Implementation of Machine Learning Applications in Healthcare Organizations: A Systematic Review of Empirical Studies

Luigi Maria Preti[1]; Vittoria Ardito[1]; Amelia Compagni[1, 2]; Francesco Petracca[1]; Giulia Cappellaro[1,2]

1Center for Research on Health and Social Care Management (CERGAS) SDA Bocconi School of Management Milan IT
2Department of Social and Political Science Bocconi University Milan IT

Corresponding Author:
Giulia Cappellaro
Department of Social and Political Science
Bocconi University
Via Sarfatti 25 Milan IT

Abstract

Background:

A growing enthusiasm for machine learning (ML) has been noted among academics and healthcare practitioners. Despite the transformative potential of ML-based applications for patient care, their uptake and implementation in healthcare organizations is still sporadic. Numerous challenges currently impede or delay widespread implementation of ML in clinical practice, and limited knowledge is available regarding how these challenges have been addressed so far.

Objectives:

The aim of this work is twofold: i) to examine the characteristics of the ML-based applications and the implementation process in clinical practice, using the Consolidated Framework for Implementation Research (CFIR) as theoretical guidance; ii) to synthesize the strategies adopted by healthcare organizations to foster successful implementation of ML.

Methods:

A systematic literature review was conducted based on the PRISMA guidelines. The search was conducted using three databases (PubMed, Scopus, and Web of Science) over a 10-year time frame (2013-2023). The search strategy was built around four blocks of keywords (artificial intelligence, implementation, health care, and study type). Only empirical studies documenting the implementation of ML applications in clinical settings were considered. The implementation process was investigated using a thematic analysis and coding procedure. The study protocol was registered in PROSPERO with registration number 403873.

Results:

Thirty-four studies were selected for data synthesis. Selected papers were relatively recent, with only 9% of the records published before 2019. ML-based applications were implemented mostly within hospitals (29/34, 85%). In terms of clinical workflow, ML-based applications supported mostly prognosis and diagnosis, as observed in 20 (59%) and 10 (29%) studies, respectively. The implementation efforts were analyzed using the CFIR domains. As for the inner setting, access to knowledge and information (12/34, 35%), IT infrastructure (11/34, 32%), and organizational culture (9/34, 26%) were amongst the most observed dimensions influencing the success of implementation. As for the ML innovation itself, factors deemed relevant were its design (15/34, 41%), relative advantage with respect to existing clinical practice (14/34, 41%), and perceived complexity (14/34, 41%). As for the other domains – i.e., processes, roles, and outer setting, stakeholder engagement (12/34, 35%), reflecting and evaluating practices (11/34, 32%) and the presence of implementation leaders (9/34, 26%) were the main factors identified as salient.

Conclusions:

This study contributes to shed some light on the factors that are relevant and that should be accounted for in an implementation process of ML-based applications in healthcare. While the relevance of ML-specific dimensions, like trust, emerges clearly across several implementation domains, the evidence from this study highlighted that relevant implementation factors are not necessarily specific for ML, but rather transversal for digital health technologies. More research is needed to further clarify the factors that are relevant to implementing ML-based applications at the organizational level, and to support their uptake within healthcare organizations.

Keywords: Artificial Intelligence; Machine Learning; Implementation; Healthcare Organization; Barriers; Facilitators

Introduction

Background

Artificial intelligence (AI) has been unquestionably acknowledged as a game changer in health care [1], even more so after technological advances in the field of machine learning (ML) have contributed to further expand the frontiers of its possible applications [2]. Compared to knowledge- or rule-based systems that automate established human clinical reasoning methods through a series of "if-then" statements [3], ML encompasses all the non-knowledge-based models that automatically (or semi-automatically) learn from the exposure to abundant quantities of data and detect patterns through explicit or latent recognition rather than conventional programming. ML is expected to serve primarily as a decision support tool to enhance rather than a replacement of human work [4], thereby providing healthcare professionals (HCPs) with improved predictions and rendering their decision-making process more accurate [5]. Despite some AI systems having already shown to be equal or even superior in performance to HCPs [6], full automation of a broad range of human tasks is expected to occur only at later stages.

Whether ML is intended to provide inputs to human decision making or to act autonomously, these technological advancements do not automatically translate into clinical practice. The road to implementing ML applications in patient care is indeed ridden with several challenges, creating an inevitable chasm between ML and its clinical integration [7], [8].

Challenges for the implementation of AI systems, without an exclusive focus on ML, have been previously outlined, with a breadth of interdependent factors at different stakeholder group levels [9], [10]. For HCPs, core considerations pertain to the need for ML outputs to be meaningful inputs in their decision making and be explainable. ML algorithms are often associated with the so-called "black box" effect [11], [12]. The lack of transparency in data and outputs can be a significant concern for HCPs, as it hampers model interpretability (i.e., the possibility to understand or interpret how a given output has been produced), and explainability (i.e., the capacity of a model to be explained, even if not totally interpretable) [13]. ML applications and outputs are therefore likely to clash with the principles of evidence-based medicine, which instead lies upon the highest possible standards of interpretability and explainability. Concerns about the potential implications for accountability and personal responsibility regarding mistakes or computational misdiagnosis by ML applications present additional implementation challenges.

At the patient level, fair implementation of ML applications necessitates continuous supply of standardized data to train, validate and incessantly improve performance and prevent algorithmic bias [9]. Notions of patient confidentiality and privacy should be reimagined entirely as data must be shared across multiple institutions to maximize their value and allow for improved algorithms [14].

Lastly, distinctive implementation challenges have been identified at the level of healthcare organizations, dealing with financial challenges and funding mechanisms, as well as issues related to the computational resources that are necessary to support the implementation of ML.

Several implementation frameworks for healthcare technologies are on hand, but no widely recognized model addresses all the specific issues that are relevant to ML applications [15], [16], [17]. To date, research on ML implementation has been predominantly conceptual in nature, with an underreporting of empirical investigations into the specifics and consequences of implementation processes in real-life settings [18], [19]. Available studies have primarily focused on the quantitative impact of ML algorithms on health outcomes or accuracy, without examining the corresponding implementation processes [20]. Recently, Chomutare et al. conducted a scoping review to identify barriers and facilitators to the implementation of ML from empirical studies [21], while Tricco et al. focused on the strategies adopted to implement ML tools in hospital settings [22]. However, additional inquiry is needed to determine whether the literature on the implementation of ML applications in healthcare adequately acknowledges the unique challenges encountered along the implementation process, as well as the strategies adopted to overcome them.

Research Objectives

This systematic literature review primarily aims to identify studies on the real-life implementation of ML applications in clinical practice and to synthesize insights about the features of these innovations and the processes deployed to facilitate their effective implementation. We set out to address the following research questions:

1. What are the characteristics of ML applications implemented in clinical practice as reported in the scientific literature?
2. What processes and strategies do healthcare organizations employ to foster the successful implementation of ML applications in clinical practice? Which factors are recognized as more relevant for the (un)successful implementation of ML applications?

Methods

Overview

This systematic review adopted the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) 2020 guidelines [23]. The review was previously registered within the International Prospective Register of Systematic Reviews (PROSPERO) with registration number 403873. All the methodological details are provided in the published research protocol [24]. The most relevant aspects are summarized hereafter, with any deviations from the protocol duly noted.

Positionality of the research team

Positionality refers to how individuals identify with, and relate to, different social dimensions such as gender, race, and ethnicity [25], and as such is a relevant aspect to consider in qualitative research. To that end, the research team comprises five Italian white researchers (LMP, VA, AC, FP, GC). Broadly speaking, the team as a whole, composed of 60% females and 40% males, shares a common background in management studies, with focus on healthcare management. LMP is a PhD student who is working in the area of AI and ML under the perspective of the organizational implementation of AI tools in healthcare organizations. VA is a PhD student who has conducted prior research at the intersection between digital health and implementation science. AC has multi-annual experience in organizational studies and qualitative research focusing on issues related to innovations in healthcare and professional dynamics in healthcare organizations. FP is a young adult expert in digital health technologies, focusing on their regulation and value assessment. GC has multi-annual experience in organizational studies and qualitative research focusing on institutional dynamics, novel technologies, and professions.

Eligibility criteria

This review focused on empirical studies investigating aspects related to the implementation of ML applications within healthcare organizations. We adopted the definition of implementation as an "active and planned effort to mainstream innovation within an organization" [26], while healthcare organizations encompass all entities delivering health services, including hospitals, outpatient centers, primary care facilities, and public health institutions. Studies were selected based on the eligibility criteria defined in the research protocol [24] and summarized in Table 1. The recently updated version of the Consolidated Framework for Implementation Research (CFIR), a commonly used model to assess factors influencing implementation and to explain barriers and facilitators to implementation effectiveness [27], [28], was used as one criterion for inclusion. Specifically, only studies that explicitly reported factors related to the CFIR domains of inner setting or processes were considered eligible for inclusion (Textbox 1).

Textbox 1. Domains of the Consolidated Framework for Implementation Research (CFIR)

I.	**Innovation -** Domain that collects the characteristics of the implemented object from a multi-faceted point of view.
II.	**Outer setting** - Domain designed to capture factors that are inherent with the context where the organization exists.

III.	**Inner setting** - Domain which encompasses the characteristics of the organization in which the innovation is implemented. It includes both structural attributes, which characterize the inner setting regardless of the implementation, and features that are specific to the implementation.
IV.	**Roles** - Domain which refers to the individuals who have significantly contributed to the implementation and their characteristics.
V.	**Implementation process** - Domain that collects all the information on the activities and strategies adopted to concretely implement the innovation.

Table 1. Eligibility criteria

	Inclusion criteria	**Exclusion criteria**
Study design	Empirical studies illustrating the implementation of ML-based applications (e.g., experimental/quasi-experimental, observational, hybrid, simulation study designs, qualitative designs, case studies, etc.)	Effectiveness research study designs, literature reviews, commentaries, editorials, opinion articles, study protocols, studies collecting perceptions on implementation and unrelated to specific ML-based applications
Intervention	Analysis of the implementation of ML-based applications by at least covering factors related to the inner setting or process domains based on the CFIR	Analysis of the implementation of logic- or knowledge-based applications (e.g., expert systems), or of ML-based applications with no considerations related to the inner setting or process domain
Stakeholder groups	ML-based applications used at least by HCPs	ML-based applications targeting patients and other non-clinical stakeholders (e.g., caregivers, policy makers, regulators) only
Setting	Hospitals, outpatients and other community care settings	All other settings, including home care
Timeframe	Studies published from 2013 until March 2023	Studies published before 2013

Abbreviations: CRIF: Consolidated Framework on Implementation Research; ML: machine learning; HCP: healthcare professionals

Information sources
Literature searches were conducted in Medline (PubMed), Scopus, and Web of Science and replicated in top-tier management journal databases. In addition, the reference lists of all included studies and of the reviews identified were scanned to ensure comprehensive coverage of relevant literature. Grey literature was not considered.

Search strategy
The search strategy was developed by the research team through an iterative process and is based on four main concepts: (1) artificial intelligence; (2) implementation; (3) healthcare; (4) study design. Multimedia Appendix 1 contains the search strings used for each database. The general term "artificial intelligence" was used broadly to encompass studies that address AI and ML as synonymous terms. The search was performed in April 2023.

Study selection and data collection process
Two researchers (VA, LMP) screened the first 100 retrieved studies based on titles and abstracts. Once alignment over the inclusion/exclusion criteria was reached, the remaining records were independently screened by the two reviewers in equal parts based on title and abstract. Disagreements over final inclusions were solved with a third researcher (FP). Studies deemed eligible for full-text reading were

assessed in-depth (VA, LMP, FP). Disagreements were resolved by dialogue with two additional researchers (GC, AC). The entire research team read all the studies included in the analysis. The data collection process was performed by three reviewers (VA, LMP, FP) who extracted data using an *ad hoc* Microsoft Excel sheet preliminarily developed by the research team. To ensure consistency across reviewers, the extraction sheet was tested by each reviewer and re-calibrated before starting the data collection process. Any disagreements were resolved by discussion with the research team, with final decisions reached by consensus.

Data items

Data items were extracted based on established classifications or schemes, when applicable [24]. These encompassed information on the papers (e.g., journal of publication, publication year), the ML application (e.g., name, brief description, main practice of use, level of autonomy, degree of integration with other technologies), and the implementation process (e.g., stage of implementation, geographical location, care setting, specific unit of implementation). Furthermore, factors influencing the implementation process were assessed following the five domains of the updated version of the CFIR.

Quality assessment

The critical appraisal of the studies selected for data synthesis was performed using the Mixed Method Appraisal Tool (MMAT; [29]), a tool designed specifically for systematic reviews that include heterogeneous studies, as it allows to assess the methodological quality of five types of study designs (i.e., qualitative studies, RCTs, non-randomized studies, quantitative descriptive studies, mixed methods studies). Quality appraisal was performed by two researchers (VA, LMP), and disagreements were discussed and solved. The quality assessment represents a deviation from the protocol, which did not include this step.

Data synthesis

Given the significant heterogeneity across study designs, research objectives, and outcomes observed, as well as the expected predominance of qualitative studies, we opted for a thematic synthesis approach to capture and synthesize the salient attributes of the implementation process based on the CFIR constructs [30], [31]. The analysis considers findings from the data extraction process as qualitative data and includes summaries and interpretation of findings from the authors of the reviewed studies. Hence, direct quotes from participants were excluded in cases where the study employed qualitative data collection methods (e.g., interviews).

We used both an inductive and a deductive approach. Following the three thematic synthesis steps, we initially reviewed each paper and highlighted relevant aspects through line-by-line coding to capture and collect key data. The coding process involved three reviewers (VA, LMP, FP). To identify recurring topics, primary codes were then compared, organized and labeled to derive descriptive themes reflecting their meaning. Descriptive themes were used to develop higher-level analytical themes. The formulation of descriptive themes and the following assignation to analytical themes was initially proposed by a researcher (LMP) and iteratively refined through discussion with two other researchers (VA, FP).

The higher-level analytical themes were subsequently deductively redefined by the entire research team within the constructs of the CFIR, which served as the final theoretical framework guiding our analysis.


Results

Study selection

We retrieved 3,520 unique records that were initially screened based on titles and abstracts. A total of 67 records were deemed eligible for full-text screening (67/3,520, 2%). Additionally, we identified 36 eligible records from manual search of reference lists of excluded literature reviews and full-text screened records. Out of the 103 papers analyzed in full-text, 69 were excluded and 34 were included

in the review (34/103, 33%). The primary reason for exclusion was the focus of the intervention analyzed in the papers (53/69, 77%), as they either had a clinical or technical purpose without addressing factors related to implementation in an organizational setting or regarded non-ML-based applications. Figure 1 provides an overview of the selection process and the reasons for exclusion.

Figure 1. PRISMA flow diagram
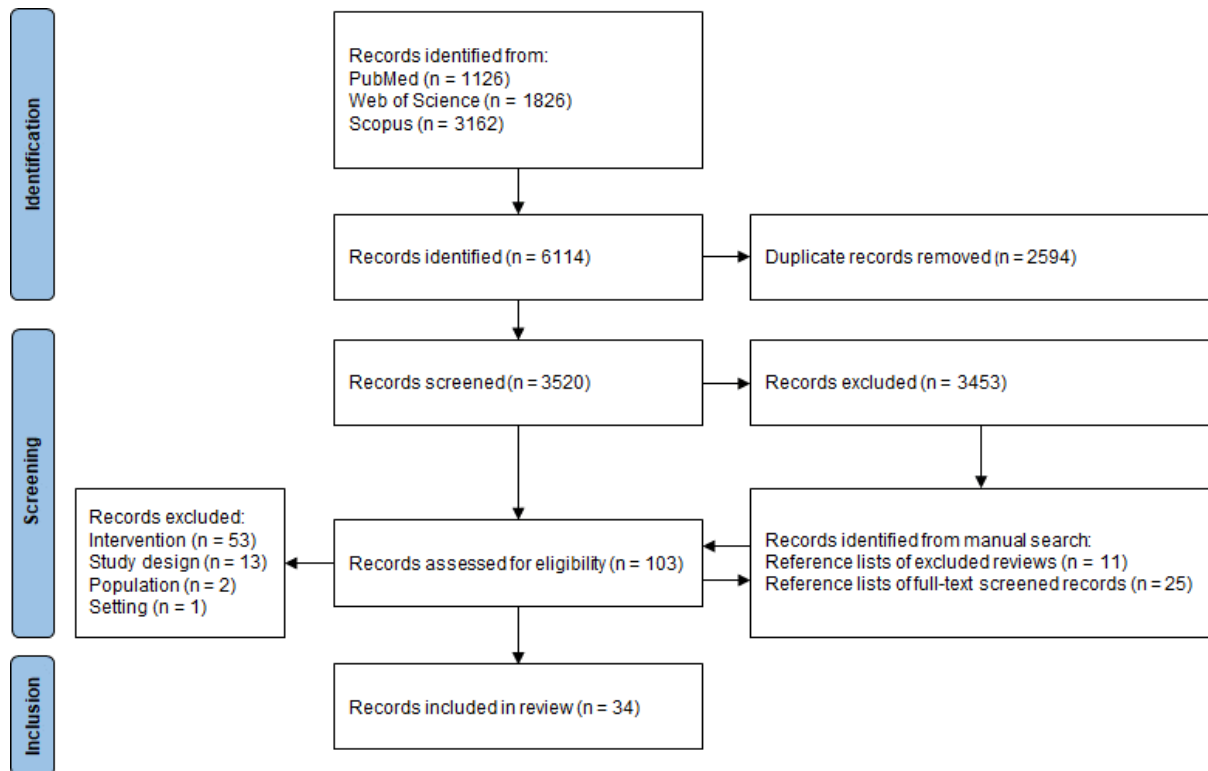
Study characteristics
Most of the studies documenting the implementation of ML-based applications were set in the USA (18/34, 53%). Other locations included China (4/34, 12%), Canada, Brazil, and the Netherlands (each 2/34, 6%), Italy, Spain, Norway, Korea, India, and Austria (each 1/34, 3%). Papers selected for data synthesis were relatively recent, with only three out of 34 (9%) published before 2019. Outlets were mostly clinical and/or in the field of information technology (IT) (30/34, 88%), while the remaining four focused on managerial or organizational studies (12%). Most of the selected studies followed qualitative or mixed methods designs (22/34, 65%), often relying on methods such as interviews and case studies.

Quality assessment in studies
Quality appraisal of the selected studies was performed using the MMAT tool. The studies were heterogeneous in terms of study design, and different MMAT questions were used to assess their quality. Overall, 18 studies leveraged the questions of qualitative studies, seven of quantitative non-randomized, five of quantitative descriptive studies, three of mixed methods studies, and one of quantitative randomized studies. Overall, the quality assessment suggests a medium-good quality of the studies, with only 12.5% of the assessment questions uncertain or unclear ("Can't tell"). The detailed output of the quality appraisal can be found in Appendix 2.

Characteristics of ML applications

Table 2 provides a general description of the ML-based applications implemented in the selected studies, while Multimedia Appendix 3 contains more detailed information on the characteristics of these applications. The most recurrent applications comprised predictive modeling algorithms, visualization tools and alert-delivering mechanisms. All the applications identified by our search were clinical practice applications, according to the definitions from the European Parliamentary Research Service [32]. Moreover, none of the applications had decisional autonomy; therefore, all systems could be classified as clinical decision support systems (CDSS).

In terms of settings, ML-based applications were mostly implemented within hospitals (29/34, 85%), including general, university or teaching hospitals, academic medical centers, and research centers. A few studies (4/34, 12%) were based in a community or primary care setting. Within hospital settings, the most recurring implementation units were emergency departments (ED) (11/34, 32%) and critical care units such as intensive care unit (ICU) (4/34, 12%), while in some studies implementation occurred in multiple units or at the hospital level (5/34, 15%).

The clusters identified by Rajkomar were used as a theoretical guide to classify the clinical workflow activities in which the ML-based applications were used [33]. In 20 studies (59%), the ML tools supported prognosis. Many of these applications were designed to predict the risk of developing specific conditions such as sepsis (8/34, 24%), in-hospital deterioration (3/34, 9%), intracranial hemorrhage (1/34, 3%), or hearth failure (1/34, 3%). Other applications predicted the risk of unplanned hospital admission or re-admission (4/34, 12%). Ten papers (29%) illustrated applications for diagnosis, either as standalone computer vision tools to detect diseases from diagnostic imaging (e.g., pneumonia from CT scans, large vessel occlusions from CT angiogram, child maturation from x-rays), or as diagnostic supports in emergency physician triage; three papers (9%) illustrated applications for treatment optimization and personalization.

ML capabilities relate to the clinical workflow activities, with forecasting (i.e., the ability to find complex patterns in data and make predictions) being the most prevalent capability (19/34, 56%), as this function is typical of tools that predict the risk of an adverse event (12/34 forecasting tools are for prognosis). Instead, computer vision was exclusively included in the algorithms with diagnostic purposes, with all six computer vision tools intended for diagnosis.

As for the level of integration with existing technologies, 17 ML-based algorithms (50%) were embedded in electronic health records (EHRs) or similar platforms (i.e., add-ons to the EHR software in use). Fourteen algorithms (41%) were standalone applications, fed either with internal or external data, including images or text. One application (3%) was embedded in the hospital hardware technology, namely scanner machines [36]. Computer vision applications were always standalone applications provided as software to be installed within existing hardware (i.e., hospital computers) and integrated with local picture archiving and communication systems (PACS).

The ownership of the algorithms was also assessed, revealing a division between applications purchased from commercial vendors (14/34, 41%) and those internally developed (12/34, 35%). The latter were often linked with the organizational setting, as six of these studies were carried out in teaching hospitals, academic medical centers or research centers. Externally purchased applications were more common in other settings and exhibited greater diversity in terms of purposes, while homegrown tools were generally intended for prognostic purposes. In eight studies, information on the name or the development process of the application was irretrievable, preventing determination of algorithm ownership.

Details on the specific ML models employed were often missing, although it is possible to infer that 20 of the analyzed studies (59%) were based on supervised learning models such as random forests, decision trees, or logistic regressions.

Table 2. Overview of ML applications

N.	Authors	Year	Application name	Application output description	Implementation setting (unit)	Clinical workflow activity
1	Lee et al. [34]	2015	-	Prediction of patient characteristics, complaint types, and admission and readmission patterns	Hospital (ED)	Prognosis
2	Hengstler et al. [35]	2016	IBM Watson	Building hypotheses and evidence on cancer diagnosis	Hospital (oncology)	Diagnosis
3	McCoy & Das [36]	2017	InSight	Prediction of risk of developing severe sepsis	Community hospital (ED, ICU)	Prognosis
4	Bhattacharya et al. [37]	2019	Niramai Thermalytix and iBreastExam (iBE)	Earlier detection of breast cancer	Hospital (radiology)	Diagnosis
5	Cruz NP et al. [38]	2019	Savana	Recommendations for improving adherence to healthcare pathways	Primary care	Clinical/Organizational workflow
6	Ginestra et al. [39]	2019	EWS 2.0	Prediction of risk of developing sepsis	Teaching hospital (non-ICU settings)	Prognosis
7	Gonçalves et al. [40]	2020	Laura	Prediction of risk of developing sepsis	Hospital (several units)	Prognosis
8	Sun & Medaglia [41]	2019	IBM Watson for Oncology	Decision-making support for personalized treatment planning	Hospital (oncology)	Treatment
9	Baxter et al. [42]	2020	-	Prediction of unplanned readmission	Teaching and research hospital (unspecified)	Prognosis
10	Cho et al. [43]	2020	DEWS (Deep-Learning-based Early Warning System)	Prediction of in-hospital cardiac events	Hospital (cardiology)	Prognosis
11	Frontoni et al. [44]	2020	-	Production of indicators for quality-of-care processes of T2D	Primary care	Clinical/Organizational workflow
12	Hassan et al. [45]	2020	Viz.ai	Detection of large vessel occlusions	Hospital (stroke unit)	Diagnosis
13	Romero-Brufau et al. [46]	2020	-	Prediction of hospital readmission and formulation of targeted recommendations	Hospital (all units)	Prognosis Treatment
14	Sandhu et al. [47]	2020	Sepsis Watch	Prediction of risk of developing sepsis	Teaching hospital (ED)	Prognosis
15	Sendak et al. [48]	2020	Sepsis Watch	Prediction of risk of developing sepsis	Teaching hospital (ED)	Prognosis
16	Strohm et al. [49]	2020	BoneXpert	Assessment of child maturation and bone age and prediction of adult height	Hospital (radiology)	Diagnosis Prognosis
17	Xu et al. [50]	2020	SensEcho	Classification of sleep stage, detection of sleep apnea and recognition of	Hospital (general and respiratory)	Diagnosis

			abnormal ECG signals from a multi-sensor wearable device		
18	Jauk et al. [51]	2021	-	Prediction of risk of developing delirium	Hospital (surgery, internal Medicine)	Prognosis
19	Morales et al. [52]	2021	Laura Digital ER	Detection of COVID-19 symptoms	Community	Diagnosis
20	Murphree DH et al. [53]	2021	-	Treatment optimization and identification of likely-to-benefit patients for palliative care	Hospital (all inpatient units)	Treatment
21	Yao et al. [54]	2021	3D CSAC-Net	Detection of mild COVID-19 pneumonia	Hospital (unspecified)	Diagnosis
22	Davis MA et al. [55]	2022	Aidoc	Prediction of risk of developing intracranial hemorrhage	Research hospital (radiology)	Prognosis
23	Henry et al. [56]	2022	TREWS	Prediction of risk of developing sepsis	Acute care hospital (inpatient acute units and ED)	Prognosis
24	Joshi et al. [57]	2022	-	Prediction of risk of developing sepsis	Community and teaching hospitals (several units)	Prognosis
25	Lebovitz et al. [58]	2022	-	Image processing, segmentation and classification for imaging diagnostics	Teaching hospital (radiology)	Diagnosis
26	Rushlow et al. [59]	2022	-	Prediction of risk of low left ventricular ejection fraction	Primary care	Prognosis
27	Schwartz et al. [60]	2022	CONCERxN	Prediction of risk of in-hospital deterioration	Teaching hospital (acute units and ICU)	Prognosis
28	Sibbald et al. [61]	2022	Isabel	Differential diagnosis	Teaching hospital (ED)	Diagnosis
29	Singer et al. [62]	2022	Low Bed Tool and Readmission Risk Tool	Prediction of reduced bed availability and prediction of risk of readmission	Hospital (ICU, surgery, pediatrics)	Clinical/Organizational workflow Prognosis
30	Wijnhoven [63]	2022	Sepsis Identification Speed	Prediction of risk of developing sepsis	Teaching hospital (neonatology)	Prognosis
31	Zhai et al. [64]	2022	Nu-CDSS	Formulation of recommendations for nurses' diagnosis, interventions and outcome evaluations	Teaching hospital (unspecified)	Clinical/Organizational workflow
32	Pou-Prom et al. [65]	2022	CHARTwatch	Early warning system designed to predict patient risk of clinical deterioration	Teaching hospital (general internal medicine)	Prognosis
33	Hinson et al. [66]	2022	-	Estimation of the short-term risk for clinical deterioration in patients with or under investigation for COVID-19	Teaching hospital (ED)	Prognosis
34	Berge et al. [67]	2023	Information System for Clinical	Detection and classification of patient allergies	Hospital (anesthesia, ICU)	Diagnosis

Abbreviations: ECG: electrocardiogram; ED: emergency department; ICU: intensive care unit; T2D: type 2 diabetes
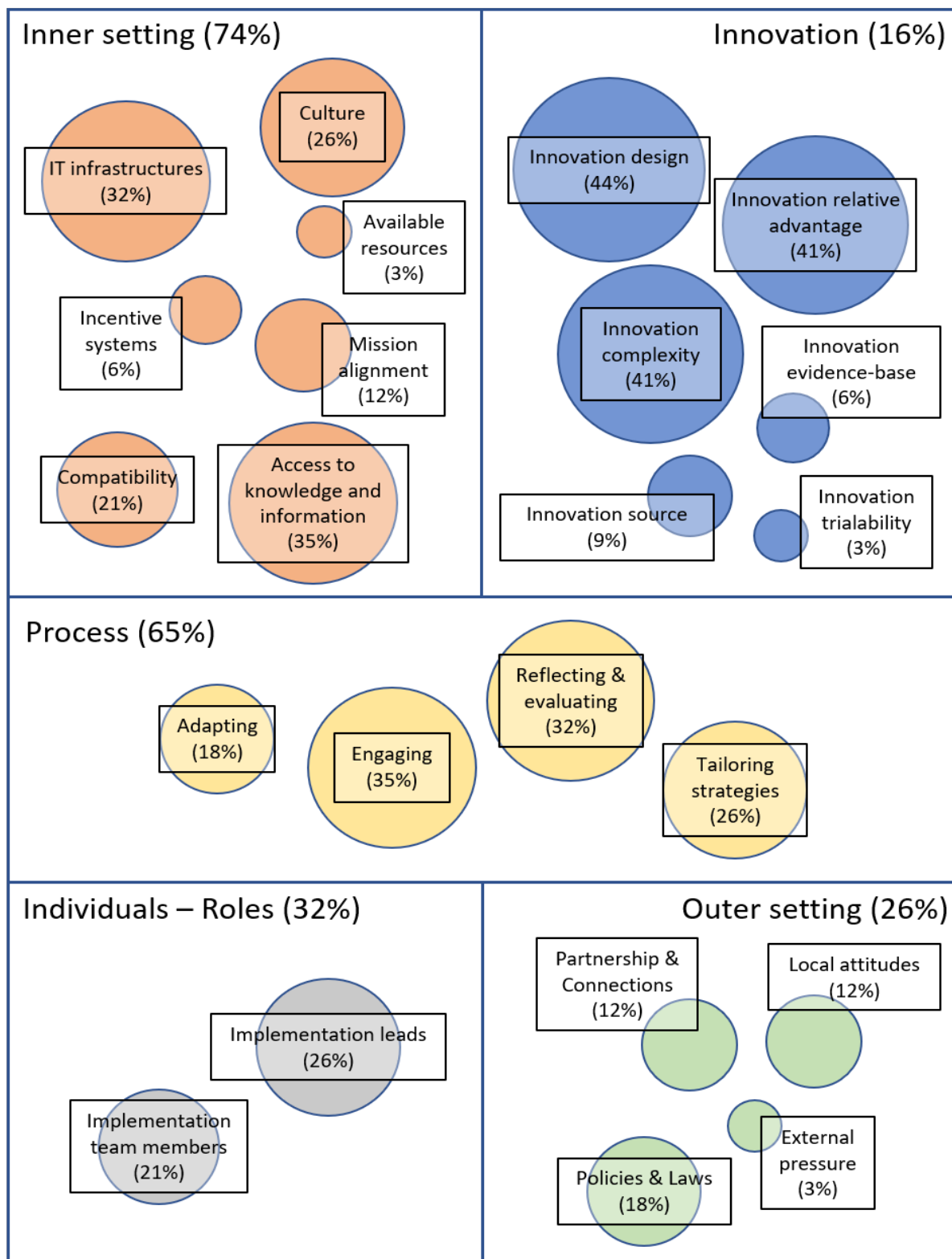
Implementation process characteristics

This section presents the results of the thematic analysis, discussed following the five domains of the CFIR, namely innovation, outer setting, inner setting, roles, and implementation process. From the 34 studies analyzed, 222 quotes were extracted. Quotes were organized in 167 descriptive themes and 35 analytical themes. Analytical themes were finally embedded into 23 CFIR constructs. The detailed result of the coding process can be found in Appendix 4. To provide a simplified overview of the coding process, Table 3 summarizes the analytical themes, their correspondence with CFIR constructs, and relative frequencies. Results are reported according to the frequency of information extracted on CFIR domains.

Table 3. Analytical themes, constructs and domains of CFIR

Construct	Analytical themes	Papers, n (%)	References
Inner setting domain (n = 25, 74%)			
A. Structural characteristics (A.2 IT infrastructure)	Integration with existing IT	11 (32%)	[35], [38], [40], [41], [49], [50], [52], [59], [60], [62], [63]
	Data governance		
	System infrastructure		
D. Culture	Professional habits	9 (26%)	[39], [40], [41], [42], [47], [49], [61], [64], [68]
	User perceptions		
F. Compatibility	Local workflow adaptation	7 (21%)	[42], [46], [49], [51], [64], [65], [67]
H. Incentive systems	Economic incentives	2 (6%)	[44], [58]
	Organizational incentives		
I. Mission alignment	Organizational strategy	4 (12%)	[41], [49], [57], [64]
	Organizational support		
J. Available resources	Resource reallocation	1 (3%)	[34]
K. Access to knowledge & information	Skills	12 (35%)	[36], [40], [41], [42], [47], [50], [51], [57], [59], [65], [66], [68]
Innovation domain (n = 22, 65%)			
A. Innovation source	Trust in the innovation source	3 (9%)	[35], [41], [46]
B. Innovation evidence base	Empirical evidence on added value	2 (6%)	[49], [51]
C. Innovation relative advantage	Performance trust	14 (41%)	[39], [41], [42], [45], [46], [47], [49], [57], [58], [60], [61], [63], [64], [67]
	Perceived cons		
	Perceived benefits		
E. Innovation trialability	Testing period	1 (3%)	[65]
F. Innovation complexity	Explainability	14 (41%)	[34], [35], [39], [41], [46], [47], [51], [56], [57], [58], [60], [61], [63], [67]
G. Innovation design	Complementarity	15 (44%)	[35], [41], [47], [51], [54], [55], [56], [57], [58], [60], [61], [64], [66], [67], [68]
	Ease of use		
	Risks		

Process domain (n = 22, 65%)			
E. Tailoring strategies	Framing	9 (26%)	[35], [40], [47], [48], [49], [56], [57], [59], [64]
	Tailoring		
F. Engaging	Early involvement of end-users	12 (35%)	[35], [40], [47], [48], [53], [57], [59], [60], [62], [63], [64], [65]
	Professional buy-in		
	Iterative development		
H. Reflecting & evaluating	Feedback	11 (32%)	[36], [38], [40], [47], [48], [50], [51], [62], [63], [64], [67]
I. Adapting	Local data	6 (18%)	[38], [41], [43], [46], [53], [60]
	Adaptability		
Individuals domain – Roles sub-domain (n = 11, 32%)			
E. Implementation leads	Implementation lead	9 (26%)	[36], [42], [47], [48], [49], [57], [63], [64], [65]
	Implementation team		
F. Implementation team members	Interdisciplinary teams	7 (21%)	[47], [48], [53], [57], [63], [64], [66]
Outer setting domain (n = 9, 26%)			
B. Local attitudes	Patient acceptance	4 (12%)	[35], [37], [41], [52]
	Public attitude		
D. Partnership & connections	Inter-institutional partnerships	4 (12%)	[44], [49], [52], [63]
	Public-private partnerships		
E. Policies & laws	Medicolegal issues	6 (18%)	[35], [41], [48], [49], [52], [63]
	MD regulation		
	Guidelines		
	Data protection		
G. External pressure	Peer influence	1 (3%)	[56]

Figure 2. Relative importance of CFIR constructs (bubble size represents frequency)

Inner setting (74%)

- Culture (26%)
- IT infrastructures (32%)
- Available resources (3%)
- Incentive systems (6%)
- Mission alignment (12%)
- Compatibility (21%)
- Access to knowledge and information (35%)

Innovation (16%)

- Innovation design (44%)
- Innovation relative advantage (41%)
- Innovation complexity (41%)
- Innovation evidence-base (6%)
- Innovation source (9%)
- Innovation trialability (3%)

Process (65%)

- Adapting (18%)
- Engaging (35%)
- Reflecting & evaluating (32%)
- Tailoring strategies (26%)

Individuals – Roles (32%)

- Implementation leads (26%)
- Implementation team members (21%)

Outer setting (26%)

- Partnership & Connections (12%)
- Local attitudes (12%)
- Policies & Laws (18%)
- External pressure (3%)

Note to Figure 2: the percentages represent the proportion of papers in which each construct and domain appear, out of the 34 included in the review. The size of bubbles corresponds to the frequency of occurrence of each construct.

Inner setting

The inner setting domain was the most frequently described, with 25 studies mentioning at least one construct from this domain as relevant to explaining the implementation process of the ML application. The most recurrent constructs were access to knowledge and information (12/34, 35%), IT infrastructures (11/34, 32%), and culture (9/34, 26%).

First, the *access to knowledge* construct aligns with the topic of skills. Studies emphasized the importance or providing end-users with access to training programs on both hard and soft skills before implementation [50], [51], [57], [66], including computer and technical literacy linked with the complexity of the application's functioning [59], [65], and the medical domain that the application addresses [47]. The latter referred to dimensions such as communication, empathy, and ability to listen, especially when different HCPs are involved in the implementation process [40], [41], [47].

Second, the *IT infrastructure* construct encompassed two prominent themes. The first broadly concerns data management and data governance. Themes such as data collection and quality [41], [60], [61], security [35], availability [38], [40], and sharing [41], [63] were highly described as challenges for adoption of the application. There were also significant references to building IT infrastructures [63] and to the need to integrate new technologies with existing IT systems (e.g., EHR). While integration promoted ease of use by reducing the need for manual inputs [49], [50], [52], some argued for the ML application not to directly populate EHRs to preserve HCP autonomy and prevent medicolegal accountability [61].

Finally, the construct of *culture* was articulated into the themes of professional habits and alignment of perceptions among stakeholders. The impact of introducing ML applications on professional habits was significant as this affected how professionals work, interact, and make decisions [64]. For instance, the habit of working without technological support was considered a barrier to implementation [41], [42]. The need for gradual changes in professional habits was seen as a factor that could hinder the adoption of ML applications in settings with high job rotations (e.g., teaching hospitals) [34]. Additionally, ML applications often are not tailored to local workflows, and do not consider the different approaches of professionals in diverse contexts [41]. The other theme related to culture was that of perceptions and expectations among different internal stakeholders (e.g., management, physicians, nurses, and technical staff). Misalignments among these stakeholders were common, particularly regarding trust in ML in general [47], [49], or the expected target users (e.g., residents vs. expert physicians) [47], [61].

Innovation

The innovation and its characteristics were among the most frequently mentioned domains (22/34, 65%), with three constructs absorbing a significant portion of relevant descriptive themes: innovation design (15/34, 41%), relative advantage (14/34, 41%), and complexity (14/34, 41%).

First, *innovation design* encompassed themes related to the applications' design and functioning, including the types of human-machine interaction, as well as the associated risks. The most recurrent themes within this construct revolved around ease of use and intuitive design [35], [47], [51], [64]. The former was often linked to minimizing manual intervention, such as data input [51], [54], and was also associated with dimensions of trust in the applications, such as trust in the process and the cognitive burden for HCPs, in the form of fatigue from (over-)alerting [35], [64], which could be a barrier to professional buy-in [57]. Some studies explicitly cited the theme of human-centered design as a development framework that starts with the assessment of end-users' needs and the environment in which the ML application will be used [56], [66], [69]. Another recurrent theme was the human-machine complementarity. For HCPs, it was often important to maintain a sense of control over the application and not perceive it as an attempt of uncontrolled substitution and automation [34], [35], [58], [67]. Human-machine complementarity was also associated with fewer disruptions to established workflows, enhancing the overall benefits associated to the use of a ML application [47], [55]. Moreover, complementarity could increase trust in the application, both from a micro-perspective (e.g., its functioning) [35], [56], and from a macro-perspective (e.g., the purpose of the application, and the

reasons for choosing to integrate ML within a clinical context) [58], [60], [67]. The risks of ML use in decision-making processes also emerged. These included the risk of automation, in terms of overreliance on ML recommendations [41], and the risk of bias, tied to the underlying data and training model of the ML application [61]. Moreover, potential negative consequences of automation risk on clinical ability were mentioned [56].

Second, the *relative advantage* revolved around the perception of benefits and costs associated with the use of ML, as well as factors influencing trust in its performance. The most frequently perceived benefits were related to the organizational dimension, in terms of optimization of the workflow resulting from the elimination of unnecessary steps [45], increased attention from end-users to all cases managed by the application [47] and enhanced interactions among physicians and other HCPs [64]. Conversely, references to the economic impact were ambiguous. On one hand, faster decision-making could be considered a potential advantage [49]; on the other hand, human-machine interaction could also lead to a loss of efficiency compared to human intervention only [42], [58]. Another barrier to professional buy-in is that the perceived poor ability of the application to take contextual factors into account calls into question its clinical relevance Among the perceived advantages, trust in the application's performance and its determinants were often commented on. For the analysis, the concept of trust was declined as already done by Hengstler et al. [35] who distinguished between trust in technology and trust in those who produce it (i.e., the source of innovation). Trust in technology is further divided into three dimensions: trust in performance, focusing on the accuracy and consistency of the output; trust in process, concerning the understanding of the reasoning behind a given output, and trust in the purpose of the innovation to be implemented [16], [46]. Concordance significantly influences trust in performance: the greater the difference between human judgment and machine recommendation, the lower the level of trust in the recommendation [46], [49], [60]. Similarly, recommendations that did not arrive in a timely or adequate manner negatively influenced performance trust [39], [46]. Additionally, trust in performance could be fueled by experience, the application's ability to identify cases missed by humans, and the consistency over time of recommendations [49].

Third, *innovation complexity* highlighted the concepts of explainability and opacity as distinctive features of ML models. Many studies agreed in identifying algorithm complexity as the primary barrier to trust in the process underlying the generation of an ML output. This is even more true when non-medical professionals (e.g., nurses) interact directly with the ML application [47]. Facilitating interpretability, explainability or cognitive compatibility were mentioned as ways to promote transparency, HCPs' trust, and professional buy-in [51], [57], [58], [60], [61], [67].

Implementation Process

Reviewed articles often mentioned the characteristics of the implementation process (22/34, 65%), with a particular emphasis on the constructs of stakeholders engaging (12/34, 35%), reflecting and evaluating (11/34, 32%), and tailoring strategies (9/34, 26%).

Attracting and encouraging the *participation of different stakeholders* in the implementation process emerged as a recurring theme. The practice of early involvement of end-users was frequently cited, not only during the implementation process but also throughout the development phase [35], [40], [47], [48], [60], [63], [65]. This was positively associated with trust in the innovation's purpose [48], the application's functioning [60], [63], [64], and the ease of use of its design [47]. During the implementation phase, stakeholder engagement was linked to evident benefits, such as improvement in the implementation climate [48], greater willingness to adopt the role of implementation leader [40], greater professional buy-in [53], [57], and better iterative collection of information and feedback [62]. Conversely, the absence of engagement was seen as a barrier to successful implementation, potentially leading to increased resistance toward the innovation among end-users [64].

In the construct of *reflecting and evaluating*, feedback and feedback loops emerged as recurring topics, with many studies underscoring the importance for both ML developers and the implementation teams to incorporate end-users' feedback on either technical issues, system design, or clinical needs [36], [38],

[40], [51], [64], [67]. Some studies noted that feedback collection extended beyond implementation, with structure feedback loop processes integrated into routine use [48], [50]. Regardless, feedback collection was described as an iterative activity [48], [50], [62], [63], which also positively influenced professional buy-in [49]. However, a critical point raised was that end-users may lack the necessary technical skills to provide feedback conducive to improvement [64].

Two additional recurring constructs were tailoring strategies and adapting. The former referred to actions addressing barriers and leveraging facilitators, while the latter involved modifying the innovation itself to best fit the context in which it was inserted.

Among *tailoring strategies*, the importance of effectively communicating the implementation efforts was often highlighted. Some works referred to the need for clearly framing communication around the expected benefits, positively affecting trust in ML-based innovations [35], [40], [47], [48], [49], [56], trust in the innovation source [35], and fostering greater professional buy-in [57]. Another aspect of framing was related to the terminology used, asserting that using terms supporting concepts such as "assistant" or "support" had a favorable impact on end-users' trust toward ML-based innovations [48], [49] and the innovation source [35].

In terms of *adapting*, the first theme involved the need for collected feedback to be effectively incorporated in the application, adapting systems to the local context of implementation [38], [60]. The second involves the issue of data, emphasizing the importance that the model is effectively trained and adapted to the cases treated in the clinical context in which the application will be used before deployment. The absence of this aspect was perceived as a barrier to trust in the ML application's performance [41], [43], [46], [53].

Individuals: roles
The sub-domain of roles was less frequently observed (11/34, 32%) and encompassed two constructs: implementation leads (9/34, 26%) and implementation teams (7/34, 21%).

The former referred to the individual or group that guided and oversaw the implementation process, and their presence was generally considered a positive factor for implementation as it contributed to establishing a favorable implementation climate [49]. Individual implementation leaders were often referred to as champions. Although it may theoretically involve figures that emerge from bottom-up processes, all works referring to this role mentioned a top-down identification [47], [49], [64], [65]. Implementation teams were observed as well, in the form of quality improvement teams [36], AI governance committees [42], [48], [65], or interdisciplinary teams of HCPs, software engineers, developers, IT specialists, and other figures [48], [53], [57], [66].

Outer setting
The outer setting domain emerged poorly in the revised studies (9/34, 26%), particularly in the form of three constructs: policy and laws (6/34, 18%), local attitudes (4/34, 12%), partnership and connections (4/34, 12%).

In the *policy and laws* construct, three main themes emerged. The first concerned the medicolegal responsibility for decisions made using a ML application [41], [63]. The second pertained to regulatory and certification aspects, with recognition of the application as a medical device (MD) seen both as a factor positively influencing trust in the application [35] and as a barrier for utilization [48]. Regulations on personal data protection were also considered implementation challenges [49]. Regarding policies, the only theme mentioned was the relevance of national policies and guidelines to create a common framework for the implementation of ML applications [52].

Local attitudes were societal expectations and beliefs on the use of ML applications. Cultural aspects, innovation attitude, and public expectations could influence the acceptability of ML [37], [41], [52]. Equally relevant for acceptance was the visibility of the application, – i.e., how noticeable and observable an innovation is to the public – that influences how organizations foster innovation trust [35], [70].

Within the *partnership and connections* construct, building partnerships with scientific societies and professional communities was considered a facilitator for implementation, as these can act as knowledge platforms/hubs [44], [49]. Professional communities and peers could also trigger external pressure that may positively impact the willingness to implement ML applications [56]. Establishing development networks across hospitals and healthcare facilities was a relevant factor for the increased reliability of the application, providing the opportunity to leverage larger datasets which are known to end-users [63]. Moreover, forging public-private partnerships was deemed a useful step for implementation, to leverage expertise not always available within (public) healthcare organizations [52], [63].

Discussion

Review of the main findings

This work aimed at synthetizing extant academic knowledge on the implementation of ML-based applications in clinical practice, focusing specifically on the characteristics of the innovation and on the processes and strategies employed by healthcare organizations to ensure their successful implementation.

We identified 34 studies reporting on the implementation process of as many ML applications, all of which were CDSS frequently based on supervised learning models in the form of predictive algorithms, visualization and alert-delivering tools. Overall, half of the observed applications were integrated in hospital information systems as add-ons to the EHR infrastructure. ML-based applications were mainly implemented in hospital settings and supported prognostic activities, although a relevant portion was intended for diagnosis. Among the diagnostic applications, those based on computer vision were either standalone software or embedded in the hospital hardware technology. Algorithms could be clustered in two groups: those internally developed, prevalently by university hospitals and academic medical centers and typically with a prognostic purpose, and those purchased from commercial vendors, more heterogeneous in terms of purposes and functions.

Furthermore, our analysis enabled us to scrutinize the characteristics of the implementation processes of ML-based applications, gathering pertinent insights relevant for their successful integration within healthcare organizations. Through the theoretical lens of the CFIR, we identified a predominant emphasis on three key domains: the inner setting, innovation characteristics and the process dimension. First, evidence from the inner setting domain highlighted the importance of addressing IT infrastructure and data management challenges, as well as the necessity of fostering an organizational culture that favors the implementation of ML-based applications. Second, in terms of innovation design, the concept of human-machine complementarity was recurrent, highlighting the importance of integrating ML-based applications into existing workflows to enhance overall benefits and foster trust by ensuring HCPs maintain a sense of overall control. In the process domain, studies emphasized the importance of fostering early stakeholder engagement during the development and pre-implementation phases, adapting strategies to local contexts, and initiating reflection and evaluation activities to support continuous improvement based on feedback loops. Conversely, while the complexity inherent in ML models in terms of algorithm opacity was largely acknowledged, we found limited investigation into effective mitigation strategies to tackle these challenges.

Comparison with prior work

Differently from prior work encompassing also logic-based and rule-based applications [10], [71], [72], our study focused exclusively on ML-based applications. While the frequency and relative significance of the various application types are not directly comparable with those observed in the cited works, other recent reviews have adopted a similar approach to ours. In their scoping review, Chomutare et al. identified 19 studies on the implementation of AI applications powered by ML, highlighting a variety of solutions across medical fields and tasks within the clinical workflow [21]. Similarly, Tricco et al. explored how implementation science strategies can facilitate the implementation of ML tools [22], but their work also included studies with effectiveness research designs, thereby adopting a partially

different approach from that of this work. Our review expanded the number of included studies, confirming the multitude of diverse applications of ML in clinical practice. The only condition for which we observed a conspicuous number of studies was sepsis, a dysfunction accounting for around 20% of deaths worldwide [73], for which ML-based applications are proliferating [74], although no definitive causal link with reduced mortality has been demonstrated to date [75]. Our search identified eight studies on sepsis, showcasing the potential attributed to ML-based applications in supporting the timely identification of hospital-acquired conditions. On a similar note, a recent review encompassing over 10,000 ML applications in healthcare settings corroborates the relevance of prognostic algorithms among those in use [76].

Consistently with previous research [21], [71], most of the included papers presented cases of real-world implementation rather than being proper implementation studies on the later phases of roll-out, often covering only a few aspects of the implementation process.

While we hypothesized that distinct implementation strategies would be prevalent based on the characteristics of the ML-based application, we only observed limited distinctions based on the type of clinical applications (prognostic, diagnostic, or therapeutic purposes), or their development process (internal development vis-à-vis external acquisition and adaptation).

For instance, the integration with existing IT infrastructures introduced ambiguity in the context of diagnostic applications, where such integration may be perceived as a risk with medicolegal implications [52], [61]. On the other hand, for applications with non-diagnostic purposes, integration with existing IT systems was viewed as a positive factor for ease of use [49], [64].

Other elements appeared relatively more pronounced in applications provided by external providers. This includes perceived risks associated with application design (e.g., overreliance, automation, bias) [41], [56], [61], considerations regarding complementarity with HCPs [35], [55], [56], [58], and aspects related to explainability. As such, exploring whether and how different application types entail different implications for their effective integration into clinical practice might be a valuable suggestion for future research.

Just like Chomutare et al. [21], our work confirmed that the outer setting domain was largely overlooked, although prior studies have highlighted the importance of external factors such as data privacy and security laws, ethical issues, regulatory frameworks, and medical liability in implementing ML applications in clinical practice [5], [77], [78], [79]. The limited relevance of such domain in our sample may stem from two reasons. First, due to the nature of the included studies, only a few used frameworks accounted for elements beyond the organizational setting in which the implementation occurred. Factors associated with the outer setting may be more frequently highlighted in implementation processes perceived as unsuccessful, which are less often reported in the scientific literature. Second, since the primary studies predominantly involved HCPs, they did not incorporate managerial and policymaker perspectives. In fact, when the outer domain perspective was explored, non-clinical stakeholders were often involved [41], [49], [63]. Furthermore, Hogg et al. suggested prioritizing the perspective of non-HCP stakeholders in primary studies to enhance understanding of implementation processes at a broader level [10], which may serve as further valuable suggestion for future studies.

Implications for the implementation of ML-based applications: a focus on trust

The importance of trust, particularly within the physician-patient relationship, has been heightened by the advent of digital health, especially with innovations such as ML applications that heavily rely on data [80]. ML applications based on computational models are often characterized as opaque (i.e., black boxes), introducing an extra layer of complexity to the trust relationship between end-users and technological innovations [81]. A recent review by Adjekum et al. categorized factors influencing trust in digital health systems into personal, technological, and institutional elements [80]. Building upon the concept of trust as articulated by Hengstler et al. [35], our work contributes to understanding the determinants of trust in facilitating the implementation of ML-based applications in healthcare organizations.

We observed that the characteristics of the innovation itself significantly challenge trust in the performance of ML-based applications. The complexity and opacity of the underlying models constitute primary barriers to trust, with trust in performance further influenced by system design elements such as ease of use, the nature of HCP-machine interaction, and the timeliness and consistency over time of recommendations. Additionally, considerations regarding data governance for internally-developed applications and the reputation of the technology provider for procured solutions further influence trust in the performance of these applications. However, as trust primarily remains a human-led process, factors beyond mere technical and mechanical characteristics influence trust in ML.

While most of the observed implementation strategies were essentially ML-agnostic, addressing the issue of clinician trust should theoretically require dedicated, ML-specific processes. Our review highlights potential ways to enhance the application perceived reliability of ML applications. On the one hand, tailoring and adaptation strategies, early end-user engagement, and appropriate framing of ML-based applications as decision-support tools might favor HCP trust in both the application's performance and its purpose [21]. On the other, specific tailoring strategies should be adopted to increase the explainability of the non-totally interpretable models [13]. For instance, Jauk et al. enhanced clinical reasoning using a web application presenting relevant features from ML-modelling [51], Davis et al. allowed radiologists to interact with the ML system by showing the types and locations of the abnormalities identified by the algorithm [55], and Henry et al. decided to delay alerts until the first verifiable symptoms were present in an attempt to increase acceptance [56].

However, these tailoring strategies may not be practicable when ML systems reach opacity levels that render the interpretation of their outputs impracticable. In such cases, other contributions have emphasized the need to highlight the level of actionability of ML models, in terms of their ability to enhance medical decision-making compared to clinical judgement alone, to power trust [82].

An additional contribution to enhancing trust may be achieved through continuous HCPs' involvement. This involvement, which generates engagement and professional buy-in, is equally significant for the successful implementation of these innovations. In the realm of digital health interventions, while there is frequent emphasis on patient engagement in the design of solutions, the empowerment of HCPs is often overlooked [83], [84]. Active involvement of HCPs and frequent communications to raise awareness have been unambiguously identified among the most common enablers of trust in previous reviews on the implementation of ML applications [21], [22], [75]. This may facilitate the implementation of innovations by improving the implementation climate reducing resistance to change, and mitigating specific barriers associated with the complexity of ML models and the reliability of the recommendations they produce.

Limitations

This study has some limitations that should be considered when interpreting our findings. Firstly, the rapidly evolving nature of the field of ML and the exponential growth of newly published studies posed a challenge in managing the vast volume of retrieved records. To address this, our search strategy incorporated a supplementary block of keywords focused on "study designs", which may have excluded certain relevant articles. Additionally, our emphasis on peer-reviewed studies introduces a potential bias, as ML-based applications reported in the scientific literature may only represent a subset of implemented systems. This could impact the generalizability of our findings, as acknowledged in similar studies such as Sharma et al. [71]. Lastly, the decision to include only papers published in English might have led to the exclusion of valuable sources published in other languages, limiting the comprehensiveness of our review.

Conclusions

Despite a relative dearth of primary studies on the implementation of ML applications in healthcare organizations, the available evidence reveals the abundance and heterogeneity of factors involved when ML applications are introduced in routine clinical practice. While certain elements, such as complexity

and trust, tend to emerge as distinctive factors for ML applications, many other aspects reflect what is already known about the implementation of digital technologies, particularly traditional CDSS.

Further research is needed to bridge the gap between the theoretical potential of ML and its actual use in healthcare organizations. Identifying the distinctive factors that can facilitate its implementation will build theoretical and practical knowledge for healthcare practitioners, ultimately promoting the uptake of ML in routine clinical practice.

List of abbreviations
AI: artificial intelligence
CDS: clinical decision support
CFIR; Consolidated Framework for Implementation Research
ED: emergency department
EHR: electronic health records
HCP: healthcare professionals
ICU: intensive care unit
IT: information technology
ML: machine learning
PACS: picture archiving and communication system
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analysis
PROSPERO: International Prospective Register of Systematic Reviews

Data availability
Data collected as part of the current review will be made available by the corresponding author upon request.

Authors' contributions
GC and AC conceived the initial concept for the review. VA and LMP carried out the search strategy and initial screening of the records. VA, LMP and FP performed the full-text analysis and data extraction. VA, FP, and LMP generated the first draft of the manuscript. All authors contributed to the analytical process and approved the final version of this manuscript.

Conflicts of interest
None declared.

Bibliography

[1] S. Secinaro, D. Calandra, A. Secinaro, V. Muthurangu, and P. Biancone, "The role of artificial intelligence in healthcare: a structured literature review," *BMC Med Inform Decis Mak*, vol. 21, no. 1, p. 125, Dec. 2021, doi: 10.1186/s12911-021-01488-9.

[2] A. L. Beam and I. S. Kohane, "Big Data and Machine Learning in Health Care," *JAMA*, vol. 319, no. 13, p. 1317, Apr. 2018, doi: 10.1001/jama.2017.18391.

[3] "Generating Evidence for Artificial Intelligence Based Medical Devices: A Framework for Training Validation and Evaluation.," World Health Organization, Geneva, 2021.

[4] M. Mittermaier, M. Raza, and J. C. Kvedar, "Collaborative strategies for deploying AI-based physician decision support systems: challenges and deployment approaches," *npj Digit. Med.*, vol. 6, no. 1, pp. 137, s41746-023-00889–6, Aug. 2023, doi: 10.1038/s41746-023-00889-6.

[5] J. Shaw, F. Rudzicz, T. Jamieson, and A. Goldfarb, "Artificial Intelligence and the Implementation Challenge," *J Med Internet Res*, vol. 21, no. 7, p. e13659, Jul. 2019, doi: 10.2196/13659.

[6] X. Liu *et al.*, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," *The Lancet Digital Health*, vol. 1, no. 6, pp. e271–e297, Oct. 2019, doi: 10.1016/S2589-7500(19)30123-2.

[7] A. Aristidou, R. Jena, and E. J. Topol, "Bridging the chasm between AI and clinical implementation," *The Lancet*, vol. 399, no. 10325, p. 620, Feb. 2022, doi: 10.1016/S0140-6736(22)00235-5.

[8] E. Coiera, "The Last Mile: Where Artificial Intelligence Meets Reality," *J Med Internet Res*, vol. 21, no. 11, p. e16323, Nov. 2019, doi: 10.2196/16323.

[9] J. He, S. L. Baxter, J. Xu, J. Xu, X. Zhou, and K. Zhang, "The practical implementation of artificial intelligence technologies in medicine," *Nat Med*, vol. 25, no. 1, pp. 30–36, Jan. 2019, doi: 10.1038/s41591-018-0307-0.

[10] H. D. J. Hogg *et al.*, "Stakeholder Perspectives of Clinical Artificial Intelligence Implementation: Systematic Review of Qualitative Evidence," *J Med Internet Res*, vol. 25, p. e39742, Jan. 2023, doi: 10.2196/39742.

[11] D. S. Watson *et al.*, "Clinical applications of machine learning algorithms: beyond the black box," *BMJ*, p. l886, Mar. 2019, doi: 10.1136/bmj.l886.

[12] D. Castelvecchi, "Can we open the black box of AI?," *Nature*, vol. 538, no. 7623, pp. 20–23, Oct. 2016, doi: 10.1038/538020a.

[13] A. Barragán-Montero *et al.*, "Towards a safe and efficient clinical implementation of machine learning in radiation oncology by exploring model interpretability, explainability and data-model dependency," *Phys. Med. Biol.*, vol. 67, no. 11, p. 11TR01, Jun. 2022, doi: 10.1088/1361-6560/ac678a.

[14] D. S. Char, N. H. Shah, and D. Magnus, "Implementing Machine Learning in Health Care — Addressing Ethical Challenges," *N Engl J Med*, vol. 378, no. 11, pp. 981–983, Mar. 2018, doi: 10.1056/NEJMp1714229.

[15] F. Gama, D. Tyskbo, J. Nygren, J. Barlow, J. Reed, and P. Svedberg, "Implementation Frameworks for Artificial Intelligence Translation Into Health Care Practice: Scoping Review," *J Med Internet Res*, vol. 24, no. 1, p. e32215, Jan. 2022, doi: 10.2196/32215.

[16] P. Nilsen *et al.*, "Realizing the potential of artificial intelligence in healthcare: Learning from intervention, innovation, implementation and improvement sciences," *Front.Health Serv.*, vol. 2, p. 961475, Sep. 2022, doi: 10.3389/frhs.2022.961475.

[17] P. Nilsen *et al.*, "A Framework to Guide Implementation of AI in Health Care: Protocol for a Cocreation Research Project," *JMIR Res Protoc*, vol. 12, p. e50216, Nov. 2023, doi: 10.2196/50216.

[18] Y. Guo, Z. Hao, S. Zhao, J. Gong, and F. Yang, "Artificial Intelligence in Health Care: Bibliometric Analysis," *J Med Internet Res*, vol. 22, no. 7, p. e18228, Jul. 2020, doi: 10.2196/18228.

[19] R. C. Li, S. M. Asch, and N. H. Shah, "Developing a delivery science for artificial intelligence in healthcare," *npj Digit. Med.*, vol. 3, no. 1, p. 107, Aug. 2020, doi: 10.1038/s41746-020-00318-y.

[20] A. K. Triantafyllidis and A. Tsanas, "Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature," *J Med Internet Res*, vol. 21, no. 4, p. e12286, Apr. 2019, doi: 10.2196/12286.

[21] T. Chomutare *et al.*, "Artificial Intelligence Implementation in Healthcare: A Theory-Based Scoping Review of Barriers and Facilitators," *IJERPH*, vol. 19, no. 23, p. 16359, Dec. 2022, doi: 10.3390/ijerph192316359.

[22] A. C. Tricco *et al.*, "Implemented machine learning tools to inform decision-making for patient care in hospital settings: a scoping review," *BMJ Open*, vol. 13, no. 2, p. e065845, Feb. 2023, doi: 10.1136/bmjopen-2022-065845.

[23] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.

[24] V. Ardito, G. Cappellaro, A. Compagni, F. Petracca, and L. M. Preti, "Implementation of Machine Learning Applications in Healthcare Organizations: Protocol for a Systematic Review of Empirical Studies (Preprint)," *JMIR Research Protocols*, Apr. 2023, doi: 10.2196/47971.

[25] S. Secules *et al.*, "Positionality practices and dimensions of impact on equity research: A collaborative inquiry and call to the community," *J of Engineering Edu*, vol. 110, no. 1, pp. 19–43, Jan. 2021, doi: 10.1002/jee.20377.

[26] T. Greenhalgh, G. Robert, F. Macfarlane, P. Bate, and O. Kyriakidou, "Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations," *Milbank Quarterly*, vol. 82, no. 4, pp. 581–629, Dec. 2004, doi: 10.1111/j.0887-378X.2004.00325.x.

[27] L. J. Damschroder, D. C. Aron, R. E. Keith, S. R. Kirsh, J. A. Alexander, and J. C. Lowery, "Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science," *Implementation Sci*, vol. 4, no. 1, p. 50, Dec. 2009, doi: 10.1186/1748-5908-4-50.

[28] L. J. Damschroder, C. M. Reardon, M. A. O. Widerquist, and J. Lowery, "The updated Consolidated Framework for Implementation Research based on user feedback," *Implementation Sci*, vol. 17, no. 1, p. 75, Oct. 2022, doi: 10.1186/s13012-022-01245-0.

[29] Q. N. Hong *et al.*, "The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers," *EFI*, vol. 34, no. 4, pp. 285–291, Dec. 2018, doi: 10.3233/EFI-180221.

[30] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, Jan. 2006, doi: 10.1191/1478088706qp063oa.

[31] J. Thomas and A. Harden, "Methods for the thematic synthesis of qualitative research in systematic reviews," *BMC Med Res Methodol*, vol. 8, no. 1, p. 45, Dec. 2008, doi: 10.1186/1471-2288-8-45.

[32] European Parliament. Directorate General for Parliamentary Research Services., *Artificial intelligence in healthcare: applications, risks, and ethical and societal impacts.* LU: Publications Office, 2022. Accessed: Mar. 10, 2023. [Online]. Available: https://data.europa.eu/doi/10.2861/568473

[33] A. Rajkomar, J. Dean, and I. Kohane, "Machine Learning in Medicine," *N Engl J Med*, vol. 380, no. 14, pp. 1347–1358, Apr. 2019, doi: 10.1056/NEJMra1814259.

[34] E. K. Lee *et al.*, "Transforming Hospital Emergency Department Workflow and Patient Care," *Interfaces*, vol. 45, no. 1, pp. 58–82, Feb. 2015, doi: 10.1287/inte.2014.0788.

[35] M. Hengstler, E. Enkel, and S. Duelli, "Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices," *Technological Forecasting and Social Change*, vol. 105, pp. 105–120, Apr. 2016, doi: 10.1016/j.techfore.2015.12.014.

[36] A. McCoy and R. Das, "Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and

hospital floor units," *BMJ Open Qual*, vol. 6, no. 2, p. e000158, Oct. 2017, doi: 10.1136/bmjoq-2017-000158.

[37] S. Bhattacharya, N. Sharma, and A. Singh, "Designing culturally acceptable screening for breast cancer through artificial intelligence-two case studies," *J Family Med Prim Care*, vol. 8, no. 2, pp. 760–762, Feb. 2019, doi: 10.4103/jfmpc.jfmpc_391_18.

[38] N. P. Cruz, L. Canales, J. G. Muñoz, B. Pérez, and I. Arnott, "Improving Adherence to Clinical Pathways Through Natural Language Processing on Electronic Medical Records," vol. Volume 264: MEDINFO 2019: Health and Wellbeing e-Networks for All, 2019, doi: 10.3233/SHTI190285.

[39] J. C. Ginestra *et al.*, "Clinician Perception of a Machine Learning–Based Early Warning System Designed to Predict Severe Sepsis and Septic Shock*:," *Critical Care Medicine*, vol. 47, no. 11, pp. 1477–1484, Nov. 2019, doi: 10.1097/CCM.0000000000003803.

[40] L. S. Gonçalves, M. L. D. M. Amaro, A. D. L. M. Romero, F. K. Schamne, J. L. Fressatto, and C. W. Bezerra, "Implementation of an Artificial Intelligence Algorithm for sepsis detection," *Rev. Bras. Enferm.*, vol. 73, no. 3, p. e20180421, 2020, doi: 10.1590/0034-7167-2018-0421.

[41] T. Q. Sun and R. Medaglia, "Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare," *Government Information Quarterly*, vol. 36, no. 2, pp. 368–383, Apr. 2019, doi: 10.1016/j.giq.2018.09.008.

[42] S. L. Baxter, J. S. Bass, and A. M. Sitapati, "Barriers to Implementing an Artificial Intelligence Model for Unplanned Readmissions," *ACI Open*, vol. 04, no. 02, pp. e108–e113, Jul. 2020, doi: 10.1055/s-0040-1716748.

[43] K.-J. Cho *et al.*, "Detecting Patient Deterioration Using Artificial Intelligence in a Rapid Response System:," *Critical Care Medicine*, vol. 48, no. 4, pp. e285–e289, Apr. 2020, doi: 10.1097/CCM.0000000000004236.

[44] E. Frontoni *et al.*, "A Decision Support System for Diabetes Chronic Care Models Based on General Practitioner Engagement and EHR Data Sharing," *IEEE J. Transl. Eng. Health Med.*, vol. 8, pp. 1–12, 2020, doi: 10.1109/JTEHM.2020.3031107.

[45] A. E. Hassan, V. M. Ringheanu, R. R. Rabah, L. Preston, W. G. Tekle, and A. I. Qureshi, "Early experience utilizing artificial intelligence shows significant reduction in transfer times and length of stay in a hub and spoke model," *Interv Neuroradiol*, vol. 26, no. 5, pp. 615–622, Oct. 2020, doi: 10.1177/1591019920953055.

[46] S. Romero-Brufau, K. D. Wyatt, P. Boyum, M. Mickelson, M. Moore, and C. Cognetta-Rieke, "A lesson in implementation: A pre-post study of providers' experience with artificial intelligence-based clinical decision support," *International Journal of Medical Informatics*, vol. 137, p. 104072, May 2020, doi: 10.1016/j.ijmedinf.2019.104072.

[47] S. Sandhu *et al.*, "Integrating a Machine Learning System Into Clinical Workflows: Qualitative Study," *Journal of Medical Internet Research*, vol. 22, no. 11, p. e22421, Nov. 2020, doi: 10.2196/22421.

[48] M. P. Sendak *et al.*, "Real-World Integration of a Sepsis Deep Learning Technology Into Routine Clinical Care: Implementation Study," *JMIR Med Inform*, vol. 8, no. 7, p. e15182, Jul. 2020, doi: 10.2196/15182.

[49] L. Strohm, C. Hehakaya, E. R. Ranschaert, W. P. C. Boon, and E. H. M. Moors, "Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors," *Eur Radiol*, vol. 30, no. 10, pp. 5525–5532, Oct. 2020, doi: 10.1007/s00330-020-06946-y.

[50] H. Xu *et al.*, "Construction and Application of a Medical-Grade Wireless Monitoring System for Physiological Signals at General Wards," *J Med Syst*, vol. 44, no. 10, p. 182, Oct. 2020, doi: 10.1007/s10916-020-01653-z.

[51] S. Jauk, D. Kramer, A. Avian, A. Berghold, W. Leodolter, and S. Schulz, "Technology Acceptance of a Machine Learning Algorithm Predicting Delirium in a Clinical Setting: a

Mixed-Methods Study," *J Med Syst*, vol. 45, no. 4, p. 48, Apr. 2021, doi: 10.1007/s10916-021-01727-6.

[52] H. M. P. Morales, M. Guedes, J. S. Silva, and A. Massuda, "COVID-19 in Brazil—Preliminary Analysis of Response Supported by Artificial Intelligence in Municipalities," *Frontiers in Digital Health*, vol. 3, 2021, Accessed: Apr. 06, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fdgth.2021.648585

[53] D. H. Murphree *et al.*, "Improving the delivery of palliative care through predictive modeling and healthcare informatics," *Journal of the American Medical Informatics Association*, vol. 28, no. 6, pp. 1065–1073, Jun. 2021, doi: 10.1093/jamia/ocaa211.

[54] J.-C. Yao *et al.*, "AI detection of mild COVID-19 pneumonia from chest CT scans," *Eur Radiol*, vol. 31, no. 9, pp. 7192–7201, Sep. 2021, doi: 10.1007/s00330-021-07797-x.

[55] M. A. Davis, B. Rao, P. A. Cedeno, A. Saha, and V. M. Zohrabian, "Machine Learning and Improved Quality Metrics in Acute Intracranial Hemorrhage by Noncontrast Computed Tomography," *Current Problems in Diagnostic Radiology*, vol. 51, no. 4, pp. 556–561, Jul. 2022, doi: 10.1067/j.cpradiol.2020.10.007.

[56] K. E. Henry *et al.*, "Human–machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system," *npj Digit. Med.*, vol. 5, no. 1, p. 97, Jul. 2022, doi: 10.1038/s41746-022-00597-7.

[57] M. Joshi, K. Mecklai, R. Rozenblum, and L. Samal, "Implementation approaches and barriers for rule-based and machine learning-based sepsis risk prediction tools: a qualitative study," *JAMIA Open*, vol. 5, no. 2, p. ooac022, Apr. 2022, doi: 10.1093/jamiaopen/ooac022.

[58] S. Lebovitz, H. Lifshitz-Assaf, and N. Levina, "To Engage or Not to Engage with AI for Critical Judgments: How Professionals Deal with Opacity When Using AI for Medical Diagnosis," *Organization Science*, vol. 33, no. 1, pp. 126–148, Jan. 2022, doi: 10.1287/orsc.2021.1549.

[59] D. R. Rushlow *et al.*, "Clinician Adoption of an Artificial Intelligence Algorithm to Detect Left Ventricular Systolic Dysfunction in Primary Care.," *Mayo Clinic Proceedings*, vol. 97, no. 11, pp. 2076–2085, Nov. 2022, doi: 10.1016/j.mayocp.2022.04.008.

[60] J. M. Schwartz *et al.*, "Factors Influencing Clinician Trust in Predictive Clinical Decision Support Systems for In-Hospital Deterioration: Qualitative Descriptive Study," *JMIR Human Factors*, vol. 9, no. 2, p. e33960, May 2022, doi: 10.2196/33960.

[61] M. Sibbald, B. Abdulla, A. Keuhl, G. Norman, S. Monteiro, and J. Sherbino, "Electronic Diagnostic Support in Emergency Physician Triage: Qualitative Study With Thematic Analysis of Interviews," *JMIR Human Factors*, vol. 9, no. 3, p. e39234, Sep. 2022, doi: 10.2196/39234.

[62] S. J. Singer, K. C. Kellogg, A. B. Galper, and D. Viola, "Enhancing the value to users of machine learning-based clinical decision support tools: A framework for iterative, collaborative development and implementation," *Health Care Manage Rev*, vol. 47, no. 2, pp. E21–E31, Apr. 2022, doi: 10.1097/HMR.0000000000000324.

[63] F. Wijnhoven, "Organizational Learning for Intelligence Amplification Adoption: Lessons from a Clinical Decision Support System Adoption Project," *Inf Syst Front*, vol. 24, no. 3, pp. 731–744, Jun. 2022, doi: 10.1007/s10796-021-10206-9.

[64] Y. Zhai, Z. Yu, Q. Zhang, and Y. Zhang, "Barriers and facilitators to implementing a nursing clinical decision support system in a tertiary hospital setting: A qualitative study using the FITT framework," *International Journal of Medical Informatics*, vol. 166, p. 104841, Oct. 2022, doi: 10.1016/j.ijmedinf.2022.104841.

[65] C. Pou-Prom, J. Murray, S. Kuzulugil, M. Mamdani, and A. A. Verma, "From compute to care: Lessons learned from deploying an early warning system into clinical practice," *Front. Digit. Health*, vol. 4, p. 932123, Sep. 2022, doi: 10.3389/fdgth.2022.932123.

[66] J. S. Hinson *et al.*, "Multisite implementation of a workflow-integrated machine learning system to optimize COVID-19 hospital admission decisions," *npj Digit. Med.*, vol. 5, no. 1, p. 94, Jul. 2022, doi: 10.1038/s41746-022-00646-1.

[67] G. T. Berge, O. C. Granmo, T. O. Tveit, B. E. Munkvold, A. L. Ruthjersen, and J. Sharma, "Machine learning-driven clinical decision support system for concept-based searching: a field trial in a Norwegian hospital," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, p. 5, Jan. 2023, doi: 10.1186/s12911-023-02101-x.

[68] T. C. Lee, N. U. Shah, A. Haack, and S. L. Baxter, "Clinical Implementation of Predictive Models Embedded within Electronic Health Record Systems: A Systematic Review," *Informatics*, vol. 7, no. 3, p. 25, Jul. 2020, doi: 10.3390/informatics7030025.

[69] A. M. Polhemus *et al.*, "Human-Centered Design Strategies for Device Selection in mHealth Programs: Development of a Novel Framework and Case Study," *JMIR Mhealth Uhealth*, vol. 8, no. 5, p. e16043, May 2020, doi: 10.2196/16043.

[70] E. M. Rogers, *Diffusion of innovations*. Simon and Schuster, 2003.

[71] M. Sharma, C. Savage, M. Nair, I. Larsson, P. Svedberg, and J. M. Nygren, "Artificial Intelligence Applications in Health Care Practice: Scoping Review," *J Med Internet Res*, vol. 24, no. 10, p. e40238, Oct. 2022, doi: 10.2196/40238.

[72] J. Yin, K. Y. Ngiam, and H. H. Teo, "Role of Artificial Intelligence Applications in Real-Life Clinical Practice: Systematic Review," *J Med Internet Res*, vol. 23, no. 4, p. e25759, Apr. 2021, doi: 10.2196/25759.

[73] K. E. Rudd *et al.*, "Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study," *The Lancet*, vol. 395, no. 10219, pp. 200–211, Jan. 2020, doi: 10.1016/S0140-6736(19)32989-7.

[74] Md. M. Islam, T. Nasrin, B. A. Walther, C.-C. Wu, H.-C. Yang, and Y.-C. Li, "Prediction of sepsis patients using machine learning approach: A meta-analysis," *Computer Methods and Programs in Biomedicine*, vol. 170, pp. 1–9, Mar. 2019, doi: 10.1016/j.cmpb.2018.12.027.

[75] A. H. Van Der Vegt, I. A. Scott, K. Dermawan, R. J. Schnetler, V. R. Kalke, and P. J. Lane, "Deployment of machine learning algorithms to predict sepsis: systematic review and application of the SALIENT clinical AI implementation framework," *Journal of the American Medical Informatics Association*, vol. 30, no. 7, pp. 1349–1361, Jun. 2023, doi: 10.1093/jamia/ocad075.

[76] K. Kolasa, B. Admassu, M. Hołownia-Voloskova, K. J. Kędzior, J.-E. Poirrier, and S. Perni, "Systematic reviews of machine learning in healthcare: a literature review," *Expert Review of Pharmacoeconomics & Outcomes Research*, pp. 1–53, Nov. 2023, doi: 10.1080/14737167.2023.2279107.

[77] L. Petersson *et al.*, "Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden," *BMC Health Serv Res*, vol. 22, no. 1, p. 850, Dec. 2022, doi: 10.1186/s12913-022-08215-8.

[78] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nat Med*, vol. 25, no. 1, pp. 44–56, Jan. 2019, doi: 10.1038/s41591-018-0300-7.

[79] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," *BMC Med*, vol. 17, no. 1, p. 195, Dec. 2019, doi: 10.1186/s12916-019-1426-2.

[80] A. Adjekum, A. Blasimme, and E. Vayena, "Elements of Trust in Digital Health Systems: Scoping Review," *J Med Internet Res*, vol. 20, no. 12, p. e11254, Dec. 2018, doi: 10.2196/11254.

[81] E. Glikson and A. W. Woolley, "Human Trust in Artificial Intelligence: Review of Empirical Research," *ANNALS*, vol. 14, no. 2, pp. 627–660, Jul. 2020, doi: 10.5465/annals.2018.0057.

[82] D. E. Ehrmann, S. Joshi, S. D. Goodfellow, M. L. Mazwi, and D. Eytan, "Making machine learning matter to clinicians: model actionability in medical decision-making," *npj Digit. Med.*, vol. 6, no. 1, p. 7, Jan. 2023, doi: 10.1038/s41746-023-00753-7.

[83] S. O'Connor, P. Hanlon, C. A. O'Donnell, S. Garcia, J. Glanville, and F. S. Mair, "Understanding factors affecting patient and public engagement and recruitment to digital health

interventions: a systematic review of qualitative studies," *BMC Med Inform Decis Mak*, vol. 16, no. 1, p. 120, Dec. 2016, doi: 10.1186/s12911-016-0359-3.

[84] B. Mesko and Z. Győrffy, "The Rise of the Empowered Physician in the Digital Health Era: Viewpoint," *J Med Internet Res*, vol. 21, no. 3, p. e12490, Mar. 2019, doi: 10.2196/12490.