Review

Application of Machine Learning Algorithms to Health Systems Planning and Public Policy Using Real-World Data: Scoping Review

Deusa Maria Arruda Rocha[1,2], MPH; Gentil Veloso Barbosa[2], PhD; Paulo Henrique De Souza Bermejo[3], PhD

[1]Department of Regulation, Evaluation, and Control (DRAC), Secretariat of Specialized Healthcare (SAES), Ministério da Saúde, Brasília, Federal District, Brazil

[2]Professional Postgraduate Program in Governance and Digital Transformation, Federal University of Tocantins, Palmas, Tocantins, Brazil

[3]Sinerji Institute for Technological and Business Research and Development, Brasília, Federal District, Brazil

Corresponding Author:

Deusa Maria Arruda Rocha, MPH
Department of Regulation, Evaluation, and Control (DRAC)
Secretariat of Specialized Healthcare (SAES), Ministério da Saúde
Esplanada dos Ministérios, Bloco G, Edifício Anexo
Brasília, Federal District 70058-900
Brazil
Phone: +55 61 98295-2606
Email: deusa.rocha@saude.gov.br

Abstract

Background: The rapid advancement of information technology and machine learning (ML) has garnered significant interest from health care managers and policymakers. However, incorporating reliable and transparent predictions into health systems remains challenging, necessitating an investigation into gaps in digital health interventions within real-world settings characterized by resource constraints and reliability issues.

Objective: This study aimed to map and synthesize the available evidence on the application of ML algorithms to real-world data in health systems and identify their operational targets and clinical implications for health planning and public policy formulation.

Methods: Following the Joanna Briggs Institute (JBI) methodology for scoping reviews and PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines, we searched the PubMed, Embase, Scopus, and LILACS databases for empirical studies published between 2016 and 2025 that evaluated the application of ML to health management, surgical queues, and public policy using real-world data. The study protocol was retrospectively registered with the OSF.

Results: Of the studies identified through the database searches (2016-2025), only 6 empirical studies met all inclusion criteria, revealing that the real-world deployment of ML in health management and public policy remains at an early stage. The included models addressed micro-operational tasks, such as surgical blood transfusion forecasting, early warning system mortality prediction, postsurgical discharge planning, laboratory error detection, patient-reported outcome measure (PROM) prediction, and emergency care classification. Although high discriminatory performance was reported (area under the receiver operating characteristic curve [AUROC] 0.76-0.99), the studies were limited by single-center retrospective designs, a lack of independent external validation, and inadequate reporting of model calibration.

Conclusions: Current evidence on ML for macrolevel health planning and public policy formulation is limited and micro-operationally focused. Translating these algorithms into public health governance (eg, waiting list regulation in universal health systems such as the Sistema Único de Saúde [SUS]) requires moving beyond "black-box" predictions toward explainable AI (XAI), natural language processing (NLP) to analyze unstructured data, transparent model calibration, and rigorous evaluations of health equity.

Interact J Med Res 2026;15:e96779; doi: 10.2196/96779

Keywords: artificial intelligence; AI; machine learning; digital health; health management; waiting lists; elective surgical procedures; public health systems; Brazil's unified health system; Sistema Único de Saúde; SUS

Background

The emergence of AI as a tool for enhancing health care offers unprecedented opportunities to optimize clinical outcomes, reduce costs, and positively impact population health [1]. Among its subfields, machine learning (ML) stands out for its use of algorithmic structures and statistical modeling to identify patterns in real-world data (RWD), encompassing clinical, behavioral, and demographic variables. This facilitates the establishment of functional relationships that significantly improve predictive accuracy [2-4]. AI-driven interventions provide robust decision support for medical professionals and create new opportunities for therapeutic monitoring, particularly in specialized care [5,6].

This technological evolution is part of a data-driven digital revolution, marking the transition from Industry 4.0 to Industry 5.0. In the health care sector, this phase —known as "Health 5.0" (H5.0)—explores technological potential to reshape service delivery, promoting personalized medicine and, most importantly, patient-centered care [2,7-10]. Interdisciplinary interest in these technologies grows alongside the global need to optimize health care spending, which accounted for approximately 11% of global gross domestic product (GDP) in 2023. This requires IT infrastructures and algorithms capable of generating clinically reliable and actionable predictions [11,12].

However, despite the exponential growth of ML research, a gap persists in both the scientific literature and government agendas regarding the effective integration of these techniques into service delivery and public policy [2,6]. This lack of standardized guidelines leaves many nations in the early stages of technological adoption [13].

A critical example is the management of waiting lists for elective surgeries, which represents a growing challenge because of population aging and the increased demand for complex procedures [3]. In this scenario, it is imperative for managers to adopt ML models to streamline workflows, ensure patient safety, and expedite care delivery. As noted by Lungu et al [14], waiting times are a central public concern and a priority for health policymakers. Yet, planning effective interventions to address waiting times remains challenging because of demographic shifts and increasing public confidence in surgical treatments.

Consequently, there is an urgent need for research addressing the intervention gap regarding the application of these tools in real-world settings. A critical limitation of the current literature is the predominance of idealized, laboratory-based simulations that fail to reflect the operational friction, missing data, and resource constraints inherent to day-to-day health system management. Demonstrating the true utility of ML in clinical practice and health governance requires evaluating models deployed within these complex real-world environments [2,15].

To address these gaps, this scoping review was guided by the following research questions:

1. How are ML algorithms currently being applied in real-world health care settings to support health systems planning and public policy formulation?
2. What are the main methodological characteristics, predictive performance measures, and clinical or operational targets of these deployed ML models?
3. What are the key gaps, methodological limitations, and barriers (eg, model calibration, external validation, and explainability) that prevent the translation of micro-operational ML models into macrolevel public health policies?

Objective

Given the need for evidence that transcends theoretical models and addresses the complexity of health management, this scoping review aims to map and synthesize the application of ML algorithms to RWD. Specifically, we aim to identify how these tools are used in health systems planning, elective surgical flows, and public policy formulation, while characterizing the methodological landscape and identifying the limitations of the current literature.

Methods

Overview

This scoping review was conducted in accordance with the Joanna Briggs Institute (JBI) methodology for scoping reviews [16] and reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines [17]. The completed PRISMA-ScR checklist is provided in Checklist 1. The study focused on pragmatic, nonpharmacological health interventions incorporating ML tools and algorithms to address gaps in clinical and public health contexts.

Protocol Registration and Open Science Transparency

The protocol for this scoping review was registered retrospectively on OSF (ucefv) [18]. Search syntaxes and documentation for both the initial search and the 2025 update were deposited and are publicly accessible via searchRxiv [19-22] and OSF repositories, ensuring full auditability and adherence to open science practices and PRISMA-ScR standards.

Search Strategy

Overview

The search syntax and definition of the search time frame were built and expanded upon the methodological benchmark framework established by Triantafyllidis and Tsanas [2] for ML in digital health. To capture the recent decade's acceleration in the development of health management algorithms

while maintaining methodological continuity, the search time frame was defined from 2016 through 2025. The search was also expanded to include LILACS to cover regional public health systems.

The systematic search was conducted in 2 distinct phases: an initial search in August 2020 and a comprehensive update in June 2025.

2020 Search Phase

The initial search strategy was developed by an experienced health librarian and structured as follows: (1) search string: "machine learning" AND "patient monitoring;" (2) period: 2016 to 2020; and (3) databases: MEDLINE (via PubMed), Embase, and Scopus. Specific details and results for this phase (n=333) are available in Table 1 and Multimedia Appendix 1.

Table 1. Quality assessment of included studies using the Effective Public Health Practice Project (EPHPP) quality assessment tool[a].

Studies	Selection bias	Study design	Confusion factors	Blinding	Data collection methods	Withdrawals and dropouts	Global assessment
Campagner et al [23], 2020	Moderate	Moderate	Weak	Weak	Moderate	Weak	Weak
Ye et al [24], 2019	Moderate	Weak	Moderate	Weak	Moderate	Weak	Weak
Fontana et al [25], 2019	Moderate	Weak	Moderate	Weak	Strong	Weak	Moderate
Karhade et al [26], 2018	Weak	Weak	Weak	Weak	Strong	Weak	Weak
Hayn et al [27], 2017	Weak	Moderate	Weak	Weak	Strong	Weak	Weak
Krämer et al [28], 2017	Weak	Weak	Weak	Weak	Strong	Weak	Weak

[a]This table was adapted from the study by Triantafyllidis and Tsanas [2]. Ratings are based on the EPHPP tool, where "strong" indicates no weak ratings, "moderate" indicates 1 weak rating, and "weak" indicates 2 or more weak ratings in the specific domains.

2025 Update Phase

The 2025 update expanded the scope to capture the latest advancements. The search strategy, developed by a health librarian, used the following 2 main conceptual blocks:

- Block 1 (AI and ML) included the MeSH term "Machine Learning" and keywords such as "artificial intelligence," "deep learning," "neural network," and "predictive analytics."
- Block 2 (health management) focused on organizational aspects and included MeSH terms such as "Health Planning" and "Organizational Policy," combined with keywords such as "integrated care" and "health systems."

This phase was conducted in June 2025 across 4 databases: MEDLINE (via PubMed), Embase, Scopus, and LILACS. The expanded strings are available via searchRxiv [19-22] and are presented in Table 2 and Multimedia Appendix 2.

Table 2. Characteristics and predictive performance of the included studies on machine learning applications in health care management[a].

Items	Campagner et al [23], 2020 (Italy)	Ye et al [24], 2019 (United States)	Fontana et al [25], 2019 (United States)	Karhade et al [26], 2018 (United States)	Hayn et al [27], 2017 (Austria)	Krämer et al [28], 2019 (Germany)
Application (objective)	Predict surgical invasiveness and recommend optimal procedures	Validate an ML[b]-based early warning system for in-hospital mortality	Predict the risk of not achieving MCID[c] in PROMs[d] after total knee arthroplasty	Predict nonroutine discharge after lumbar surgery to reduce hospital stay duration	Predict red blood cell transfusion volume for PBM[e]	Predict emergency department vs elective care urgency
ML algorithms	Random forest (best-performing model), ordinal regression, SVM[f], LASSO[g], KNN[h], and decision trees	Random forest (best-performing model), XGBoost[i], gradient boosting, SVM, LASSO, and KNN	Logistic LASSO, random forest, and linear SVM	ANN[j] (best-performing model), SVM, BPM[k], and decision tree; random forest used for variable selection	Random forest, multivariate logistic regression, linear models, and the Mercuriali algorithm	Random forest and KNN (for data imputation)
Key results (performance)	AUC[l] of 0.87 for the invasiveness score	Strong discrimination (AUC 0.884); mean	Fair-to-good ability to predict	ANN outperformed the other models	Best-performing algorithm showed a	96% accuracy and excellent

Items	Campagner et al [23], 2020 (Italy)	Ye et al [24], 2019 (United States)	Fontana et al [25], 2019 (United States)	Karhade et al [26], 2018 (United States)	Hayn et al [27], 2017 (Austria)	Krämer et al [28], 2019 (Germany)
	and 0.76 for procedure invasiveness	lead time of 40.8 hours	postsurgical MCID achievement	(PPV[m] 33%; sensitivity 54%)	moderate correlation (correlation coefficient=0.61)	discrimination (AUROC[n]>0.99)
Statistical significance and evaluation metrics	Statistically significant (*P*=.04)	High AUC, indicating robust model validity	AUROC used as the primary evaluation metric	Superior AUROC, calibration, and Brier score	High statistical significance (*P*<.001)	High precision and AUROC demonstrate validity
Clinical implications	High physician acceptance and consistent predictive variables	Improves patient safety through early detection of clinical deterioration	Focus on MCID strengthens practical clinical applicability	Open-access tool available to support replicability and validation	Optimize PBM; applicable to various medical fields	Improves patient flow management in emergency services

[a]This table was adapted from the study by Triantafyllidis and Tsanas [2].
[b]ML: machine learning.
[c]MCID: minimal clinically important difference.
[d]PROM: patient-reported outcome measure.
[e]PBM: patient blood management.
[f]SVM: support vector machine.
[g]LASSO: least absolute shrinkage and selection operator.
[h]KNN: k-nearest neighbors.
[i]XGBoost: extreme gradient boosting.
[j]ANN: artificial neural network.
[k]BPM: Bayes point machine.
[l]AUC: area under the curve.
[m]PPV: positive predictive value.
[n]AUROC: area under the receiver operating characteristic curve.

Eligibility Criteria

To capture the full spectrum of relevant literature, our search strategy encompassed 4 major databases (PubMed, Embase, Scopus, and LILACS) and covered the period from 2016 through 2025. The inclusion criteria required empirical studies that applied ML algorithms specifically to health services management, care regulation, surgical queue prioritization, or public policy formulation using RWD.

The inclusion and exclusion criteria were operationalized to exclude broad, nonapplied research. Studies were excluded if they (1) evaluated ML exclusively in silico or using synthetic or benchmark datasets, without validation against real-world health system records; (2) focused exclusively on clinical diagnostic accuracy (eg, radiological image classification or genomic sequencing) without linking algorithmic outputs to operational planning, bed allocation, or administrative decision-making; (3) consisted of conceptual frameworks, reviews, editorials, or conference abstracts without full empirical implementation data; or (4) examined closed commercial software without transparent algorithmic reporting.

Study Selection and Screening

The selection process was conducted as follows: for 2020 records, a total of 333 were identified and imported into EndNote Basic (Clarivate Analytics) [29], of which 4 duplicate records were removed. For 2025 records, a total of 611 were identified and imported into the Rayyan (Rayyan Systems Inc) web application (software as a service) [30], of which 14 duplicate records were removed.

Consolidation and Final Selection

A total of 944 records were identified across both phases. After removing all duplicate records (n=18), 926 records underwent independent title and abstract screening. Of these, 891 were excluded. For the remaining 35 reports, full-text retrieval was successfully achieved (100%). During the eligibility assessment, 29 reports were excluded (eg, because they were reviews, lacked technical detail, or inappropriate publication date). Ultimately, 6 studies met all the criteria and were included. The complete process is illustrated in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.

Data Extraction and Methodological Quality Assessment

Data extraction was performed using a standardized charting form inspired by the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) statement [31], focusing on key predictive performance metrics such as area under the receiver operating characteristic curve (AUROC), sensitivity, and specificity. Methodological quality was independently assessed using the Effective Public Health Practice Project (EPHPP) quality assessment tool [32,33] by a multidisciplinary team of specialists in computer science, dentistry, psychology, and statistics.

Data Synthesis and Mapping

The data synthesis and mapping process followed a structured approach. Detailed information was extracted regarding the study objectives, methodological designs, sample sizes, and follow-up durations. Additionally, primary outcomes,

reported levels of statistical significance, specific ML algorithms used, and their practical implications for clinical practice and public policy formulation were systematically mapped.

Ethical Considerations

As this study is a scoping review of previously published literature and publicly available data, it did not involve direct interaction with human participants or the collection of identifiable private information. Consequently, formal institutional review board approval was not required. The research was conducted in accordance with the PRISMA-ScR guidelines to ensure transparency, reproducibility, and ethical scientific conduct (Checklist 1).

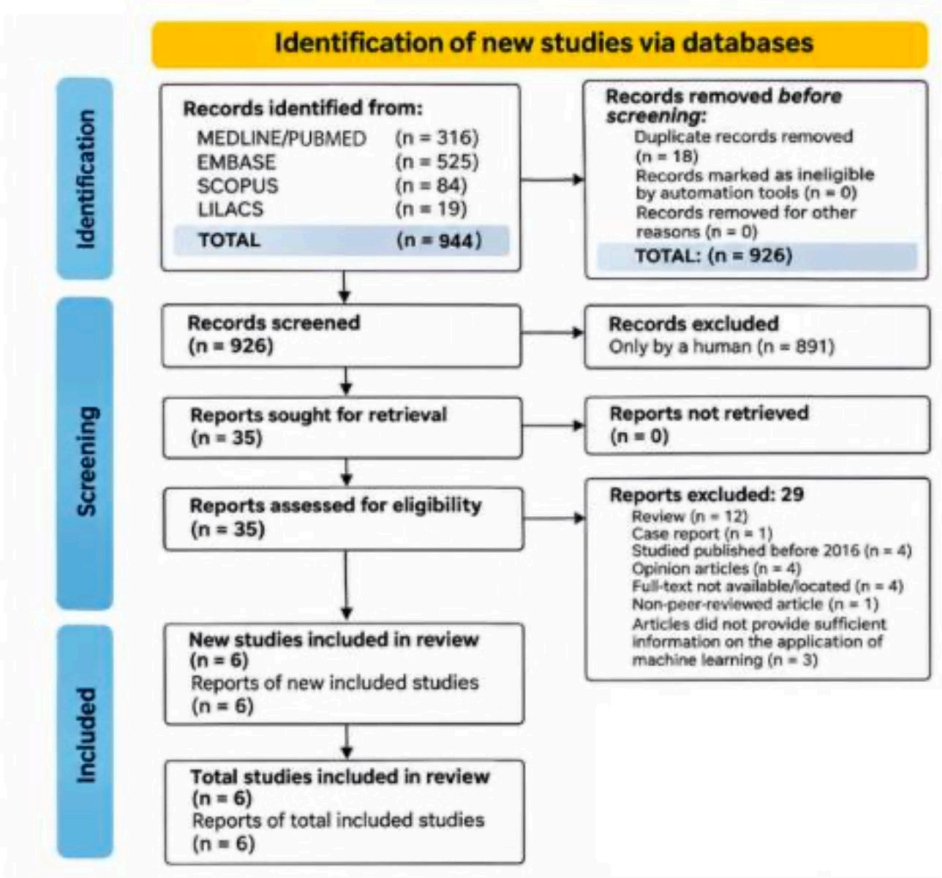
Results

Literature Search

The systematic search, conducted in 2 phases (2020 and 2025), yielded a total of 944 records across the MEDLINE (via PubMed), Embase, Scopus, and LILACS databases. After removing 18 duplicate records, 926 records underwent an initial title and abstract screening. Of these, 35 articles were selected for full-text review, with a 100% retrieval success rate.

Following the eligibility assessment, 29 reports were excluded for specific reasons (eg, reviews, case reports, or lack of technical detail regarding ML implementation). Ultimately, 6 studies met all the criteria and were included in the review. These studies originated from the United States (n=3), Austria (n=1), Germany (n=1), and Italy (n=1). The complete selection process is illustrated in the PRISMA flow diagram (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram for updated systematic reviews, which included searches of databases. Adapted from the study by Page et al [34], reflecting the structure of the PRISMA 2020 flow diagram adapted for this study's database search.



Quality Assessment

Study quality was assessed using the EPHPP tool. It was observed that 83% (5/6) [23,24,26-28] of the studies presented "weak" methodological quality, while only 17% (n=1) [25] achieved a "moderate" rating. The primary weaknesses identified were selection bias and a lack of detailed information regarding participant recruitment. Regarding study design, observational cohorts prevailed (5/6, 83%) [30-34], with only 1 study characterized as a non-

randomized experimental design [23]. These findings are detailed in Table 1.

Intervention Type and Target Population

The samples varied significantly, using sources ranging from retrospective electronic health records to administrative insurance billing data. Interventions focused on patients undergoing elective surgeries (spine, hip, and knee), emergency care, red blood cell transfusion volume management, and urgency or emergency triage classification. The average follow-up duration across the studies was approximately 5 years.

ML Applications and Clinical Outcomes

The identified ML applications were predominantly directed toward the development of predictive models and clinical decision support systems (CDSS). As synthesized in Table 2, key application objectives encompassed risk prediction, such as modeling in-hospital mortality and nonroutine discharge following lumbar surgeries, and surgical planning, which focused on evaluating invasiveness levels and estimating required red blood cell transfusion volumes. Furthermore, technologies were implemented in triage and diagnosis, including emergency department urgency classification and predicting minimal clinically important differences (MCID) following joint arthroplasties. Algorithms used across studies included artificial neural networks (ANNs), random forests (RF), support vector machines (SVMs), and other predictive prognostic models. Although 33% of the studies reported high statistical significance (P<.001), a notable lack of metric standardization was evident. Regarding performance, the models demonstrated robust results, with area under the curve (AUC) values ranging from 0.76 to 0.99, indicating excellent discriminatory performance. These clinical implications highlight the expanding role of AI tools as modalities complementary to medical judgment, fostering data-driven and evidence-based practice.

Discussion

Principal Findings

A critical reflection on the yield of this scoping review reveals that identifying only 6 empirical studies meeting all the strict inclusion criteria is highly informative. Rather than indicating a limitation of the search strategy, this restricted yield directly reflects an emerging evidence base in digital health. Although the theoretical literature on ML in health care is vast, true real-world implementation of ML for macrolevel health planning and public policy remains in its infancy. Most existing algorithms remain confined to laboratory simulations or isolated micro-operational tasks without integrating administrative, regulatory, or policy end points. The scarcity of published empirical deployments highlights a current structural gap between technical algorithmic development and its actual integration into public health governance, such as national waiting list regulation in universal health systems such as Brazil's Sistema Único de Saúde (SUS). Recognizing this field as an emerging area of research underscores the imperative to shift future research frameworks toward implementation science, model generalizability, and transparent calibration [6,35,36].

The empirical evidence reveals that ML use in health planning is predominantly anchored in micro-operational, clinical, and process management decision support. In elective surgical settings, models were used to optimize resource allocation and anticipate operational bottlenecks. Hayn et al [27] demonstrated that ML significantly outperformed routine practice in predicting surgical blood transfusion volumes for inventory management, while Karhade et al [26] showed that preoperative identification of nonroutine discharges mitigates financial waste and bed turnover. For long-term outcomes, Fontana et al [25] applied logistic least absolute shrinkage and selection operator (LASSO) regression (AUROC 0.60-0.89) to predict the preoperative achievement of the MCID in patient-reported outcome measures (PROMs), supporting shared decision-making. In diagnostic and systemic regulation, Campagner et al [23] applied RF models (AUROC 0.76-0.87) to administrative and laboratory data to identify sample identification errors, potentially reducing unnecessary retesting and operational costs. At the macropolicy level, Krämer et al [28] used an RF (accuracy=96%; AUROC>0.99) with data from Germany's national inpatient database (which covered 99.84% of inpatient cases) to classify emergency and elective admissions. They revealed that emergency department growth was concentrated in low-urgency cases driven by hospital reimbursement incentives (diagnosis-related groups), proving ML's utility for regulatory auditing and financial incentive design. Additionally, Ye et al [24] investigated in-hospital mortality prediction to support early warning systems for acute patient flow management.

These micro-operational ML outputs aggregate directly into 3 primary macrogovernance mechanisms: surgical queue regulation, by aggregating patient-level discharge predictions to optimize bed turnover; supply chain planning, by forecasting regional blood component demand (Hayn et al [27]); and emergency care streamlining, by stabilizing patient flow management through automated error detection and risk triage (Krämer et al [28]). To align with evidence-informed policy frameworks, ML deployment in universal health systems must address social determinants of health (SDOH) and health equity [37]. As shown by Fontana et al [25], socioeconomic variables influence predictive outcomes, posing risks that algorithms might unfairly prioritize patients with higher baseline probabilities of success over socially vulnerable individuals. Furthermore, data integrity at the source, including within systems such as e-SUS primary health care (PHC) [38], must be prioritized to avoid automation bias (Campagner et al [23]). Without continuous bias auditing, regulatory impact assessments, and robust data governance, including compliance with the General Data Protection Regulation (GDPR) [39] or Lei Geral de Proteção de Dados Pessoais (LGPD) [40], predictive algorithms risk amplifying historical disparities in access to care.

Limitations

Methodological Vulnerabilities and Performance Limits

Interpreting these findings in the context of public policy requires caution. A TRIPOD-aligned appraisal revealed high discriminatory performance (AUROC 0.76-0.99) but a pervasive omission of formal model calibration measures, such as calibration slopes, intercepts [35], or decision curve analysis. The upper bound (AUROC 0.99 in Krämer et al [28]) reflects a narrow binary triage task, whereas complex health management tasks exhibited more modest discrimination (AUROC 0.76-0.84 in Karhade et al [26]). In public health systems such as SUS, deploying miscalibrated models introduces severe risks of inappropriate queue prioritization or resource misallocation. Furthermore, 83% (5/6) of the studies exhibited overall weak methodological quality under the EPHPP tool [32,33], primarily due to high selection bias in single-center retrospective samples and unadjusted confounding. Only Campagner et al [23] achieved a moderate rating through internal validation, although still limited to a single institution. Current evidence thus represents preliminary proof-of-concept evidence of feasibility rather than evidence supporting generalizable decision-making. Scalability requires prospective multicenter validation, adherence to TRIPOD-AI [41] and PROBAST (Prediction Model Risk of Bias Assessment Tool)–AI guidelines [42], and compliance with FAIR (findable, accessible, interoperable, and reusable) data-sharing principles [43].

Implementation Barriers Across the Nonadoption, Abandonment, and Scale-up, Spread, and Sustainability Framework

Analyzing implementation barriers through the nonadoption, abandonment, and scale-up, spread, and sustainability (NASSS) framework [6] shows that ML operates primarily as a decision support technology rather than an autonomous system. Regarding the technology and value proposition [44,45], the lack of external validation and calibration transparency [35] restricts technical trust. Small sample sizes relative to high feature dimensionality and single-institution coding biases further compromise transferability. Concerning adopters, the "black-box" nature [46] of complex models—such as neural networks in Karhade et al [26] or RF in Ye et al [24], Campagner et al [23], and Krämer et al [28]—fosters clinical skepticism. Multicollinearity further obscures coefficient interpretation (Fontana et al [25]), underscoring the need for explainable AI (XAI) tools [47,48] such as SHAP [49] (Shapley additive explanations) or LIME (local interpretable model-agnostic explanations), at the point of care. At the organizational level, data fragmentation across electronic health records and administrative systems hinders seamless integration. Furthermore, continuous real-time monitoring faces an "intervention paradox," as described by Ye et al [24] and Campagner et al [23], wherein successful clinical interventions avert adverse outcomes, thereby dynamically altering the model's performance metrics into statistical false positives.

Future Directions

Bridging the current implementation gap requires 2 key future research priorities. First, adopting XAI [47,48] is critical to move beyond global discrimination metrics such as AUROC toward interpretable feature attributions, thereby ensuring ethical accountability and clinical trust. Second, integrating natural language processing (NLP) [50] is necessary to overcome the limitation of current studies, which relied exclusively on structured tabular data while ignoring rich clinical narratives. Deploying domain-adapted NLP architectures, such as ClinicalBERT, enables the automated parsing of unstructured free-text records, including referral letters, emergency intake notes, and cancellation logs in regulation systems such as SisCNRAC. Structuring unstructured text converts administrative logs into actionable intelligence, optimizing waitlist regulation, syndromic surveillance, and resource allocation across public health networks.

Conclusions

This scoping review synthesizes the empirical baseline of ML applications deployed for health system management and public policy execution. The evidence demonstrates that real-world applications remain in an incipient phase, operating predominantly as micro-operational prediction tools for surgical queue optimization, emergency triage, postdischarge destination planning, and blood component supply management. Although these operational algorithms offer preliminary feasibility for local resource optimization, the predominance of single-center retrospective cohorts and overall weak methodological quality—as appraised via EPHPP—prevent immediate, uncritical scalability to macrolevel health governance. Bridging the gap between micro-operational predictions and systemic health policy requires transitioning to prospective multicenter validations, adopting transparent reporting guidelines (TRIPOD-AI and PROBAST-AI), enforcing strict ethical data governance, and integrating domain-specific NLP to structure real-world health records.

Acknowledgments

The authors express their gratitude to the Secretariat of Specialized Healthcare (SAES) and the Department of Healthcare Regulation, Evaluation and Control (DRAC) for their institutional support, with special thanks to Juliana F Lujan and Débora Spalding Verdi for providing the necessary resources and support for this research. We are grateful to Departamento de Informática do Sistema Único de Saúde, Secretaria de Informação e Saúde Digital, and Ministério da Saúde for the

development and support of the National High Complexity Regulation Center Information System (SisCNRAC). We also thank the Professional Doctorate Program in Governance and Digital Transformation at the Federal University of Tocantins (Universidade Federal do Tocantins), in partnership with the University of Brasília. Finally, the authors thank the Library of the Ministry of Health (Biblioteca Virtual em Saúde do Ministério da Saúde) and the National School of Public Health (Escola Nacional de Saúde Pública/Fiocruz) for their expert technical assistance with literature searching and bibliographic normalization.

Generative AI Use

During the preparation of this work, the authors used a generative AI model, Gemini (Google) solely for technical language translation, syntactic refinement, and vocabulary optimization when translating the manuscript from Portuguese into academic English. The AI tool was not used to generate scientific concepts, synthesize data, conduct statistical analyses, or formulate research conclusions. Following translation and refinement, the authors thoroughly reviewed, edited, and validated all content to ensure scientific accuracy, and they take full responsibility for the final content of the manuscript.

Funding

The Brazilian Ministry of Health supported this study through the professional doctorate program. The funder had no role in the study design, data collection, analysis, interpretation of data, or writing of the manuscript. The authors had full access to all the data in the study and take responsibility for data integrity and the accuracy of the data analysis.

Authors' Contributions

DMAR contributed to the study's conceptualization, methodology, and formal analysis of the results. She led the literature screening and data extraction and drafted the original manuscript. GVB acted as the primary supervisor, contributing to the study's conceptualization and methodological oversight and providing critical revisions to the intellectual content. PHDSB contributed to the validation of the findings and the critical review of the final manuscript, offering strategic insights into governance and digital transformation policies. All authors have read and approved the final version of the manuscript and are accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Initial search strategy and search strings.
[DOCX File (Microsoft Word File), 114 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Search strategies deposited in the searchRxiv repository [19-22]; this appendix also contains the final MEDLINE search strategy.
[DOCX File (Microsoft Word File), 52 KB-Multimedia Appendix 2]

Checklist 1

PRISMA-ScR checklist.
[DOCX File (Microsoft Word File), 110 KB-Checklist 1]

References

1. National Academy of Medicine. Whicher D, Ahmed M, Israni ST, Matheny M, editors. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. National Academies Press (US); 2023. [doi: 10.17226/27111]
2. Triantafyllidis AK, Tsanas A. Applications of machine learning in real-life digital health interventions: review of the literature. J Med Internet Res. Apr 5, 2019;21(4):e12286. [doi: 10.2196/12286] [Medline: 30950797]
3. Russell SJ, Norvig P. Artificial Intelligence: A Modern Approach. Prentice Hall; 1995. ISBN: 9780136042594
4. Witten IH, Frank E, Hall MA, Pal CJ. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann; 2017. [doi: 10.1016/C2015-0-02071-8]
5. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. Lancet. May 16, 2020;395(10236):1579-1586. [doi: 10.1016/S0140-6736(20)30226-9] [Medline: 32416782]
6. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res. Jul 10, 2019;21(7):e13659. [doi: 10.2196/13659] [Medline: 31293245]
7. Vishnoi S, Yadav SL, Saini AK, Vats P. Transforming healthcare delivery: assessing the role of e-governance strategies in IOMT and AIML-based medical system administration during the pandemic in the Indian subcontinent. Biomed Pharmacol J. 2025;18(December Spl Edition):331-346. [doi: 10.13005/bpj/3091]

8. Kaladharan S, Manayath D, Gopalakrishnan R. Regulatory challenges in AI/ML-enabled medical devices: a scoping review and conceptual framework. J Med Device. Dec 2024;18(4):040801. [doi: 10.1115/1.4066280]
9. Basulo-Ribeiro J, Teixeira L. The future of healthcare with industry 5.0: preliminary interview-based qualitative analysis. Future Internet. 2024;16(3):68. [doi: 10.3390/fi16030068]
10. Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. J R Soc Med. Jan 2019;112(1):22-28. [doi: 10.1177/0141076818815510] [Medline: 30507284]
11. Saha E, Rathore P. The impact of healthcare 4.0 technologies on healthcare supply chain performance: extending the organizational information processing theory. Technol Forecast Soc Change. Apr 2024;201:123256. [doi: 10.1016/j.techfore.2024.123256]
12. Collins GS, Moons KG. Reporting of artificial intelligence prediction models. Lancet. Apr 20, 2019;393(10181):1577-1579. [doi: 10.1016/S0140-6736(19)30037-6] [Medline: 31007185]
13. Hazarika I. Artificial intelligence: opportunities and implications for the health workforce. Int Health. Jul 1, 2020;12(4):241-245. [doi: 10.1093/inthealth/ihaa007] [Medline: 32300794]
14. Lungu DA, Grillo Ruggieri T, Nuti S. Decision making tools for managing waiting times and treatment rates in elective surgery. BMC Health Serv Res. Jun 11, 2019;19(1):369. [doi: 10.1186/s12913-019-4199-6] [Medline: 31185989]
15. Agboola S, Hale TM, Masters C, Kvedar J, Jethwani K. "Real-world" practical evaluation strategies: a review of telehealth evaluation. JMIR Res Protoc. Dec 17, 2014;3(4):e75. [doi: 10.2196/resprot.3459] [Medline: 25524892]
16. Peters MD, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Scoping reviews. In: Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. JBI Manual for Evidence Synthesis. JBI; 2024. [doi: 10.46658/JBIMES-24-09]
17. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. Oct 2, 2018;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
18. Rocha DM, Barbosa GV, De Souza Bermejo PH. Application of machine learning algorithms in health systems planning and public policy: a scoping review. Open Science Framework. Preprint posted online on Jul 10, 2026. [doi: 10.17605/OSF.IO/VWBHX]
19. Rocha DMA. Search strategy about applications of machine learning algorithms in the context of real-life interventions and resulting clinical implications to guide planning and public policy formulation actions for the health system (MEDLINE). searchRxiv. Preprint posted online on Sep 18, 2025. [doi: 10.1079/searchRxiv.2025.01103]
20. Rocha DM. Search strategy about applications of machine learning algorithms in the context of real-life interventions and resulting clinical implications to guide planning and public policy formulation actions for the health system (EMBASE). searchRxiv. Preprint posted online on Sep 18, 2025. [doi: 10.1079/searchRxiv.2025.01104]
21. Rocha DM. Search strategy about applications of machine learning algorithms in the context of real-life interventions and resulting clinical implications to guide planning and public policy formulation actions for the health system (Scopus). searchRxiv. Preprint posted online on Sep 18, 2025. [doi: 10.1079/searchRxiv.2025.01105]
22. Rocha DM. Search strategy about applications of machine learning algorithms in the context of real-life interventions and resulting clinical implications to guide planning and public policy formulation actions for the health system (LILACS). searchRxiv. Preprint posted online on Sep 18, 2025. [doi: 10.1079/searchRxiv.2025.01106]
23. Campagner A, Berjano P, Lamartina C, Langella F, Lombardi G, Cabitza F. Assessment and prediction of spine surgery invasiveness with machine learning techniques. Comput Biol Med. Jun 2020;121:103796. [doi: 10.1016/j.compbiomed.2020.103796] [Medline: 32568677]
24. Ye C, Wang O, Liu M, et al. A real-time early warning system for monitoring inpatient mortality risk: prospective study using electronic medical record data. J Med Internet Res. Jul 5, 2019;21(7):e13719. [doi: 10.2196/13719] [Medline: 31278734]
25. Fontana MA, Lyman S, Sarker GK, Padgett DE, MacLean CH. Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? Clin Orthop Relat Res. Jun 2019;477(6):1267-1279. [doi: 10.1097/CORR.0000000000000687] [Medline: 31094833]
26. Karhade AV, Ogink P, Thio Q, et al. Development of machine learning algorithms for prediction of discharge disposition after elective inpatient surgery for lumbar degenerative disc disorders. Neurosurg Focus. Nov 1, 2018;45(5):E6. [doi: 10.3171/2018.8.FOCUS18340] [Medline: 30453463]
27. Hayn D, Kreiner K, Ebner H, et al. Development of multivariable models to predict and benchmark transfusion in elective surgery supporting patient blood management. Appl Clin Inform. Jun 14, 2017;8(2):617-631. [doi: 10.4338/ACI-2016-11-RA-0195] [Medline: 28850152]
28. Krämer J, Schreyögg J, Busse R. Classification of hospital admissions into emergency and elective care: a machine learning approach. Health Care Manag Sci. Mar 2019;22(1):85-105. [doi: 10.1007/s10729-017-9423-5] [Medline: 29177993]
29. EndNote. Clarivate. URL: https://clarivate.com/accessibility/endnote/ [Accessed 2026-08-23]

30. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. Dec 5, 2016;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
31. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. Jan 6, 2015;162(1):55-63. [doi: 10.7326/M14-0697] [Medline: 25560714]
32. EPHPP quality assessment tool for quantitative studies 2026. Elite Providers Hub for Progressive Play. URL: https://www.ephpp.ca/quality-assessment-tool-for-quantitative-studies/ [Accessed 2026-08-23]
33. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. J Eval Clin Pract. Feb 2012;18(1):12-18. [doi: 10.1111/j.1365-2753.2010.01516.x] [Medline: 20698919]
34. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. Mar 29, 2021;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
35. Stevens RJ, Poppe KK. Validation of clinical prediction models: what does the “calibration slope” really measure? J Clin Epidemiol. Feb 2020;118:93-99. [doi: 10.1016/j.jclinepi.2019.09.016] [Medline: 31605731]
36. Toma TS, Pereira TV, Vanni T, Barreto JO. Avaliação de tecnologias de saúde & políticas informadas por evidências [Book in Portuguese]. Instituto de Saúde; 2017. ISBN: 9788588169319
37. Ramezani M, Mobinizadeh M, Bakhtiari A, et al. Agenda setting for health equity assessment through the lenses of social determinants of health using machine learning approach: a framework and preliminary pilot study. BioData Min. Feb 10, 2025;18(1):14. [doi: 10.1186/s13040-025-00428-x] [Medline: 39930525]
38. E-SUS primary health care strategy (e-SUS APS). Ministry of Health, Brasil. 2021. URL: https://www.gov.br/saude/pt-br/assuntos/esua-aps [Accessed 2026-10-29]
39. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). European Union. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng [Accessed 2026-10-29]
40. General Personal Data Protection Law (LGPD) - law no. 13.709/2018. Federal Government of Brazil. 2018. URL: https://www2.senado.leg.br/bdsf/handle/id/545155 [Accessed 2026-10-29]
41. Collins GS, Moons KG, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. Apr 16, 2024;385:e078378. [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
42. Moons KG, Damen JA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. Mar 24, 2025;388:e082505. [doi: 10.1136/bmj-2024-082505] [Medline: 40127903]
43. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data. Mar 15, 2016;3:160018. [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
44. Porter ME. A strategy for health care reform--toward a value-based system. N Engl J Med. Jul 9, 2009;361(2):109-112. [doi: 10.1056/NEJMp0904131] [Medline: 19494209]
45. Porter ME. What is value in health care? N Engl J Med. Dec 23, 2010;363(26):2477-2481. [doi: 10.1056/NEJMp1011024] [Medline: 21142528]
46. Qinggui H, Xiuqiong T, Hui C, Jinsong W. The ethical risks and solutions of the algorithm black box in medical artificial intelligence. Acta Bioeth. 2025;31(1):24-34. [doi: 10.4067/S1726-569X2025000100024]
47. Huo Z, Booth J, Monks T, et al. Dynamic mortality prediction in critically ill children during interhospital transports to PICUs using explainable AI. NPJ Digit Med. Feb 17, 2025;8(1):108. [doi: 10.1038/s41746-025-01465-w] [Medline: 39962177]
48. Gu F, Meyer AJ, Ježek F, et al. Identification of digital twins to guide interpretable AI for diagnosis and prognosis in heart failure. NPJ Digit Med. Feb 18, 2025;8(1):110. [doi: 10.1038/s41746-025-01501-9] [Medline: 39966509]
49. Cammarata R, Ruffini F, Catamerò A, et al. Machine learning for predicting the low risk of postoperative pancreatic fistula after pancreaticoduodenectomy: toward a dynamic and personalized postoperative management strategy. Cancers (Basel). May 31, 2025;17(11):1846. [doi: 10.3390/cancers17111846] [Medline: 40507327]
50. Singh NP, Mujawar MA, Golani A. Role of artificial intelligence in enhancing mechanical ventilation - a peek into the future. Indian J Anaesth. Jul 2025;69(7):722-728. [doi: 10.4103/ija.ija_995_24] [Medline: 40657139]

Abbreviations

ANN: artificial neural network
AUC: area under the curve
AUROC: area under the receiver operating characteristic curve
CDSS: clinical decision support system
EPHPP: Effective Public Health Practice Project
FAIR: findable, accessible, interoperable, and reusable
GDP: gross domestic product
GDPR: General Data Protection Regulation
H5.0: Health 5.0
JBI: Joanna Briggs Institute
LASSO: least absolute shrinkage and selection operator
LGPD: Lei Geral de Proteção de Dados Pessoais
LIME: local interpretable model-agnostic explanations
MCID: minimal clinically important difference
ML: machine learning
NASSS: nonadoption, abandonment, and scale-up, spread, and sustainability
NLP: natural language processing
PHC: primary health care
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
PROBAST: Prediction Model Risk of Bias Assessment Tool
PROM: patient-reported outcome measure
RF: random forest
RQ: research question
RWD: real-world data
SDOH: social determinants of health
SHAP: Shapley additive explanations
SUS: Sistema Único de Saúde
SVM: support vector machine
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
XAI: explainable AI

Edited by Matthew Balcarras; peer-reviewed by Claire Nierva Herrera, Francesco Andrea Causio, Jean-Christophe Boucher, Joshua Odeniya; submitted 20.Apr.2026; final revised version received 01.Sep.2026; accepted 14.Sep.2026; published 16.Nov.2026

Please cite as:
Rocha DMA, Barbosa GV, De Souza Bermejo PH
Application of Machine Learning Algorithms to Health Systems Planning and Public Policy Using Real-World Data: Scoping Review
Interact J Med Res 2026;15:e96779
URL: https://www.i-jmr.org/2026/1/e96779
doi: 10.2196/96779

© Deusa Maria Arruda Rocha, Gentil Veloso Barbosa, Paulo Henrique De Souza Bermejo. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 16.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.