to AI systems were not presented in the results (n = 17, 94.4 %), nor were the requirements for AI implementation discussed (n = 18, 100.0 %). The detailed table containing the item-specific references is available as Appendix 2.

4. Discussion

4.1. Main findings

This review synthesizes current evidence on the economic aspects of ML-based technologies implemented in healthcare. The technologies (n = 18) included a variety of systems used for diagnostic support, detection, or predictive analytics, predominantly employed by clinicians in outpatient settings. The utilized evaluation methods were primarily complete economic evaluations, including cost-effectiveness, cost-utility and cost-minimization analyses. Finally, the study reports were compared against the CHEERS-AI [28] statement, indicating limitations in reporting details related to both economic evaluations and the AI nature of the technologies.

4.2. Comparison with prior work

In line with past literature covering the economic evaluation of AI technologies in healthcare [23,24,25,26,27], the findings of this review identified significant gaps in reporting essential details related to economic evaluations. It must, however, be noted that half of the included articles did not recognize the study as an economic evaluation in the title, nor was economic evaluation the main aim of the studies. Nevertheless, the theoretical foundation or evidence base used to guide the evaluations in general was not entirely clear or transparent. It has been argued that, despite borrowing from the theoretical and methodological foundations of economics, economic evaluations in health are complicated by overlapping terminology, ambiguity, and interpretive confusion [54]. Furthermore, the availability of established guidelines for conducting economic evaluations on digital health technologies is limited, with the majority of health economic evaluation frameworks focusing on pharmaceuticals and medical devices [55]. As the significance of AI in healthcare grows, so does the need for consistent economic evaluations, which urges the development or complementation of existing technology development frameworks with unified and established economic evaluation and reporting guidelines.

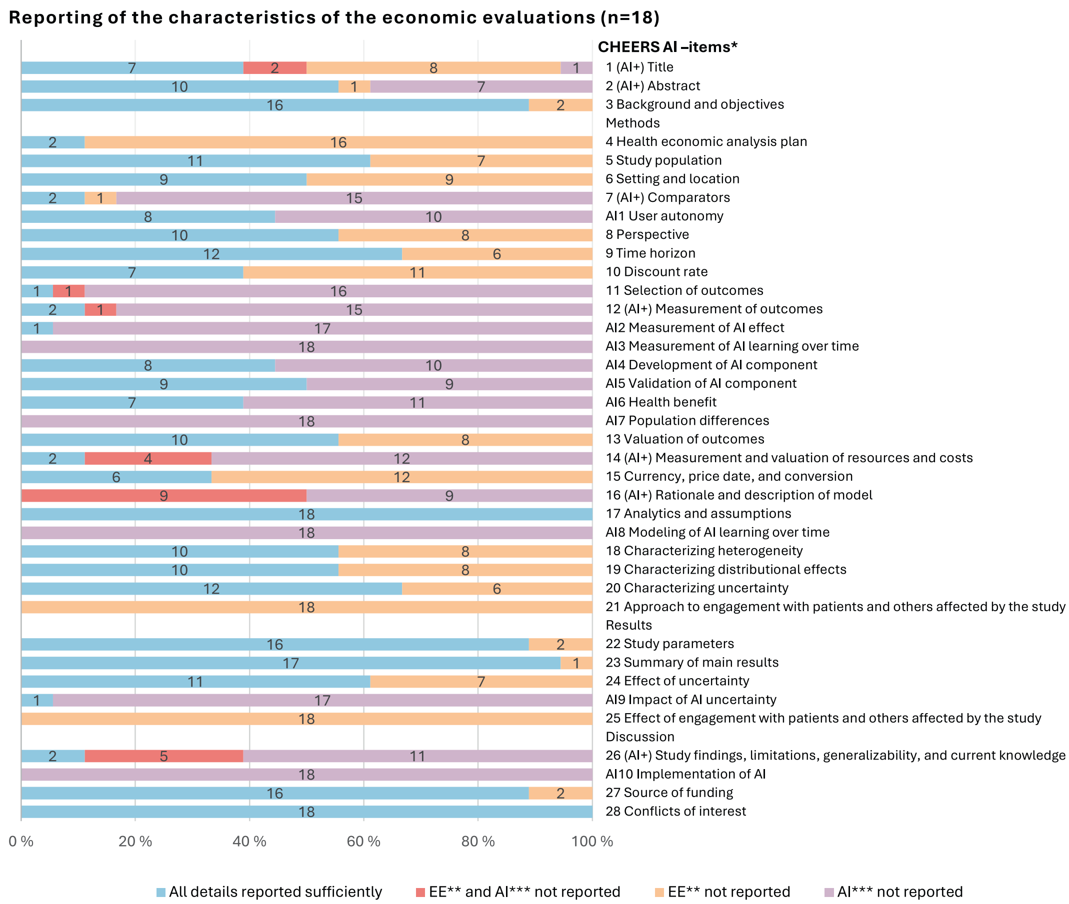

Fig. 4. Reporting of the characteristics of the economic evaluations (n = 18) of the machine learning based technologies, where **CHEERS-AI Items*** 1–28 represent the original CHEERS-2022 statement items, (AI +) the elaborations to the original items of potential nuances related to artificial intelligence (AI), and items AI1-AI10 the AI-specific items. **EE**** refers to not reporting some or any key economic evaluation details according to the original statement, whereas **AI***** refers to not reporting some or any key details related to the AI nature of the evaluated technology.

Likewise, previous reviews have reported insufficient and superficial reporting of technological details of the evaluated systems [23,24,25,26,27]. More than half (n = 10, 55.6 %) of the studies included in this review reported the methods used as simply ML or deep learning, and the comparison of the studies to the CHEERS-AI statement indicated a lack of key details related to specific characteristics or impacts of AI. Overlooking these details is problematic for several reasons. Firstly, AI, ML, or deep learning are not specific methods, but rather umbrella terms for a myriad of methods that utilize different features, encompassing a variety of model architectures and complexities. Failing to provide sufficient details poses significant limitations to assessing the reliability, generalizability, comparability or replicability of the model evaluations. This, in turn, hinders the use of these studies in strengthening the knowledge base of the documented impacts of using AI in healthcare.

Secondly, assuming the implemented technologies are static and not accounting for their possible adaptability or learning abilities may inadvertently lead to incomplete, inadequate or even misleading economic evaluations. AI-based systems are not isolated or standalone tools, but rather involve complex interdependencies between the technologies, healthcare providers, and healthcare users [24]. The performance of learning systems progressively changes in time, affecting not only the outcomes provided by them but also the behavior of the healthcare professionals using them [56]. Further, their economic and clinical value is highly dependent on the availability of good quality data, their integration into existing healthcare infrastructures and human-centered considerations regarding their usability and acceptability [57]. Complementing the economic evaluations with specific features and characteristics of the used systems is vital to assess and truly understand their real impacts [58]. AI is commonly framed as a cost and resource-saving tool, overlooking the support and maintenance it requires to operate successfully [59], including regular evaluations of the system's performance or the associated costs for integrating it into novel or existing healthcare systems. Most studies in this review did not cover the purchase or other cost components related to the technologies. The implementation costs might include expensive hardware, and maintaining the systems requires frequent updates, continuous training and data, infrastructure and knowledge related to data management, data security and privacy regulations [60,61]. Conversely, frozen systems, i.e. systems that cannot be altered or trained after their initial deployment, risk becoming outdated and losing their clinical value, or they may yield unforeseen outcomes or reveal unanticipated bias when used with populations that differ from the training data [62]. These factors can increase the need to renew systems with high costs regularly.

Novel and updated guidelines, such as the CHEERS-AI statement, have been developed to better account for the distinctive features associated with AI in economic evaluations. However, system evaluators cannot use the guidelines to enhance the reporting of specific characteristics if those characteristics are not within the scope of their current understanding. For example, it is unclear how well the costs related to system development, or their implementation, have been evaluated, calculated, or reported during the development phase, or whether the pre-development project plans include details pertaining to post-implementation evaluation, maintenance, or updates. The dialogue between different stakeholders, including developers, policymakers, and regulators, regarding the impact assessment and key economic and clinical variables should begin during the earliest planning stages of system development [63].

Issues in reporting AI-specific details in economic evaluations might partly stem from a lack of comprehensive understanding of the possible impacts of these systems [24]. This could reflect a lack of engagement between the system developers, who understand the technical details and requirements of the AI systems, and the system users and evaluators, who understand the healthcare environments and specific features related to the favorable outcomes of the systems. The advantages of developing systems on a solid foundation of existing AI evaluation frameworks are well established, ensuring more effective implementation, adoption, and scaling while mitigating possible bias caused by these systems [64]. However, the relationship between using broad terms like AI or ML to describe method-specific details and the risk of misinterpretation or missing critical nuances related to specific systems should be further investigated. Moving forward, frameworks guiding the development and evaluation of AI in healthcare could gradually shift from ambiguous presentations aiming to include the whole wide scope of AI methods to highlighting distinctive attributes of individual methods. Furthermore, it is essential to advance the common understanding and language to promote multidisciplinary system development, whilst including experts with comprehensive economic knowledge in the early phases of system development. Future research could further investigate how to support better preparation for and account for more in-depth system evaluations, starting from the early planning phases of the system development life cycle.

4.3. Strengths and limitations

The strengths of this review include the broad mapping of evidence using established databases that focus on biomedical and health sciences, complemented by databases representing technological and multidisciplinary research. Furthermore, the research team included experts in the field of health informatics, representing a variety of domain-specific knowledge related to health sciences, computer engineering, and economics.

The limitations of this review relate to the scoping review methodology, as the quality or risk of bias of the included studies were not evaluated. The search terms selected for the review, as well as limiting the search to titles, abstracts and keywords, could potentially result in not identifying all literature touching on the subject. Inclusion of diverse study types across multiple clinical domains increases the risk of limited comparability or overgeneralization of the review results. Lastly, the reporting of the studies was examined by one researcher, increasing the risk of error in the analysis.

5. Conclusions

Current research evaluating the economic aspects of ML-based technologies implemented in healthcare is relatively sparse, revealing gaps in coherence and coverage. Furthermore, the issues in reporting these evaluations risk reducing their clarity and usefulness to support decision making. Issues may arise from the rapid development of novel AI technologies in healthcare, where system developers, evaluators, users, policymakers, and regulators struggle to adapt their existing guidelines and theoretical knowledge to meet the demands of the rapidly advancing field. They also reflect the need to enhance and promote multidisciplinary collaboration, calling for a wide variety of expert knowledge representing different stakeholders, including the healthcare professionals and healthcare users affected by these technologies. AI development frameworks should be refined to encourage formulating evaluation plans in early technology development phases, as well as account for a more comprehensive understanding of the post-implementation period.

CRediT authorship contribution statement

Hanna von Gerich: Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mikael Helenius:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Iiris Hörhammer:** Writing – review & editing, Methodology, Investigation, Formal analysis, Conceptualization. **Hans Moen:** Writing – review & editing, Methodology, Investigation, Formal analysis, Conceptualization. **Laura-Maria Peltonen:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Strategic Research Council at the Academy of Finland (Project #352501 and #352503), Business Finland (3964/31/2024), Research Council of Finland (Flagship programme: Finnish Center for Artificial Intelligence FCAI, and grants 352986, 358246) and EU (H2020 grant 101016775 and NextGenerationEU).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2025.106103.

References

[1] A. Alanazi, Using machine learning for healthcare challenges and opportunities, Inform Med Unlocked. 23 (30) (2022 Mar) 100924, https://doi.org/10.1016/j.imu.2022.100924.

[2] L. Rubinger, A. Gazendam, S. Ekhtiari, M. Bhandari, Machine learning and artificial intelligence in research and healthcare, Injury 54 (Suppl 3) (2023 May) S69–S73, https://doi.org/10.1016/j.injury.2022.01.046.

[3] S. Shamshirband, M. Fathi, A. Dehzangi, A.T. Chronopoulos, H. Alinejad-Rokny, A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues, J. Biomed. Inform. 113 (2021 Jan) 103627, https://doi.org/10.1016/j.jbi.2020.103627.

[4] A. Al-Kababji, F. Bensaali, S.P. Dakua, Y. Himeur, Automated liver tissues delineation techniques: a systematic survey on machine learning current trends and future orientations, Eng. Appl. Artif. Intell. 21 (117) (2023 Nov) 105532, https://doi.org/10.1016/j.engappai.2022.105532.

[5] A. Barragán-Montero, U. Javaid, G. Valdés, D. Nguyen, P. Desbordes, B. Macq, S. Willems, L. Vandewinckele, M. Holmström, F. Löfman, S. Michiels, K. Souris, E. Sterpin, J.A. Lee, Artificial intelligence and machine learning for medical imaging: a technology review, Phys. Med. 83 (2021 Mar) 242–256, https://doi.org/10.1016/j.ejmp.2021.04.016.

[6] D.M.S. Barros, J.C.C. Moura, C.R. Freire, A.C. Taleb, R.A.M. Valentim, P.S. G. Morais, Machine learning applied to retinal image processing for glaucoma detection: review and perspective, Biomed. Eng. Online 19 (1) (2020 Apr 15) 20, https://doi.org/10.1186/s12938-020-00767-2.

[7] M.P. Hosseini, A. Hosseini, K. Ahi, A Review on Machine Learning for EEG Signal Processing in Bioengineering, IEEE Rev. Biomed. Eng. 14 (2021) 204–218, https://doi.org/10.1109/RBME.2020.2969915.

[8] D. Velazquez-Diaz, J.E. Arco, A. Ortiz, V. Pérez-Cabezas, D. Lucena-Anton, J. A. Moral-Munoz, A. Galán-Mercant, Use of Artificial Intelligence in the Identification and Diagnosis of Frailty Syndrome in older adults: Scoping Review, J. Med. Internet Res. 20 (25) (2023 Oct) e47346, https://doi.org/10.2196/47346. PMID: 37862082; PMCID: PMC10625070.

[9] S. Patil, S. Albogami, J. Hosmani, S. Mujoo, M.A. Kamil, M.A. Mansour, H. N. Abdul, S. Bhandi, S.S.S.J. Ahmed, Artificial Intelligence in the Diagnosis of Oral Diseases: applications and pitfalls, Diagnostics (basel). 12 (5) (2022 Apr 19) 1029, https://doi.org/10.3390/diagnostics12051029. PMID: 35626185; PMCID: PMC9139975.

[10] M. Jiang, Y. Ma, S. Guo, L. Jin, L. Lv, L. Han, N. An, Using Machine Learning Technologies in Pressure Injury Management: Systematic Review, JMIR Med. Inform. 9 (3) (2021 Mar 10) e25704, https://doi.org/10.2196/25704.

[11] Q. Xie, X. Wang, J. Pei, Y. Wu, Q. Guo, Y. Su, H. Yan, R. Nan, H. Chen, X. Dou, Machine Learning-based Prediction Models for Delirium: a Systematic Review and Meta-Analysis, J. Am. Med. Dir. Assoc. 23 (10) (2022 Oct) 1655–1668.e6, https://doi.org/10.1016/j.jamda.2022.06.020.

[12] N.H. Alhumaidi, D. Dermawan, H.F. Kamaruzaman, N. Alotaiq, The use of Machine Learning for Analyzing Real-World Data in Disease Prediction and Management: Systematic Review, JMIR Med. Inform. 19 (13) (2025 Jun) e68898, https://doi.org/10.2196/68898.

[13] S.S.W. Yip, S. Ning, N.Y.K. Wong, et al., Leveraging machine learning in nursing: innovations, challenges, and ethical insights, Front Digit Health. 7 (2025) 1514133, https://doi.org/10.3389/fdgth.2025.1514133.

[14] S.D. Khan, Z. Hoodbhoy, M.H.R. Raja, J.Y. Kim, H.D.J. Hogg, A.A.A. Manji, F. Gulamali, A. Hasan, A. Shaikh, S. Tajuddin, N.S. Khan, M.R. Patel, S. Balu, Z. Samad, M.P. Sendak, Frameworks for procurement, integration, monitoring, and evaluation of artificial intelligence tools in clinical settings: a systematic review, PLOS Digit Health. 3 (5) (2024 May 29) e0000514, https://doi.org/10.1371/journal.pdig.0000514.

[15] S. Joshi, I. Urteaga, W.A.C. van Amsterdam, G. Hripcsak, P. Elias, B. Recht, N. Elhadad, J. Fackler, M.P. Sendak, J. Wiens, K. Deshpande, Y. Wald, M. Fiterau, Z. Lipton, D. Malinsky, M. Nayan, H. Namkoong, S. Park, J.E. Vogt, R. Ranganath, AI as an intervention: improving clinical outcomes relies on a causal approach to AI development and validation, J. Am. Med. Inform. Assoc. 32 (3) (2025 Mar 1) 589–594, https://doi.org/10.1093/jamia/ocae301.

[16] B. Khan, H. Fatima, A. Qureshi, S. Kumar, A. Hanan, J. Hussain, S. Abdullah, Drawbacks of Artificial Intelligence and their potential Solutions in the Healthcare Sector, Biomed. Mater. Devices 8 (2023 Feb) 1–8, https://doi.org/10.1007/s44174-023-00063-2.

[17] Y. Wang, N. Li, L. Chen, M. Wu, S. Meng, Z. Dai, Y. Zhang, M. Clarke, Guidelines, Consensus statements, and Standards for the use of Artificial Intelligence in Medicine: Systematic Review, J. Med. Internet Res. 22 (25) (2023 Nov) e46089, https://doi.org/10.2196/46089.

[18] B.V. Seixas, F. Dionne, C. Mitton, Practices of decision making in priority setting and resource allocation: a scoping review and narrative synthesis of existing frameworks, Health Econ Rev. 11 (1) (2021 Jan 7) 2, https://doi.org/10.1186/s13561-020-00300-0.

[19] K.M. Rabarison, C.L. Bish, M.S. Massoudi, W.H. Giles, Economic Evaluation Enhances Public Health Decision making, Front. Public Health 24 (3) (2015 Jun) 164, https://doi.org/10.3389/fpubh.2015.00164.

[20] K. Zisis, P. Naoum, K. Athanasakis, Qualitative comparative analysis of health economic evaluation guidelines for health technology assessment in European countries, Int. J. Technol. Assess. Health Care 10 (37) (2020 Dec) e2.

[21] A. Bhardwaj, Promise and Provisos of Artificial Intelligence and Machine Learning in Healthcare, J Healthc Leadersh. 20 (14) (2022 Jul) 113–118, https://doi.org/10.2147/JHL.S369498.

[22] N.N. Khanna, M.A. Maindarkar, V. Viswanathan, J.F.E. Fernandes, S. Paul, M. Bhagawati, P. Ahluwalia, Z. Ruzsa, A. Sharma, R. Kolluri, I.M. Singh, J.R. Laird, M. Fatemi, A. Alizad, L. Saba, V. Agarwal, A. Sharma, J.S. Teji, M. Al-Maini, V. Rathore, S. Naidu, K. Liblik, A.M. Johri, M. Turk, L. Mohanty, D.W. Sobel, M. Miner, K. Viskovic, G. Tsoulfas, A.D. Protogerou, G.D. Kitas, M.M. Fouda, S. Chaturvedi, M.K. Kalra, J.S. Suri, Economics of Artificial Intelligence in Healthcare: Diagnosis vs, Treatment. Healthcare (basel) 10 (12) (2022 Dec 9) 2493, https://doi.org/10.3390/healthcare10122493.

[23] W. Jiao, X. Zhang, F. D'Souza, The Economic Value and Clinical Impact of Artificial Intelligence in Healthcare: a Scoping Literature Review, IEEE Access 11 (2023) 123445–123457, https://doi.org/10.1109/ACCESS.2023.3327905.

[24] N. Kastrup, A.W. Holst-Kristensen, J.B. Valentin, Landscape and challenges in economic evaluations of artificial intelligence in healthcare: a systematic review of methodology, BMC Digit Health. 2 (2024) 39, https://doi.org/10.1186/s44247-024-00088-7.

[25] J. Vithlani, C. Hawksworth, J. Elvidge, L. Ayiku, D. Dawoud, Economic evaluations of artificial intelligence-based healthcare interventions: a systematic literature review of best practices in their conduct and reporting, Front. Pharmacol. 8 (14) (2023 Aug) 1220950, https://doi.org/10.3389/fphar.2023.1220950.

[26] M.M. Voets, J. Veltman, C.H. Slump, S. Siesling, H. Koffijberg, Systematic Review of Health Economic evaluations Focused on Artificial Intelligence in Healthcare: the Tortoise and the Cheetah, Value Health 25 (3) (2022 Mar) 340–349, https://doi.org/10.1016/j.jval.2021.11.1362.

[27] J. Wolff, J. Pauling, A. Keck, J. Baumbach, The Economic Impact of Artificial Intelligence in Health Care: Systematic Review, J. Med. Internet Res. 22 (2) (2020 Feb 20) e16866, https://doi.org/10.2196/16866.

[28] J. Elvidge, C. Hawksworth, T.S. Avşar, A. Zemplenyi, A. Chalkidou, S. Petrou, Z. Petykó, D. Srivastava, G. Chandra, J. Delaye, A. Denniston, M. Gomes, S. Knies, P. Nousios, P. Siirtola, J. Wang, D. Dawoud, CHEERS-AI Steering Group. Consolidated Health Economic Evaluation Reporting Standards for Interventions that Use Artificial Intelligence (CHEERS-AI), Value Health 27 (9) (2024 Sep) 1196–1205, https://doi.org/10.1016/j.jval.2024.05.006.

[29] Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, Augustovski F, Briggs AH, Mauskopf J, Loder E; ISPOR Health Economic Evaluation Publication Guidelines-CHEERS Good Reporting Practices Task Force. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)–explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value Health. 2013 Mar-Apr;16(2):231-50. doi: 10.1016/j.jval.2013.02.002.

[30] M.D.J. Peters, C. Marnie, A.C. Tricco, D. Pollock, Z. Munn, L. Alexander, P. McInerney, C.M. Godfrey, H. Khalil, Updated methodological guidance for the conduct of scoping reviews, JBI Evid Synth. 18 (10) (2020 Oct) 2119–2126, https://doi.org/10.11124/JBIES-20-00167.

[31] JBI. Appendix 10.1 JBI template source of evidence details, characteristics and results extraction instrument [Internet]. 2024 Mar 26 [cited 2025 Jul 24]. Available from: https://jbi-global-wiki.refined.site/space/MANUAL/355863340/Appendix+10.1+JBI+template+source+of+evidence+details%252C+characteristics+and+results+extraction+instrument.

[32] H.C. Turner, R.A. Archer, L.E. Downey, W. Isaranuwatchai, K. Chalkidou, M. Jit, Y. Teerawattananon, An Introduction to the Main Types of Economic evaluations used for Informing Priority setting and Resource Allocation in Healthcare: Key Features, Uses, and Limitations, Front. Public Health 25 (9) (2021 Aug) 722927, https://doi.org/10.3389/fpubh.2021.722927.

[33] A. Pandit, Health economic evaluation-methods and techniques, IOSR J Econ Financ. 7 (1) (2016) 1–9, https://doi.org/10.9790/5933-07120109.

[34] A.C. Tricco, E. Lillie, W. Zarin, K.K. O'Brien, H. Colquhoun, D. Levac, D. Moher, M. D.J. Peters, T. Horsley, L. Weeks, S. Hempel, E.A. Akl, C. Chang, J. McGowan, L. Stewart, L. Hartling, A. Aldcroft, M.G. Wilson, C. Garritty, S. Lewin, C. M. Godfrey, M.T. Macdonald, E.V. Langlois, K. Soares-Weiser, J. Moriarty, T. Clifford, Ö. Tunçalp, S.E. Straus, PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation, Ann. Intern. Med. 169 (7) (2018 Oct 2) 467–473, https://doi.org/10.7326/M18-0850.

[35] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, L.A. McGuinness, L.A. Stewart, J. Thomas, A.C. Tricco, V.A. Welch, P. Whiting, D. Moher, The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, BMJ 29 (372) (2021 Mar) n71, https://doi.org/10.1136/bmj.n71.

[36] Y. Bian, Y. Xiang, B. Tong, B. Feng, X. Weng, Artificial Intelligence-Assisted System in Postoperative follow-up of Orthopedic patients: Exploratory Quantitative and Qualitative Study, J. Med. Internet Res. 22 (5) (2020 May 26) e16896, https://doi.org/10.2196/16896.

[37] M.A.K. Brix, J. Järvinen, M.K. Bode, M. Nevalainen, M. Nikki, J. Niinimäki, E. Lammentausta, Financial impact of incorporating deep learning reconstruction into magnetic resonance imaging routine, Eur. J. Radiol. 175 (2024 Jun) 111434, https://doi.org/10.1016/j.ejrad.2024.111434.

[38] S.E. Chin, F.T. Wan, J. Ladlad, K.M. Chue, S.K.H. Endoscopy Centre, E.K. Teo, C. L. Lin, F.J. Foo, F.H. Koh, One-year review of real-time artificial intelligence (AI)-aided endoscopy performance, Surg. Endosc. 37 (8) (2023 Aug) 6402–6407, https://doi.org/10.1007/s00464-023-09979-8.

[39] L. Frassanito, R. Di Bidino, F. Vassalli, K. Michnacs, P.P. Giuri, B.A. Zanfini, S. Catarci, N. Filetici, C. Sonnino, A. Cicchetti, G. Arcuri, G. Draisci, Personalized Predictive Hemodynamic Management for Gynecologic Oncologic Surgery: Feasibility of Cost-Benefit Derivatives of Digital Medical Devices, J Pers Med. 14 (1) (2023 Dec 30) 58, https://doi.org/10.3390/jpm14010058.

[40] R. Geilleit, Z.Q. Hen, C.Y. Chong, A.P. Loh, N.L. Pang, G.M. Peterson, K.C. Ng, A. Huis, D.F. de Korne, Feasibility of a real-time hand hygiene notification machine learning system in outpatient clinics, J. Hosp. Infect. 100 (2) (2018 Oct) 183–189, https://doi.org/10.1016/j.jhin.2018.04.004.

[41] N.R. Hill, L. Groves, C. Dickerson, R. Boyce, S. Lawton, M. Hurst, K.G. Pollock, D. M. Sugrue, S. Lister, C. Arden, D.W. Davies, A.C. Martin, B. Sandler, J. Gordon, U. Farooqui, D. Clifton, C. Mallen, J. Rogers, A.J. Camm, A.T. Cohen, Identification of undiagnosed atrial fibrillation using a machine learning risk prediction algorithm and diagnostic testing (PULsE-AI) in primary care: cost-effectiveness of a screening strategy evaluated in a randomized controlled trial in England, J. Med. Econ. (2022 Jan-Dec;25(1):974–983.), https://doi.org/10.1080/13696998.2022.2102355.

[42] F. Li, P. Wang, L.X. Wang, Deep learning-based regional ECG diagnosis platform, Pacing Clin. Electrophysiol. 47 (1) (2024 Jan) 139–148, https://doi.org/10.1111/pace.14891.

[43] H. Li, G. Li, N. Li, C. Liu, Z. Yuan, Q. Gao, S. Hao, S. Fan, J. Yang, Cost-effectiveness analysis of artificial intelligence-based diabetic retinopathy screening in rural China based on the Markov model, PLoS One 18 (11) (2023 Nov 16) e0291390, https://doi.org/10.1371/journal.pone.0291390.

[44] W.T. Liu, P.H. Hsieh, C.S. Lin, W.H. Fang, C.H. Wang, C.S. Tsai, Y.J. Hung, C. B. Hsieh, C. Lin, D.J. Tsai, Opportunistic Screening for Asymptomatic Left Ventricular Dysfunction with the use of Electrocardiographic Artificial Intelligence: a Cost-Effectiveness Approach, Can. J. Cardiol. 40 (7) (2024 Jul) 1310–1321, https://doi.org/10.1016/j.cjca.2023.11.044.

[45] Y. Mori, S.E. Kudo, J.E. East, A. Rastogi, M. Bretthauer, M. Misawa, M. Sekiguchi, T. Matsuda, Y. Saito, H. Ikematsu, K. Hotta, K. Ohtsuka, T. Kudo, K. Mori, Cost savings in colonoscopy with artificial intelligence-aided polyp diagnosis: an add-on analysis of a clinical trial (with video), Gastrointest. Endosc. 92 (4) (2020 Oct) 905–911.e1, https://doi.org/10.1016/j.gie.2020.03.3759.

[46] F. Schwendicke, S. Mertens, A.G. Cantu, A. Chaurasia, H. Meyer-Lueckel, J. Krois, Cost-effectiveness of AI for caries detection: randomized trial, J. Dent. 119 (2022 Apr) 104080, https://doi.org/10.1016/j.jdent.2022.104080.

[47] A.M. Smak Gregoor, T.E. Sangers, L.J. Bakker, L. Hollestein, C.A. Uyl-de Groot, T. Nijsten, M. Wakkee, An artificial intelligence based app for skin cancer detection evaluated in a population based setting, NPJ Digit Med. 6 (1) (2023 May 20) 90, https://doi.org/10.1038/s41746-023-00831-w.

[48] M.Y. Sun, Y. Wang, T. Zheng, X. Wang, F. Lin, L.Y. Zheng, M.Y. Wang, P.H. Zhang, L.Y. Chen, Y. Yao, J. Sun, Z.N. Li, H.Y. Hu, H. Jiang, H.Y. Yue, Q. Zhao, H.Y. Wang, L. Han, X. Ma, M.T. Ji, H.X. Xu, S.Y. Luo, Y.H. Liu, Y. Zhang, T. Han, Y.S. Li, P.

P. Hou, W. Chen, Health economic evaluation of an artificial intelligence (AI)-based rapid nutritional diagnostic system for hospitalised patients: a multicentre, randomised controlled trial, Clin. Nutr. 43 (10) (2024 Oct) 2327–2335, https://doi.org/10.1016/j.clnu.2024.08.030.

[49] N.R. Thiruvengadam, P. Solaimani, M. Shrestha, S. Buller, R. Carson, B. Reyes-Garcia, R.D. Gnass, B. Wang, N. Albasha, P. Leonor, M. Saumoy, R. Coimbra, A. Tabuenca, W. Srikureja, S. Serrao, The Efficacy of Real-time Computer-aided Detection of Colonic Neoplasia in Community Practice: a Pragmatic Randomized Controlled Trial, Clin. Gastroenterol. Hepatol. 22 (11) (2024 Nov) 2221–2230.e15, https://doi.org/10.1016/j.cgh.2024.02.021.

[50] C. Turino, I.D. Benítez, X. Rafael-Palou, A. Mayoral, A. Lopera, L. Pascual, R. Vaca, A. Cortijo, A. Moncusí-Moix, M. Dalmases, E. Vargiu, J. Blanco, F. Barbé, J. de Batlle, Management and Treatment of patients with Obstructive sleep Apnea using an Intelligent monitoring System based on Machine Learning Aiming to Improve Continuous positive Airway pressure Treatment Compliance: Randomized Controlled Trial, J. Med. Internet Res. 23 (10) (2021 Oct 18) e24072, https://doi.org/10.2196/24072.

[51] C. van Dooijeweert, R.N. Flach, N.D. Ter Hoeve, C.P.H. Vreuls, R. Goldschmeding, J.E. Freund, P. Pham, T.Q. Nguyen, E. van der Wall, G.W.J. Frederix, N. Stathonikos, P.J. van Diest, Clinical implementation of artificial-intelligence-assisted detection of breast cancer metastases in sentinel lymph nodes: the CONFIDENT-B single-center, non-randomized clinical trial, Nat Cancer. 5 (8) (2024 Aug) 1195–1205, https://doi.org/10.1038/s43018-024-00788-z.

[52] Y. Wang, D. Shi, Z. Tan, Y. Niu, Y. Jiang, R. Xiong, G. Peng, M. He, Screening Referable Diabetic Retinopathy using a Semi-automated Deep Learning Algorithm Assisted Approach, Front Med (lausanne). 25 (8) (2021 Nov) 740987, https://doi.org/10.3389/fmed.2021.740987.

[53] R. Zhang, L. Dong, R. Li, K. Zhang, Y. Li, H. Zhao, J. Shi, X. Ge, X. Xu, L. Jiang, X. Shi, C. Zhang, W. Zhou, L. Xu, H. Wu, H. Li, C. Yu, J. Li, J. Ma, W. Wei, Automatic retinoblastoma screening and surveillance using deep learning, Br. J. Cancer 129 (3) (2023 Aug) 466–474, https://doi.org/10.1038/s41416-023-02320-z.

[54] Lauer JA, Morton A, Culyer AJ, Chalkidou K. What counts in economic evaluations in health? Benefit-cost analysis compared to other forms of economic evaluations. Geneva: World Health Organization; 2020. (Health systems financing and governance, working paper no. 18). Licence: CC BY-NC-SA 3.0 IGO.

[55] M. Gomes, E. Murray, J. Raftery, Economic Evaluation of Digital Health Interventions: Methodological Issues and Recommendations for Practice, Pharmacoeconomics 40 (4) (2022 Apr) 367–378, https://doi.org/10.1007/s40273-022-01130-0.

[56] G.P. Jackson, E.H. Shortliffe, Understanding the evidence for artificial intelligence in healthcare, BMJ Qual. Saf. 34 (7) (2025 Jun 19) 421–424, https://doi.org/10.1136/bmjqs-2025-018559.

[57] S.S. Dhanda, D. Panwar, C.C. Lin, T.K. Sharma, D. Rastogi, S. Bindewari, A. Singh, Y.H. Li, N. Agarwal, S. Agarwal, Advancement in public health through machine learning: a narrative review of opportunities and ethical considerations, J Big Data. 04 (12) (2025 Jul) 154, https://doi.org/10.1186/s40537-025-01201-x.

[58] J. Elvidge, D. Dawoud, Reporting standards to support cost-effectiveness evaluations of AI-driven health care, Lancet Digit Health. 6 (9) (2024 Sep) e602–e603, https://doi.org/10.1016/S2589-7500(24)00171-7.

[59] L. Farah, J. Davaze-Schneider, T. Martin, P. Nguyen, I. Borget, N. Martelli, Are current clinical studies on artificial intelligence-based medical devices comprehensive enough to support a full health technology assessment? A Systematic Review. Artif Intell Med. 140 (2023 Jun) 102547 https://doi.org/10.1016/j.artmed.2023.102547.

[60] Y.Y.M. Aung, D.C.S. Wong, D.S.W. Ting, The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare, Br. Med. Bull. 139 (1) (2021 Sep 10) 4–15, https://doi.org/10.1093/bmb/ldab016.

[61] S. Mizna, S. Arora, P. Saluja, G. Das, W.A. Alanesi, An analytic research and review of the literature on practice of artificial intelligence in healthcare, Eur. J. Med. Res. 30 (1) (2025 May 14) 382, https://doi.org/10.1186/s40001-025-02603-6.

[62] J.C. Bélisle-Pipon, V. Couture, M.C. Roy, I. Ganache, M. Goetghebeur, I.G. Cohen, What Makes Artificial Intelligence Exceptional in Health Technology Assessment? Front. Artif. Intell. 2 (4) (2021 Nov) 736697 https://doi.org/10.3389/frai.2021.736697.

[63] R. Tarricone, G. Callea, M. Ogorevc, R.V. Prevolnik, Improving the Methods for the Economic Evaluation of Medical Devices, Health Econ. 26 (Suppl 1) (2017 Feb) 70–92, https://doi.org/10.1002/hec.3471.

[64] K. Cresswell, N. de Keizer, F. Magrabi, R. Williams, M. Rigby, M. Prgomet, P. Kukhareva, Z.S. Wong, P. Scott, C.K. Craven, A. Georgiou, S. Medlock, J. Brender McNair, E. Ammenwerth, Evaluating Artificial Intelligence in Clinical Settings-Let Us not Reinvent the Wheel, J. Med. Internet Res. 7 (26) (2024 Aug) e46407, https://doi.org/10.2196/46407.

Glossary

AI: Artificial intelligence
CHEERS: Consolidated Health Economic Evaluation Reporting Standards
CHEERS-AI: The Consolidated Health Economic Evaluation Reporting Standards for Interventions that use AI
ML: Machine learning