ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Systematic Literature Review

Predicting Healthcare Utilization Outcomes With Artificial Intelligence: A Large Scoping Review

Carlos Gallego-Moll, MSc, Lucía A. Carrasco-Ribelles, PhD, Marc Casajuana, MSc, Laia Maynou, PhD, Pablo Arocena, PhD, Concepción Violán, MD, PhD, Edurne Zabaleta-del-Olmo, PhD

ABSTRACT

Objectives: To broadly map the research landscape to identify trends, gaps, and opportunities in data sets, methodologies, outcomes, and reporting standards for artificial intelligence (AI)-based healthcare utilization prediction.

Methods: We conducted a scoping review following the Joanna Briggs Institute methodology. We searched 3 major international databases (from inception to January 2025) for studies applying AI in predictive healthcare utilization. Extracted data were categorized into data sets characteristics, AI methods and performance metrics, predicted outcomes, and adherence to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) + AI reporting guidelines.

Results: Among 1116 records, 121 met inclusion criteria. Most were conducted in the United States (62%). No study incorporated all 6 relevant variable groups: demographic, socioeconomic, health status, perceived need, provider characteristics, and prior utilization. Only 7 studies included 5 of these groups. The main data sources were electronic health records (60%) and claims (28%). Ensemble models were the most frequently used (66.9%), whereas deep learning models were less common (16.5%). AI methods were primarily used to predict future events (90.1%), with hospitalizations (57.9%) and visits (33.1%) being the most predicted outcomes. Adherence to general reporting standards was moderate; however, compliance with AI-specific TRIPOD + AI items was limited.

Conclusions: Future research should broaden predicted outcomes to include process- and logistics-oriented events, extend applications beyond prediction—such as cohort selection and matching—and explore underused AI methods, including distance-based algorithms and deep neural networks. Strengthening adherence to TRIPOD-AI reporting guidelines is also essential to enhance the reliability and impact of AI in healthcare planning and economic evaluation.

Keywords: artificial intelligence, healthcare utilization outcomes, health economics, resource allocation, review.

VALUE HEALTH. 2026; 29(1):159–171

Highlights

- Healthcare utilization outcomes remain underexplored within health economics and outcomes research (HEOR). This review explores artificial intelligence (AI) applications predicting healthcare utilization outcomes, identifying gaps and opportunities in datasets, methodologies, outcomes, and reporting standards.
- Most AI models focused on predicting hospitalizations or visits, whereas process-oriented outcomes—such as ambulance arrivals—remain underrepresented. None of the studies incorporated all the relevant variable groups. The use of AI was limited to prediction, with its potential contribution for causal analysis often overlooked. Deep learning algorithms were rarely used.
- This review provides a foundation for future research on specific outcomes, settings, methods, and theory-informed variable selection. These elements represent critical steps toward promoting equitable, efficient, and evidence-based decision making in healthcare planning and economic evaluation.

Introduction

Health economics and outcomes research (HEOR) is a discipline designed to complement traditional clinical development information—such as efficacy, safety, and quality—by guiding decision makers on patient access to specific drugs and services.[1] HEOR encompasses various outcomes, including clinical events, disease incidence, treatment outcomes, healthcare utilization, disease progression, and symptoms. Among these, healthcare utilization refers to "the quantification or description of the use of services by persons for the purpose of preventing and curing health problems, promoting maintenance of health and well-being, or obtaining information about one's health status and prognosis."[2]

Developing research in the field of predicting healthcare utilization offers potential benefits for policymakers, researchers, providers and health managers. It underpins the projection of future healthcare needs, such as facilities, personnel, or supplies,[2] and fosters a more thorough understanding of healthcare utilization patterns, thereby allocating resources to those uses that have the greatest impact on health.[3]

Studies on healthcare utilization are diverse. Theoretical studies develop a conceptual framework to understand the factors determining healthcare utilization levels, such as Andersen's healthcare utilization model.[4,5] It incorporates the complex interplay of individual, societal, and system-level factors that determine healthcare utilization into previous models. These factors are categorized into predisposing, enabling, and need and include the process of health care as a facet of health behavior, alongside the use of health services and personal health practices. Similarly, Anderson's framework described 5 distinct approaches for examining health services utilization: sociocultural, sociodemographic, social-psychological, organizational, and social systems.[6] Empirical studies on healthcare utilization generally fall into 2 main categories: those analyzing disparities in the use of health services and those focused on predicting utilization levels. Disparities have been examined across a range of dimensions, including race/ethnicity,[7-9] sex,[10,11] age,[12] geography,[13] financial constraints,[14,15] lack of insurance,[16] language barriers,[17] and experiences of discrimination.[18] Meanwhile, predictive studies have gained increasing prominence with the rise of artificial intelligence (AI) methods, which often outperform traditional statistical techniques in processing large and complex data sets.[19,20] Although prediction remains the primary application of AI, these methods are increasingly being adapted for causal inference tasks. However, as Athey cautions, relying solely on off-the-shelf AI models is insufficient for guiding policy decisions or resource allocation. To generate actionable insights, it is essential to integrate AI predictions with domain knowledge and rigorous theoretical frameworks, including causal analysis where appropriate.[21,22]

Despite the advancements, predicting healthcare utilization outcomes using AI remains an area of ongoing research. Fewer studies focus on this topic compared with other areas of HEOR, such as clinical events or disease incidence.[23] Healthcare utilization prediction present different challenges and potential benefits,[20] including the need to integrate multiple data sources. Studies from the provider perspective may encounter further barriers, such as the reluctance to share institutional data and the methodological requirement for a multilevel approach, often requiring the combination of multiple sources. Consequently, researchers face increased administrative and time burdens in preparing data for analysis.[24] Additionally, the sample size may be smaller, due to reliance on aggregated, may limit AI algorithm performance, whereas the type and frequency of available data can reduce model complexity.

Given these challenges, it is crucial to map the characteristics of studies on healthcare utilization and assess reporting standards, such as the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD + AI) guidelines, to promote accuracy, reproducibility, and trustworthiness, thereby supporting the effective evaluation, validation, and implementation of AI models.[25]

However, we have not identified any comprehensive reviews in scientific literature focusing mainly on the prediction of healthcare utilization by AI methods. Therefore, our aim was to broadly map the research landscape to identify trends, gaps, and opportunities in data sets, methodologies, outcomes, and reporting standards for AI-based healthcare utilization prediction. Particularly, we addressed 5 questions: (1) which data sets are used, and what are their characteristics? (2) which AI methods are used, and what are their characteristics? (3) which performance metrics are used? (4) which healthcare utilization outcomes are predicted? and (5) what is the degree of adherence to the TRIPOD + AI guidelines?[25]

Methods

This review followed the Joanna Briggs Institute methodology for scoping reviews,[26] and it is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews.[27] The protocol was registered prospectively on the Open Science Framework platform (available at https://osf.io/udm76). An additional assessment of adherence to the TRIPOD + AI guidelines was conducted for included studies, which was not described in the original protocol.

Eligibility Criteria

We defined the following eligibility criteria according to the Joanna Briggs Institute's approach:

- Participants: no specific criteria for participants were applied.
- Concept: studies were eligible if they used AI methods at any stage to predict healthcare utilization outcomes. We defined healthcare utilization outcomes as "the quantification or description of the use of services by persons to preventing and curing health problems, promoting maintenance of health and well-being, or obtaining information about one's health status and prognosis."[2]

We define AI as the field focused on the development of computer systems that mimic human intelligence. According to the National Library of Medicine, AI involves programs capable of adaptively improving their performance over time by processing and analyzing large data sets, recognizing patterns, and using those patterns to enhance problem solving and task execution (National Library of Medicine, 2025).

- Context: we included studies conducted in any healthcare setting.
- Type of evidence sources: eligible sources included journal publications, reviews, dissertations and theses, conference abstracts, and ongoing studies. No language restrictions were applied during the search or screening process.

Studies were excluded if they did not meet one or more of the above eligibility criteria (eg, they did not apply an AI method as defined, did not report a relevant healthcare utilization outcome, or were conducted outside healthcare contexts).

Information Sources

We searched the following databases from inception to January 2025: EconLit (via EBSCOhost), MEDLINE (Via PubMed), and Scopus. Reference lists of included studies and relevant reviews were also manually scanned to identify additional studies.

We applied no date restrictions, as our aim was to capture the full breadth of published evidence on the use of AI in predicting healthcare utilization. Although we anticipated that most relevant studies would be from recent years, we considered it important to include earlier research to ensure completeness and to identify potentially foundational work within this emerging field.

Search Strategy

We conducted a preliminary search of MEDLINE (via PubMed) to identify relevant terms. A comprehensive search strategy was created using the text words identified in the titles and abstracts of relevant reports and the index terms. The final search strategy is provided in Appendix 1 in Supplemental Materials.

Selection of Sources of Evidence

Search results were imported to Rayyan software[28] for storage, duplicate removal, and screening. Duplicates were automatically removed in Rayyan and manually verified by a reviewer. Two reviewers independently screened the titles and abstracts against the eligibility criteria. Full texts of potentially relevant reports were obtained and examined by 2 reviewers. Reasons for exclusion of ineligible studies were documented.[29] Disagreements were resolved by consensus or by consulting a third reviewer.

Data Charting and Items

Before data extraction, a data charting form was developed in Microsoft Excel and piloted on a small sample of included studies to assess its feasibility and suitability. The form was iteratively refined before final approval by all reviewers. Adherence to TRIPOD + AI guidelines was systematically assessed for each study. Two reviewers independently extracted and charted data, resolving disagreements through discussion or, if needed, a third reviewer.

Detailed definitions of each data item and their potential values are provided in Appendix 2 in Supplemental Materials.

Risk of Bias Across Studies

Because this scoping review aims to describe the existing research, studies were not excluded based on methodological quality standards. Therefore, critical appraisal or risk of bias assessment was not performed on the included studies.[26]

Synthesis of Results

The extracted data were compiled into a unified spreadsheet and imported into Microsoft Excel for discrepancy resolution and validation. Fields were scrutinized to homogenize vocabulary and detect implausible values. These data were then exported to RStudio 4.3.0 for analysis. Descriptive statistics were used to summarize study characteristics. Results are presented in tables and figures, with a narrative summary outlining main findings. Note that studies could fall into multiple categories—data collection types, groups of variables, groups of AI algorithms, intended uses of AI, or predicted healthcare utilization outcomes—therefore, *N* may differ from 121 in these items. All items were described using frequencies and percentages of studies.

Results

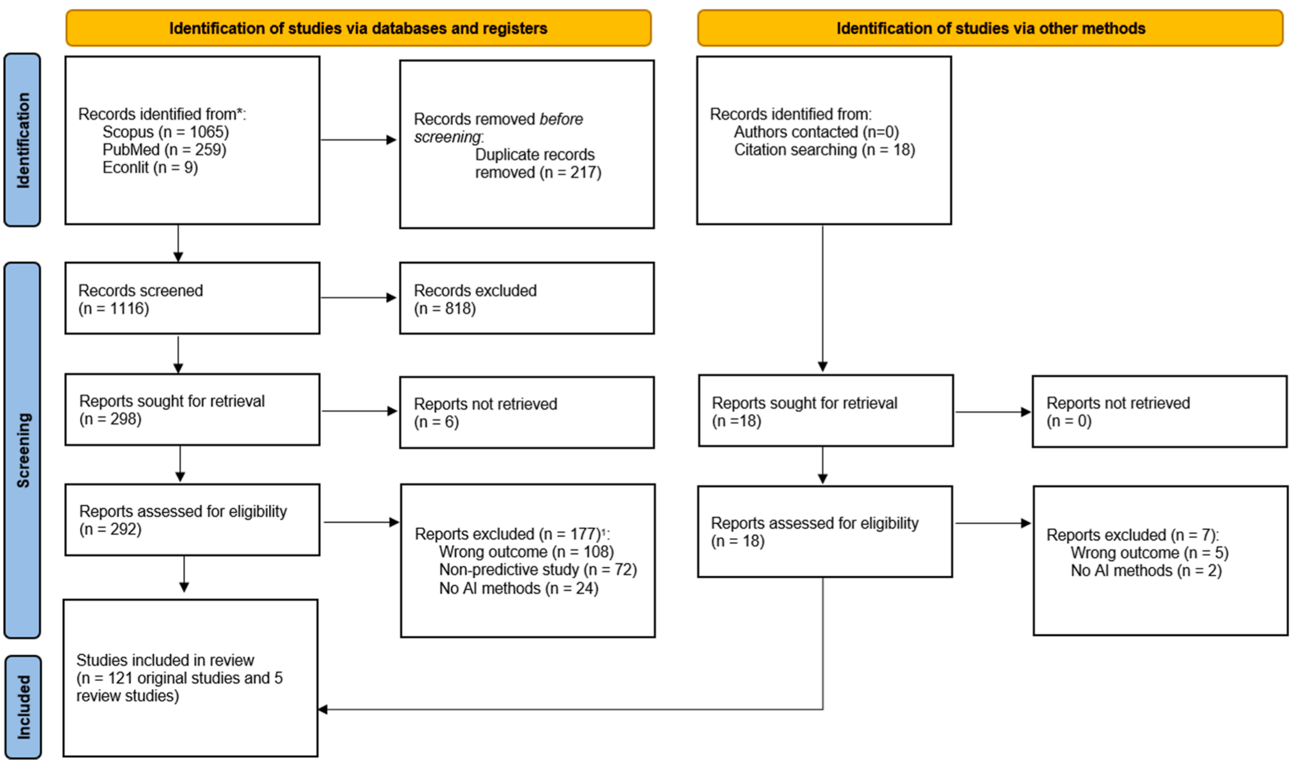
After removing duplicates, 1116 references were identified from searches of electronic databases and citation searching. Based on titles and abstracts, 818 references were excluded, leaving 298 full-text articles for eligibility assessment. Of these, 183 studies were excluded—177 for not meeting the eligibility criteria and 6 because of being unretrievable (details in Appendix 3 in Supplemental Materials). The remaining 126 studies—121 original and 5 reviews—were included in this scoping review. Data extraction focused only on the original 121 studies. The review studies were used to identify additional studies not captured in the searches, labeled as "identification of studies via other methods" (see Fig. 1).

Description of Included Studies

Research activity was sparse between 1994 and 2015. A marked increase emerged from 2020 onward, reaching its peak in the most recent years, with 22 and 23 publications reported in 2023 and 2024, respectively.

Geographically, 78 studies were conducted in North American countries (75 in the United States[30-104] and 3 in Canada[105-107]), 20 in Asian countries (4 each in Taiwan[108-111] and Singapore,[112-115] 3 in

Figure 1. PRISMA-ScR flow-diagram.[27] *The same study can be excluded for multiple reasons, so the sum of excluded studies for each reason does not add up to the total number of excluded estudies.



China,[116-118] 2 each in Indonesia[119,120] and India,[121,122] and 1 each in Qatar,[123] Saudi Arabia,[124] South Korea,[125] Malaysia,[126] and Israel[127]), 15 in European countries (3 each in The Netherlands,[128-130] the United Kingdom[131-133] and Germany,[134-136] and 1 each in France,[137] Denmark,[138] Italy,[139] Finland,[140] Sweden,[141] and Switzerland[142]), and 7 in other countries (3 in Australia,[143-145] 2 in Brazil,[146,147] and 1 each in Tanzania,[148] New Zealand,[149] and Chile[150]).

Follow-up ranged from 2 months to 20 years, and sample sizes varied from 83 to 14 422 233. participants. Healthcare settings were notably diverse, with 37% conducted in hospitals, followed by the general healthcare system (16%), home and community care (12%), and emergency departments (9.1%).

Key charted data from each included study, aligned with the scoping review's questions and aims, are described below (Table 1). The raw parameters of these studies are available in Appendix 4 in Supplemental Materials.

Data Sets Characteristics

Table 2 shows that nearly every study included health status (94%) and demographic (83%) variables, whereas provider (14%) and perceived-need (5.8%) factors were largely ignored. No study covered all 6 relevant variable groups (demographic, socioeconomic, health status, perceived need for healthcare, provider characteristics, and prior healthcare utilization, which were derived from the Andersen's Model, see Appendix 2 in Supplemental Materials for details), and only 7 included 5; just 2 of those 7 justified their choices with a theoretical framework.[82,101]

The primary data source was electronic health records (EHRs), followed by claims and surveys, the latter being the only reported method that captured perceived need for healthcare. Some studies collected data specifically for the study, whereas other studies used other sources for monitoring purposes, such as predicting hospitalizations due to COVID-19[53,139,150] or social media data.[46] Studies using official statistics often combined this source with surveys,[148] claims,[97,113] surveillance,[139] or even multiple sources.[82] In contrast, most studies (75%) relied on a single data source.

Nine studies focused on aggregate-level predictions. Of these, 5 combined 2 or more data sources, compared with only 16.1% of the individual-level studies. Among the aggregate-level analyses, 3 relied on survey data and 2 used official statistics. Across all studies, the number of variables per data set ranged from 3 to 5624, with a median of 51.

AI Methods and Performance Metrics

Figure 2 summarizes the AI methods and metrics reported. Most studies (90.1%) applied AI for predictive purposes. In 27.3% of studies AI was used for feature selection, predominantly alongside predictive modeling, with only 7 studies using it as a stand-alone aim. Other studies focused on cohort selection by clustering patients with similar characteristics.[64,81,92] In other cases, AI was used for matching patients with similar covariates by generating propensity scores,[63] sometimes integrating these approaches with difference-in-difference designs to strengthen causal inference by addressing unobserved confounding and temporal trends.[90,135]

No data transformation was required or reported in 13.2% of the studies. Among transformation processes, handling outliers was the least common (9.9%), whereas dimensionality reduction was the most frequent (53.7%). Additionally, 84.3% of the studies reported the software used for analysis, with RStudio (37.2%) and Python (29.8%) being the most common. Since 2020, the evolution of AI methods highlights a greater on the number of transformation processes and metrics reported, dimensionality reduction, internal validation, hyperparameter tuning and feature importance (details in Appendix 5 in Supplemental Materials).

Table 1. Study details.

Parameter	*n* = 121*
Publication year	
2023-2025 (until January)	47 (39%)
2020-2022	50 (41%)
2015-2019	18 (15%)
1994-2014	6 (5.0%)
Country of study	
North America	78 (64.5%)
USA	75 (62%)
Canada	3 (2.5%)
Asia	20
Singapore	4 (3.3%)
Taiwan	4 (3.3%)
China	3 (2.5%)
Others	9 (7.4%)
Europe	
Germany	3 (2.5%)
The Netherlands	3 (2.5%)
UK	3 (2.5%)
Others	6 (5.0%)
Oceania	
Australia	3 (2.5%)
New Zealand	1 (0.8%)
South America	
Brazil	2 (1.7%)
Chile	1 (0.8%)
Africa	
Tanzania	1 (0.8%)
Publication type	
Journal article	114 (94%)
Conference proceedings	6 (5.0%)
Ongoing study	1 (0.8%)
Follow-up time, in years	
From 0 to 2	39 (35%)
From 2 to 4	21 (19%)
From 4 to 8	30 (27%)
>8	20 (18%)
Not reported	11
Sample size, both training and testing samples (*n*)	
From 0 to 1000	20 (17%)
From 1001 to 10 000	35 (30%)
From 10 001 to 100 000	35 (30%)
>100 000	27 (23%)
Not reported	4
Healthcare setting	
Hospital care	45 (37%)
Healthcare system	19 (16%)
Home and community care	14 (12%)
Emergency care	11 (9.1%)
Intensive care	10 (8.3%)
Primary/ambulatory care	7 (5.8%)
Others	15 (12.4%)

*n (%).

AI Algorithms and Predicted Healthcare Utilization Outcomes

We identified 10 groups of AI algorithms across the included studies (Fig. 3). Ensemble models (eg, random forest, bagging, and boosting) were the most common (66.9%), followed by logistic regressions and linear models (54.5%), including LASSO and Enet, and tree-based models (32.2%). Deep learning models and distance-based algorithms appeared in only 16.5% and 9.1%, respectively. Nearly half of the studies (48.8%) used only 1 or 2 algorithm groups.

Hospitalizations (57.9%) were the most frequently predicted healthcare utilization outcomes, followed by visits (33.1%) (see Fig. 3). Hospitalization outcomes included frequency, length of stay, and readmissions, either for any cause or for a specific reason. Visits outcomes mainly focused on unplanned visits to emergency departments, which were predicted alongside hospitalizations in 20 studies. However, we also found visits related to pediatrics,[49] imaging utilization,[48] oncology,[63] low-back pain,[134] and psychotherapy,[73] among others. Devices or equipment for treatment included renal replacement therapy,[103,108] mechanical ventilation,[70,72,77,84,123,146] blood transfusion,[122] and intubation.[78] Diagnostic tests included radiology resource utilization,[37] medical tests recommended for the management and monitoring of diabetes,[125] and low-dose computed tomography.[47] Surgical procedures included

surgery duration,[120] elective surgery,[76] and bypass of healthcare facilities.[148] From our predefined list of outcomes, there were no studies on immunization/vaccinations and screening, and only 1 study on waiting times.[57] Some studies predicted outcomes that did not fit into predefined categories due to their specificity, which is reflected in their low frequency. Examples include predictions of nonattendance,[40,71,149] ambulance arrivals,[113] digital intervention utilization,[90,102,136,144] maternal healthcare utilization,[121,147] and mental healthcare utilization.[62,79,91,95,101,129]

Hospitalizations and discharges were primarily studied in hospitals, intensive care, home and community care, and broader healthcare system settings. In contrast, visits were studied across a wider range of settings, including diagnostic imaging services, emergency care, healthcare facilities, mental healthcare, and primary/ambulatory care. Studies on devices and equipment for treatment were mainly concentrated in intensive care settings,[70,72,77,78] but there were also studies conducted in hospital settings[123,146] and the entire healthcare system.[108]

Adherence to TRIPOD + AI Guidelines

Adherence to core TRIPOD items was generally high, but AI-specific items were poorly reported. This discrepancy resulted in a remarkable difference in adherence rates between the original TRIPOD items (n = 24) and the newly introduced AI-specific items (n = 27).

Specifically, adherence statistics for the original items and the AI items were as follows: mean adherence (50.6% vs 24.1%), median adherence (54.1% vs 13.2%), first quartile (38.6% vs 1.7%), and third quartile (72.7% vs. 43.3%), respectively. Some critical AI-related items, such as predictors measurement, model updating and evaluation, participant distribution, handling of poor-quality data, and the use of interaction and expertise, were not reported across all studies. These results are detailed in Figure 4.

Discussion

This scoping review presents a comprehensive synthesis of how AI research has been applied to predict healthcare utilization outcomes. The findings underscored the rapid expansion of

Table 2. Data set characteristics.

Parameter	n = 121*
Group of variables used	
Health status	114 (94%)
Demographic	101 (83%)
Healthcare utilization	73 (60%)
Socioeconomic	52 (43%)
Provider characteristics	17 (14%)
Perceived need for healthcare	7 (5.8%)
Number of variable groups used	
From 1 to 2	32 (27%)
From 3 to 4	82 (67%)
From 5 to 6	7 (5.8%)
Data collection type	
Electronic health records	73 (60%)
Claims	34 (28%)
Surveys	17 (14%)
Official statistics	6 (5.0%)
Primary data collection	5 (4.1%)
Surveillance data	4 (3.3%)
Social media data	1 (0.8%)
Number of data collection types	
1	91 (92%)
2	21 (17%)
3	1 (0.8%)
4	1 (0.8%)
Not reported	7 (5.8%)
Level of analysis	
Individual data	111 (92%)
Aggregated data	9 (7.4%)
Not applicable	1 (0.8%)
Total variables before feature selection	84 (69.4%) 386 (1166), 50 [3, 7476]
Combined databases (Yes)	35 (29%)
Data set availability (Yes)	37 (31%)

*n (%); mean (SD), median (minimum, maximum).

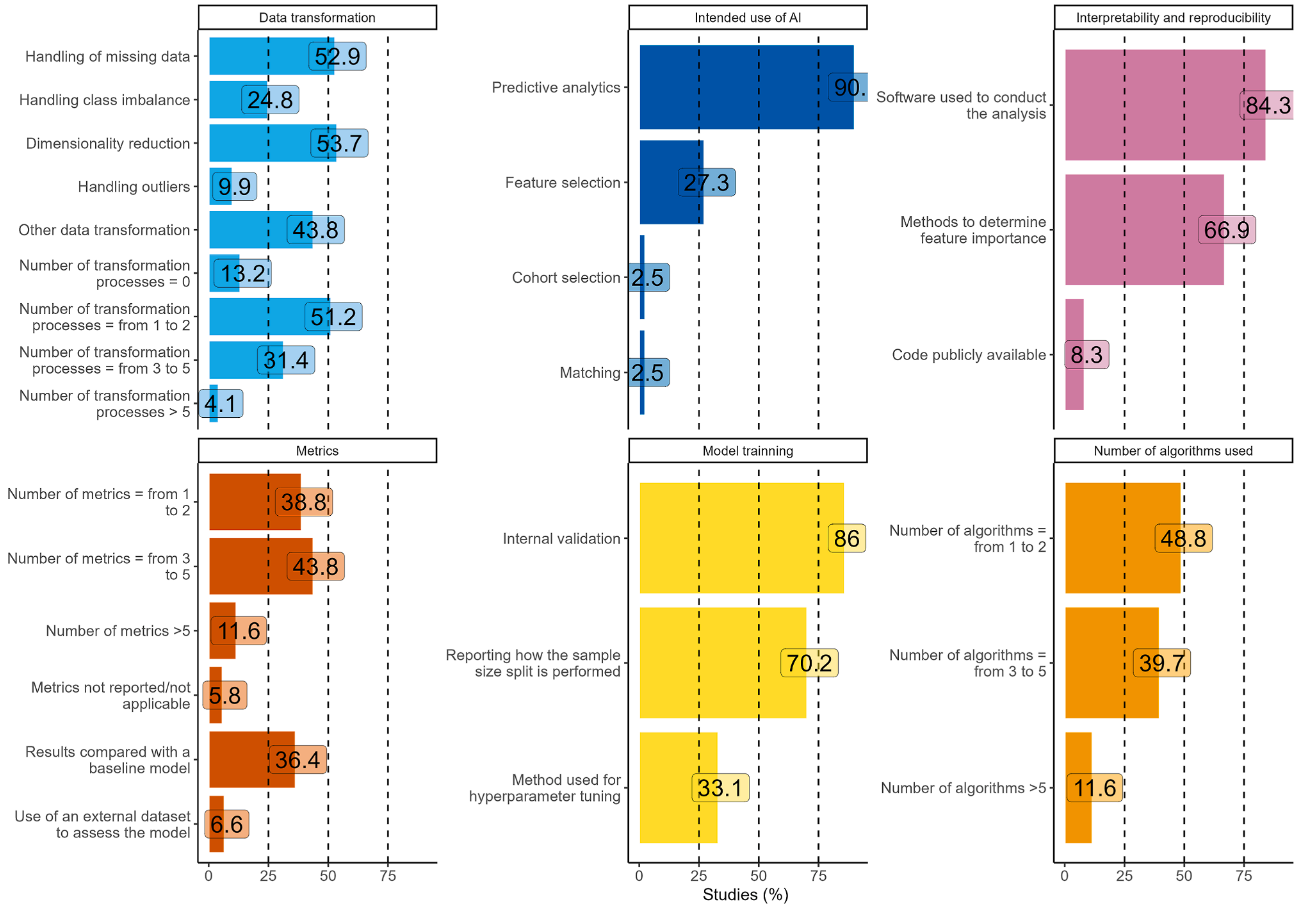
research in this area, which encompassed a wide range of healthcare settings, data sets, methodological approaches, and predicted outcomes. Despite this variability, certain patterns emerged. Most studies focused on predicting hospitalizations or visits, often using health status data from EHRs. However, data sets frequently lacked comprehensiveness, omitting broader factors influencing healthcare utilization. The primary AI application was predicting future events, with ensemble techniques most used. Metric reporting requires improvement to enhance robustness. Although adherence to original TRIPOD items was strong, reporting on

AI-specific aspects—such as transparency, fairness, and public involvement—was weaker, with critical omissions.

There was a notable surge in publications from 2020 onward, reflecting the rapid growth of AI in HEOR,[20,23] likely driven by increased popularity of AI methodologies and the availability of larger data sets. The majority of studies were conducted in the United States (62%), likely due to broader access to healthcare data.[151] The predominance of hospital-based studies suggests easier data access but underscores gaps in primary care and preventive services, which remain underexplored.

Predicting healthcare utilization outcomes requires considering factors affecting demand for services and resources but also

Figure 2. Methodological practices in included studies (in %).



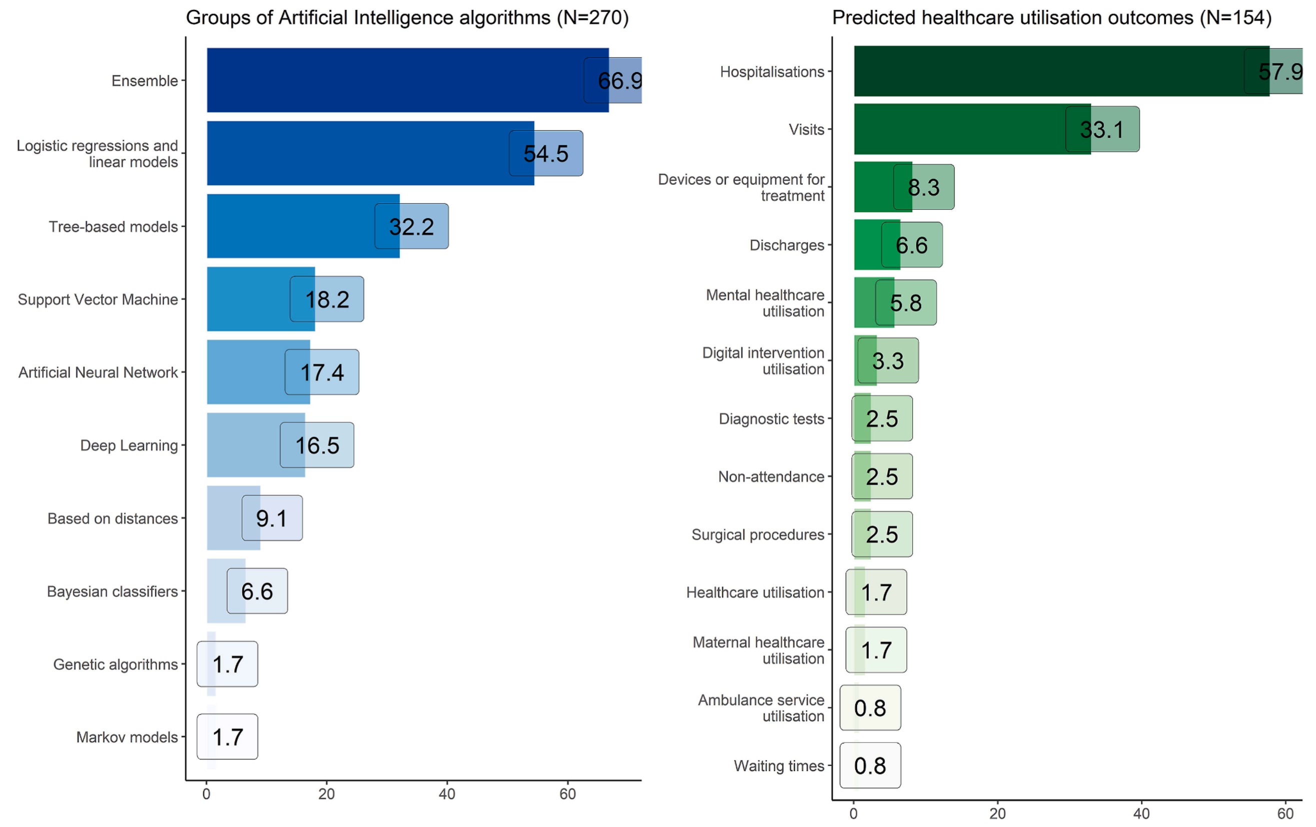
considering the supply side: the availability of these resources and aspects such as complementarity and substitution.[152] In our review, variables related to the demand side showed high variability. Most studies included health status variables (94%) and demographic data (83%), whereas socioeconomic variables (43%) and perceived healthcare need (5.8%) were underrepresented. The socioeconomic factors were often census based (eg, residence quintiles), overlooking aspects such as urbanization, employment status, neighborhood safety, and pollution. In contrast, supply-side factors captured through provider characteristics group, such as management type (public or private), available resources, or quality of care, were notably underrepresented (14%). These findings highlight the importance of careful variable selection when modeling healthcare utilization. As Athey and Imbens (2019) emphasize, AI models reward data sets adaptation: exploiting domain-specific context by sample splitting, orthogonalization, and theory-guided variable selection typically produces models that both explain and predict better than relying on off-the-shelf machine learning models.[153]

Data set combination was used in 29% of studies but mainly to increase sample sizes by incorporating additional territories rather than to enable aggregate-level predictions. Similar to our findings, Lee et al[23] (2023) identified EHRs as the predominant data source, although we found greater use of claims data (28% vs. 9%), likely due to differences in inclusion criteria and temporal scope. The focus on hospital-based settings highlights gaps in primary and community care, limiting progress in these healthcare settings. Addressing these gaps requires systematic data collection across healthcare settings, integrating diverse sources, and expanding theoretically grounded variables to enhance the accuracy of predictive models.

Consistent with Athey (2018), AI applications predominately focused on predictive purposes, often overlooking their potential contribution for causal studies.[22] However, AI alone cannot establish causal relationships or replace the strong statistical and econometric assumptions required for credible causal claims; rather, it assists in strengthening key technical components of the causal process. First, it supports feature selection: in high-dimensional settings, techniques such as LASSO, decision trees, boosting, or random forests help identify the most relevant covariates for treatment assignment or outcome prediction, aiding hypothesis generation.[154,155] Second, AI enhances cohort selection by identifying patterns that group individuals with similar characteristics, also supporting hypothesis development. Third, it facilitates matching procedures, particularly in large datasets, by efficiently pairing individuals with similar covariates and improving propensity score models than traditional methods, thereby reducing omitted variable bias.[156]

Regarding AI techniques, Ensemble and logistic/linear regression models dominated, whereas distance-based algorithms were less common, reflecting the limited application of AI for cohort selection. This aligns with the findings by Jiang et al[157] (2017) on the underuse of unsupervised learning in healthcare. Advanced methods, including deep learning, were also

Figure 3. Distribution of AI models (left; N = 270) and target outcomes (right; N = 154) (in % of included studies).



infrequently used. Similarly, Lee et al[20] (2022) report the dominance of tree-based models and logistic/linear regression, with limited use of neural networks. In contrast, some researchers found deep learning models are more prominent than machine learning methods for broader HEOR outcomes, such as disease progression, health status, or subsequent events.[158,159] This distribution suggests that the structure and complexity of data were adequately handled by simpler algorithms or reflects a shortage of AI expertise among healthcare managers and researchers. It may also reflect a preference for interpretable and user-friendly methods, especially when compared with studies predicting clinical outcomes, which often involve irregular time intervals and higher data complexity. Comparing studies before and after 2020 revealed advancements in handling class imbalance, dimensionality reduction, and transformation processes, reflecting increased sophistication in preprocessing. Greater internal validation and hyperparameter tuning suggest more rigorous model training, whereas the rise of feature importance methods emphasized interpretability.

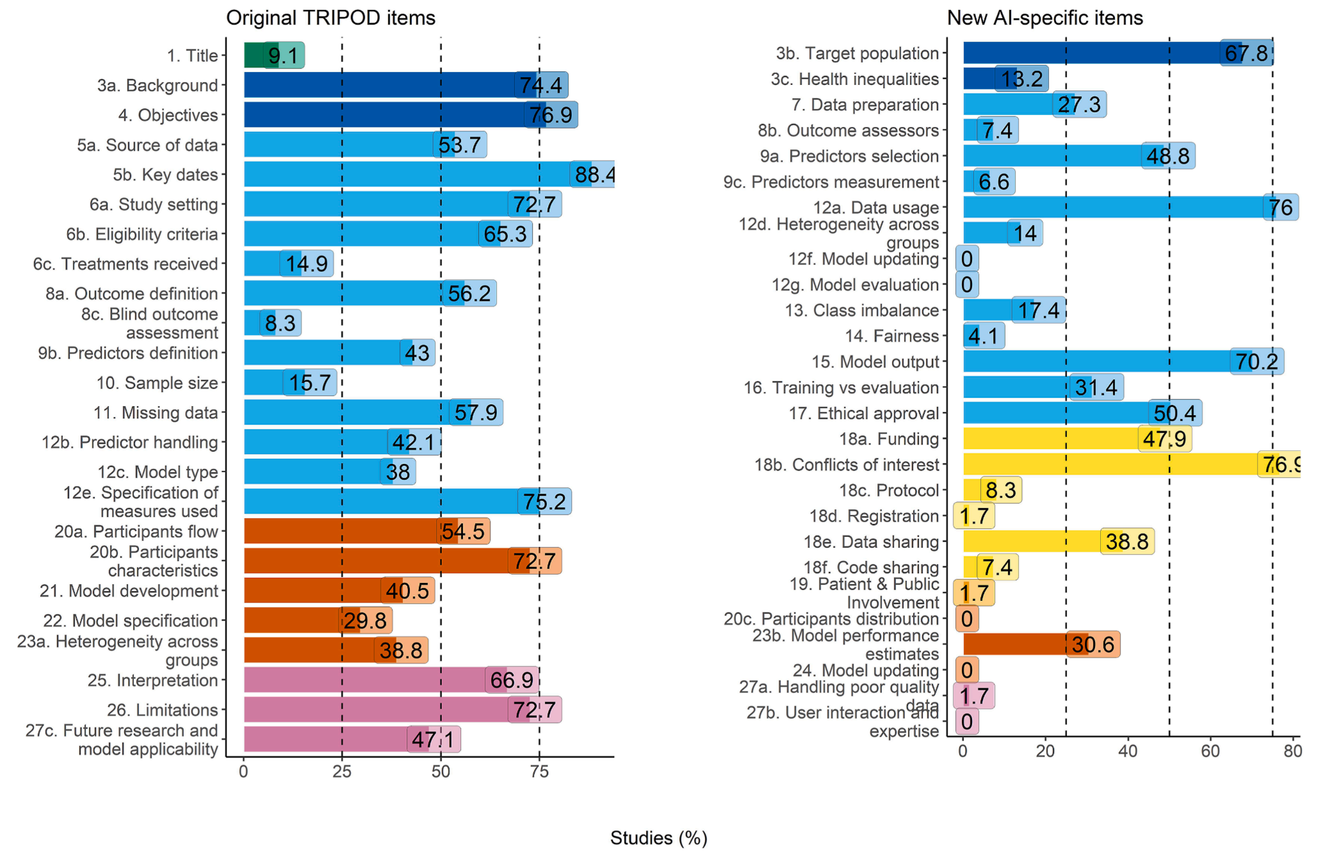
Hospitalizations and visits were the most frequently predicted healthcare utilization outcomes, whereas other areas—screenings, diagnostic tests, surgical procedures, treatment equipment and devices, nonattendance, ambulance arrivals, and waiting times metrics—were rarely modeled. These underrepresented outcomes often reflect process- or logistics-oriented events rather than direct clinical results. Their frequency prediction are further shaped by contextual factors: staffing ratios, clinic capacity, public holiday calendars, transport links, digital literacy, and policy shifts—requiring features beyond standard demographic or morbidity data. Because they are typically recorded in ancillary systems rather than core EHRs or claims data, labeled data sets for these outcomes are harder to obtain. This highlights the need for broader data integration and access to diverse variables to better capture nonclinical influences and improve service delivery. The absence of studies predicting vaccination uptake—even during the COVID-19—illustrates how data availability and contextual complexity, rather than clinical relevance, continue to drive research priorities.

Adherence to TRIPOD + AI guidelines revealed notable differences between original TRIPOD items and new AI-specific ones. Only a small proportion of studies shared code (7.4%) or data (38.8%), highlighting transparency and reproducibility gaps. Overall, these findings suggest that, although the original items of TRIPOD were generally covered, AI-specific aspects, particularly those enhancing transparency and ethical integrity, were underreported.

Strengths and Limitations

A key strength of this scoping review is its comprehensive and theory-informed scope, offering what is, to our knowledge, the first broad synthesis of studies applying AI to predict healthcare utilization outcomes across all service types, populations, and healthcare settings. Whereas earlier reviews have addressed broad HEOR topics—occasionally touching on healthcare utilization among other outcomes but without specific focus or discussion of its unique challenges and opportunities[20,23]—this scoping review uses a targeted conceptual framework, integrates diverse data sources, and systematically evaluates adherence to both general and AI-specific reporting standards (TRIPOD + AI),

Figure 4. Adherence of reporting original TRIPOD items versus new AI-specific extensions (in % of included studies).



thereby providing novel insights into current trends, methodological gaps, and underexplored areas in healthcare-utilization prediction. However, several limitations should be acknowledged. Despite dissertations and theses were

considered eligible sources, a systematic search for gray literature was not conducted. This may have modestly constrained the comprehensiveness of the review, particularly regarding unpublished studies or technical reports. Furthermore, although we selected MEDLINE, EconLit, and Scopus for their extensive coverage of biomedical, health services, and multidisciplinary research, the exclusion of other databases may have led to the omission of studies indexed elsewhere. However, MEDLINE and Scopus rank among the largest and most widely used bibliographic databases globally, and we believe that this combination offers a robust and practical approach to capturing literature at the intersection of AI methods and healthcare utilization. This approach was complemented by manual screening of reference lists to identify additional relevant sources. Our search strategy was designed to balance sensitivity and specificity while preserving conceptual relevance. It combined both controlled vocabulary (eg, MeSH terms) and free-text keywords. Although some expressions—such as "readmissions" or "ensemble learning"—were not explicitly included as free-text terms, their underlying concepts may have been retrieved through broader controlled vocabulary indexing (eg, MeSH hierarchies). Nonetheless, we acknowledge the possibility that a small number of relevant studies may have been missed due to these omissions. Finally, there is an inherent bias in the intended uses of AI methods in our review because we specifically focused on predictive studies. This focus might have led to the inclusion of studies in which AI was used during some stage of the prediction process but not necessarily for the prediction task itself.

Implications and Future Directions

The findings of this scoping review highlight key opportunities to advance the development and application of AI-based predictive models for healthcare utilization outcomes. The recent surge in publications suggests that AI in this field of HEOR remains in an exploratory phase, but with a rising interest, driven by greater data availability and AI advances. However, the predominance of US-based studies (62%) limits generalizability to other healthcare systems, underscoring the need for research across more diverse settings. Expanding data collection in other regions could promote more equitable HEOR approaches and enable AI applications tailored to a wider range of healthcare settings.

The limited inclusion of key variables, particularly those affecting the supply side, suggests that current AI models rarely adopt an integrated approach, instead restricting inputs to 1 or a few variable groups. Given likely within-group collinearity, especially as the number of variables increases, it is plausible that breadth—ensuring at least minimal representation from each group—is more informative than adding many variables from a single group, for which the marginal returns diminish. These omissions expose important gaps that may undermine model accuracy and restrict the scope of predicted outcomes. Furthermore, the underrepresentation of equity-relevant variables and population groups raises concerns about algorithmic bias. This includes critical dimensions, such as ethnicity, income, language, or geographic deprivation, whose omission may compromise fairness and limit the external validity of predictions across underserved populations. Models trained on incomplete or biased data risk perpetuating structural inequities, especially when used to inform resource allocation or service planning. Integrating equity considerations in both model development and validation is therefore essential to ensure more just and inclusive AI applications in healthcare.[160]

Unified databases integrating diverse data sources will be required to address these limitations. This will demand coordinated, multidisciplinary efforts, aligning theoretical frameworks on healthcare utilization with AI expertise. These efforts should also include training for researchers and policymakers to encourage the adoption of integrated approaches of emerging literature at the intersection of AI and causal inference, which aims to harness the strengths of AI to solve causal inference processes. Moreover, the reliance on surveys as the main source for capturing user data reveals a gap that could hinder person-centered care approaches. Strengthening data integration and increasing variable diversity are critical steps to improve the robustness and applicability of AI-based predictive models.

Although ensemble models were the most used AI methods, advanced techniques, such as deep learning, were underutilized. This preference for simpler, more interpretable methods may reflect the relatively straightforward structure of data sets or user comfort with established techniques. However, exploring innovative AI methods could yield deeper insights and further improve model performance. Initially, we extracted the AI models from each study exactly as reported by their authors (see Appendix 4 in Supplemental Materials for detailed information). Given the considerable heterogeneity across studies, we subsequently grouped these models into broader families based on similarities in their methodological approach, thereby facilitating synthesis and comparability. Nonetheless, as methods and reporting standards evolve, future reviews may consider finer classifications.

Inconsistencies in reporting, particularly regarding hyperparameter tuning, model calibration, and performance metrics, highlight the need for stricter methodological standards. In addition, the lack of systematic reporting on data quality indicators, such as completeness, missingness, or data provenance, limited our ability to assess how source limitations might affect model performance and generalizability. Encouraging authors to transparently report these attributes will be essential to understand the robustness of AI-based models and guide their appropriate application in real-world healthcare settings.

Adherence to the original TRIPOD guideline (first published in 2015)[161] was relatively high, but compliance with TRIPOD + AI remains limited, likely because of its recent introduction. Promoting TRIPOD + AI could enhance both methodological rigor and ethical standards, aligning AI research with best practices in open science and person-centered care. Although the guideline was primarily designed for clinical prediction models, we found that many of its principles are equally relevant for studies focused on healthcare utilization. At the same time, some items may require contextual interpretation when applied to nonclinical outcomes. This experience highlights the need to assess the broader applicability of TRIPOD + AI to health services research and may inform the development of complementary guidance tailored to these types of predictive models.

Studies primarily focused on predicting hospitalizations and visits, important for resource planning, but overlooked other important aspects of healthcare utilization, including screenings, diagnostic tests, surgical procedures, treatment equipment and devices, nonattendance, and waiting times. Expanding the scope of predicted outcomes would provide a more holistic understanding of healthcare demand. Notably, the lack of studies predicting vaccination uptake, even during the COVID-19 pandemic, highlights a missed opportunity to inform public health preparedness and response. Incorporating preventive services into predictive models is essential to address these gaps. This review identified only 1 study reporting the real-world implementation of AI models,[94] revealing a critical gap between model development and practical application. Addressing this gap is essential to fully realize the potential of AI to improve healthcare delivery and system efficiency. As AI-based prediction tools become more robust and transparent, their integration into health technology assessment processes and resource planning frameworks could support more informed budgeting and policy decisions. Achieving this will require transparent methods, context-specific validation, and alignment with established standards of cost-effectiveness and equity.

Moreover, looking ahead, ensuring the long-term value of predictive models will also require mechanisms for continuous monitoring, recalibration, and governance. As healthcare systems evolve, predictive tools must adapt to shifting population needs, care practices, and data ecosystems. Finally, this review could serve as a foundation for future systematic reviews focused on specific outcomes, healthcare settings, methodological advancements, or predictor theoretical frameworks. Such efforts could further refine and shape future investigations, including scoping reviews in the area of healthcare utilization outcomes beyond predictive studies, and detailed temporal analyses of model characteristics over time. Future reviews could also explore how predictive AI models are distributed across clinical areas or disease groups, offering valuable insights into research priorities and unmet needs in disease-specific service planning.

Conclusions

By mapping the current use of AI in predicting healthcare utilization, this review identifies methodological trends and evidence gaps in this HEOR field. Although AI is increasingly used to predict hospitalizations and visits, important areas such as diagnostic tests and surgical procedures remain underexplored. The findings highlight the need for diverse and integrated datasets, with stronger adherence to TRIPOD + AI guidelines to improve transparency, fairness, and reproducibility. Limited compliance with AI-specific items reflects ongoing challenges in adapting to recent methodological and ethical developments. Future research should broaden predicted outcomes to include process- and logistics-oriented events, such as ambulance arrivals and waiting times, extend applications beyond prediction, such as cohort selection and matching, and explore underused AI methods, including distance-based algorithms and deep neural networks. Strengthening adherence to TRIPOD-AI reporting guidelines is also essential to enhance the reliability and impact of AI in healthcare planning and economic evaluation. By addressing these gaps, this review establishes a foundation for further investigations, including reviews focused on specific healthcare utilization outcomes, healthcare settings, methodological advances, and theory-informed variable selection.

Strengthening these areas will be key to leveraging the full potential of AI in advancing equitable, efficient, and evidence-based healthcare decision making.

Author Disclosures

Author disclosure forms can be accessed below in the Supplemental Material section.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2025.08.007.

Article and Author Information

Accepted for Publication: August 4, 2025

Published Online: September 17, 2025

doi: https://doi.org/10.1016/j.jval.2025.08.007

Author Affiliations: Institute for Advanced Research in Business and Economics, Public University of Navarre, Pamplona, Spain (Gallego-Moll, Arocena, Zaleta-del-Olmo); Department of Econometrics, Statistics and Applied Economics, Universitat de Barcelona, Barcelona, Spain (Gallego-Moll, Maynou); Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain (Gallego-Moll, Carrasco-Ribelles, Casajuana, Violán); Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Bellaterra, Spain (Casajuana, Violán); Department of Health Policy, London School of Economics and Political Science, London, England, UK (Maynou); Center for Research in Health and Economics, Universitat Pompeu Fabra, Barcelona, Spain (Maynou); Direcció d'Atenció Primària Metropolitana Nord, Institut Català de Salut, Badalona, Spain (Violán); Gerència Territorial Barcelona, Institut Català de la Salut, Barcelona, Spain (Zabaleta-del-Olmo); Nursing Department, Faculty of Nursing, Universitat de Girona, Girona, Spain (Zabaleta-del-Olmo).

Correspondence: Carlos Gallego-Moll, MSc, Department of Econometrics, Statistics and Applied Economics, Universitat de Barcelona, John Maynard Keynes Street, 1-11, Faculty of Economics, Third floor, Spain 08034. Email: carlosgallego@ub.edu

Authorship Confirmation: All authors certify that they meet the ICMJE criteria for authorship.

Funding/Support: This research was supported by the Spanish Ministry of Science, Innovation and Universities (MCIU), under grant agreement no. CNS2023-144351.

Acknowledgment: The authors appreciate the review of the English text by Patryk Bialoskorski, MA.

Data Availability: All data supporting the results and conclusions of this study are available in the supplementary information. Specifically, Appendix 4 in Supplemental Materials contains the raw extracted parameters of the included studies, whereas Appendix 6 in Supplemental Materials provides the TRIPOD + Artificial Intelligence (AI) extraction. The data have not been archived in external repositories because they do not originate from a single data set but rather consist of extracted information from multiple published studies. Given that the data set is a synthesis of previously published research, it does not constitute original raw data that can be deposited in a dedicated data repository. However, the supplementary materials provide full transparency for replicating and verifying our findings.

REFERENCES

1. Holtorf AP, Brixner D, Bellows B, Keskinaslan A, Dye J, Oderda G. Current and future use of HEOR data in healthcare decision-making in the United States and in emerging markets. *Am Health Drug Benefits*. 2012;5(7):428–438.
2. Carrasquillo O. Health care utilization. In: Gellman MD, Turner JR, eds. *Encyclopedia of Behavioral Medicine*. New York, NY: Springer; 2013:909–910.
3. Sarma S, Simpson W. A microeconometric analysis of Canadian health care utilization. *Health Econ*. 2006;15(3):219–239.
4. Andersen R, Newman JF. Societal and individual determinants of medical care utilization in the United States. *Milbank Mem Fund Q Health Soc*. 1973;51(1):95–124.
5. Andersen RM. National health surveys and the behavioral model of health services use. *Med Care*. 2008;46(7):647–653.
6. Anderson JG. Health services utilization: framework and review. *Health Serv Res*. 1973;8(3):184–199.
7. Ahmedani BK, Stewart C, Simon GE, et al. Racial/ethnic differences in health care visits made before suicide attempt across the United States. *Med Care*. 2015;53(5):430–435.
8. Manuel JI. Racial/ethnic and gender disparities in health care use and access. *Health Serv Res*. 2018;53(3):1407–1429.
9. Canedo JR, Miller ST, Schlundt D, Fadden MK, Sanderson M. Racial/ethnic disparities in diabetes quality of care: the role of healthcare access and socioeconomic status. *J Racial Ethn Health Disparities*. 2018;5(1):7–14.
10. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. *J Fam Pract*. 2000;49(2):147–152.
11. Koopmans GT, Lamers LM. Gender and health care utilization: the role of mental distress and help-seeking propensity. *Soc Sci Med*. 2007;64(6):1216–1230.
12. Freid VM, Bernstein AB, Bush MA. Multiple chronic conditions among adults aged 45 and over: trends over the past 10 years. *NCHS Data Brief*. 2012;100:1–8.
13. Douthit N, Kiv S, Dwolatzky T, Biswas S. Exposing some important barriers to health care access in the rural USA. *Public Health*. 2015;129(6):611–620.
14. Davis K, Ballreich J. Equitable access to care– how the United States ranks internationally. *N Engl J Med*. 2014;371(17):1567–1570.
15. Squires D, Anderson C. U.S. health care from a global perspective: spending, use of services, prices, and health in 13 countries. *Issue Brief (Commonw Fund)*. 2015;15:1–15.
16. Ashton CM, Haidet P, Paterniti DA, et al. Racial and ethnic disparities in the use of health services: bias, preferences, or poor communication? *J Gen Intern Med*. 2003;18(2):146–152.
17. Flores G. Language barriers to health care in the United States. *N Engl J Med*. 2006;355(3):229–231.
18. Sealy-Jefferson S, Vickers J, Elam A, Wilson MR. Racial and ethnic health disparities and the Affordable Care Act: a status update. *J Racial Ethn Health. Disparities*. 2015;2(4):583–588.
19. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare [Journal]. *Future Healthc J*. 2019;6(2):94–98.
20. Lee W, Schwartz N, Bansal A, et al. A scoping review of the use of machine learning in health economics and outcomes research: part 2-data from nonwearables. *Value Health*. 2022;25(12):2053–2061.
21. Athey S. Beyond prediction: using big data for policy problems. *Science*. 2017;355(6324):483–485.
22. Athey S. *The impact of machine learning on economics*. University of Chicago Press; 2018:507–547.
23. Lee W, Schwartz N, Bansal A, et al. A scoping review of the use of machine learning in health economics and outcomes research: part 1-data from wearable devices. *Value Health*. 2023;26(2):292–299.
24. Romstorfer G, Parragh S, Schneckenreither G, Landsiedl M, Einzinger P, Scheuringer M. Integration of GIS data in health care utilization. *Simul Notes Eur*. 2011;21(3-4):141–146.
25. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378.
26. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. *JBI Evid Synth*. 2020;18(10):2119–2126.
27. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467–473.
28. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210.
29. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*. 2021;372:n160.
30. AlSaad R, Malluhi Q, Janahi I, Boughorbel S. Predicting emergency department utilization among children with asthma using deep learning models. *Healthc Anal*. 2022;2:100050.
31. Alloghani M, Aljaaf A, Hussain A, et al. Implementation of machine learning algorithms to create diabetic patient re-admission profiles. *BMC Med Inform Decis Mak*. 2019;19(suppl 9):253.
32. Alturki L, Aloraini K, Aldughayshim A, Albahli S. *Predictors of readmissions and length of stay for diabetes related patients*. Abu Dhabi, United Arab Emirates: Paper presented at: IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA); 2019. https://ieeexplore.ieee.org/document/9035280. Accessed September 8, 2025.
33. Berger JS, Haskell L, Ting W, et al. Evaluation of machine learning methodology for the prediction of healthcare resource utilization and healthcare costs in patients with critical limb ischemia-is preventive and personalized approach on the horizon? *EPMA J*. 2020;11(1):53–64.
34. Bhavsar NA, Gao A, Phelan M, Pagidipati NJ, Goldstein BA. Value of neighborhood socioeconomic status in predicting risk of outcomes in studies that use electronic health record data. *JAMA Netw Open*. 2018;1(5):e182716.

35. Bory C, Schmutte T, Davidson L, Plant R. Predictive modeling of service discontinuation in transitional age youth with recent behavioral health service use. *Health Serv Res*. 2022;57(1):152–158.
36. Bose S, Kenyon CC, Masino AJ. Personalized prediction of early childhood asthma persistence: a machine learning approach. *PLoS One*. 2021;16(3):e0247784.
37. Brown AD, Kachura JR. Natural language processing of radiology reports in patients with hepatocellular carcinoma to predict radiology resource utilization. *J Am Coll Radiol*. 2019;16(6):840–844.
38. Buchman TG, Kubos KL, Seidler AJ, Siegforth MJ. A comparison of statistical and connectionist models for the prediction of chronicity in a surgical intensive care unit. *Crit Care Med*. 1994;22(5):750–762.
39. Chen S, Bergman D, Miller K, Kavanagh A, Frownfelter J, Showalter J. Using applied machine learning to predict healthcare utilization based on socioeconomic determinants of care. *Am J Manag Care*. 2020;26(1):26–31.
40. Coppa K, Kim EJ, Oppenheim MI, Bock KR, Zanos TP, Hirsch JS. Application of a machine learning algorithm to develop and validate a prediction model for ambulatory non-arrivals. *J Gen Intern Med*. 2023;38(10):2298–2307.
41. DeCenso B, Duber HC, Flaxman AD, Murphy SM, Hanlon M. Improving hospital performance rankings using discrete patient diagnoses for risk adjustment of outcomes. *Health Serv Res*. 2018;53(2):974–990.
42. Edgcomb J, Shaddox T, Hellemann G, Brooks 3rd JO. High-risk phenotypes of early psychiatric readmission in bipolar disorder with comorbid medical illness. *Psychosomatics*. 2019;60(6):563–573.
43. Greenwald S, Chamoun GF, Chamoun NG, et al. Risk stratification Index 3.0, a broad set of models for predicting adverse events during and after hospital admission. *Anesthesiology*. 2022;137(6):673–686.
44. Grinspan ZM, Shapiro JS, Abramson EL, Hooker G, Kaushal R, Kern LM. Predicting frequent ED use by people with epilepsy with health information exchange data. *Neurology*. 2015;85(12):1031–1038.
45. Grinspan ZM, Patel AD, Hafeez B, Abramson EL, Kern LM. Predicting frequent emergency department use among children with epilepsy: a retrospective cohort study using electronic health data from 2 centers. *Epilepsia*. 2018;59(1):155–169.
46. Guntuku SC, Schwartz HA, Kashyap A, et al. Variability in Language used on Social Media prior to Hospital Visits. *Sci Rep*. 2020;10(1):4346.
47. Guo Y, Yin S, Chen S, Ge Y. Predictors of underutilization of lung cancer screening: a machine learning approach. *Eur J Cancer Prev*. 2022;31(6):523–529.
48. Hassanpour S, Langlotz CP. Predicting high imaging utilization based on initial radiology reports: a feasibility study of machine learning. *Acad Radiol*. 2016;23(1):84–89.
49. Heyming TW, Knudsen-Robbins C, Feaster W, Ehwerhemuepha L. Criticality index conducted in pediatric emergency department triage. *Am J Emerg Med*. 2021;48:209–217.
50. Jorge AM, Smith D, Wu Z, et al. Exploration of machine learning methods to predict systemic lupus erythematosus hospitalizations. *Lupus*. 2022;31(11):1296–1305.
51. Jung D, Pollack HA, Konetzka RT. Predicting hospitalization among Medicaid home- and community-based services users using machine learning methods. *J Appl Gerontol*. 2023;42(2):241–251.
52. Karnuta JM, Golubovsky JL, Haeberle HS, et al. Can a machine learning model accurately predict patient resource utilization following lumbar spinal fusion? *Spine J*. 2020;20(3):329–336.
53. Kasturi SN, Park J, Wild D, Khan B, Haggstrom DA, Grannis S. Predicting COVID-19–related health care resource utilization across a statewide patient population: model development study. *J Med Internet Res*. 2021;23(11):e31337.
54. Keller MS, Qureshi N, Albertson E, et al. Comparing risk prediction models aimed at predicting hospitalizations for adverse drug events in community dwelling older adults: a protocol paper [Preprint]. *Research Square*; 2023. https://www.researchsquare.com/article/rs-2429369/v1. Accessed September 8, 2025.
55. Khazanchi R, Bajaj A, Shah RM, et al. Using machine learning and deep learning algorithms to predict postoperative outcomes following anterior cervical discectomy and fusion. *Clin Spine Surg*. 2023;36(3):143–149.
56. LaFaro RJ, Pothula S, Kubal KP, et al. Neural network prediction of ICU length of stay following cardiac surgery based on pre-incision variables. *PLoS One*. 2015;10(12):e0145395.
57. Lin WC, Goldstein IH, Hribar MR, Sanders DS, Chiang MF. *Predicting wait times in pediatric ophthalmology outpatient clinic using machine learning*. Washington, DC: Paper presented at: AMIA Annual Symposium Proceedings; 2019. https://ohsu.elsevierpure.com/en/publications/predicting-wait-times-in-pediatric-ophthalmology-outpatient-clini. Accessed September 8, 2025.
58. Lu Y, Khazi ZM, Agarwalla A, Forsythe B, Taunton MJ. Development of a machine learning algorithm to predict nonroutine discharge following unicompartmental knee arthroplasty. *J Arthroplasty*. 2021;36(5):1568–1576.
59. Morel D, Yu KC, Liu-Ferrara A, Caceres-Suriel AJ, Kurtz SG, Tabak YP. Predicting hospital readmission in patients with mental or substance use disorders: a machine learning approach. *Int J Med Inform*. 2020;139:104136.
60. Mueller L, Berhanu P, Bouchard J, et al. Application of machine learning models to evaluate hypoglycemia risk in type 2 diabetes. *Diabetes Ther*. 2020;11(3):681–699.
61. Nguyen NH, Patel S, Gabunilas J, et al. Simplified machine learning models can accurately identify high-need high-cost patients with inflammatory bowel disease. *Clin Transl Gastroenterol*. 2022;13(7):e00507.
62. Park Y, Hu J, Singh M, et al. Comparison of methods to reduce bias from clinical prediction models of postpartum depression. *JAMA Netw Open*. 2021;4(4):e213909.
63. Phillips-Wren G, Sharkey P, Dy SM. Mining lung cancer patient data to assess healthcare resource utilization. *Expert Syst Appl*. 2008;35(4):1611–1619.
64. Rajkomar A, Yim JW, Grumbach K, Parekh A. Weighting Primary Care Patient Panel size: a novel electronic health record-derived measure using machine learning. *JMIR Med Inform*. 2016;4(4):e29.
65. Ricket IM, MacKenzie TA, Emond JA, Ailawadi KL, Brown JR. Can diverse population characteristics be leveraged in a machine learning pipeline to predict resource intensive healthcare utilization among hospital service areas? *BMC Health Serv Res*. 2022;22(1):847.
66. Ricket IM, Matheny ME, MacKenzie TA, Emond JA, Ailawadi KL, Brown JR. Novel integration of governmental data sources using machine learning to identify super-utilization among U.S. counties. *Intell-Based Med*. 2023;7:100093.
67. Riester MR, McAuliffe L, Collins C, Zullo AR. Development and validation of the Tool for Pharmacists to Predict 30-day hospital readmission in patients with Heart Failure (ToPP-HF). *Am J Health Syst Pharm*. 2021;78(18):1691–1700.
68. Rutledge R, Osler T. The ICD-9-Based illness severity score: a new model that outperforms both DRG and APR-DRG as predictors of survival and resource utilization. *J Trauma*. 1998;45(4):791–799.
69. Sarthak SS, Tripathi SP. *EmbPred30: Assessing 30-Days Readmission for Diabetic Patients Using Categorical Embeddings*. Republic of India: Paper presented at: Smart Innovations in Communication and Computational Sciences: Proceedings of ICSICCS; January 2021. Dr R.M.L. Avadh University; 2020. https://doi.org/10.48550/arXiv.2002.11215. Accessed September 8, 2025.
70. Sayed M, Riaño D, Villar J. Predicting duration of mechanical ventilation in acute respiratory distress syndrome using supervised machine learning. *J Clin Med*. 2021;10(17):3824.
71. Schlairet MC, Heddon MA, Randolph J. Predicting survivorship appointment nonattendance in a community cancer center: a machine-learning approach. *West J Nurs Res*. 2023;45(7):607–617.
72. Shashikumar SP, Wardi G, Paul P, et al. Development and prospective validation of a deep learning algorithm for predicting need for mechanical ventilation. *Chest*. 2021;159(6):2264–2273.
73. Shiner B, D'Avolio LW, Nguyen TM, et al. Measuring use of evidence based psychotherapy for posttraumatic stress disorder. *Adm Policy Ment Health*. 2013;40(4):311–318.
74. Smith D, Mutic A, Mac VVT, Hertzberg VS, McCauley LA. Analyzing the predictors of health care utilization in the agricultural worker population using decision tree analysis: does language matter? *Public Health Nurs*. 2021;38(1):56–63.
75. Tariq A, Celi LA, Newsome JM, et al. Patient-specific COVID-19 resource utilization prediction using fusion AI model. *NPJ Digit Med*. 2021;4(1):94.
76. Thornblade LW, Flum DR, Flaxman AD. Predicting future elective colon resection for diverticulitis using patterns of health care utilization. *EGEMS (Wash DC)*. 2018;6(1):1.
77. Toltzis P, Soto-Campos G, Shelton CR, et al. Evidence-based pediatric outcome predictors to guide the allocation of critical care resources in a mass casualty event. *Pediatr Crit Care Med*. 2015;16(7):e207–e216.
78. Varghese BA, Shin H, Desai B, et al. Predicting clinical outcomes in COVID-19 using radiomics on chest radiographs. *Br J Radiol*. 2021;94(1126):20210221.
79. Vázquez AL, Chou T, Navarro Flores CM, Barrett TS, Villodas MT, Domenech Rodríguez MM. High value correlates of caregiver reported counseling service need and utilization for adolescents at-risk for childhood maltreatment and neglect. *PLoS One*. 2021;16(10):e0258082.
80. Walczak S, Velanovich V. Predicting elective surgical patient outcome destination based on the preoperative modified frailty index and laboratory values. *J Surg Res*. 2022;275:341–351.
81. Wilson FA, Zallman L, Pagán JA, et al. Comparison of use of health care services and spending for unauthorized immigrants vs authorized immigrants or US citizens using a machine learning model. *JAMA Netw Open*. 2020;3(12):e2029230.
82. Wong ES, Schuttner L, Reddy A. Does machine learning improve prediction of VA primary care reliance? *Am J Manag Care*. 2020;26(1):40–44.
83. Yang Y, Yu J, Liu S, Wang H, Dresden S, Luo Y. Predicting avoidable emergency department visits using the NHAMCS dataset. *AMIA Jt Summits Transl Sci Proc*. 2022;2022:514–523.
84. Alser O, Dorken-Gallastegi A, Proaño-Zamudio JA, et al. Using the Field Artificial Intelligence Triage (FAIT) tool to predict hospital critical care resource utilization in patients with truncal gunshot wounds. *Am J Surg*. 2023;226(2):245–250.
85. Amsalu R, Oltman SP, Medvedev MM, et al. Predicting the risk of 7-day readmission in late preterm infants in California: a population-based cohort study. *Health Sci Rep*. 2023;6(1):e994.
86. Bensken WP, Vaca GF, Williams SM, et al. Disparities in adherence and emergency department utilization among people with epilepsy: a machine learning approach. *Seizure*. 2023;110:169–176.

87. Evans L, Wu Y, Xi W, et al. Risk stratification models for predicting preventable hospitalization in commercially insured late middle-aged adults with depression. *BMC Health Serv Res*. 2023;23(1):621.
88. Fanconi C, de Hond A, Peterson D, Capodici A, Hernandez-Boussard T. A Bayesian approach to predictive uncertainty in chemotherapy patients at risk of acute care utilization. *EBioMedicine*. 2023;92:104632.
89. Gonzalez-Suarez AD, Rezaii PG, Herrick D, et al. Using machine learning models to identify factors associated with 30-day readmissions after posterior cervical fusions: a longitudinal cohort study. *Neurospine*. 2024;21(2):620–632.
90. Jamal A. Effect of telemedicine use on medical spending and health care utilization: a machine learning approach. *AJPM Focus*. 2023;2(3):100127.
91. Janczewski CE, Nitkowski J. Predicting mental and behavioral health service utilization among child welfare-involved caregivers: a machine learning approach. *Child Youth Serv Rev*. 2023;155:107150.
92. Limketkai BN, Maas L, Krishna M, et al. Machine learning-based characterization of longitudinal health care utilization among patients with inflammatory bowel diseases. *Inflam Bowel Dis*. 2024;30(5):697–703.
93. McClellan CB. Health care utilization and Expenditures in Health Professional Shortage Areas. *Med Care Res Rev*. 2024;81(4):335–345.
94. Ming DY, Zhao C, Tang X, et al. Predictive modeling to identify children with complex health needs at risk for hospitalization. *Hosp Pediatr*. 2023;13(5):357–369.
95. Panaite V, Finch DK, Pfeiffer P, et al. Predictive modeling of initiation and delayed mental health contact for depression. *BMC Health Serv Res*. 2024;24(1):529.
96. Peacock J, Stanelle EJ, Johnson LC, et al. Using atrial fibrillation burden trends and machine learning to predict near-term risk of cardiovascular hospitalization. *Circ Arrhythm Electrophysiol*. 2024;17(11):e012991.
97. Rezaeiahari M, Brown CC, Eyimina A, et al. Predicting pediatric severe asthma exacerbations: an administrative claims-based predictive model. *J Asthma*. 2024;61(3):203–211.
98. Schuch HS, Furtado M, Silva GFDS, Kawachi I, Chiavegatto Filho ADP, Elani HW. Fairness of machine learning algorithms for predicting foregone preventive dental care for adults. *JAMA Netw Open*. 2023;6(11):e2341625.
99. Seger DL, Amato MG, Frits M, et al. A machine learning technology for addressing medication-related risk in older, multimorbid patients. *Am J Manag Care*. 2024;30(8):e233–e239.
100. Sheffer HF, Bruce M, McLeod C, et al. High risk populations for unplanned healthcare utilization following ostomy construction. *Am J Surg*. 2025;239:115799.
101. Shepherd-Banigan M, Shapiro A, Stechuchak KM, et al. Exploring the importance of predisposing, enabling, and need factors for promoting Veteran engagement in mental health therapy for post-traumatic stress: a multiple methods study. *BMC Psychiatry*. 2023;23(1):372.
102. Shilane D, Lu TH, Zheng Z. In: *Machine learning methods to predict telehealth utilization*. New York, NY, USA: IEEE; 2023:24–27. https://ieeexplore.ieee.org/document/10316021. Accessed September 8, 2025.
103. Tan Y, Dede M, Mohanty V, et al. Forecasting Acute Kidney Injury and Resource Utilization in ICU patients using longitudinal, multimodal models. *J Biomed Inform*. 2024;154:104648.
104. Wu Q, Pajor NM, Lu Y, et al. A latent transfer learning method for estimating hospital-specific post-acute healthcare demands following SARS-CoV-2 infection. *Patterns (N Y)*. 2024;5(11):101079.
105. Moslemi A, Makimoto K, Tan WC, et al. Quantitative CT lung imaging and machine learning improves prediction of emergency room visits and hospitalizations in COPD. *Acad Radiol*. 2023;30(4):707–716.
106. Moslemi A, Hague CJ, Hogg JC, Bourbeau J, Tan WC, Kirby M. Classifying future healthcare utilization in COPD using quantitative CT lung imaging and two-step feature selection via sparse subspace learning with the CanCOLD study. *Acad Radiol*. 2024;31(10):4221–4230.
107. Sidra M, Pietrosanu M, Zwicker J, Johnson DW, Round J, Ohinmaa A. Clinical and socioeconomic predictors of hospital use and emergency department visits among children with medical complexity: a machine learning approach using administrative data. *PLoS One*. 2024;19(10):e0312195.
108. Dovgan E, Gradišek A, Luštrek M, et al. Using machine learning models to predict the initiation of renal replacement therapy among chronic kidney disease patients. *PLoS One*. 2020;15(6):e0233976.
109. Hsu CN, Liu CL, Tain YL, Kuo CY, Lin YC. Machine learning model for risk prediction of community-acquired acute kidney injury hospitalization from electronic health records: development and validation study. *J Med Internet Res*. 2020;22(8):e16903.
110. Huang JS, Chen YF, Hsu JC. *Design of a clinical decision support model for predicting pneumonia readmission*. Taichung, Taiwan: Paper at: International Symposium on Computer Consumer and Control; 2014. https://ieeexplore.ieee.org/document/8268310. Accessed September 8, 2025.
111. Lee JD, Lee TH, Huang YC, et al. Prediction model of early return to hospital after discharge following acute ischemic stroke. *Curr Neurovasc Res*. 2019;16(4):348–357.
112. Goh KH, Wang L, Yeow AYK, et al. Prediction of readmission in geriatric patients from clinical notes: retrospective text mining study. *J Med Internet Res*. 2021;23(10):e26486.
113. Lin AX, Ho AFW, Cheong KH, et al. Leveraging machine learning techniques and engineering of multi-nature features for national daily regional ambulance demand prediction. *Int J Environ Res Public Health*. 2020;17(11):4179.
114. Lo JJ, Tromp J, Ouwerkwerk W, et al. Examining predictors for 6-month mortality and healthcare utilization for patients admitted for heart failure in the acute care setting. *Int J Cardiol*. 2023;390:131237.
115. Tan JK, Quan L, Salim NNM, et al. Machine learning–based prediction for high health care utilizers by using a multi-institutional diabetes registry: model training and evaluation. *JMIR AI*. 2024;3:e58463.
116. Cui L, Xie X, Shen Z. Prediction task guided representation learning of medical codes in EHR. *J Biomed Inform*. 2018;84:1–10.
117. Gu Q, Zheng Q, Zhang X, et al. Trends in health service use for dry eye disease from 2017 to 2021: a real-world analysis of 369,755 outpatient visits. *Transl Vis Sci Technol*. 2024;13(1):17.
118. Wen Z, Wang Y, Chen S, et al. Construction of a predictive model for postoperative hospitalization time in colorectal cancer patients based on interpretable machine learning algorithm: a prospective preliminary study. *Front Oncol*. 2024;14:1384931.
119. Barsasella D, Gupta S, Malwade S, et al. Predicting length of stay and mortality among hospitalized patients with type 2 diabetes mellitus and hypertension. *Int J Med Inform*. 2021;154:104569.
120. Yuniartha DR, Masruroh NA, Herliansyah MK. An evaluation of a simple model for predicting surgery duration using a set of surgical procedure parameters. *Inform Med Unlocked*. 2021:25100633.
121. Dey A, Hay K, Afroz B, et al. Understanding intersections of social determinants of maternal healthcare utilization in Uttar Pradesh, India. *PLoS One*. 2018;13(10):e0204810.
122. Ansari MS, Jain D, Budhiraja S. Machine-learning prediction models for any blood component transfusion in hospitalized dengue patients. *Hematol Transfus Cell Ther*. 2024;46(suppl 5):S13–S23.
123. Abujaber A, Fadlalla A, Gammoh D, Abdelrahman H, Mollazehi M, El-Menyar A. Using trauma registry data to predict prolonged mechanical ventilation in patients with traumatic brain injury: machine learning approach. *PLoS One*. 2020;15(7):e0235231.
124. Alzeer AH, Althemery A, Alsaawi F, et al. Using machine learning to reduce unnecessary rehospitalization of cardiovascular patients in Saudi Arabia. *Int J Med Inform*. 2021;154:104565.
125. Byeon H. Factors influencing the utilization of diabetes complication tests under the COVID-19 pandemic: machine learning approach. *Front Endocrinol (Lausanne)*. 2022;13:925844.
126. Loo WK, Voon W, Suhaimi A, et al. Predictive modeling of COVID-19 readmissions: insights from machine learning and deep learning approaches. *Diagnostics (Basel)*. 2024;14(14):1511.
127. Shadmi E, Flaks-Manov N, Hoshen M, Goldman O, Bitterman H, Balicer RD. Predicting 30-day readmissions with preadmission electronic health record data. *Med Care*. 2015;53(3):283–289.
128. de Korte MH, Verhoeven GS, Elissen AMJ, Metzelthin SF, Ruwaard D, Mikkers MC. Using machine learning to assess the predictive potential of standardized nursing data for home healthcare case-mix classification. *Eur J Health Econ*. 2020;21(8):1121–1129.
129. Kwakernaak S, van Mens K, GROUP Investigators, Cahn W, Janssen R. Using machine learning to predict mental healthcare consumption in non-affective psychosis. *Schizophr Res*. 2020;218:166–172.
130. Weil LI, Zwerwer LR, Chu H, Verhoeff M, Jeurissen PPT, van Munster BC. Identifying future high healthcare utilization in patients with multimorbidity—development and internal validation of machine learning prediction models using electronic health record data. *Health Technol*. 2024;14(3):433–449.
131. Harper PR. A review and comparison of classification algorithms for medical decision making. *Health Policy*. 2005;71(3):315–331.
132. Beaney T, Jha S, Alaa A, et al. Comparing natural language processing representations of coded disease sequences for prediction in electronic health records. *J Am Med Inform Assoc*. 2024;31(7):1451–1462.
133. Davillas A, Jones AM. Biological age and predicting future health care utilisation. *J Health Econ*. 2025;99:102956.
134. Richter A, Truthmann J, Chenot JF, Schmidt CO. Predicting physician consultations for low back pain using claims data and population-based cohort data-an interpretable machine learning approach. *Int J Environ Res Public Health*. 2021;18(22):12013.
135. Ress V, Wild EM. The impact of integrated care on health care utilization and costs in a socially deprived urban area in Germany: a difference-in-differences approach within an event-study framework. *Health Econ*. 2024;33(2):229–247.
136. Wähnke L, Plück J, Bodden M, et al. Acceptance and utilization of web-based self-help for caregivers of children with externalizing disorders. *Child Adolesc Psychiatry Ment Health*. 2024;18(1):40.
137. Launay CP, Rivière H, Kabeshova A, Beauchet O. Predicting prolonged length of hospital stay in older emergency department users: use of a novel analysis method, the Artificial Neural Network. *Eur J Intern Med*. 2015;26(7):478–482.
138. Riis AH, Kristensen PK, Lauritsen SM, Thiesson B, Jørgensen MJ. Using explainable artificial intelligence to predict potentially preventable hospitalizations: a population-based cohort study in Denmark. *Med Care*. 2023;61(4):226–236.
139. Sciannameo V, Goffi A, Maffeis G, et al. A deep learning approach for Spatio-Temporal forecasting of new cases and new hospital admissions of COVID-19 spread in Reggio Emilia, Northern Italy. *J Biomed Inform*. 2022;132:104132.
140. Kumar Y, Ilin A, Salo H, Kulathinal S, Leinonen MK, Marttinen P. Self-supervised forecasting in electronic health records with attention-free models. *IEEE Trans Artif Intell*. 2024;5(8):3926–3938.

141. Henriksson A, Pawar Y, Hedberg P, Nauclér P. Multimodal fine-tuning of clinical language models for predicting COVID-19 outcomes. *Artif Intell Med*. 2023;146:102695.
142. Lo SB, Huber CG, Meyer A, et al. The relationship between psychological characteristics of patients and their utilization of psychiatric inpatient treatment: a cross-sectional study, using machine learning. *PLoS One*. 2022;17(4):e0266352.
143. Betts KS, Kisely S, Alati R. Predicting postpartum psychiatric admission using a machine learning approach. *J Psychiatr Res*. 2020;130:35–40.
144. Linardon J, Fuller-Tyszkiewicz M, Shatte A, Greenwood CJ. An exploratory application of machine learning methods to optimize prediction of responsiveness to digital interventions for eating disorder symptoms. *Int J Eat Disord*. 2022;55(6):845–850.
145. Gould DJ, Bailey JA, Spelman T, Bunzli S, Dowsey MM, Choong PFM. Predicting 30-day readmission following total knee arthroplasty using machine learning and clinical expertise applied to clinical administrative and research registry data in an Australian cohort. *Arthroplasty*. 2023;5(1):30.
146. Wollenstein-Betech S, Silva AAB, Fleck JL, Cassandras CG, Paschalidis IC. Physiological and socioeconomic characteristics predict COVID-19 mortality and resource utilization in Brazil. *PLoS One*. 2020;15(10):e0240346.
147. Soares FM, da Rocha Carvalho Rosa LO, Cecatti JG, et al. Design, construction, and validation of obstetric risk classification systems to predict intensive care unit admission. *Int J Gynecol Obstet*. 2024;167(3):1243–1254.
148. Bezu S, Binyaruka P, Mæstad O, Somville V. Pay-for-performance reduces bypassing of health facilities: evidence from Tanzania. *Soc Sci Med*. 2021;268:113551.
149. Yang Y, Madanian S, Parry D. Enhancing health equity by predicting missed appointments in health care: machine learning study. *JMIR Med Inform*. 2024;12:e48273.
150. Goic M, Bozanic-Leal MS, Badal M, Basso LJ. COVID-19: short-term forecast of ICU beds in times of crisis. *PLoS One*. 2021;16(1):e0245272.
151. Squires DA. Explaining high health care spending in the United States: an international comparison of supply, utilization, prices, and quality. *Issue Brief (Commonw Fund)*. 2012;10:1–14.
152. Phillips KA, Morrison KR, Andersen R, Aday LA. Understanding the context of healthcare utilization: assessing environmental and provider-related variables in the behavioral model of utilization. *Health Serv Res*. 1998;33(3 Pt 1):571–596.
153. Athey S, Imbens GW. Machine learning methods that economists should know about. *Annu Rev Econ*. 2019;11:685–725.
154. de Jong VMT, Moons KGM, Eijkemans MJC, Riley RD, Debray TPA. Developing more generalizable prediction models from pooled studies and large clustered data sets. *Stat Med*. 2021;40(15):3533–3559.
155. Cheng L, Guo R, Moraffah R, Sheth P, Candan KS, Liu H. Evaluation methods and measures for causal learning algorithms. *IEEE Trans Artif Intell*. 2022;3(6):924–943.
156. Lourenço L, Weber L, Garcia L, Ramos V, Souza J. Machine learning algorithms to estimate propensity scores in health policy evaluation: a scoping review. *Int J Environ Res Public Health*. 2024;21(11):1484.
157. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2(4):230–243.
158. Carrasco-Ribelles LA, Llanes-Jurado J, Gallego-Moll C, et al. Prediction models using artificial intelligence and longitudinal data from electronic health records: a systematic methodological review. *J Am Med Inform Assoc*. 2023;30(12):2072–2082.
159. Kolasa K, Admassu B, Hołownia-Voloskova M, Kędzior KJ, Poirrier JE, Perni S. Systematic reviews of machine learning in healthcare: a literature review. *Expert Rev Pharmacoecon Outcomes Res*. 2024;24(1):63–115.
160. Siddique SM, Tipton K, Leas B, et al. The impact of health care algorithms on racial and ethnic disparities: a systematic review. *Ann Intern Med*. 2024;177(4):484–496.
161. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med*. 2015;13:1.