

Review

Explainable Machine Learning Predictive Models for Surgical Site Infections: Scoping Review

Rui Sun¹, BM; Yao Liu¹, BM; Shuya Lu², MCM; Ermira Tartari³, PhD; Gabriel Birgand^{4,5,6}, PhD; Lin Yang^{2,7}, PhD; Lei Zhou^{1,8}, PhD

¹Public Health Emergency Center, Chinese Center for Disease Control and Prevention, Beijing, China

²School of Nursing, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region, China (Hong Kong)

³Faculty of Health Sciences, University of Malta, Msida, Malta

⁴Cibles et Médicaments des Infections et de l'Immunité – UR 1155, Nantes Université, Nantes, France

⁵Centre d'Appui à la Prévention des Infections Associées aux Soins des Pays de la Loire, Centre Hospitalier Universitaire (CHU) - Le Tourville, Nantes, France

⁶National Institute for Health Research Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, Imperial College London, London, United Kingdom

⁷Research Centre of Textiles for Future Fashion, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region, China (Hong Kong)

⁸National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, Chinese Center for Disease Control and Prevention, Beijing, China

Corresponding Author:

Lin Yang, PhD
School of Nursing
The Hong Kong Polytechnic University
11 Yuk Choi Rd
Hong Kong Special Administrative Region
China (Hong Kong)
Phone: +85227666398
Email: l.yang@polyu.edu.hk

Abstract

Background: Surgical site infections (SSIs) remain a major cause of health care–associated infections, and early prediction is essential for improving patient outcomes. Machine learning (ML) has shown potential for SSI prediction; however, clinical implementation requires models that are both accurate and explainable. Despite recent progress in explainable ML, its clinical application to SSI prediction remains limited.

Objective: This study aimed to map explainable ML models for SSI prediction from a clinical perspective and examine their use of structured and unstructured data across the dimensions of data, methodology, and explanation output.

Methods: We conducted a scoping review following PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) and Joanna Briggs Institute (JBI) guidance, and registered the protocol in PROSPERO. Six databases were searched for eligible studies published from January 2010 onward, without language restrictions. The search was conducted on August 9, 2025, and updated on July 14, 2026. We included studies that developed or validated an explainable ML model for predicting SSI in adults. Two reviewers (RS and YL) independently screened studies and extracted data. Findings were narratively synthesized and presented in evidence maps. Methodological quality was assessed using the PROBAST+AI (Prediction model Risk Of Bias Assessment Tool for prediction models using regression or artificial intelligence) methods.

Results: Overall, 77 studies reporting 98 ML models were included. Most models were prognostic (72/98, 73.5%), whereas 26 focused on postoperative SSI diagnosis. A total of 81.8% (63/77) of studies addressed a single surgical specialty, most commonly gastrointestinal surgery (27/63, 42.9%). Overall, 51.9% (40/77) of studies addressed composite SSI predictions. Among all models, 40% (40/98) were black-box models explained by post hoc methods; SHAP combined with ensemble learning was the leading approach (18/40, 45%). Regression models accounted for half of the inherently interpretable models, interpreted using coefficients. Prognostic models commonly included health and lifestyle (60/72, 83.3%), individual characteristics, and surgical process details (both 57/72, 79.2%); health and lifestyle factors were most frequently important across SSI types.

Diagnostic models commonly included surgical process details (13/26, 50%), administrative codes, and individual characteristics (both 10/26, 38.5%). Key diagnostic predictors varied by SSI types: postoperative clinical interventions predominated for composite SSI; vital signs, postoperative interventions, and administrative codes for superficial SSI; postoperative recovery status for deep SSI; and vital signs for organ-space SSI.

Conclusions: Extending previous reviews focusing on model performance, this review mapped explainability methods and important features in SSI prediction, identifying recurring predictor patterns and substantial methodological heterogeneity across prognostic and diagnostic settings. Incomplete reporting of feature definitions and explanatory rationale, together with limited clinical relevance, constrained clinical interpretation and actionability. Clinician-informed reporting frameworks and validation of explanation fidelity and clinical relevance are needed to improve the trustworthiness and utility of SSI prediction models.

Trial Registration: PROSPERO CRD420251124760; <https://www.crd.york.ac.uk/PROSPERO/view/CRD420251124760>

J Med Internet Res 2026;28:e94617; doi: [10.2196/94617](https://doi.org/10.2196/94617)

Keywords: surgical site infection; machine learning; explainable artificial intelligence; predictive models; scoping review

Introduction

Surgical site infection (SSI) is defined as an infection occurring at or near the surgical incision within 30 days after surgery, or within 1 year in the presence of implants [1]. SSIs are classified as superficial, deep, or organ-space infections according to the depth and anatomical location of the affected tissue [2]. These infections can substantially worsen postoperative outcomes by prolonging hospital stay [3], increasing readmission risk [4], and raising mortality [5]. In addition, the need for further treatment imposes a considerable economic burden on both patients and health care systems [6]. Despite major advances in medicine, SSI remains one of the most common health care-associated infections worldwide [7], with reported incidence ranging from 0.5% to 11% [8-12]. This highlights the need for earlier and more comprehensive SSI surveillance [13-16].

As a pivotal subset of AI, machine learning (ML) leverages iterative algorithms to extract intricate patterns from data, ultimately automating clinical prediction [17]. In earlier predictive modeling studies, generalized linear models [18], particularly logistic regression [19], were widely used. However, because disease-related predictors often span multiple domains, traditional regression-based approaches may not adequately capture the complexity of real-world data, especially when relationships are nonlinear [20]. As a result, more complex models, such as deep neural networks and ensemble learning methods, have emerged and often achieve superior performance on high-dimensional data compared with conventional statistical models [21]. Both generalized regression models and nonlinear models can fall within the broader scope of ML [22].

ML models have been increasingly applied to early SSI detection and prediction, including semiautomated or automated surveillance systems [23,24] and risk calculators [25,26]. These models use large and complex health data sources, including electronic health records [27,28], clinical notes [29,30], and thermal images [31,32]. By reducing the need for manual chart review, which is time-consuming and labor-intensive, ML methods may improve the efficiency and

accuracy of SSI identification and assessment [33,34] and support more precise clinical decision-making.

Several reviews have shown that ML models can achieve strong and robust performance for SSI detection and prediction. A systematic review reported that ML models for general SSI prediction showed good performance, with a median area under the receiver operating characteristic curve (AUC) of approximately 0.79 in external validation [35]. A meta-analysis found that AI methods for detecting SSI from wound images had excellent discriminative ability and may support automated postdischarge follow-up [36]. Two additional reviews also reported improved performance of ML-based prediction models [37,38].

However, these reviews primarily focused on model development and validation and paid limited attention to how predictions are explained. In clinical practice, clinicians need to understand why a model identifies a patient as high risk to reduce errors and guide appropriate interventions [39-41]. Transparency is therefore essential for improving the clinical validity, trustworthiness, and acceptability of ML-based SSI prediction tools and for supporting their implementation in practice [42].

Explainable methods can improve the transparency of ML models [43]. Under the concept of self-explainability, ML models can be broadly divided into inherently interpretable models and post hoc explainable models. Generalized linear models, decision tree models [44], and Bayesian models [45] are typically considered inherently interpretable because their structures and feature effects are relatively transparent and can be visualized [46]. The ability to understand how these models generate predictions is often referred to as interpretability. By contrast, advanced nonlinear models are usually regarded as black-box models because their internal prediction processes are opaque [47]. To address this limitation without sacrificing predictive performance, post hoc explanation techniques have been developed and are now widely used [46]. When combined with these techniques, black-box models can provide human-readable explanations of their predictions [48]. This is often referred to as explainability. Post hoc explainers include model-specific methods, such as Gini importance for random

forests [44], and model-agnostic methods, such as SHAP values [49] and permutation feature importance (PFI) [50]. Together, black-box models and post hoc explainers form post hoc explainable models [46]. However, terminology in this area remains inconsistent, and the terms interpretability and explainability are often used interchangeably [51–53]. To ensure consistency and clarity, this review uses explainability as an umbrella term that includes both explainability and interpretability, consistent with previous studies [52,54].

The explainability approaches described above mainly focus on technical explainability. However, previous studies have emphasized that clinically meaningful explainability should also consider the context of model inputs, the alignment between algorithmic architecture and explanation methods, and the format and interpretive meaning of explanation outputs [55,56]. Using this multidimensional perspective, this scoping review aimed to systematically map how explainability has been implemented and reported in ML-based SSI prediction across three complementary and observable domains: predictors, methods, and outputs. We also sought to identify key evidence and reporting gaps. These evidence maps may provide foundational dimensions and practical entry points for developing a more comprehensive framework to guide the implementation and evaluation of explainability in ML-based SSI prediction.

Methods

Protocol Registration

This scoping review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [57] and the Joanna Briggs Institute (JBI) methodological guidance for scoping reviews [58]. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist is provided in [Checklist 1](#). The protocol was prospectively registered in PROSPERO (CRD420251124760) on September 19, 2025, and subsequently updated on July 14, 2026, to include additional databases and an updated search date. One deviation from the registered protocol occurred: the planned secondary analysis comparing predictor frequency between regression models and black-box ML approaches was not performed because some SSI categories contained too few models for meaningful comparison. All other procedures were conducted as originally specified.

Information Sources

Reporting of the search strategy followed the PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension) [59]. We searched the following electronic databases: MEDLINE (PubMed), Embase, Scopus, Web of Science Core Collection, CINAHL (EBSCOhost), and CNKI. Gray literature was searched separately in Google Scholar (first 200 records screened) and ProQuest Dissertations & Theses Global. Each database was searched independently through its native interface rather than through a single multidatabase platform. Clinical trial

registries such as ClinicalTrials.gov did not apply to this study, as this review focused on ML explainability, which is typically described in both the methods and results sections of predictive modeling studies. We restricted the search to studies published from January 2010 onwards to capture the contemporary era of data-driven ML-based infection prediction. No language or other restrictions were applied. The search strategy was developed without validated filters to maximize sensitivity in this emerging field. A comprehensive search was initially run on August 9, 2025, and then rerun on July 14, 2026, using a refined search strategy to identify recent eligible studies. Reference lists of all retrieved reviews were manually screened to identify additional studies not captured by the database search. Conference proceedings identified were reviewed in full text when available. When full texts were unavailable, corresponding authors were contacted by email.

Search Strategy

An initial limited search of Google Scholar and PubMed was performed to identify relevant records. Text words from titles and abstracts, together with index terms, were used to develop the full search strategy. The final strategy combined controlled vocabulary terms (eg, MeSH terms such as “artificial intelligence,” “machine learning,” and “surgical wound infection”) with free-text terms searched in titles and abstracts (eg, “deep learning,” “neural network*,” and “surgical site infection*”). Terms were organized into 3 concept blocks: AI and machine learning, SSI, and prediction. The PubMed strategy was developed first using MeSH terms and free-text terms, then adapted for Embase (Emtree) and other databases by modifying controlled vocabulary and syntax as appropriate. The initial search strategies of these databases and the subsequent update are provided in [Multimedia Appendix 1](#).

Eligibility Criteria

Because children and adults differ in risk profiles, exposure patterns, and effect sizes [60–62], this review was restricted to adult populations. The definition of adult was based on the age threshold used in each study’s country. The central concept was the explanation of ML models used to predict SSIs. SSIs included superficial, deep, and organ-space infections according to the criteria of the Centers for Disease Control and Prevention (CDC) and National Healthcare Safety Network (NHSN) [2]. To ensure comprehensive coverage, we also explicitly included specific anatomical variants, such as deep sternal wound infections (DSWI). Other wound types, such as chronic wounds, and other postoperative infectious complications, such as pneumonia and urinary tract infection, were excluded.

We included studies that developed or validated at least one modern explainable ML-based prediction model. ML models were defined as models capable of learning predictive patterns from data through iterative, data-driven training and optimization, including feature selection and hyperparameter tuning [17]. Conventional statistical models primarily used for association testing or hypothesis testing were excluded [63]. All included models had to provide

a global explanation using at least one explanatory technique, including both inherently interpretable models and post hoc explainable approaches. Regression-based methods were considered carefully because they may function either as statistical inference tools or as ML prediction models. Following previous work [64], regression models that were knowledge-driven, based on strict assumptions and fixed hyperparameters, were treated as conventional statistical models and excluded. In contrast, regression approaches that incorporated ML characteristics, such as hyperparameter tuning, automated feature selection, or sparsity learning [65], were retained.

We included studies that used structured data or free text to develop predictive models. Studies focused only on nonpredictive tasks or on data types other than structured data or free text were excluded. Purely algorithm-focused studies were also excluded.

Study Screening and Selection

All retrieved records were imported into EndNote for automated deduplication, after which the remaining records were manually reviewed to identify and remove residual duplicates. Two reviewers (RS and YL) independently screened titles and abstracts to identify potentially relevant studies, followed by full-text screening against the eligibility criteria. Any disagreements were resolved by discussion, and a third reviewer (SYL) was consulted when necessary.

Data Charting Process

Based on the Checklist for the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies (CHARMS), a predesigned data charting form was augmented to capture additional predictor and explanation variables specific to machine learning explainability [66]. The form was piloted on the first 10 included studies and revised iteratively. Extracted information included author, year of publication, study objective, setting, participant characteristics (sample size, study period, and surgical procedure), outcome details (classification, definition, and time point of measurement), predictors (source, selection method, collection time window, and model predictors), model development details (model type, algorithm, and performance including discrimination, calibration, and clinical utility), and model explanation details (explainability techniques, visualizations, and outputs).

Because no standardized framework exists for categorizing predictors in SSI prediction models, we developed a structured classification framework based on SSI prevention guidance from the World Health Organization (WHO) and CDC [67,68]. Predictors were organized into three dimensions: patient, surgical, and hospital and contextual factors, across three periods: preoperative, intraoperative, and postoperative. The time point of measurement referred to the reported collection time of candidate predictors.

Models developed using different predictor sets or designed to predict different SSI outcomes were regarded as distinct models. Accordingly, when multiple models developed using the same predictor set or designed to predict

the same SSI outcome were reported within a study, only the explainability information from the best-performing model identified in the original study was extracted to balance predictive performance and explainability. Two reviewers (RS and YL) independently extracted the data in Microsoft Excel, and disagreements were resolved by a third reviewer (SYL).

Synthesis of Results

To summarize the evidence and identify common patterns, we redefined and recategorized model types, predictor categories, and SSI categories. Model type was defined primarily by model purpose. When the original study did not explicitly state the type, two reviewers (RS and YL) classified the model as prognostic or diagnostic based on the temporal relationship between predictor collection and the outcome prediction period, following the operational definition in CHARMS [66] and previous research [69]. Prognostic models estimated the risk of future SSI, whereas diagnostic models assessed the risk of existing SSI. Disagreements were resolved by the third author (SYL).

To better summarize heterogeneous predictors, all candidate predictors were regrouped into clinically meaningful categories based on their definitions and clinical implications, informed by American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) indicator definitions [70] and risk factor taxonomies from previous reviews [13,71,72]. Two reviewers (RS and YL) independently categorized the predictors, and interrater agreement was assessed using Cohen kappa. Disagreements were resolved by the third author (SYL).

Because SSI definitions and risk profiles differ across infection types [73], outcomes were grouped into separate infections and composite infections. Separate infections included superficial, deep, and organ-space infections according to criteria from CDC and NHSN [2]. Composite infections referred to composite SSIs in studies that predicted multiple SSI types within a single ML model.

The review synthesized findings across three dimensions: data-level explainability, methodological explainability, and output-level explainability. An evidence gap map was used to visualize the distribution of evidence across these dimensions. Study characteristics were summarized descriptively and presented in tables. For data-level explainability, the distribution of collection timing and predictor categories across model types was summarized and visualized using bubble matrix plots to assess the clinical interpretability of model inputs.

For methodological explainability, ML models were grouped by algorithm type, and the frequency of explainability techniques was summarized within each algorithm category and displayed in a tree map to show common pairings between ML methods and explainability approaches.

Because ML predictions are generated through learned relationships between input features and outcomes, explainability research often focuses on predictive variables as the

central objects of interpretation [74]. Accordingly, output-level explainability focused on the most important features identified across models. For this purpose, the top 10 features in the global feature-importance rankings were extracted and summarized qualitatively across prognostic and diagnostic models and across SSI types using heatmaps, with color intensity reflecting how often each predictor category appeared among the top-ranked features across all models. All visualizations were created using R (version 4.5.2; R Foundation for Statistical Computing). Patterns identified in the evidence map were interpreted narratively to highlight gaps and inform future research priorities.

Quality and Applicability Assessment

Two reviewers (RS and YL) independently assessed the methodological quality, risk of bias (ROB), and applicability of the included models using the updated PROBAST+AI (Prediction model Risk Of Bias Assessment Tool for prediction models using regression or artificial intelligence) tool [75]. Interrater agreement was evaluated via Cohen kappa (κ), with values interpreted according to Landis and Koch criteria [76]. Disagreements were resolved by a third reviewer (SYL). PROBAST+AI tool comprises 34 signaling

questions across two phases: model development (16 items) and validation (18 items). Studies containing both components were evaluated across both phases. For both phases and applicability domains, overall judgments were summarized as low, high, or unclear concern or bias, where low indicates optimal quality and high denotes significant methodological limitations. All analyses were conducted using R version 4.5.2.

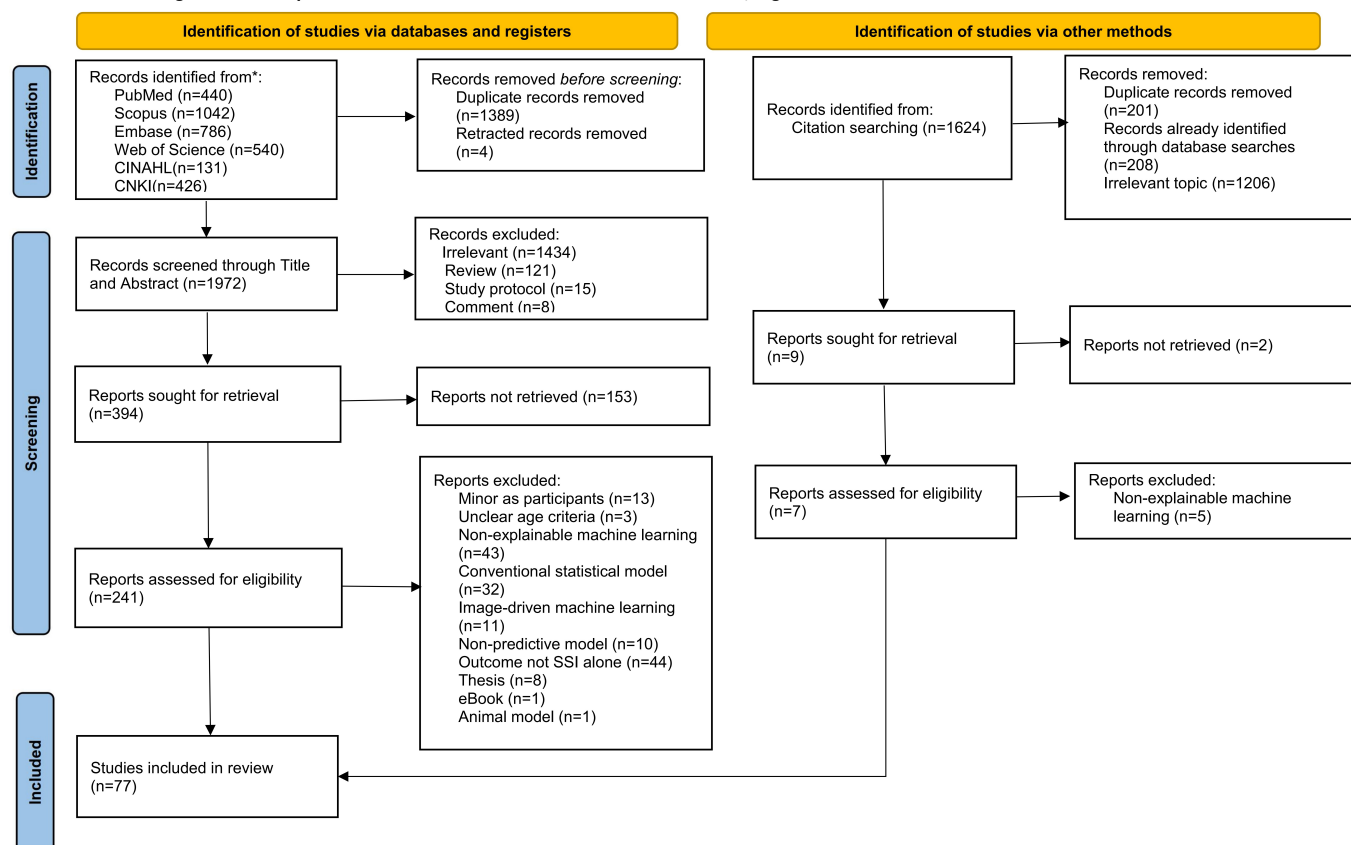
Results

Study Selection

Figure 1 presents the study selection process. The database search yielded 3365 records. After removal of duplicates and retracted articles, 1972 unique records remained for title and abstract screening, and 1571 were excluded. We assessed 394 full-text articles, of which 75 met the inclusion criteria. An additional 2 studies were identified through citation searching of 19 relevant reviews, resulting in a total of 77 included studies [29,30,39,77-150]. Because some studies developed multiple models with different feature sets or for distinct SSI subtypes, 98 ML models were included in the final analysis.

Figure 1. Literature selection process according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.

PRISMA 2020 flow diagram for new systematic reviews which included searches of databases, registers and other sources



Study and Model Characteristics

The characteristics of the 77 included studies [29,30,39,77-150] and 98 models are summarized in Table 1. The studies were conducted across 14 countries and published between 2014 and 2026, with the majority published after 2021. One

study developed both prognostic and diagnostic predictive models [77]. Among the remaining studies, 76.3% developed prognostic models for future SSI risk (58 studies, 71 models) [30,78-82,88-98,105-145], while 18 studies [29,39,83-87,99-104,146-150] developed 25 diagnostic models for detecting existing SSI. Nearly all studies were retrospective, except for

4 prospective cohort studies [78-80,151], and more than half were conducted at a single center. Most studies focused on a single surgical specialty [39,77-80,83,85,87-92,94,96-98,102-109,111-116,119-148,150] (63/77, 81.8%), with gastrointestinal surgery [77-80,83,85,87,94,96,105,108,109,111-115,119-121,126,129,134,139,142,146,148] (27/63, 42.9%) being the most common specialty. The remaining 14 studies [29, 30,81,82,84,86,93,95,99-101,117,118,149] developed general models across multiple surgical procedures. A total of 23 studies [78,83,87,88,93,94,104,105,107,108,110,123,130,133,

134,137-139,141,143-145,148] did not report their targeted SSI types. Four studies [100,118,119,146] addressed multiple SSI types, whereas the remaining 50 studies [29,30,39,77, 79-86,89,90,92,95-99,101-103,106,109,111-115,117,120-122, 124-129,131,132,134-136,140,142,147,149,150] aimed to predict a single SSI outcome, including 39 studies [29,30,39, 79-86,89,90,92,95-97,101-103,106,109,113-115,117,120,121, 124,126,128,129,131,132,135,136,142,149,150] that predicted composite SSI and 11 studies [77,92,98,111,112, 122,125,127,134,140,147] that predicted a specific SSI type.

Table 1. Evidentiary table of 77 selected publications.

Author and year	Country	Study period	Research site	Target surgery	Predictors source	Target SSI ^a	Model development	Best-performing model
Prognostic prediction (n=59)								
Mamlook et al [117] (2023)	Multiple countries	2013 to 2016	Multiple medical centers	Multiple surgical procedures	ACS-NSQIP ^b Data	Single (composite SSI ^c)	Seven algorithms	DNN ^d
Van et al [81] (2014)	Multiple countries	2005 to 2010	Multiple medical centers	Multiple surgical procedures	ACS-NSQIP Data	Single (composite SSI)	Seven algorithms	TEXT-SVM ^e
Walczak et al [95](2019)	United States	July to Dec, 2015	A single medical center	Multiple surgical procedures	ACS-NSQIP Data	Single (composite SSI)	Single algorithm	ANN ^f
Bonde et al [118] (2021)	Multiple countries	2012 to 2018	Multiple medical centers	Multiple surgical procedures	ACS-NSQIP Data	Multiple SSIs (including superficial, deep, and organ-space)	Three predictor sets per outcome	DNN
Bonde et al [119] (2024)	United States	2002 to 2018	Multiple medical centers	Gastrointestinal	ACS-NSQIP Data	Multiple SSIs (including superficial, deep, and organ-space)	Two algorithms plus three predictor sets per outcome	DNN
Chen et al [82] (2020)	China	2014 to 2019	A single medical center	Multiple surgical procedures	Institutional EMR ^g	Single (composite SSI)	Six algorithms	The Self-Attention Network
Zhuang et al [30] (2024)	United States	2013 to 2019	Multiple medical centers	Multiple surgical procedures	Institutional EHR ^h	Single (composite SSI)	Single algorithm	LASSO ⁱ Regression
Ohno et al [114] (2022)	Japan	2000 to 2018	A single medical center	Gastrointestinal	Institutional EMR	Single (composite SSI ^j)	Single algorithm	An Ensemble Model (Gradient Boosting Tree and Neural Network)
Piebpjen et al [109] (2024)	Thailand	2013 to 2019	A single medical center	Gastrointestinal	Hospital and operation databases	Single (composite SSI)	Four algorithms	NB ^k
Yang et al [78] (2024)	China	2021 to 2022	Multiple medical centers	Gastrointestinal	Chinese SSI surveillance cohort	Not reported	Two algorithms	LASSO Regression
Chuyang et al [80] (2017)	Netherlands	2007 to 2009	A single medical center	Gastrointestinal	An open abdominal surgery cohort in the Netherlands	Single (composite SSI) Time to first SSI onset	Three algorithms	An Enhanced Regression Model
Grass et al [120] (2021)	United States	2006 to 2014	A single medical center	Gastrointestinal	ACS-NSQIP data	Single (composite SSI ^l)	Three algorithms	BPMI ^m
Julien et al [79] (2022)	France	2010 to 2017	Multiple medical centers	Gastrointestinal	The REMIND cohort	Single (composite SSI) within 90 days	Single algorithm	RF ⁿ
Chen et al [121] (2023)	Multiple countries	2012 to 2019	Multiple medical centers	Gastrointestinal	ACS-NSQIP data	Single (composite SSI)	Single algorithm	NN ^o
Isbell et al [111] (2021)	United States	2011 to 2017	A single medical center	Gastrointestinal	Institutional EMR	Single (superficial SSI)	Single algorithm	Regularized Bayesian Multilevel Logistic Regression
Chen et al [113] (2024)	China	2012 to 2022	Multiple medical centers	Gastrointestinal	Institutional EMR	5 perioperative complications (including Composite SSI)	Five algorithms	RF
Salimy et al [122] (2025)	Multiple countries	2013 to 2020	Multiple medical centers	Orthopedics	ACS-NSQIP data	Single (periprosthetic Joint Infection [PJI])	Five algorithms	HGB ^p
Wang et al [123] (2021)	China	2012 to 2019	A single medical center	Orthopedics	Institutional EMR	Not reported	Six algorithms	NB
Wang et al [88] (2025)	China	2018 to 2020	A single medical center	Orthopedics	Institutional EMR	Not reported	Six algorithms	XGBoost ^q
Xiong et al [124] (2022)	China	2019 to 2021	A single medical center	Orthopedics	Institutional EHR; Surgical anesthesia system; Mobile nursing system	Single (composite SSI) within 90 days	Seven algorithms	AdaBoost classification trees
Zhang et al [110] (2024)	China	2015 to 2022	A single medical center	Orthopedics	Institutional EMR	Not reported	Seven algorithms	NB
Golinelli et al [97] (2025)	Italy	2017 to 2021	Multiple medical centers	Orthopedics	Routinely collected health care database	Single (composite SSI)	Four algorithms	XGBoost

Author and year	Country	Study period	Research site	Target surgery	Predictors source	Target SSI ^a	Model development	Best-performing model
Orfanoudaki et al [125] (2022)	Multiple countries	2008 to 2017	Multiple medical centers	Cardiac	STS ACSDF data	5 Surgical outcomes (including deep sternal wound infection)	Five algorithms	XGBoost
Kilic et al [98] (2021)	Multiple countries	2007 to 2017	Multiple medical centers	Cardiac	STS ACSF data	7 Surgical outcomes (including deep sternal wound infection)	Single algorithm	XGBoost
McLean et al [126] (2024)	Multiple countries	2014 and 2016	Multiple medical centers	Gastrointestinal	Prospectively secondary analysis of GlobalSurg-1 cohort and GlobalSurg-2 cohort	Single (composite SSI)	Two algorithms	LASSO logistic regression
Wei et al [112] (2020)	United States	2011 to 2016	A single medical center	Gastrointestinal	Institutional EMR	Single (organ-space SSI)	Single algorithm	Regularized Bayesian multilevel logistic regression
Betts et al [89] (2019)	Australia	2009 to 2015	Multiple Medical Centers	Obstetric	Perinatal data collection and Queens Hospital admitted patient data collection	Single (composite SSI)	Single algorithm	XGBoost
Bülow et al [127] (2022)	Sweden	2008 to 2015	Multiple Medical Centers	Orthopedics	The Swedish hip arthroplasty register data	Single (periprosthetic joint infection [PJI])	Single algorithm	LASSO Regression
Cui et al [128] (2025)	China	2013 to 2024	Multiple Medical Centers	Orthopedics	Institutional EHR	Single (composite SSI)	Five algorithms	GBM ^S
Hopkins et al [90] (2020)	United States	2000 to 2015	A single medical center	Orthopedics	Institutional EHR	Single (composite SSI)	Single algorithm	DNN
Li and Yan [129] (2024)	China	2018 to 2021	A single medical center	Gastrointestinal	Institutional EHR	Single (composite SSI)	Three algorithms	DL ^I
Liu et al [130] 2022	China	2010 to 2019	A single medical center	Orthopedics	Institutional EHR	Not reported	Six algorithms	XGBoost
Yeo et al [131] (2023)	United States	2016 to 2019	A single medical center	Orthopedics	Institutional EHR	Single (composite SSI)	Five algorithms	ANN
Hui et al [132] (2023)	China	2018 to 2022	A single medical center	Orthopedics	Institutional EHR	Single (composite SSI)	Ten algorithms	ET ^U
Liang [134] et al (2025)	China	2021 to 2022	A single medical center	Gastrointestinal	Institutional EHR	Single (superficial SSI)	Eight algorithms	RF
Liao et al [91] (2018)	China	2017	A single medical center	Neurosurgery	Institutional EHR	Not reported	Single algorithm	ANN
Hu et al [108] (2024)	China	2018 to 2023	A single medical center	Gastrointestinal	Institutional EHR	Not reported	Two algorithms	DT ^V
Cheng et al [135] (2024)	China	2018 to 2020	A single medical center	Thoracic	Institutional EHR	Single (composite SSI)	Six algorithms	Meta - LASSO Regression
Gutierrez-Naranjo et al [106] (2024)	United States	2014 to 2020	A single medical center	Orthopedics	Institutional EHR	Single (composite SSI)	Four algorithms	Not specified
Wang et al [96] (2025)	China	2020 to 2022	A single medical center	Gastrointestinal	Institutional EHR	Single (composite SSI)	Single algorithm	LASSO Regression
Song and Wei [133] (2025)	China	2023 to 2024	A single medical center	Orthopedics	Institutional EHR	Not reported	Single algorithm	RF
An et al [136] (2023)	China	2017 to 2021	A single medical center	Orthopedics	Institutional EHR	Single (composite SSI)	Single algorithm	LASSO Regression
Choi et al [137] (2024)	Korea	2010 to 2021	Multiple Medical Centers	Orthopedic	Health insurance review and assessment service data	Not reported	Single algorithm	RF
Pang et al [138] (2025)	China	2017 to 2025	Multiple Medical Centers	Orthopedics	Institutional EHR	Not reported	Ten algorithms	SVM
Ghabisha et al [139] (2026)	Yemen	2018 to 2023	A single medical center	Gastrointestinal	Institutional EHR	Not reported	Four algorithms	XGBoost
Cao et al ^W [77] (2026)	China	2020 to 2024	Multiple medical centers	Gastrointestinal	Institutional EHR	Single (organ-space SSI)	Seven algorithms	Stacking ensemble models
De et al [93] (2025)	India	Not reported	Not reported	Multiple surgical procedures	Not reported	Not reported	Four algorithms	RF
Michael et al [115] (2025)	United States	2017 to 2021	Multiple medical centers	Gastrointestinal	Trauma Quality Improvement Program (TQIP) dataset	Single (composite SSI ^X)	Five algorithms	XGBoost
Haider et al [116] (2025)	Japan	2010 to 2024	Multiple medical centers	Orthopedics	Institutional EHR	Single (composite SSI)	Seven algorithms	Stacking ensemble model
Orlandi et al [140] (2022)	Brazil	2017 to 2019	Multiple medical centers	Cardiac	REPLICCAR II database	Single (organ-space SSI)	Two algorithms	LASSO Regression
Kocbek et al [94] (2019)	Norway	2004 to 2012	A single medical center	Gastrointestinal	Institutional EHR	Not reported	Three algorithms	LASSO Regression

Author and year	Country	Study period	Research site	Target surgery	Predictors source	Target SSI ^a	Model development	Best-performing model
Gowd et al [107] (2019)	United States	2005 to 2017	Multiple medical centers	Orthopedics	ACS-NSQIP data	Not reported	Six algorithms	Logistic Regression
Nie et al [141] (2026)	China	2020 to 2024	Two medical centers	Orthopedics	Institutional EMR	Not reported	Eight algorithms	RF
Jha et al [105] (2026)	India	Not reported	A single medical center	Gastrointestinal	Institutional EMR	Not reported	Five algorithms	SVM
Rahimi et al [142](2025)	Iran	2017 to 2024	A single medical center	Gastrointestinal	Institutional EHR	Single (Composite SSI)	Four algorithms	RF
Ying et al [143] (2026)	China	2011 to 2014	A single medical center	Orthopedics	Institutional EHR	Not reported	Six algorithms	LightGBM
Zhang et al [144] (2025)	China	2023 to 2024	A single medical center	Orthopedics	Institutional EHR	Not reported	Eight algorithms	GBM
Zhou et al [145] (2025)	China	2015 to 2020	A single medical center	Head and neck surgery	Institutional EHR	Not reported	Two algorithms	RF
Li et al [92] (2025)	China	2022 to 2023	A single medical center	Orthopedics	Institutional EHR	Single (deep SSI)	Single algorithm	DT
Diagnostic Prediction (n=19)								
Verberk et al [85] (2023)	Sweden	2015 to 2020	A single medical center	Gastrointestinal	Institutional EHR	Single (composite SSI)	Two algorithms	DL
Petrosyan et al [29] (2021)	Canada	2010 to 2015	A single medical center	Multiple surgical procedures	The discharged abstract database and same day surgery database; The physician services database; The Ontario health insurance plan database	Single (composite SSI)	One algorithm plus three predictor sets per outcome	RF
Zhu et al [100] (2021)	United States	2011 to 2017	Two medical centers	Multiple surgical procedures	ACS-NSQIP data	Multiple SSIs (including composite SSI, superficial SSI, and organ-space SSI)	Single algorithm	LASSO Regression
Colborn et al [99] (2023)	United States	2013 to 2019	Multiple medical centers	Multiple surgical procedures	Institutional EHR	Multiple postoperative infections (including composite SSI ^X)	Single algorithm	Regularized Logistic Regression
Kiser et al [84] (2024)	United States	2016 to 2021	A single medical center	Multiple surgical procedures	Enterprise data warehouse (only using EHR)	Single (composite SSI)	Six algorithms	LSTM ^Y
Colborn et al [99] (2018)	United States	2013 to 2016	A single medical center	Multiple surgical procedures	Health data compass (a data warehouse composed of EHR)	Postoperative infection (including composite SSI)	Nine algorithms	LASSO Regression
Ruan et al [146] (2022)	United States	2006 to 2018	A single medical center	Gastrointestinal	ACS-NSQIP data	Multiple SSIs (including superficial, wound infection, and organ-space SSI)	Four algorithms plus three predictor sets per outcome	Multiple GRU-D Model
Weller et al [83] (2018)	United States	2010 to 2013	A single medical center	Gastrointestinal	Institutional EHR	Multiple postoperative complications (The type of SSI is not reported)	Five algorithms	LASSO Regression
Rennert-May et al [103] (2022)	Canada	2013 to 2019	Multiple medical centers	Cardiac	Alberta Health Services System	Single (composite SSI)	Two algorithms	Regularized Logistic Regression
Flores-Balado et al [39] (2023)	Spain	2014 to 2021	Multiple medical centers	Orthopedics	Institutional EHR; Clinical notes	Single (composite SSI)	Single algorithm	XGBoost
Kalishnik et al [147] (2025)	Germany	2007 to 2022	A single medical center	Cardiac	Quality management SAP; THG-QIMS database	Single (deep sternal wound infection [DSWI])	Two algorithms	XGBoost
Yu et al [104] (2014)	China	2005 to 2008	Two medical centers	Cardiac	National health insurance claims data; Health care-associated infection Surveillance data	Not reported	Three algorithms	DT
Xu et al [148] (2020)	China	2015 to 2016	Multiple medical centers	Gastrointestinal	Duke Infection Control Outpatient Surveillance Network (DICONN)	Not reported	Five algorithms	RF
Cao et al ^W [77] (2026)	China	2020 to 2024	Multiple medical centers	Gastrointestinal	Institutional EHR	Single (organ-space SSI)	Seven algorithms	Stacking ensemble model
Li et al [102] (2025)	China	2011 to 2024	A single medical center	Orthopedics	Institutional EMR	Single (composite SSI)	Single algorithm	LASSO regression
Celik et al [87] (2025)	United States	2018 to 2023	A single medical center	Gastrointestinal	Institutional EHR	Not reported	Three algorithms	XGBoost

Author and year	Country	Study period	Research site	Target surgery	Predictors source	Target SSI ^a	Model development	Best-performing model
Perkins et al [149] (2024)	United States	2020 to 2022	A single medical center	Multiple surgical procedures	Institutional EHR	Single (composite SSI)	Single algorithm	LASSO regression
Phuyal et al [150] (2026)	United States	2016 to 2022	Multiple medical centers	Breast surgery	ACS-NSQIP Data	Single (composite SSI)	Single algorithm	XGBoost
Agostinho et al [86] (2025)	Sweden	2016 to 2022	A single medical center	Multiple surgical procedures	Institutional EHR	Single (composite SSI)	Six algorithms	NB and DNN

^aSSI: surgical site infection.
^bACS-NSQIP: American College of Surgeons National Surgical Quality Improvement Program.
^cComposite SSI including superficial, deep, and organ-space SSIs.
^dDNN: deep neural network.
^eSVM: support vector machine.
^fANN: artificial neural network.
^gEMR: electronic medical record.
^hEHR: electronic health record.
ⁱLASSO: least absolute shrinkage and selection operator.
^jComposite SSI including superficial and deep SSIs
^kNB: native Bayes.
^lComposite SSI including deep and organ SSIs
^mBPMI: Bayesian-Probit regression model with multiple-imputation.
ⁿRF: random forest.
^oNN: neural network.
^pHBGT: histogram-based gradient boosting.
^qXGBoost: extreme gradient boosting.
^rSTS ACSD: The Society of Thoracic Surgeons Adult Cardiac Surgery Database.
^sGMB: gradient boosting machine.
^tDL: deep learning.
^uET: extra trees classifier.
^vDT: decision tree.
^wthe two records are from the same study, which developed both prognostic and diagnostic prediction
^xComposite SSI including superficial, deep, organ-space SSIs, and wound disruption.
^yLSTM: long short-term memory.

Data-Level Explainability

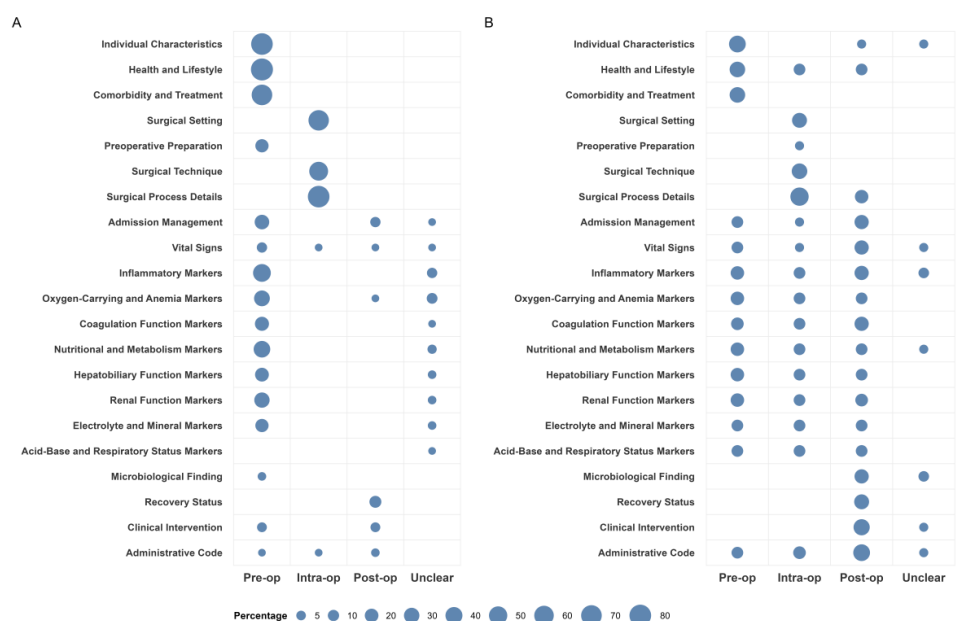
All included models used structured data, typically extracted from electronic health records or registry databases. These predictors were organized into 21 clinically meaningful categories spanning patient, surgery, laboratory, and clinical intervention factors across preoperative, intraoperative, and postoperative phases. Interrater agreement for this classification was high (Cohen κ =0.868). The operational definitions for each category are provided in Multimedia Appendix 2. Only 8 models included free-text data from intraoperative and postoperative clinical records [39,81-87].

Figure 2 illustrates the distribution of variable categories and corresponding collection periods. Among prognostic models, the most frequently represented feature categories were health status and lifestyle collected preoperatively (60/72, 83.3%), followed by individual demographics (57/72, 79.2%), and surgical process details recorded during the surgeries (57/72, 79.2%). Surgery setting information and technique were also regularly used (49/72, 68.1%; 39/72, 54.2%). Other commonly used preoperative features included comorbidity situation (50/72, 69.4%) and inflammatory laboratory indicators (33/72, 45.8%). Features collected postoperatively were generally less frequently represented, which were only included in 11 models across 9 studies [29,88-92,152], and mainly involved recovery status after surgery (9/72, 12.5%) and admission status (5/72, 6.9%). Two studies [93,94] developed predictive models using laboratory

test results alone. One incorporated laboratory test results obtained at three distinct time points during the perioperative period [93], while the other included only preoperative laboratory results [94]. Nevertheless, 6 models constructed in 2 studies [95,96] did not specify the timing of laboratory testing and vital-sign measurement.

Among the diagnostic models, the frequently represented categories included surgical process details (13/26, 50%), administrative codes obtained after surgery (10/26, 38.5%), individual demographics (10/26, 38.5%), and postoperative clinical diagnostic orders and treatments (9/26, 34.6%). Across the eight laboratory test categories, the proportion of diagnostic models incorporating intraoperative or postoperative laboratory results ranged from 11.5% to 23.1%, whereas none of the prognostic models incorporated such data. Five studies relied exclusively on postoperative predictors, most commonly laboratory tests and clinical interventions [29, 99-102]. Two studies developed models based on ICD-9 (*International Classification of Diseases, Ninth Revision*) or ICD-10 (*International Statistical Classification of Diseases, Tenth Revision*) administrative codes alone [84,103]. Additionally, one study developed diagnostic models using longitudinal laboratory data collected from 4 days before surgery to 14 days after surgery [99]. The timing of predictor collection could not be determined for three models because of insufficient reporting [39,87,104].

Figure 2. Bubble plot of predictor categories and collection period in prognostic and diagnostic models. Predictor categories are displayed on the y-axis and data collection periods on the x-axis for (A) prognostic models and (B) diagnostic models. Bubble size represents the percentage of models within the corresponding model type that incorporated each category-period combination, as indicated in the legend. Because individual models could include predictors from multiple categories and collection periods, the percentages are not mutually exclusive. “Unclear” indicates that the collection period was not reported or could not be determined. Pre-op: preoperative period; Intra-op: intraoperative period; Post-op: postoperative period.



Methodological Explainability

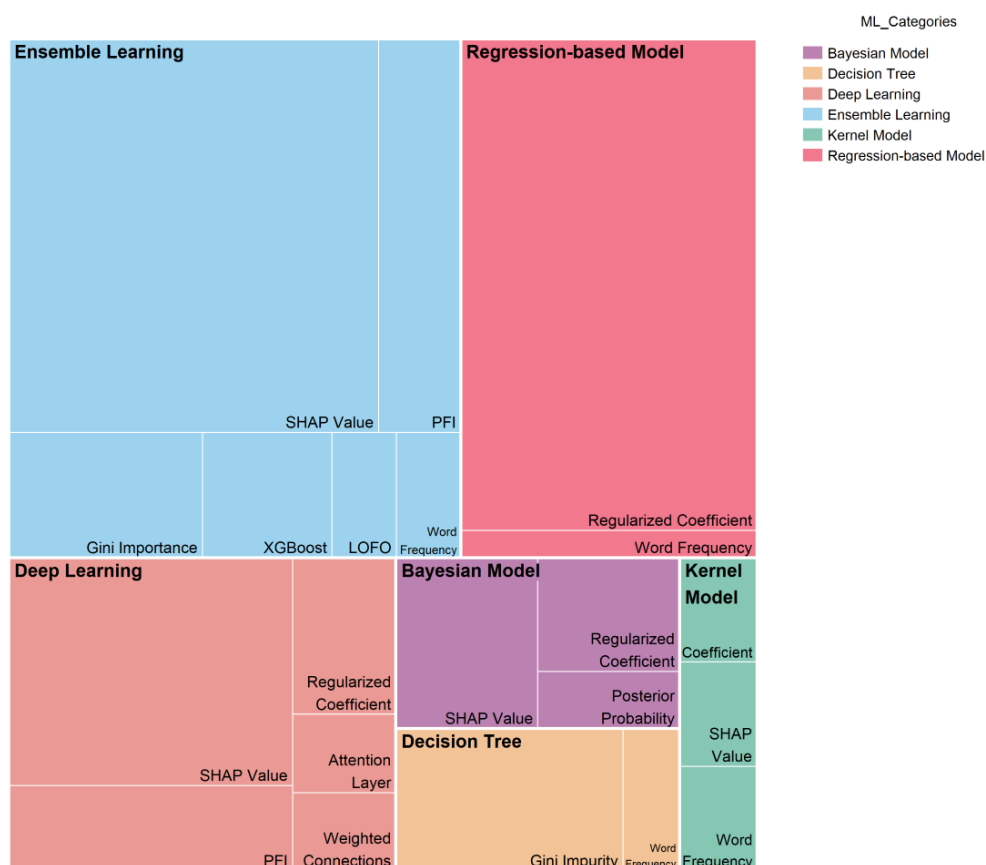
Figure 3 shows the frequency of explainability techniques used across different ML model categories. Overall, 63.3% (62/98) of predictive models were black-box models, with ensemble learning models ranking first (38/62, 61.3%), followed by deep learning (20/62, 32.3%). Ensemble learning models include tree-based bagging, boosting, and heterogeneous stacking. Only 4 black-box models were based on kernel structure. Among these black-box models, 15 models did not specify the explicit explanatory approaches used to produce their explanatory outcomes. Among the 47 models that reported explanatory approaches, the Shapley additive explanation (SHAP) was the most frequently applied method ($n=27$), followed by PFI ($n=7$). Specifically, the combination of SHAP values and ensemble learning models was the most common combination among black-box models reporting specific explanatory approaches (18/47, 38.3%), followed by SHAP values combined with deep learning models (8/47, 17%) and PFI used in ensemble learning models (4/47, 8.5%).

In addition to post hoc explanation methods, two ensemble learning studies used the built-in interpretability of extreme gradient boosting (XGBoost): one assessed feature importance based on how often each feature was used for splitting across trees [97], whereas the other did not specify an importance metric [98]. Three studies explained their models using feature coefficients estimated by regression models [82, 90,105]. Two studies used frequency-based measures: one

kernel model counted the occurrences of free-text terms in the model [81], while the other ensemble learning model recorded the frequency of each feature identified as important across repeated random forest models [79]. One deep learning study incorporated an attention layer to enhance the model's intrinsic explainability [84].

Among the remaining 36 inherently interpretable models, 61.1% (22/36) were regression-based models, followed by Bayesian models (8/36, 22.2%) and decision tree models (6/39, 16.7%). Six models did not report any explanatory approaches or explainability metrics. Among the regression-based models, 18 provided the coefficient of each feature as an explanation, while one model was explained based on the frequency with which each variable was included across 100 cross-validation runs [106]. Three regression models did not report the specific explanation methods or importance metric used [106,107]. Four decision tree models used Gini impurity values or feature splitting location to explain feature importance [92,102,104,108], and one calculated the frequency of free text words [85,108]. Additionally, 3 Bayesian models were explained using SHAP plots, which allowed their feature importance to be compared with other ML models developed in the same paper [86,109,110]. Two Bayesian models were combined with a regression model and hence used coefficients to illustrate the feature importance [111,112].

Figure 3. Distribution of explainable techniques used in different machine learning algorithms. The tree map shows the corresponding relationships between explainable techniques and ML models. Segment size indicates the frequency of each relationship, and the colors differentiate machine learning categories, as indicated by the legend. LOFO: leave-one-feature-out; PFI: permutation feature importance; DT: decision tree; ML: machine learning; SHAP: Shapley additive explanations; XGBoost: extreme gradient boosting.



Output-Level Explainability

Explanations were usually presented as static visualizations (95/101, 94.1%) to show the rank of feature importance or model weight, such as bar charts, heatmaps, word clouds, and tables. Six models established interactive tools, such as nomograms or risk calculators [78,111,113,152].

Figure 4 shows how frequently each predictor category was identified as important in models using multiple variable categories, stratified by SSI type. Eight models using variables from a single category are described separately. In prognostic models, 317 important features were identified for composite SSI, 38 each for superficial and deep SSI, and 55 for organ-space SSI. The most frequently reported important features were related to health status and lifestyle across all four SSI types, with 14.8% (47/317) in composite SSI, 23.7% (9/38) in superficial SSI, 31.6% (12/38) in deep SSI, and 20% (11/55) in organ-space SSI. Beyond this common pattern, the distribution of other important predictor categories varied across SSI types. For composite SSI, surgical process details accounted for 13.8% (44/317) of important features, followed by surgical technique (42/317, 13.2%) and comorbidity and treatment intervention (33/317, 10.4%). For superficial SSI, the second most common top predictors were individual characteristics and comorbidities and treatment (both 6/38, 15.8%), followed by surgical process details and

nutritional and metabolism markers (both 4/38, 10.5%). The critical features for deep SSI prediction consistently focused on comorbidity and treatment (7/38, 18.4%), individual characteristics (5/38, 13.2%), and surgical setting (3/38, 7.9%). For organ-space SSI, the second most common key predictor category was surgical setting (10/55, 18.2%), followed by comorbidities and treatment (7/55, 12.7%). Among those models for unclassified SSI type, inflammatory markers were the most common key predictor classification (20/107, 17.7%), followed by surgical process details (17/107, 15.9%) and nutritional and metabolism markers (13/107, 11.5%). Zero-importance predictor categories varied across 5 SSI types and predominantly involved laboratory indicators.

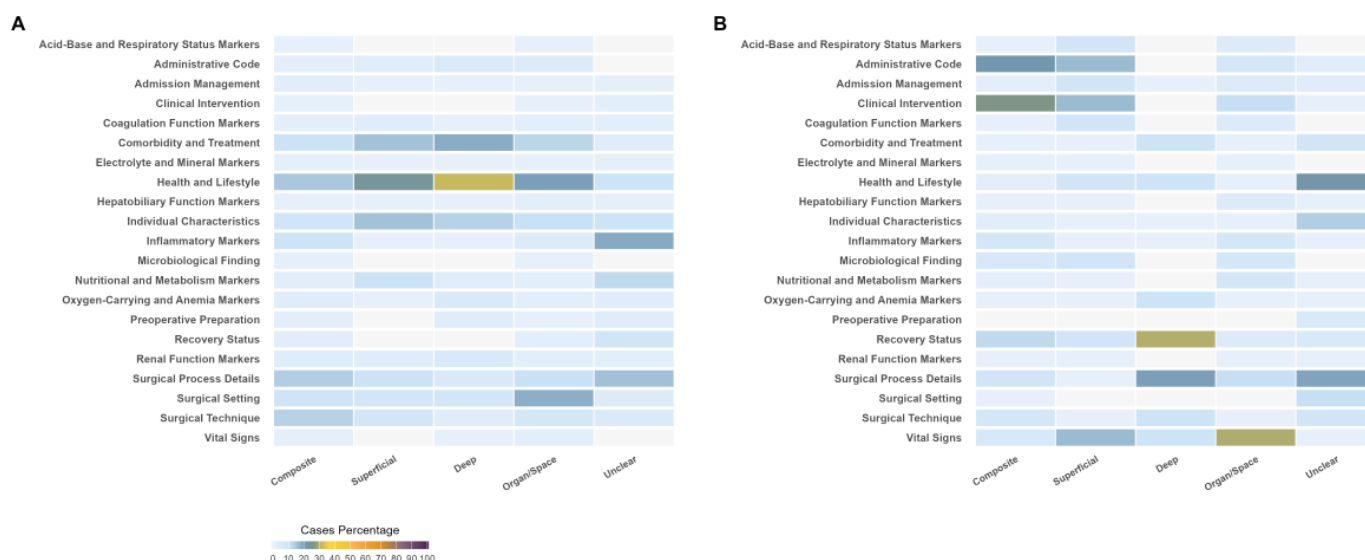
In diagnostic models, 83 features were identified as important for composite SSI, 12 for superficial SSI, 10 for deep SSI, and 27 for organ-space SSI. Postoperative features were more prominent. For composite SSI, clinical interventions were the most frequently reported category, comprising 26.5% (22/83) of important features, followed by administrative codes (18/83, 21.7%) and recovery status (10/83, 12%). For superficial SSI, vital signs, clinical interventions, and administrative codes were most common (2/12, 16.7% each). For deep SSI, recovery status ranked first (3/10, 30%), followed by surgical process details (2/10, 20%). For organ-space SSI, vital signs (8/27, 29.6%)

were widely identified as important predictors, followed by surgical process details and clinical interventions (3/27, 11.1% each). For SSI without clear classification, preoperative and intraoperative metrics, especially health and lifestyle (8/36, 22.2%) and surgical process details (7/36, 19.4%), were the most common predictors. The distribution of zero-importance feature categories varied across SSI types, with patient demographics, health and lifestyle, comorbidities, and laboratory indicators most frequently assigned zero importance.

Two studies [29,103] developed diagnostic models using administrative codes alone. A total of 53 important codes were used and could be grouped into 10 clinically meaningful categories, involving postoperative complications, iatrogenic

injuries, laboratory biomarkers, inpatient services, therapeutic or nutritional interventions, surgical procedures, gastrointestinal diseases, malignant neoplasms, and other systemic diseases. Postoperative complications were the most frequent category, followed by surgical procedures and abdominal diseases. Additionally, one study comprehensively assessed the importance of the preoperative collection period and variables by examining selection frequencies and found that the leukocyte count measured during the medium interval was most important [93]. Another study compared the importance of inflammatory markers measured at 3 collection times during the perioperative period and found that those measured 3 days after surgery were more critical [94].

Figure 4. Frequency distribution of critical features across surgical site infections categories in prognostic model and diagnostic model. Heat maps show the distribution of predictive features in the included studies, stratified by surgical site infections categories, across prognostic model and diagnostic model. The left one (A) indicates the figure of prognostic model, while the right one (B) indicates the diagnostic model. The color gradient denotes the percentage of cases, increasing from light blue (0%) to dark purple (100%), while light grey represents corresponding categories that are not initial input features.



Risk of Bias and Applicability

The interrater agreement between the 2 researchers (RS and YL) was high, with the overall Cohen κ of 0.863 ($P < .01$), and domain-level values ranging from 0.319 to 1. Detailed results are shown in Figure 5 and Multimedia Appendix 3.

The 77 included studies [29,30,39,77-150] comprised 98 eligible models for PROBAST+AI assessment. As one study reported model evaluation alone [114], 98 model development processes and 97 model evaluation processes were ultimately assessed [153].

For model development, 92.8% (91/98) of models were judged to have an overall high-quality concern, and only 7 models had an overall unclear rating [78,80,86,115,116]. Overall, 88.8% of models (87/98) were assessed as high-quality concerns in domain 1 (participants), which was the primary driver of great concern, followed by domain 4 (analysis, 76.5%, 75/98). In domain 1, concerns were mainly related to inappropriate or unclear study design. In domain 4, concerns are often derived from small development datasets

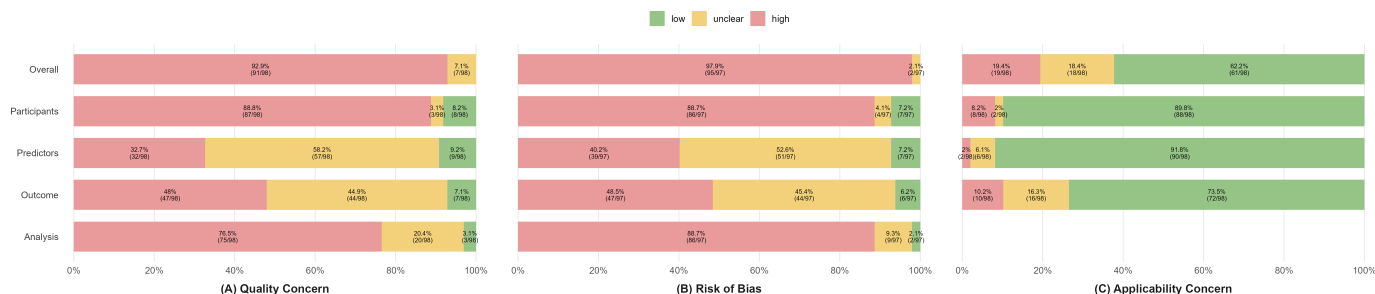
relative to model complexity and the absence of a procedure for handling missing data. Most of domain 2 (predictors) and domain 3 (outcomes) were rated as unclear, with 58.2% (57/98) and 44.9% (44/98), respectively. The main issues were unclear definitions of predictors and outcomes, uncertain prediction horizons, and incomplete reporting of measurement and assessment procedures.

For model evaluation, 97.9% (95/97) of models were classified as high risk of bias, and 2 models were evaluated as having an unclear risk [115]. In development, domains 1 (high risk: 86/97, 88.7%) and 4 (high risk: 86/97, 88.7%) were the main contributors to elevated risk. Most models were internally evaluated, carrying over limitations in design and data preprocessing. Additionally, model performance assessment was largely limited to discrimination, with calibration and clinical utility rarely evaluated, resulting in an incomplete evaluation of model performance. The rating distributions and underlying reasons for domains 2 and 3 during model evaluation were identical to those during model development.

Approximately 62.2% (61/98) of models had low applicability concerns in model applicability, whereas 18

models presented unclear concerns due to ambiguous predictor measurement and outcome assessment procedures.

Figure 5. Summary of risk of bias, quality concern, and applicability concern assessed with PROBAST+AI (Prediction model Risk Of Bias Assessment Tool for prediction models using regression or artificial intelligence). Percentage stacked bar charts summarize ratings of model risk of bias, quality concern, and applicability concern. Numbers on the x-axis (0-100) indicate the proportion (%) of models in each category. Green denotes low risk or concern; yellow denotes risk or concern; red denotes high risk or concern.



Discussion

Principal Findings

This scoping review mapped explainability in ML models for SSI prediction across the data, methodological, and output levels. Previous reviews have shown that explainable ML has the potential to improve SSI surveillance and support earlier clinical action by integrating large and complex health data [35,36]. However, the clinical implementation of ML models depends not only on predictive performance but also on whether clinicians can understand, trust, and act on the explanations they provide.

Overall, explainability practices in ML-based SSI prediction remain in their early stages, with important information gaps across all three levels. At the data level, existing models incorporated predictors from multiple perioperative domains. However, unclear definitions and collection timing limited the assessment of their clinical meaning and suitability for the intended prediction point. At the method level, black-box models commonly use post-hoc techniques such as SHAP and PFI. Inherently interpretable models generally relied on their internal structures, while some studies supplemented them with post-hoc techniques to form hybrid explanation strategies. Nevertheless, some studies did not clearly report the explanation methods used. Available reports also provided limited information on method-selection rationales and intended explanatory purposes, making it difficult to assess compatibility with model architectures and application contexts. At the output level, explanations were dominated by static visualizations of global feature importance, with limited evidence regarding causality, actionability, or user-level effectiveness. Many top-ranked predictors represented inherent patient characteristics or clinical conditions rather than clearly modifiable factors, further limiting their direct translation into clinical interventions. The generally high risk of bias further restricted the credibility and generalizability of these explanations. Thus, current ML explanations help describe model behavior, but available evidence remains insufficient to establish their clinical context and practical utility.

Predictor Selection and Clinical Relevance

The pattern of predictor selection differed meaningfully between prognostic and diagnostic models, with important clinical implications. Prognostic models were dominated by preoperative and intraoperative variables, especially patient characteristics, lifestyle factors, and surgical process details. This is appropriate for models intended to estimate risk when preventive strategies are still being implemented. Diagnostic models, by contrast, more often rely on postoperative variables, particularly recovery status and clinical interventions, which is consistent with their purpose of identifying SSI after it has developed.

Even so, the current evidence suggests that prognostic models may be missing clinically informative postoperative signals. SSI commonly occurs 10 to 20 days after surgery, and many cases emerge after discharge [154,155]. Early postoperative trends in vital signs [156], inflammatory markers [153,157], and recovery status [158] may therefore improve the timeliness and precision of prediction. This suggests that future prognostic models should move beyond static perioperative snapshots and incorporate dynamic postoperative trajectories when clinically appropriate. In parallel, important contextual factors such as operating room environment remain underrepresented, despite evidence that they may contribute to SSI risk [159,160].

Ambiguity in Predictor Definitions and Timing

Several studies did not clearly report the definitions and collection timing of model predictors, which may pose a major barrier to clinical interpretability and actionability. This lack of chronological clarity regarding multidomain features introduces serious uncertainty about whether these predictors would be available at the intended point of clinical use and whether they are clinically plausible drivers of SSI risk, ultimately leading to an unclear definition of model deployment windows and limiting their real-world clinical application [161,162]. This methodological problem is particularly important for potential longitudinal variables such as laboratory indicators and vital signs, whose clinical

meaning and normal ranges can vary substantially depending on the exact time of measurement [157,159,163].

Administrative codes were relatively common in diagnostic models since they are readily available in retrospective datasets. However, they primarily reflect billing practices, institutional workflows, and reimbursement incentives rather than straightforward clinical observations [164,165]. Our findings revealed that most models incorporating administrative codes rarely report their context, including the exact temporal alignment of each medical order. Without explicit temporal records, the codes may include medical interventions targeted at SSI diagnosis and treatment, thereby introducing information leakage or circular reasoning [166] and weakening both reproducibility and cross-institutional interpretability. These findings underscore the need for more standardized reporting of predictor semantics and collection timing.

Selection and Alignment of Explainability Methods

The review found that combinations of ML models and explanatory approaches are considerably diverse, with no consistent correspondence. Among inherently interpretable models, most explanations matched the models' own structures, which are more transparent because they directly reflect their internal logic [47]. A few models adopted hybrid explanation strategies by applying post hoc tools to inherently interpretable models. This strategy facilitates cross-model comparison within a study; however, post hoc explanations may not fully capture the internal logic of inherently interpretable models, thereby obscuring some of their intrinsic transparency [47,167].

For black-box models, aside from a few that attempted to reform the model structure with interpretable modules, most used post hoc explanatory approaches. SHAP and PFI were the most common methods. However, it is noteworthy that these methods answer different questions, and that distinction is often not made explicit. SHAP estimates how a feature contributes to a specific prediction [168], whereas PFI measures how model performance changes when a feature is shuffled [169]. Clinically, SHAP is often more meaningful because it links predictors to individual risk estimates, whereas PFI mainly reflects the model's dependence on a feature. In our review, many studies reported "feature importance" without clarifying the questions the explanation method was intended to answer, which can mislead clinicians and limit practical use. Future studies should therefore consider explainability techniques based on the clinical purpose of the explanation.

Furthermore, the model's intrinsic structure influences the fidelity of post hoc explanatory methods. This issue is especially important for SHAP. Although SHAP is widely used in SSI prediction, its reliability depends on the underlying model. For tree-based models, TreeSHAP can closely reflect the model's decision structure [170]. However, in deep learning, feature interactions are often nonlinear and distributed, making additive explanations less faithful and

more sensitive to assumptions such as the choice of baseline [171,172]. Because these differences are rarely discussed in SSI studies, SHAP-derived outputs are often presented as if they were interchangeable across architectures. Greater attention to method-model alignment is therefore needed to improve the fidelity of explanations.

Current explainable methods primarily provide feature importance rankings or graphical visualizations. These outputs indicate which predictors contributed more but often provide limited context on the meaning and actionability of the explanations [173]. Emerging studies have shown that transforming model outputs into human-centered narrative explanations can improve clinicians' trust and acceptance of AI-assisted decision support [174-176]. Future research should therefore move beyond technical explainability and focus on human-centered explainability. Additionally, incorporating a large language model (LLM) into explanatory approaches, which excels at generating natural and conversational language [177], could assist in translating narratively technical outputs into natural language. Additionally, co-development with clinicians and model developers to achieve participatory system design [176,178] may help bridge the gap between explainability and clinical usability [179].

Clinical Interpretation of Feature Importance

A clinically important finding was the consensus on critical predictors across SSI subtypes, which largely aligns with established SSI risk factors [180-183]. These recurrent predictors may help minimize optimal feature sets and guide future model development [169], thereby providing some reassurance about the face validity of the models. However, since most prediction models were developed using observational data, these feature-importance outputs should be interpreted as associations rather than causal effects. Without causal analysis, there is a risk that nonmodifiable correlates will be mistaken for actionable targets, which may lead to inappropriate clinical inference [184,185]. Beyond ranking important features, future studies should therefore clinically validate model-identified features by assessing their clinical plausibility, causal relevance, modifiability, and actionability in clinical decision-making [179,186]. Counterfactual inference approaches may be better suited for this purpose [187-189].

The apparent lack of importance of several laboratory indicators is somewhat counterintuitive given their clinical relevance after SSI onset [153,157]. However, this may reflect either limited independent predictive value or methodological limitations in the explanation process. Common problems such as class imbalance, collinearity, and incomplete feature engineering can distort feature rankings and obscure relevant predictors [190-192]. Additionally, it is shown that the values and meanings of biomarkers can vary substantially across the perioperative course [193-195], and some eligible studies in this review have shown that the importance of laboratory tests collected from different time windows varied significantly [93,94]. However, the

timing of the laboratory test was reported only as an implicit period rather than an exact time in most included studies, as previously discussed. These features may therefore appear insignificant simply because they were measured at an inappropriate time or evaluated within an unsuitable modeling framework. Future studies should report collection timing clearly, assess temporal sensitivity, and address collinearity explicitly to ensure that explanatory outputs are clinically credible [41].

Limitations

This review has several limitations. First, one protocol amendment should be acknowledged: the omission of one planned subgroup analysis because some SSI categories contained too few models for meaningful comparison. Second, when multiple models were reported in a study, we extracted explainability information only for the model the authors selected as best performing. Because predictive accuracy and explainability are not always aligned, this may have excluded informative alternative models. Third, the definition of “best-performing” varied across studies, which limited direct comparisons. Fourth, this is a rapidly evolving

field, so the review may not capture the most recent developments. Finally, because no established framework exists for quantitatively synthesizing feature-importance outputs across heterogeneous explainability methods, we summarized our findings qualitatively rather than meta-analyzing them.

Conclusions

To our knowledge, this review is the first to focus on how model explainability is implemented and reported in this field. We systematically mapped current practices across three complementary clinical levels: predictors, explanatory techniques, and outputs. The findings revealed incomplete explainability information across all three levels, with current practices remaining largely limited to static, global feature attribution. This review provides baseline evidence on implementation and reporting gaps in explainable ML for SSI prediction and identifies priority dimensions for future evaluation frameworks. Further studies should strengthen reporting at each level and involve clinicians and other intended users in developing and validating explainability evaluation approaches tailored to real-world settings.

Acknowledgments

Disclosure of Delegation to generative AI (GenAI)

The authors declare the use of GenAI in the research and writing process. According to the GAIDeT (Generative AI Delegation Taxonomy; 2025), the following tasks were delegated to GenAI tools under full human supervision:

- Proofreading and editing
- Translation

The GenAI tool used was ChatGPT 5.4-mini.

Responsibility for the final manuscript lies entirely with the authors.

GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Declaration submitted by: Rui Sun

Funding

This study was funded by the Public Health Talent Program (Lei Zhou, No. 01602).

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

Conceptualization: LY, ET, GB

Data curation: RS, YL, SYL

Formal analysis: RS

Funding acquisition: LZ

Investigation: RS, YL, SYL

Methodology: RS, YL, SYL

Project administration & Supervision: LY, ET, GB, LZ

Visualization: RS

Writing – original draft: RS

Writing – review & editing: YL, ET, GB

All authors reviewed and approved the final version of the manuscript.

LY and LZ contributed equally to the strategic oversight of this scoping review as co-corresponding authors. LY, as the senior corresponding author, provided the foundational theoretical framework and institutional resources, while LZ managed the methodological rigor and technical synthesis of the evidence. This collaborative model ensured both clinical depth and methodological precision throughout the review process.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategy.

[[DOCX File \(Microsoft Word File\), 38 KB-Multimedia Appendix 1](#)]

Multimedia Appendix 2

Feature categories.

[[DOCX File \(Microsoft Word File\), 28 KB-Multimedia Appendix 2](#)]

Multimedia Appendix 3

Risk of bias assessment result.

[[DOCX File \(Microsoft Word File\), 247 KB-Multimedia Appendix 3](#)]

Checklist 1

PRISMA-ScR fillable checklist.

[[DOCX File \(Microsoft Word File\), 87 KB-Checklist 1](#)]

Checklist 2

PRISMA_2020_abstract_checklist.

[[DOCX File \(Microsoft Word File\), 270 KB-Checklist 2](#)]

References

1. Horan TC, Gaynes RP, Martone WJ, Jarvis WR, Emori TG. CDC definitions of nosocomial surgical site infections, 1992: a modification of CDC definitions of surgical wound infections. *Infect Control Hosp Epidemiol*. Oct 1992;13(10):606-608. [Medline: [1334988](#)]
2. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. *Am J Infect Control*. Jun 2008;36(5):309-332. [doi: [10.1016/j.ajic.2008.03.002](#)] [Medline: [18538699](#)]
3. Shambhu S, Gordon AS, Liu Y, et al. The burden of health care utilization, cost, and mortality associated with select surgical site infections. *Jt Comm J Qual Patient Saf*. Dec 2024;50(12):857-866. [doi: [10.1016/j.jcjq.2024.08.005](#)] [Medline: [39384467](#)]
4. Eckmann C, Kramer A, Assadian O, et al. Clinical and economic burden of surgical site infections in inpatient care in Germany: a retrospective, cross-sectional analysis from 79 hospitals. *PLoS ONE*. 2022;17(12):e0275970. [doi: [10.1371/journal.pone.0275970](#)] [Medline: [36576894](#)]
5. Fowler AJ, Wan YI, Prowle JR, et al. Long-term mortality following complications after elective surgery: a secondary analysis of pooled data from two prospective cohort studies. *Br J Anaesth*. Oct 2022;129(4):588-597. [doi: [10.1016/j.bja.2022.06.019](#)] [Medline: [35989114](#)]
6. Monahan M, Jowett S, Pinkney T, et al. Surgical site infection and costs in low- and middle-income countries: a systematic review of the economic burden. *PLoS ONE*. 2020;15(6):e0232960. [doi: [10.1371/journal.pone.0232960](#)] [Medline: [32497086](#)]
7. Behnke M, Aghdassi SJ, Hansen S, Diaz LAP, Gastmeier P, Piening B. The prevalence of nosocomial infection and antibiotic use in German hospitals. *Dtsch Arztebl Int*. Dec 15, 2017;114(50):851-857. [doi: [10.3238/arztebl.2017.0851](#)] [Medline: [29271343](#)]
8. Mengistu DA, Alemu A, Abdukadir AA, et al. Global incidence of surgical site infection among patients: systematic review and meta-analysis. *INQUIRY*. 2023;60:469580231162549. [doi: [10.1177/00469580231162549](#)] [Medline: [36964747](#)]
9. Hou Y, Collinsworth A, Hasa F, Griffin L. Incidence and impact of surgical site infections on length of stay and cost of care for patients undergoing open procedures. *Surg Open Sci*. Jan 2023;11:1-18. [doi: [10.1016/j.sopen.2022.10.004](#)] [Medline: [36425301](#)]
10. Lin J, Peng Y, Guo L, et al. The incidence of surgical site infections in China. *J Hosp Infect*. Apr 2024;146:206-223. [doi: [10.1016/j.jhin.2023.06.004](#)] [Medline: [37315807](#)]
11. Seidelman JL, Mantyh CR, Anderson DJ. Surgical site infection prevention: a review. *JAMA*. Jan 17, 2023;329(3):244-252. [doi: [10.1001/jama.2022.24075](#)] [Medline: [36648463](#)]
12. Gillespie BM, Harbeck E, Rattray M, et al. Worldwide incidence of surgical site infections in general surgical patients: a systematic review and meta-analysis of 488,594 patients. *Int J Surg*. Nov 2021;95:106136. [doi: [10.1016/j.ijsu.2021.106136](#)] [Medline: [34655800](#)]

13. Marzoug OA, Anees A, Malik EM. Assessment of risk factors associated with surgical site infection following abdominal surgery: a systematic review. *BMJ Surg Interv Health Technol*. 2023;5(1):e000182. [doi: [10.1136/bmjsit-2023-000182](https://doi.org/10.1136/bmjsit-2023-000182)] [Medline: [37529828](https://pubmed.ncbi.nlm.nih.gov/37529828/)]
14. Abbas M, de Kraker MEA, Aghayev E, et al. Impact of participation in a surgical site infection surveillance network: results from a large international cohort study. *Hosp Infect*. Jul 2019;102(3):267-276. [doi: [10.1016/j.jhin.2018.12.003](https://doi.org/10.1016/j.jhin.2018.12.003)]
15. Liu Y, Liu Y, Yang Z, Wu J, Li J. Risk factors for surgical site infection (SSI) in patients undergoing hysterectomy: a systematic review and meta-analysis. *BMJ Open*. Jun 2025;15(6):e093072. [doi: [10.1136/bmjopen-2024-093072](https://doi.org/10.1136/bmjopen-2024-093072)]
16. Korol E, Johnston K, Waser N, et al. A systematic review of risk factors associated with surgical site infections among surgical patients. *PLoS ONE*. 2013;8(12):e83743. [doi: [10.1371/journal.pone.0083743](https://doi.org/10.1371/journal.pone.0083743)] [Medline: [24367612](https://pubmed.ncbi.nlm.nih.gov/24367612/)]
17. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. *Am J Epidemiol*. Dec 31, 2019;188(12):2222-2239. [doi: [10.1093/aje/kwz189](https://doi.org/10.1093/aje/kwz189)] [Medline: [31509183](https://pubmed.ncbi.nlm.nih.gov/31509183/)]
18. Arnold KF, Davies V, de Kamps M, Tennant PWG, Mbotwa J, Gilthorpe MS. Reflection on modern methods: generalized linear models for prognosis and intervention-theory, practice and implications for machine learning. *Int J Epidemiol*. Jan 23, 2021;49(6):2074-2082. [doi: [10.1093/ije/dyaa049](https://doi.org/10.1093/ije/dyaa049)] [Medline: [32380551](https://pubmed.ncbi.nlm.nih.gov/32380551/)]
19. Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. *J Thorac Dis*. Mar 2019;11(Suppl 4):S574-S584. [doi: [10.21037/jtd.2019.01.25](https://doi.org/10.21037/jtd.2019.01.25)] [Medline: [31032076](https://pubmed.ncbi.nlm.nih.gov/31032076/)]
20. Ma Q. Recent applications and perspectives of logistic regression modelling in healthcare. *TNS*. 2024;36(1):185-190. [doi: [10.54254/2753-8818/36/20240614](https://doi.org/10.54254/2753-8818/36/20240614)]
21. Ndjoko LCM, Chakraborty A, Petri F, et al. Evaluating predictive performance and generalizability of traditional and artificial intelligence models in predicting surgical site infections postspinal surgery: a systematic review. *Spine J*. Feb 2026;26(2):280-291. [doi: [10.1016/j.spinee.2025.07.032](https://doi.org/10.1016/j.spinee.2025.07.032)] [Medline: [40669754](https://pubmed.ncbi.nlm.nih.gov/40669754/)]
22. Singh A, Singh P, Tiwari AK, Amity School of Engineering and technology, Amity University, Lucknow, India. A comprehensive survey on machine learning. *J Manage Serv Sci*. Mar 2021;1(1):1-17. [doi: [10.54060/JMSS/001.01.003](https://doi.org/10.54060/JMSS/001.01.003)]
23. Verberk JDM, Aghdassi SJS, Abbas M, et al. Automated surveillance systems for healthcare-associated infections: results from a European survey and experiences from real-life utilization. *J Hosp Infect*. Apr 2022;122:35-43. [doi: [10.1016/j.jhin.2021.12.021](https://doi.org/10.1016/j.jhin.2021.12.021)] [Medline: [35031393](https://pubmed.ncbi.nlm.nih.gov/35031393/)]
24. Verberk JDM, van Rooden SM, Koek MBG, et al. Validation of an algorithm for semiautomated surveillance to detect deep surgical site infections after primary total hip or knee arthroplasty-a multicenter study. *Infect Control Hosp Epidemiol*. Jan 2021;42(1):69-74. [doi: [10.1017/ice.2020.377](https://doi.org/10.1017/ice.2020.377)] [Medline: [32856575](https://pubmed.ncbi.nlm.nih.gov/32856575/)]
25. Westercamp MD, Dudeck MA, Allen-Bridson K, et al. Performance of simplified surgical site infection (SSI) surveillance case definitions for resource limited settings: comparison to SSI cases reported to the national healthcare safety network, 2013-2017. *Infect Control Hosp Epidemiol*. May 2020;41(5):611-613. [doi: [10.1017/ice.2020.36](https://doi.org/10.1017/ice.2020.36)] [Medline: [32167439](https://pubmed.ncbi.nlm.nih.gov/32167439/)]
26. Wang Y, Shi Y, Wang L, et al. Risk prediction model for surgical site infection in patients with gastrointestinal cancer: a systematic review and meta-analysis. *World J Surg Oncol*. Mar 1, 2025;23(1):72. [doi: [10.1186/s12957-025-03726-0](https://doi.org/10.1186/s12957-025-03726-0)] [Medline: [40025565](https://pubmed.ncbi.nlm.nih.gov/40025565/)]
27. Bontekoning N, Huisman H, Ali M, et al. Artificial intelligence for the detection of surgical site infection on wound images. A systematic review and meta-analysis. *Br J Surg*. Aug 20, 2025;112(Supplement_12). [doi: [10.1093/bjs/zna159.005](https://doi.org/10.1093/bjs/zna159.005)]
28. van Boekel AM, van der Meijden SL, Arbous SM, et al. Systematic evaluation of machine learning models for postoperative surgical site infection prediction. *PLoS ONE*. 2024;19(12):e0312968. [doi: [10.1371/journal.pone.0312968](https://doi.org/10.1371/journal.pone.0312968)] [Medline: [39666725](https://pubmed.ncbi.nlm.nih.gov/39666725/)]
29. Petrosyan Y, Thavorn K, Smith G, et al. Predicting postoperative surgical site infection with administrative data: a random forests algorithm. *BMC Med Res Methodol*. Aug 28, 2021;21(1):179. [doi: [10.1186/s12874-021-01369-9](https://doi.org/10.1186/s12874-021-01369-9)] [Medline: [34454414](https://pubmed.ncbi.nlm.nih.gov/34454414/)]
30. Zhuang Y, Dyas A, Meguid RA, et al. Preoperative prediction of postoperative infections using machine learning and electronic health record data. *Ann Surg*. Apr 1, 2024;279(4):720-726. [doi: [10.1097/SLA.0000000000006106](https://doi.org/10.1097/SLA.0000000000006106)] [Medline: [37753703](https://pubmed.ncbi.nlm.nih.gov/37753703/)]
31. Bonde A, Lorenzen S, Brixen G, Troelsen A, Sillesen M. Assessing the utility of deep neural networks in detecting superficial surgical site infections from free text electronic health record data. *Front Digit Health*. 2023;5:1249835. [doi: [10.3389/fdgh.2023.1249835](https://doi.org/10.3389/fdgh.2023.1249835)] [Medline: [38259257](https://pubmed.ncbi.nlm.nih.gov/38259257/)]
32. Wu G, Cheliegeer C, Southern DA, et al. Development of machine learning models for the detection of surgical site infections following total hip and knee arthroplasty: a multicenter cohort study. *Antimicrob Resist Infect Control*. Sep 2, 2023;12(1):88. [doi: [10.1186/s13756-023-01294-0](https://doi.org/10.1186/s13756-023-01294-0)] [Medline: [37658409](https://pubmed.ncbi.nlm.nih.gov/37658409/)]

33. Tabja Bortesi JP, Ranisau J, Di S, et al. Machine learning approaches for the image-based identification of surgical wound infections: scoping review. *J Med Internet Res*. Jan 18, 2024;26:e52880. [doi: [10.2196/52880](https://doi.org/10.2196/52880)] [Medline: [38236623](https://pubmed.ncbi.nlm.nih.gov/38236623/)]
34. Prey BJ, Colburn ZT, Williams JM, et al. The use of mobile thermal imaging and machine learning technology for the detection of early surgical site infections. *Am J Surg*. May 2024;231:60-64. [doi: [10.1016/j.amjsurg.2023.04.011](https://doi.org/10.1016/j.amjsurg.2023.04.011)] [Medline: [37173166](https://pubmed.ncbi.nlm.nih.gov/37173166/)]
35. Wu G, Khair S, Yang F, et al. Performance of machine learning algorithms for surgical site infection case detection and prediction: a systematic review and meta-analysis. *Ann Med Surg*. 2022;84. [doi: [10.1016/j.amsu.2022.104956](https://doi.org/10.1016/j.amsu.2022.104956)]
36. Shan J, Bao X, Wang B, et al. The best machine learning algorithm for building surgical site infection predictive models: a systematic review and network meta-analysis. *Comput Biol Med*. Jun 2025;192(Pt A):110286. [doi: [10.1016/j.compbimed.2025.110286](https://doi.org/10.1016/j.compbimed.2025.110286)] [Medline: [40311461](https://pubmed.ncbi.nlm.nih.gov/40311461/)]
37. Antoniadi AM, Du Y, Guendouz Y, et al. Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. *Applied Sciences*. 2021;11(11):5088. [doi: [10.3390/app11115088](https://doi.org/10.3390/app11115088)]
38. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ*. Feb 17, 2025;388:r340. [doi: [10.1136/bmj.r340](https://doi.org/10.1136/bmj.r340)] [Medline: [39961614](https://pubmed.ncbi.nlm.nih.gov/39961614/)]
39. Flores-Balado Á, Castresana Méndez C, Herrero González A, et al. Using artificial intelligence to reduce orthopedic surgical site infection surveillance workload: Algorithm design, validation, and implementation in 4 Spanish hospitals. *Am J Infect Control*. Nov 2023;51(11):1225-1229. [doi: [10.1016/j.ajic.2023.04.165](https://doi.org/10.1016/j.ajic.2023.04.165)] [Medline: [37100291](https://pubmed.ncbi.nlm.nih.gov/37100291/)]
40. Guedes M, Almeida F, Andrade P, et al. Surgical site infection surveillance in knee and hip arthroplasty: optimizing an algorithm to detect high-risk patients based on electronic health records. *Antimicrob Resist Infect Control*. Aug 15, 2024;13(1):90. [doi: [10.1186/s13756-024-01445-x](https://doi.org/10.1186/s13756-024-01445-x)] [Medline: [39148104](https://pubmed.ncbi.nlm.nih.gov/39148104/)]
41. Rosenbacke R, Melhus Å, McKee M, Stuckler D. How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care: systematic review. *JMIR AI*. Oct 30, 2024;3:e53207. [doi: [10.2196/53207](https://doi.org/10.2196/53207)] [Medline: [39476365](https://pubmed.ncbi.nlm.nih.gov/39476365/)]
42. Jabbour S, Fouhey D, Shepard S, et al. Measuring the impact of AI in the diagnosis of hospitalized patients: a randomized clinical vignette survey study. *JAMA*. Dec 19, 2023;330(23):2275-2284. [doi: [10.1001/jama.2023.22295](https://doi.org/10.1001/jama.2023.22295)] [Medline: [38112814](https://pubmed.ncbi.nlm.nih.gov/38112814/)]
43. Hassija V, Chamola V, Mahapatra A, et al. Interpreting black-box models: a review on explainable artificial intelligence. *Cogn Comput*. Jan 2024;16(1):45-74. [doi: [10.1007/s12559-023-10179-8](https://doi.org/10.1007/s12559-023-10179-8)]
44. Leo Breiman JF, Olshen RA, Stone CJ. Stone. In: *Classification and Regression Trees*. 1st ed. Chapman and Hall/CRC; 1984. ISBN: 9781315139470
45. Mihaljević B, Bielza C, Larrañaga P. Bayesian networks for interpretable machine learning and optimization. *Neurocomputing*. Oct 2021;456:648-665. [doi: [10.1016/j.neucom.2021.01.138](https://doi.org/10.1016/j.neucom.2021.01.138)]
46. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Christoph Molnar (Self-published); 2025. URL: <https://christophm.github.io/interpretable-ml-book/> [Accessed 2026-09-19]
47. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. May 2019;1(5):206-215. [doi: [10.1038/s42256-019-0048-x](https://doi.org/10.1038/s42256-019-0048-x)] [Medline: [35603010](https://pubmed.ncbi.nlm.nih.gov/35603010/)]
48. Bodria F, Giannotti F, Guidotti R, Naretto F, Pedreschi D, Rinzivillo S. Benchmarking and survey of explanation methods for black box models. *Data Min Knowl Disc*. Sep 2023;37(5):1719-1778. [doi: [10.1007/s10618-023-00933-9](https://doi.org/10.1007/s10618-023-00933-9)]
49. Li M, Sun H, Huang Y, Chen H. Shapley value: from cooperative game to explainable artificial intelligence. *Auton Intell Syst*. 2024;4(1):2. [doi: [10.1007/s43684-023-00060-8](https://doi.org/10.1007/s43684-023-00060-8)]
50. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics*. May 15, 2010;26(10):1340-1347. [doi: [10.1093/bioinformatics/btq134](https://doi.org/10.1093/bioinformatics/btq134)] [Medline: [20385727](https://pubmed.ncbi.nlm.nih.gov/20385727/)]
51. Dib L, Capus L. Classifying XAI methods to resolve conceptual ambiguity. *Technologies (Basel)*. 2025;13(9):390. [doi: [10.3390/technologies13090390](https://doi.org/10.3390/technologies13090390)]
52. Vilone G, Longo L. Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*. Dec 2021;76:89-106. [doi: [10.1016/j.inffus.2021.05.009](https://doi.org/10.1016/j.inffus.2021.05.009)]
53. Graziani M, Dutkiewicz L, Calvaresi D, et al. A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences. *Artif Intell Rev*. 2023;56(4):3473-3504. [doi: [10.1007/s10462-022-10256-8](https://doi.org/10.1007/s10462-022-10256-8)] [Medline: [36092822](https://pubmed.ncbi.nlm.nih.gov/36092822/)]
54. Ali S, Akhlaq F, Imran AS, Kastrati Z, Daudpota SM, Moosa M. The enlightening role of explainable artificial intelligence in medical & healthcare domains: a systematic literature review. *Comput Biol Med*. Nov 2023;166:107555. [doi: [10.1016/j.compbimed.2023.107555](https://doi.org/10.1016/j.compbimed.2023.107555)] [Medline: [37806061](https://pubmed.ncbi.nlm.nih.gov/37806061/)]

55. Nasarian E, Alizadehsani R, Acharya UR, Tsui KL. Designing interpretable ML system to enhance trust in healthcare: a systematic review to proposed responsible clinician-AI-collaboration framework. *Information Fusion*. Aug 2024;108:102412. [doi: [10.1016/j.inffus.2024.102412](https://doi.org/10.1016/j.inffus.2024.102412)]
56. Xu Q, Xie W, Liao B, et al. Interpretability of clinical decision support systems based on artificial intelligence from technological and medical perspective: a systematic review. *J Healthc Eng*. 2023;2023(1):9919269. [doi: [10.1155/2023/9919269](https://doi.org/10.1155/2023/9919269)] [Medline: [36776958](https://pubmed.ncbi.nlm.nih.gov/36776958/)]
57. 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](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
58. Aromataris ELC, Porritt K, Pilla B, editors. JBI manual for evidence synthesis. JBI Global Wiki. 2024. URL: <https://synthesismanual.jbi.global> [Accessed 2026-09-07]
59. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst Rev*. Jan 26, 2021;10(1):39. [doi: [10.1186/s13643-020-01542-z](https://doi.org/10.1186/s13643-020-01542-z)] [Medline: [33499930](https://pubmed.ncbi.nlm.nih.gov/33499930/)]
60. Chen Y, Ge XH, Yu Q, et al. Prediction model for urinary tract infection in pediatric urological surgery patients. *Front Public Health*. 2022;10:888089. [doi: [10.3389/fpubh.2022.888089](https://doi.org/10.3389/fpubh.2022.888089)] [Medline: [35812501](https://pubmed.ncbi.nlm.nih.gov/35812501/)]
61. Matsumoto H, Franzoni JM, Sinha R, et al. A novel risk calculator predicting surgical site infection after spinal surgery in patients with cerebral palsy. *Dev Med Child Neurol*. Aug 2022;64(8):1034-1043. [doi: [10.1111/dmcn.15193](https://doi.org/10.1111/dmcn.15193)] [Medline: [35229288](https://pubmed.ncbi.nlm.nih.gov/35229288/)]
62. Subramanyam R, Schaffzin J, Cudilo EM, Rao MB, Varughese AM. Systematic review of risk factors for surgical site infection in pediatric scoliosis surgery. *Spine J*. Jun 1, 2015;15(6):1422-1431. [doi: [10.1016/j.spinee.2015.03.005](https://doi.org/10.1016/j.spinee.2015.03.005)] [Medline: [25796355](https://pubmed.ncbi.nlm.nih.gov/25796355/)]
63. Dhillon SK, Ganggayah MD, Sinnadurai S, Lio P, Taib NA. Theory and practice of integrating machine learning and conventional statistics in medical data analysis. *Diagnostics (Basel)*. Oct 18, 2022;12(10):2526. [doi: [10.3390/diagnostics12102526](https://doi.org/10.3390/diagnostics12102526)] [Medline: [36292218](https://pubmed.ncbi.nlm.nih.gov/36292218/)]
64. Hu Y, Zhang X, Slavin V, et al. Beyond comparing machine learning and logistic regression in clinical prediction modelling: shifting from model debate to data quality. *J Med Internet Res*. Nov 5, 2025;27:e77721. [doi: [10.2196/77721](https://doi.org/10.2196/77721)] [Medline: [41191908](https://pubmed.ncbi.nlm.nih.gov/41191908/)]
65. Emmert-Streib F, Dehmer M. High-dimensional LASSO-based computational regression models: regularization, shrinkage, and selection. *MAKE*. 2019;1(1):359-383. [doi: [10.3390/make1010021](https://doi.org/10.3390/make1010021)]
66. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. Oct 2014;11(10):e1001744. [doi: [10.1371/journal.pmed.1001744](https://doi.org/10.1371/journal.pmed.1001744)] [Medline: [25314315](https://pubmed.ncbi.nlm.nih.gov/25314315/)]
67. Global Guidelines for the Prevention of Surgical Site Infection. World Health Organization; 2018. URL: <https://www.who.int/publications/i/item/9789241550475> [Accessed 2026-09-19]
68. Berríos-Torres SI, Umscheid CA, Bratzler DW, et al. Centers for disease control and prevention guideline for the prevention of surgical site infection, 2017. *JAMA Surg*. Aug 1, 2017;152(8):784-791. [doi: [10.1001/jamasurg.2017.0904](https://doi.org/10.1001/jamasurg.2017.0904)] [Medline: [28467526](https://pubmed.ncbi.nlm.nih.gov/28467526/)]
69. van Smeden M, Reitsma JB, Riley RD, Collins GS, Moons KG. Clinical prediction models: diagnosis versus prognosis. *J Clin Epidemiol*. Apr 2021;132:142-145. [doi: [10.1016/j.jclinepi.2021.01.009](https://doi.org/10.1016/j.jclinepi.2021.01.009)] [Medline: [33775387](https://pubmed.ncbi.nlm.nih.gov/33775387/)]
70. User guide for the 2015 ACS NSQIP participant use data file (PUF). American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP); 2016. URL: https://www.facs.org/media/fcvlu3t0/nsqip_puf_user_guide_2015.pdf [Accessed 2026-09-19]
71. Cai W, Wang L, Wang W, Zhou T. Systematic review and meta-analysis of the risk factors of surgical site infection in patients with colorectal cancer. *Transl Cancer Res*. Apr 2022;11(4):857-871. [doi: [10.21037/tcr-22-627](https://doi.org/10.21037/tcr-22-627)] [Medline: [35571649](https://pubmed.ncbi.nlm.nih.gov/35571649/)]
72. Liu D, Zhu Y, Chen W, Li M, Liu S, Zhang Y. Multiple preoperative biomarkers are associated with incidence of surgical site infection following surgeries of ankle fractures. *Int Wound J*. Jun 2020;17(3):842-850. [doi: [10.1111/iwj.13351](https://doi.org/10.1111/iwj.13351)] [Medline: [32219994](https://pubmed.ncbi.nlm.nih.gov/32219994/)]
73. Elliott IA, Chan C, Russell TA, et al. Distinction of risk factors for superficial vs organ-space surgical site infections after pancreatic surgery. *JAMA Surg*. Nov 1, 2017;152(11):1023-1029. [doi: [10.1001/jamasurg.2017.2155](https://doi.org/10.1001/jamasurg.2017.2155)] [Medline: [28700780](https://pubmed.ncbi.nlm.nih.gov/28700780/)]
74. Marcinkevičs R, Vogt JE. Interpretable and explainable machine learning: a methods-centric overview with concrete examples. *WIREs Data Min & Knowl*. May 2023;13(3):e1493. [doi: [10.1002/widm.1493](https://doi.org/10.1002/widm.1493)]
75. Moons KGM, Damen JAA, 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](https://doi.org/10.1136/bmj-2024-082505)] [Medline: [40127903](https://pubmed.ncbi.nlm.nih.gov/40127903/)]

76. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. Mar 1977;33(1):159-174. [doi: [10.2307/2529310](https://doi.org/10.2307/2529310)] [Medline: [843571](https://pubmed.ncbi.nlm.nih.gov/843571/)]
77. Cao S, Peng HB, Wei XZ, et al. Interpretable machine learning models for pre- and postoperative prediction of early intra-abdominal infections after liver transplantation: a multicenter retrospective cohort study. *Ther Adv Infect Dis*. 2026;13:20499361261453152. [doi: [10.1177/20499361261453152](https://doi.org/10.1177/20499361261453152)] [Medline: [42369884](https://pubmed.ncbi.nlm.nih.gov/42369884/)]
78. Yang Y, Zhang X, Zhang J, et al. Prediction models of surgical site infection after gastrointestinal surgery: a nationwide prospective cohort study. *Int J Surg*. Jan 1, 2024;110(1):119-129. [doi: [10.1097/JS9.0000000000000808](https://doi.org/10.1097/JS9.0000000000000808)]
79. Julien C, Anakok E, Treton X, et al. Impact of the ileal microbiota on surgical site infections in Crohn's disease: a nationwide prospective cohort. *J Crohns Colitis*. Aug 30, 2022;16(8):1211-1221. [doi: [10.1093/ecco-jcc/jjac026](https://doi.org/10.1093/ecco-jcc/jjac026)] [Medline: [35218661](https://pubmed.ncbi.nlm.nih.gov/35218661/)]
80. Ke C, Jin Y, Evans H, et al. Prognostics of surgical site infections using dynamic health data. *J Biomed Inform*. Jan 2017;65:22-33. [doi: [10.1016/j.jbi.2016.10.021](https://doi.org/10.1016/j.jbi.2016.10.021)] [Medline: [27825798](https://pubmed.ncbi.nlm.nih.gov/27825798/)]
81. Van Esbroeck A, Rubinfeld I, Hall B, Syed Z. Quantifying surgical complexity with machine learning: looking beyond patient factors to improve surgical models. *Surgery*. Nov 2014;156(5):1097-1105. [doi: [10.1016/j.surg.2014.04.034](https://doi.org/10.1016/j.surg.2014.04.034)] [Medline: [25108343](https://pubmed.ncbi.nlm.nih.gov/25108343/)]
82. Chen W, Lu Z, You L, Zhou L, Xu J, Chen K. Artificial intelligence-based multimodal risk assessment model for surgical site infection (AMRAMS): development and validation study. *JMIR Med Inform*. Jun 15, 2020;8(6):e18186. [doi: [10.2196/18186](https://doi.org/10.2196/18186)] [Medline: [32538798](https://pubmed.ncbi.nlm.nih.gov/32538798/)]
83. Weller GB, Lovely J, Larson DW, Earnshaw BA, Huebner M. Leveraging electronic health records for predictive modeling of post-surgical complications. *Stat Methods Med Res*. Nov 2018;27(11):3271-3285. [doi: [10.1177/0962280217696115](https://doi.org/10.1177/0962280217696115)] [Medline: [29298612](https://pubmed.ncbi.nlm.nih.gov/29298612/)]
84. Kiser AC, Shi J, Bucher BT. An explainable long short-term memory network for surgical site infection identification. *Surgery*. Jul 2024;176(1):24-31. [doi: [10.1016/j.surg.2024.03.006](https://doi.org/10.1016/j.surg.2024.03.006)] [Medline: [38616153](https://pubmed.ncbi.nlm.nih.gov/38616153/)]
85. Verberk JDM, van der Werff SD, Weegar R, et al. The augmented value of using clinical notes in semi-automated surveillance of deep surgical site infections after colorectal surgery. *Antimicrob Resist Infect Control*. Oct 26, 2023;12(1):117. [doi: [10.1186/s13756-023-01316-x](https://doi.org/10.1186/s13756-023-01316-x)] [Medline: [37884948](https://pubmed.ncbi.nlm.nih.gov/37884948/)]
86. Agostinho A, Chalot E, Teixeira D, et al. Semi-automated surveillance of surgical site infections using machine learning and rule-based classification models. *NPJ Digit Med*. Oct 17, 2025;8(1):617. [doi: [10.1038/s41746-025-01989-1](https://doi.org/10.1038/s41746-025-01989-1)] [Medline: [41107441](https://pubmed.ncbi.nlm.nih.gov/41107441/)]
87. Celik U, Liu F, Kobayashi K, et al. Machine learning-enhanced surveillance for surgical site infections in patients undergoing colon surgery: model development and evaluation study. *JMIR Form Res*. Oct 1, 2025;9:e75121. [doi: [10.2196/75121](https://doi.org/10.2196/75121)] [Medline: [41032359](https://pubmed.ncbi.nlm.nih.gov/41032359/)]
88. Wang P, Liu L, Xie Z, et al. Explainable machine learning models for prediction of surgical site infection after posterior lumbar fusion surgery based on Shapley additive explanations. *World Neurosurg*. May 2025;197:123942. [doi: [10.1016/j.wneu.2025.123942](https://doi.org/10.1016/j.wneu.2025.123942)] [Medline: [40154601](https://pubmed.ncbi.nlm.nih.gov/40154601/)]
89. Betts KS, Kisely S, Alati R. Predicting common maternal postpartum complications: leveraging health administrative data and machine learning. *BJOG*. May 2019;126(6):702-709. [doi: [10.1111/1471-0528.15607](https://doi.org/10.1111/1471-0528.15607)] [Medline: [30628159](https://pubmed.ncbi.nlm.nih.gov/30628159/)]
90. Hopkins BS, Mazmudar A, Driscoll C, et al. Using artificial intelligence (AI) to predict postoperative surgical site infection: a retrospective cohort of 4046 posterior spinal fusions. *Clin Neurol Neurosurg*. May 2020;192(105718):105718. [doi: [10.1016/j.clineuro.2020.105718](https://doi.org/10.1016/j.clineuro.2020.105718)] [Medline: [32065943](https://pubmed.ncbi.nlm.nih.gov/32065943/)]
91. CX-w L chun, Qiong D, Ling Z. Logistic regression and neural network prediction of neurosurgical site infections in a grade A tertiary hospital. *Chin J Nosocomiol*. 2018;28(8):1203-1206. [doi: [10.11816/cn.ni.2018-180362](https://doi.org/10.11816/cn.ni.2018-180362)]
92. Li H, Lu Q, Zhao Y. A risk prediction model for postoperative incision infection in patients with open tibia and fibula fractures was established based on decision tree. *Chinese Evidence-based Nursing*. 2025;11(16):189777803. [doi: [10.12102/j.issn.2095-8668.2025.16.029](https://doi.org/10.12102/j.issn.2095-8668.2025.16.029)]
93. De A, Satya Sai Preeti VV, Singh M, et al. Application of AI in predicting postoperative infections using routine blood parameters. *Bioinformation*. 2025;21(12):4271-4274. [doi: [10.6026/973206300214271](https://doi.org/10.6026/973206300214271)] [Medline: [41907999](https://pubmed.ncbi.nlm.nih.gov/41907999/)]
94. Kocbek P, Fijacko N, Soguero-Ruiz C, et al. Maximizing interpretability and cost-effectiveness of surgical site infection (SSI) predictive models using feature-specific regularized logistic regression on preoperative temporal data. *Comput Math Methods Med*. 2019;2019:2059851. [doi: [10.1155/2019/2059851](https://doi.org/10.1155/2019/2059851)] [Medline: [30915154](https://pubmed.ncbi.nlm.nih.gov/30915154/)]
95. Walczak S, Davila M, Velanovich V. Prophylactic antibiotic bundle compliance and surgical site infections: an artificial neural network analysis. *Patient Saf Surg*. 2019;13(1):41. [doi: [10.1186/s13037-019-0222-4](https://doi.org/10.1186/s13037-019-0222-4)] [Medline: [31827618](https://pubmed.ncbi.nlm.nih.gov/31827618/)]
96. Wang Q, Zhu Y, Cao L, Zhang T, Chang J, Wang X. Pathogenetic characteristics and related risk factors of incisional infection after surgery for acute intestinal obstruction and construction of prediction model. *Eur J Med Res*. May 10, 2025;30(1):376. [doi: [10.1186/s40001-025-02652-x](https://doi.org/10.1186/s40001-025-02652-x)] [Medline: [40349095](https://pubmed.ncbi.nlm.nih.gov/40349095/)]

97. Golinelli D, Rosa S, Rucci P, et al. ML-predicted surgical site infections: an epidemiological study utilizing machine learning on routinely collected healthcare data to predict infection risk. *Smart Health*. Sep 2025;37:100596. [doi: [10.1016/j.smhl.2025.100596](https://doi.org/10.1016/j.smhl.2025.100596)]
98. Kilic A, Goyal A, Miller JK, Gleason TG, Dubrawski A. Performance of a machine learning algorithm in predicting outcomes of aortic valve replacement. *Ann Thorac Surg*. Feb 2021;111(2):503-510. [doi: [10.1016/j.athoracsur.2020.05.107](https://doi.org/10.1016/j.athoracsur.2020.05.107)]
99. Colborn KL, Zhuang Y, Dyas AR, et al. Development and validation of models for detection of postoperative infections using structured electronic health records data and machine learning. *Surgery*. Feb 2023;173(2):464-471. [doi: [10.1016/j.surg.2022.10.026](https://doi.org/10.1016/j.surg.2022.10.026)] [Medline: [36470694](https://pubmed.ncbi.nlm.nih.gov/36470694/)]
100. Zhu Y, Simon GJ, Wick EC, et al. Applying machine learning across sites: external validation of a surgical site infection detection algorithm. *J Am Coll Surg*. Jun 2021;232(6):963-971. [doi: [10.1016/j.jamcollsurg.2021.03.026](https://doi.org/10.1016/j.jamcollsurg.2021.03.026)] [Medline: [33831539](https://pubmed.ncbi.nlm.nih.gov/33831539/)]
101. Colborn KL, Bronsert M, Amioka E, Hammermeister K, Henderson WG, Meguid R. Identification of surgical site infections using electronic health record data. *Am J Infect Control*. Nov 2018;46(11):1230-1235. [doi: [10.1016/j.ajic.2018.05.011](https://doi.org/10.1016/j.ajic.2018.05.011)] [Medline: [29907448](https://pubmed.ncbi.nlm.nih.gov/29907448/)]
102. Li X, Shuid AN, Mohd Miswan MF, et al. Risk factors and predictive model for early surgical site infection following single-level PLIF in diabetic patients. *Front Surg*. 2025;12. [doi: [10.3389/fsurg.2025.1709831](https://doi.org/10.3389/fsurg.2025.1709831)]
103. Rennert-May E, Leal J, MacDonald MK, et al. Validating administrative data to identify complex surgical site infections following cardiac implantable electronic device implantation: a comparison of traditional methods and machine learning. *Antimicrob Resist Infect Control*. Nov 10, 2022;11(1):138. [doi: [10.1186/s13756-022-01174-z](https://doi.org/10.1186/s13756-022-01174-z)] [Medline: [36357948](https://pubmed.ncbi.nlm.nih.gov/36357948/)]
104. Yu TH, Hou YC, Lin KC, Chung KP. Is it possible to identify cases of coronary artery bypass graft postoperative surgical site infection accurately from claims data? *BMC Med Inform Decis Mak*. May 29, 2014;14:42. [doi: [10.1186/1472-6947-14-42](https://doi.org/10.1186/1472-6947-14-42)] [Medline: [24884488](https://pubmed.ncbi.nlm.nih.gov/24884488/)]
105. Lahariya R, Jha AK, Sinha M, et al. SurgiCut-AI: an AI-driven tool for predicting surgical site infections following ventral hernia repair using pre- and intra-operative parameters. *Hernia*. Jan 2, 2026;30(1):47. [doi: [10.1007/s10029-025-03498-2](https://doi.org/10.1007/s10029-025-03498-2)] [Medline: [41483341](https://pubmed.ncbi.nlm.nih.gov/41483341/)]
106. Gutierrez-Naranjo JM, Moreira A, Valero-Moreno E, Bullock TS, Ogden LA, Zelle BA. -A machine learning model to predict surgical site infection after surgery of lower extremity fractures. *Int Orthop*. Jul 2024;48(7):1887-1896. [doi: [10.1007/s00264-024-06194-5](https://doi.org/10.1007/s00264-024-06194-5)] [Medline: [38700699](https://pubmed.ncbi.nlm.nih.gov/38700699/)]
107. Gowd AK, Agarwalla A, Amin NH, et al. Construct validation of machine learning in the prediction of short-term postoperative complications following total shoulder arthroplasty. *J Shoulder Elbow Surg*. Dec 2019;28(12):e410-e421. [doi: [10.1016/j.jse.2019.05.017](https://doi.org/10.1016/j.jse.2019.05.017)] [Medline: [31383411](https://pubmed.ncbi.nlm.nih.gov/31383411/)]
108. Hu DM, Xu R, Li J, Ai Y. Influencing factors for surgical site infection after gastrointestinal perforation repair surgery: analysis based on decision tree and logistic regression model. *Chin J Infect Control*. 2024;23(7):826-832. [doi: [10.12138/j.issn.1671-9638.20245130](https://doi.org/10.12138/j.issn.1671-9638.20245130)]
109. Piebpien P, Tansawet A, Pattanaprateep O, et al. Can machine learning models improve the prediction of surgical site infection in abdominal surgery than traditional statistical models? *J Int Med Res*. Nov 2024;52(11):3000605241293696. [doi: [10.1177/03000605241293696](https://doi.org/10.1177/03000605241293696)] [Medline: [39552114](https://pubmed.ncbi.nlm.nih.gov/39552114/)]
110. Zhang Q, Chen G, Zhu Q, et al. Construct validation of machine learning for accurately predicting the risk of postoperative surgical site infection following spine surgery. *J Hosp Infect*. Apr 2024;146:232-241. [doi: [10.1016/j.jhin.2023.09.024](https://doi.org/10.1016/j.jhin.2023.09.024)] [Medline: [38029857](https://pubmed.ncbi.nlm.nih.gov/38029857/)]
111. Isbell KD, Hatton GE, Wei S, et al. Risk stratification for superficial surgical site infection after emergency trauma laparotomy. *Surg Infect (Larchmt)*. Sep 2021;22(7):697-704. [doi: [10.1089/sur.2020.242](https://doi.org/10.1089/sur.2020.242)] [Medline: [33404358](https://pubmed.ncbi.nlm.nih.gov/33404358/)]
112. Wei S, Green C, Kao LS, et al. Accurate risk stratification for development of organ/space surgical site infections after emergent trauma laparotomy. *J Trauma Acute Care Surg*. Feb 2019;86(2):226-231. [doi: [10.1097/TA.0000000000002143](https://doi.org/10.1097/TA.0000000000002143)] [Medline: [30531329](https://pubmed.ncbi.nlm.nih.gov/30531329/)]
113. Chen B, Sheng W, Wu Z, et al. Machine learning based peri-surgical risk calculator for abdominal related emergency general surgery: a multicenter retrospective study. *Int J Surg*. Jun 1, 2024;110(6):3527-3535. [doi: [10.1097/JS9.0000000000001276](https://doi.org/10.1097/JS9.0000000000001276)] [Medline: [38489557](https://pubmed.ncbi.nlm.nih.gov/38489557/)]
114. Ohno Y, Mazaki J, Udo R, et al. Preliminary evaluation of a novel artificial intelligence-based prediction model for surgical site infection in colon cancer. *Cancer Diagn Progn*. 2022;2(6):691-696. [doi: [10.21873/cdp.10161](https://doi.org/10.21873/cdp.10161)] [Medline: [36340449](https://pubmed.ncbi.nlm.nih.gov/36340449/)]
115. Cobler-Lichter MD, Delamater JM, Weiss ZM, et al. Machine learning models accurately predict surgical site infection after emergent trauma laparotomy. *J Surg Res*. Dec 2025;316:158-168. [doi: [10.1016/j.jss.2025.10.043](https://doi.org/10.1016/j.jss.2025.10.043)] [Medline: [41317593](https://pubmed.ncbi.nlm.nih.gov/41317593/)]
116. Bangash AH, Mani K, Goldman SN, et al. Predicting surgical site infection after lumbar laminectomy and discectomy: a cutting-edge algorithmic approach by incorporating ensembled stacking into the current state-of-the-art for automated machine learning. *Neurosurg Rev*. Sep 18, 2025;48(1):653. [doi: [10.1007/s10143-025-03766-w](https://doi.org/10.1007/s10143-025-03766-w)] [Medline: [40965694](https://pubmed.ncbi.nlm.nih.gov/40965694/)]

117. Mamlook REA, Wells LJ, Sawyer R. Machine-learning models for predicting surgical site infections using patient pre-operative risk and surgical procedure factors. *Am J Infect Control*. May 2023;51(5):544-550. [doi: [10.1016/j.ajic.2022.08.013](https://doi.org/10.1016/j.ajic.2022.08.013)] [Medline: [36002080](https://pubmed.ncbi.nlm.nih.gov/36002080/)]
118. Bonde A, Varadarajan KM, Bonde N, et al. Assessing the utility of deep neural networks in predicting postoperative surgical complications: a retrospective study. *Lancet Digit Health*. Aug 2021;3(8):e471-e485. [doi: [10.1016/S2589-7500\(21\)00084-4](https://doi.org/10.1016/S2589-7500(21)00084-4)] [Medline: [34215564](https://pubmed.ncbi.nlm.nih.gov/34215564/)]
119. Bonde M, Bonde A, Kaafarani H, Millarch A, Sillesen M. Assessing the value of deep neural networks for postoperative complication prediction in pancreaticoduodenectomy patients. *PLoS ONE*. 2024;19(12):e0316402. [doi: [10.1371/journal.pone.0316402](https://doi.org/10.1371/journal.pone.0316402)] [Medline: [39774499](https://pubmed.ncbi.nlm.nih.gov/39774499/)]
120. Grass F, Storlie CB, Mathis KL, et al. Challenges of modeling outcomes for surgical infections: a word of caution. *Surg Infect (Larchmt)*. Jun 2021;22(5):523-531. [doi: [10.1089/sur.2020.208](https://doi.org/10.1089/sur.2020.208)] [Medline: [33085571](https://pubmed.ncbi.nlm.nih.gov/33085571/)]
121. Chen KA, Joisa CU, Stem JM, Guillem JG, Gomez SM, Kapadia MR. Improved prediction of surgical-site infection after colorectal surgery using machine learning. *Dis Colon Rectum*. Mar 1, 2023;66(3):458-466. [doi: [10.1097/DCR.0000000000002559](https://doi.org/10.1097/DCR.0000000000002559)] [Medline: [36538699](https://pubmed.ncbi.nlm.nih.gov/36538699/)]
122. Salimy MS, Buddhiraju A, Chen TLW, Mittal A, Xiao P, Kwon YM. Machine learning to predict periprosthetic joint infections following primary total hip arthroplasty using a national database. *Arch Orthop Trauma Surg*. Jan 17, 2025;145(1):131. [doi: [10.1007/s00402-025-05757-4](https://doi.org/10.1007/s00402-025-05757-4)] [Medline: [39820648](https://pubmed.ncbi.nlm.nih.gov/39820648/)]
123. Wang H, Fan T, Yang B, Lin Q, Li W, Yang M. Development and internal validation of supervised machine learning algorithms for predicting the risk of surgical site infection following minimally invasive transforaminal lumbar interbody fusion. *Front Med*. Dec 20, 2021;8:WOS. [doi: [10.3389/fmed.2021.771608](https://doi.org/10.3389/fmed.2021.771608)]
124. Xiong C, Zhao R, Xu J, et al. Construct and validate a predictive model for surgical site infection after posterior lumbar interbody fusion based on machine learning algorithm. *Comput Math Methods Med*. 2022;2022:2697841. [doi: [10.1155/2022/2697841](https://doi.org/10.1155/2022/2697841)] [Medline: [36050996](https://pubmed.ncbi.nlm.nih.gov/36050996/)]
125. Orfanoudaki A, Giannoutsou A, Hashim S, Bertsimas D, Hagberg RC. Machine learning models for mitral valve replacement: a comparative analysis with the Society of Thoracic Surgeons risk score. *J Card Surg*. Jan 2022;37(1):18-28. [doi: [10.1111/jocs.16072](https://doi.org/10.1111/jocs.16072)] [Medline: [34669218](https://pubmed.ncbi.nlm.nih.gov/34669218/)]
126. McLean K, Knight S, Clark N, Ademuyiwa A, Adisa A, Aguilera-Arevalo M. Development and external validation of the “Global Surgical-Site Infection” (GloSSI) predictive model in adult patients undergoing gastrointestinal surgery. *British Journal of Surgery*. Jul 3, 2024;111(Supplement_6). [doi: [10.1093/bjs/znae163.082](https://doi.org/10.1093/bjs/znae163.082)]
127. Bülow E, Hahn U, Andersen IT, Rolfson O, Pedersen AB, Hailer NP. Prediction of early periprosthetic joint infection after total hip arthroplasty. *Clin Epidemiol*. 2022;14(239–53):239-253. [doi: [10.2147/CLEP.S347968](https://doi.org/10.2147/CLEP.S347968)] [Medline: [35281208](https://pubmed.ncbi.nlm.nih.gov/35281208/)]
128. Cui Y, Shi X, Wang Q, et al. Artificial intelligence-based prediction model for surgical site infection in metastatic spinal disease: a multicenter development and validation study. *Int J Surg*. Oct 1, 2025;111(10):6867-6884. [doi: [10.1097/JS9.0000000000002806](https://doi.org/10.1097/JS9.0000000000002806)] [Medline: [40576176](https://pubmed.ncbi.nlm.nih.gov/40576176/)]
129. Li J, Yan Z. Machine learning model predicting factors for incisional infection following right hemicolectomy for colon cancer. *BMC Surg*. Oct 1, 2024;24(1):279. [doi: [10.1186/s12893-024-02543-8](https://doi.org/10.1186/s12893-024-02543-8)] [Medline: [39354475](https://pubmed.ncbi.nlm.nih.gov/39354475/)]
130. Liu WC, Ying H, Liao WJ, et al. Using preoperative and intraoperative factors to predict the risk of surgical site infections after lumbar spinal surgery: a machine learning-based study. *World Neurosurg*. Jun 2022;162:e553-e560. [doi: [10.1016/j.wneu.2022.03.060](https://doi.org/10.1016/j.wneu.2022.03.060)] [Medline: [35318153](https://pubmed.ncbi.nlm.nih.gov/35318153/)]
131. Yeo I, Klemt C, Robinson MG, Esposito JG, Uzosike AC, Kwon YM. The use of artificial neural networks for the prediction of surgical site infection following TKA. *J Knee Surg*. May 2023;36(6):637-643. [doi: [10.1055/s-0041-1741396](https://doi.org/10.1055/s-0041-1741396)] [Medline: [35016246](https://pubmed.ncbi.nlm.nih.gov/35016246/)]
132. Ying H, Guo BW, Wu HJ, Zhu RP, Liu WC, Zhong HF. Using multiple indicators to predict the risk of surgical site infection after ORIF of tibia fractures: a machine learning based study. *Front Cell Infect Microbiol*. 2023;13:1206393. [doi: [10.3389/fcimb.2023.1206393](https://doi.org/10.3389/fcimb.2023.1206393)] [Medline: [37448774](https://pubmed.ncbi.nlm.nih.gov/37448774/)]
133. Song Z-hZ Y, Wei AN. The predictive value of column chart model and random forest model based on preoperative systemic inflammatory factors for postoperative infection in patients with open tibiofibular fractures. *Chinese Journal of Bone and Joint*. 2025;14(9):810-815. [doi: [10.3969/j.issn.2095-252X.2025.09.007](https://doi.org/10.3969/j.issn.2095-252X.2025.09.007)]
134. Liang SS W, Gao D, Song, Mingxue M, Miao H, Li T. Research on risk prediction model of postoperative wound infection in patients with colorectal cancer undergoing radical resection surgery. *Beijing Medical Journal*. 2025;47(1):42-50. [doi: [10.15932/j.0253-9713.2025.01.009](https://doi.org/10.15932/j.0253-9713.2025.01.009)]
135. Cheng Y, Tang Q, Li X, Ma L, Yuan J, Hou X. Meta-lasso: new insight on infection prediction after minimally invasive surgery. *Med Biol Eng Comput*. Jun 2024;62(6):1703-1715. [doi: [10.1007/s11517-024-03027-w](https://doi.org/10.1007/s11517-024-03027-w)] [Medline: [38347344](https://pubmed.ncbi.nlm.nih.gov/38347344/)]
136. An YW H, Zhao X, Zhu B, Guo X, Jiang J. Establishment of nomogram model predicting risk of surgical site infection after posterior lumbar surgery. *Journal of Clinical Anesthesiology*. 2023;39(5):461-466. [doi: [10.12089/jca.2023.05.002](https://doi.org/10.12089/jca.2023.05.002)]
137. Choi JH, Choi Y, Lee KS, Ahn KH, Jang WY. Explainable model using shapley additive explanations approach on wound infection after wide soft tissue sarcoma resection: “Big Data” analysis based on health insurance review and

- assessment service hub. *Medicina (Kaunas)*. Feb 14, 2024;60(2):327. [doi: [10.3390/medicina60020327](https://doi.org/10.3390/medicina60020327)] [Medline: [38399614](https://pubmed.ncbi.nlm.nih.gov/38399614/)]
138. Pang Z, Liang J, Chen J, et al. Systemic immune-inflammatory biomarkers combined with the CRP-albumin-lymphocyte index predict surgical site infection following posterior lumbar spinal fusion: a retrospective study using machine learning. *Front Med (Lausanne)*. 2025;12:1590248. [doi: [10.3389/fmed.2025.1590248](https://doi.org/10.3389/fmed.2025.1590248)] [Medline: [40809418](https://pubmed.ncbi.nlm.nih.gov/40809418/)]
139. Ghabisha S, Alyhari Q, Ateik A, et al. Machine learning prediction model for surgical site infections after major abdominal surgery. *Patient Saf Surg*. Mar 16, 2026;20(1):17. [doi: [10.1186/s13037-026-00481-3](https://doi.org/10.1186/s13037-026-00481-3)] [Medline: [41840617](https://pubmed.ncbi.nlm.nih.gov/41840617/)]
140. Orlandi BMM, Mejia OAV, Sorio JL, et al. Performance of a novel risk model for deep sternal wound infection after coronary artery bypass grafting. *Sci Rep*. Sep 7, 2022;12(1):15177. [doi: [10.1038/s41598-022-19473-1](https://doi.org/10.1038/s41598-022-19473-1)] [Medline: [36071086](https://pubmed.ncbi.nlm.nih.gov/36071086/)]
141. Nie X, Wang G, Ba Y, et al. Development and validation of a predictive model for surgical site infection in open hand injuries. *Infect Drug Resist*. 2026;19:589797. [doi: [10.2147/IDR.S589797](https://doi.org/10.2147/IDR.S589797)] [Medline: [41835196](https://pubmed.ncbi.nlm.nih.gov/41835196/)]
142. Rahimi M, Ansari M, Abdollahi A, et al. A comprehensive feature importance analysis of surgical site infection following colorectal cancer surgery. *Sci Rep*. Nov 14, 2025;15(1):WOS. [doi: [10.1038/s41598-025-23722-4](https://doi.org/10.1038/s41598-025-23722-4)]
143. Ying Y, Fan Z, Fu L, He C, Huang Z, Teng H. Multiclass classification of infections after cervical spine surgery in the elderly: a machine learning approach based on preoperative and perioperative data. *Eur Spine J*. Mar 2026;35(3):1377-1388. [doi: [10.1007/s00586-026-09761-z](https://doi.org/10.1007/s00586-026-09761-z)] [Medline: [41619022](https://pubmed.ncbi.nlm.nih.gov/41619022/)]
144. Zhang Q, Chen G, Zhou C, et al. Construct validation of machine learning models for predicting surgical site infection risk following ankle fracture surgery. *International Journal of Surgery*. 2025;111(12):9431-9448. [doi: [10.1097/JS9.0000000000003239](https://doi.org/10.1097/JS9.0000000000003239)]
145. Zhou H, Luo C, Xu Q, et al. Machine learning approach to predict surgical site infection in head and neck squamous cell carcinoma patients after free flap reconstruction. *J Craniomaxillofac Surg*. Dec 2025;53(12):2112-2117. [doi: [10.1016/j.jcms.2025.09.010](https://doi.org/10.1016/j.jcms.2025.09.010)] [Medline: [40987694](https://pubmed.ncbi.nlm.nih.gov/40987694/)]
146. Ruan X, Fu S, Storlie CB, Mathis KL, Larson DW, Liu H. Real-time risk prediction of colorectal surgery-related post-surgical complications using GRU-D model. *J Biomed Inform*. Nov 2022;135:104202. [doi: [10.1016/j.jbi.2022.104202](https://doi.org/10.1016/j.jbi.2022.104202)] [Medline: [36162805](https://pubmed.ncbi.nlm.nih.gov/36162805/)]
147. Kalisnik JM, Zibert J, Kamensek T, Hanuna M, Santarpino G, Fischlein T. Machine learning enhanced prediction of deep sternal wound infection after surgical myocardial revascularization. *Cardiovasc Revasc Med*. May 2026;86:81-87. [doi: [10.1016/j.carrev.2025.07.016](https://doi.org/10.1016/j.carrev.2025.07.016)] [Medline: [40753065](https://pubmed.ncbi.nlm.nih.gov/40753065/)]
148. Xu CX P, Ge M, Liu X. Establishment of predictive model for surgical site infection following colorectal surgery based on machine learning. *West China Medical Journal*. 2020;35(7):827-832. [doi: [10.7507/1002-0179.202002027](https://doi.org/10.7507/1002-0179.202002027)]
149. Perkins L, O'Keefe T, Ardill W, Potenza B. Modernizing Surgical Quality: A Novel Approach to Improving Detection of Surgical Site Infections in the Veteran Population. *Surg Infect (Larchmt)*. Sep 2024;25(7):499-504. [doi: [10.1089/sur.2024.013](https://doi.org/10.1089/sur.2024.013)] [Medline: [38973692](https://pubmed.ncbi.nlm.nih.gov/38973692/)]
150. Phuyal D, Ozmen B, Berber I, Schwarz G. Explainable AI modeling of postoperative surgical site infection risk in autologous breast reconstruction. *J Plast Reconstr Aesthet Surg*. Mar 2026;114:117-126. [doi: [10.1016/j.bjps.2025.12.029](https://doi.org/10.1016/j.bjps.2025.12.029)] [Medline: [41610823](https://pubmed.ncbi.nlm.nih.gov/41610823/)]
151. Fletcher RR, Olubeko O, Sonthalia H, et al. Application of machine learning to prediction of surgical site infection. *Annu Int Conf IEEE Eng Med Biol Soc*. Jul 2019;2019:2234-2237. [doi: [10.1109/EMBC.2019.8857942](https://doi.org/10.1109/EMBC.2019.8857942)] [Medline: [31946345](https://pubmed.ncbi.nlm.nih.gov/31946345/)]
152. Long Z, Hu X, Liu J, et al. Establishment and validation of a nomogram model for surgical site infections after posterior lumbar interbody fusion: a retrospective observational study. *Neurosurg Rev*. Jun 6, 2025;48(1):488. [doi: [10.1007/s10143-025-03645-4](https://doi.org/10.1007/s10143-025-03645-4)] [Medline: [40478467](https://pubmed.ncbi.nlm.nih.gov/40478467/)]
153. Shaukat W, Baig AM, Ali Z, et al. Prognostic value of C-reactive protein and neutrophil-to-lymphocyte ratio in predicting postoperative infections after gastrointestinal surgery: a meta-analysis. *Cureus*. Aug 2025;17(8):e91123. [doi: [10.7759/cureus.91123](https://doi.org/10.7759/cureus.91123)] [Medline: [41018345](https://pubmed.ncbi.nlm.nih.gov/41018345/)]
154. Almohrij SA. Time of developing surgical site infections and its association with patient and procedure characteristics. *J Infect Public Health*. May 2025;18(5):102734. [doi: [10.1016/j.jiph.2025.102734](https://doi.org/10.1016/j.jiph.2025.102734)] [Medline: [40154107](https://pubmed.ncbi.nlm.nih.gov/40154107/)]
155. Alemayehu MA, Azene AG, Mihretie KM. Time to development of surgical site infection and its predictors among general surgery patients admitted at specialized hospitals in Amhara region, northwest Ethiopia: a prospective follow-up study. *BMC Infect Dis*. May 17, 2023;23(1):334. [doi: [10.1186/s12879-023-08301-0](https://doi.org/10.1186/s12879-023-08301-0)] [Medline: [37198551](https://pubmed.ncbi.nlm.nih.gov/37198551/)]
156. Ejaz A, Schmidt C, Johnston FM, Frank SM, Pawlik TM. Risk factors and prediction model for inpatient surgical site infection after major abdominal surgery. *J Surg Res*. Sep 2017;217:153-159. [doi: [10.1016/j.jss.2017.05.018](https://doi.org/10.1016/j.jss.2017.05.018)] [Medline: [28595819](https://pubmed.ncbi.nlm.nih.gov/28595819/)]
157. Roch PJ, Ecker C, Jäckle K, et al. Interleukin-6 as a critical inflammatory marker for early diagnosis of surgical site infection after spine surgery. *Infection*. Dec 2024;52(6):2269-2277. [doi: [10.1007/s15010-024-02271-4](https://doi.org/10.1007/s15010-024-02271-4)] [Medline: [38709460](https://pubmed.ncbi.nlm.nih.gov/38709460/)]
158. Sanger PC, van Ramshorst GH, Mercan E, et al. A prognostic model of surgical site infection using daily clinical wound assessment. *J Am Coll Surg*. Aug 2016;223(2):259-270. [doi: [10.1016/j.jamcollsurg.2016.04.046](https://doi.org/10.1016/j.jamcollsurg.2016.04.046)] [Medline: [27188832](https://pubmed.ncbi.nlm.nih.gov/27188832/)]

159. Fu Shaw L, Chen IH, Chen CS, et al. Factors influencing microbial colonies in the air of operating rooms. *BMC Infect Dis*. Jan 2, 2018;18(1):4. [doi: [10.1186/s12879-017-2928-1](https://doi.org/10.1186/s12879-017-2928-1)] [Medline: [29291707](https://pubmed.ncbi.nlm.nih.gov/29291707/)]
160. Pokrywka M, Byers K. Traffic in the operating room: a review of factors influencing air flow and surgical wound contamination. *Infect Disord Drug Targets*. Jun 2013;13(3):156-161. [doi: [10.2174/1871526511313030002](https://doi.org/10.2174/1871526511313030002)] [Medline: [24001332](https://pubmed.ncbi.nlm.nih.gov/24001332/)]
161. Schuessler M, Fleming S, Meyer S, Seto T, Hernandez-Boussard T. Diagnostic framework to validate clinical machine learning models locally on temporally stamped data. *Commun Med (Lond)*. Jul 1, 2025;5(1):261. [doi: [10.1038/s43856-025-00965-w](https://doi.org/10.1038/s43856-025-00965-w)] [Medline: [40596645](https://pubmed.ncbi.nlm.nih.gov/40596645/)]
162. Davis SE, Matheny ME, Balu S, Sendak MP. A framework for understanding label leakage in machine learning for health care. *J Am Med Inform Assoc*. Dec 22, 2023;31(1):274-280. [doi: [10.1093/jamia/ocad178](https://doi.org/10.1093/jamia/ocad178)] [Medline: [37669138](https://pubmed.ncbi.nlm.nih.gov/37669138/)]
163. Chen T, Chen R, Zhang H, Feng Q, Cai L, Li J. Postoperative laboratory markers as predictors of early spinal surgical site infections: a retrospective cohort study. *Chin J Traumatol*. Nov 2025;28(6):412-417. [doi: [10.1016/j.cjtee.2025.08.002](https://doi.org/10.1016/j.cjtee.2025.08.002)] [Medline: [40962683](https://pubmed.ncbi.nlm.nih.gov/40962683/)]
164. Morey J, Winters R, Jones D. Artificial intelligence to predict billing code levels of emergency department encounters. *Ann Emerg Med*. Jan 2025;85(1):63-73. [doi: [10.1016/j.annemergmed.2024.07.011](https://doi.org/10.1016/j.annemergmed.2024.07.011)] [Medline: [39320279](https://pubmed.ncbi.nlm.nih.gov/39320279/)]
165. Hama T, Alsaleh MM, Allery F, et al. Enhancing patient outcome prediction through deep learning with sequential diagnosis codes from structured electronic health record data: systematic review. *J Med Internet Res*. Mar 18, 2025;27:e57358. [doi: [10.2196/57358](https://doi.org/10.2196/57358)] [Medline: [40100249](https://pubmed.ncbi.nlm.nih.gov/40100249/)]
166. Ramadan B, Liu MC, Burkhart MC, Parker WF, Beaulieu-Jones BK. Diagnostic codes in AI prediction models and label leakage of same-admission clinical outcomes. *JAMA Netw Open*. Dec 1, 2025;8(12):e2550454. [doi: [10.1001/jamanetworkopen.2025.50454](https://doi.org/10.1001/jamanetworkopen.2025.50454)] [Medline: [41632159](https://pubmed.ncbi.nlm.nih.gov/41632159/)]
167. Lipton ZC. The mythos of model interpretability. *Commun ACM*. Sep 26, 2018;61(10):36-43. [doi: [10.1145/3233231](https://doi.org/10.1145/3233231)]
168. Chen H, Covert IC, Lundberg SM, Lee SI. Algorithms to estimate Shapley value feature attributions. *Nat Mach Intell*. 2023;5(6):590-601. [doi: [10.1038/s42256-023-00657-x](https://doi.org/10.1038/s42256-023-00657-x)]
169. Khan A, Ali A, Khan J, Ullah F, Faheem M. Using permutation-based feature importance for improved machine learning model performance at reduced costs. *IEEE Access*. 2025;13:36421-36435. [doi: [10.1109/ACCESS.2025.3544625](https://doi.org/10.1109/ACCESS.2025.3544625)]
170. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. Jan 2020;2(1):56-67. [doi: [10.1038/s42256-019-0138-9](https://doi.org/10.1038/s42256-019-0138-9)] [Medline: [32607472](https://pubmed.ncbi.nlm.nih.gov/32607472/)]
171. Carmichael Z, Scheirer WJ. How well do feature-additive explainers explain feature-additive predictors? arXiv. Preprint posted online on Oct 27, 2023. [doi: [10.48550/arXiv.2310.18496](https://doi.org/10.48550/arXiv.2310.18496)]
172. Chen H, Lundberg S, Lee SI. Explaining models by propagating shapley values of local components. In: Shaban-Nejad A, Michalowski M, Buckeridge DL, editors. *Explainable AI in Healthcare and Medicine: Building a Culture of Transparency and Accountability*. Springer International Publishing; 2021:261-270. [doi: [10.1007/978-3-030-53352-6_24](https://doi.org/10.1007/978-3-030-53352-6_24)]
173. Abbas Q, Jeong W, Lee SW. Explainable AI in clinical decision support systems: a meta-analysis of methods, applications, and usability challenges. *Health Care (Don Mills)*. 2025;13(17):2154. [doi: [10.3390/healthcare13172154](https://doi.org/10.3390/healthcare13172154)]
174. Hur S, Lee Y, Park J, et al. Comparison of SHAP and clinician friendly explanations reveals effects on clinical decision behaviour. *NPJ Digit Med*. Sep 26, 2025;8(1):578. [doi: [10.1038/s41746-025-01958-8](https://doi.org/10.1038/s41746-025-01958-8)] [Medline: [41006498](https://pubmed.ncbi.nlm.nih.gov/41006498/)]
175. Panigutti C, Beretta A, Fadda D, et al. Co-design of human-centered, explainable AI for clinical decision support. *ACM Trans Interact Intell Syst*. Dec 31, 2023;13(4):1-35. [doi: [10.1145/3587271](https://doi.org/10.1145/3587271)]
176. Jung J, Kang S, Choi J, El-Kareh R, Lee H, Kim H. Evaluating the impact of explainable AI on clinicians' decision-making: a study on ICU length of stay prediction. *Int J Med Inform*. Sep 2025;201:105943. [doi: [10.1016/j.ijmedinf.2025.105943](https://doi.org/10.1016/j.ijmedinf.2025.105943)] [Medline: [40318498](https://pubmed.ncbi.nlm.nih.gov/40318498/)]
177. Martens D, Hinns J, Dams C, Vergouwen M, Evgeniou T. Tell me a story! narrative-driven XAI with large language models. *Decis Support Syst*. Apr 2025;191:114402. [doi: [10.1016/j.dss.2025.114402](https://doi.org/10.1016/j.dss.2025.114402)]
178. Sion Y, Sousa S, Lamas D. In AI we trust: human-centred AI explanations for early dementia risk prediction. Presented at: Proceedings of the 36th Annual Conference of the European Association of Cognitive Ergonomics: Association for Computing Machinery; Oct 7-10, 2025; Tallinn Estonia. Oct 7, 2025. [doi: [10.1145/3746175.3746195](https://doi.org/10.1145/3746175.3746195)]
179. Jin W, Li X, Fatehi M, Hamarneh G. Guidelines and evaluation of clinical explainable AI in medical image analysis. *Med Image Anal*. Feb 2023;84:102684. [doi: [10.1016/j.media.2022.102684](https://doi.org/10.1016/j.media.2022.102684)] [Medline: [36516555](https://pubmed.ncbi.nlm.nih.gov/36516555/)]
180. Banik S, Balasubramanian K, Manna S, Derrible S, Sankaranarayanan S. Evaluating generalized feature importance via performance assessment of machine learning models for predicting elastic properties of materials. *Comput Mater Sci*. Mar 2024;236:112847. [doi: [10.1016/j.commatsci.2024.112847](https://doi.org/10.1016/j.commatsci.2024.112847)]
181. Calu V, Pirliianu C, Miron A, Grigorean VT. Surgical site infections in colorectal cancer surgeries: a systematic review and meta-analysis of the impact of surgical approach and associated risk factors. *Life (Basel)*. Jul 5, 2024;14(7):850. [doi: [10.3390/life14070850](https://doi.org/10.3390/life14070850)] [Medline: [39063604](https://pubmed.ncbi.nlm.nih.gov/39063604/)]
182. Chen M, Liang H, Chen M, et al. Risk factors for surgical site infection in patients with gastric cancer: a meta-analysis. *Int Wound J*. Nov 2023;20(9):3884-3897. [doi: [10.1111/iwj.14264](https://doi.org/10.1111/iwj.14264)] [Medline: [37337711](https://pubmed.ncbi.nlm.nih.gov/37337711/)]

183. He L, Jiang Z, Wang W, Zhang W. Predictors for different types of surgical site infection in patients with gastric cancer: a systematic review and meta-analysis. *Int Wound J*. Apr 2024;21(4):e14549. [doi: [10.1111/iwj.14549](https://doi.org/10.1111/iwj.14549)] [Medline: [38155362](https://pubmed.ncbi.nlm.nih.gov/38155362/)]
184. Ramspek CL, Steyerberg EW, Riley RD, et al. Prediction or causality? A scoping review of their conflation within current observational research. *Eur J Epidemiol*. Sep 2021;36(9):889-898. [doi: [10.1007/s10654-021-00794-w](https://doi.org/10.1007/s10654-021-00794-w)] [Medline: [34392488](https://pubmed.ncbi.nlm.nih.gov/34392488/)]
185. Feuerriegel S, Frauen D, Melnychuk V, et al. Causal machine learning for predicting treatment outcomes. *Nat Med*. Apr 2024;30(4):958-968. [doi: [10.1038/s41591-024-02902-1](https://doi.org/10.1038/s41591-024-02902-1)] [Medline: [38641741](https://pubmed.ncbi.nlm.nih.gov/38641741/)]
186. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health*. Nov 2021;3(11):e745-e750. [doi: [10.1016/S2589-7500\(21\)00208-9](https://doi.org/10.1016/S2589-7500(21)00208-9)] [Medline: [34711379](https://pubmed.ncbi.nlm.nih.gov/34711379/)]
187. Oka S, Inoue N, Takefuji Y. Beyond predictive accuracy: statistical validation of feature importance in biomedical machine learning. *Comput Methods Programs Biomed*. Dec 2025;272:109085. [doi: [10.1016/j.cmpb.2025.109085](https://doi.org/10.1016/j.cmpb.2025.109085)] [Medline: [41016203](https://pubmed.ncbi.nlm.nih.gov/41016203/)]
188. Buijsman S. Causal scientific explanations from machine learning. *Synthese*. 2023;202(6):202. [doi: [10.1007/s11229-023-04429-3](https://doi.org/10.1007/s11229-023-04429-3)]
189. Nguyen TL, Collins GS, Landais P, Le Manach Y. Counterfactual clinical prediction models could help to infer individualized treatment effects in randomized controlled trials—an illustration with the international stroke trial. *J Clin Epidemiol*. Sep 2020;125:47-56. [doi: [10.1016/j.jclinepi.2020.05.022](https://doi.org/10.1016/j.jclinepi.2020.05.022)] [Medline: [32464321](https://pubmed.ncbi.nlm.nih.gov/32464321/)]
190. Flora ML, Potvin CK, McGovern A, Handler S. A machine learning explainability tutorial for atmospheric sciences. *Artif Intell Syst*. Jan 1, 2024;3(1):e230018. [doi: [10.1175/AIES-D-23-0018.1](https://doi.org/10.1175/AIES-D-23-0018.1)]
191. Salih AM. Explainable artificial intelligence and multicollinearity: a mini review of current approaches. *arXiv*. Preprint posted online on Jun 17, 2024. [doi: [10.48550/arXiv.2406.11524](https://doi.org/10.48550/arXiv.2406.11524)]
192. Boubekki A, Myhre JN, Luppino LT, Mikalsen KO, Revhaug A, Jenssen R. Clinically relevant features for predicting the severity of surgical site infections. *IEEE J Biomed Health Inform*. Apr 2022;26(4):1794-1801. [doi: [10.1109/JBHI.2021.3121038](https://doi.org/10.1109/JBHI.2021.3121038)] [Medline: [34665748](https://pubmed.ncbi.nlm.nih.gov/34665748/)]
193. Chen A, Zhao Z, Hou W, Singer AJ, Li H, Duong TQ. Time-to-death longitudinal characterization of clinical variables and longitudinal prediction of mortality in COVID-19 patients: a two-center study. *Front Med*. Apr 29, 2021;8:2021. [doi: [10.3389/fmed.2021.661940](https://doi.org/10.3389/fmed.2021.661940)]
194. Thorsen-Meyer HC, Nielsen AB, Nielsen AP, et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. *Lancet Digit Health*. Apr 2020;2(4):e179-e191. [doi: [10.1016/S2589-7500\(20\)30018-2](https://doi.org/10.1016/S2589-7500(20)30018-2)] [Medline: [33328078](https://pubmed.ncbi.nlm.nih.gov/33328078/)]
195. Omiya M, Takada T, Yano T, Fujii K, Fujiishi R, Fukuhara S. Time-dependent predictive performance of inflammatory markers for 30-day all-cause mortality in patients with suspected infection in a Japanese emergency department: a retrospective cohort study. *BMJ Open*. Dec 24, 2025;15(12):e103082. [doi: [10.1136/bmjopen-2025-103082](https://doi.org/10.1136/bmjopen-2025-103082)] [Medline: [41448710](https://pubmed.ncbi.nlm.nih.gov/41448710/)]

Abbreviations

ACS-NSQIP: American College of Surgeons National Surgical Quality Improvement Program

AUC: area under the receiver operating characteristic curve

CDC: Centers for Disease Control and Prevention

CHARMS: Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies

DSWI: deep sternal wound infection

ICD-10: *International Statistical Classification of Diseases, Tenth Revision*

ICD-9: *International Classification of Diseases, Ninth Revision*

JBI: Joanna Briggs Institute

LLM: large language model

LSTM: long short-term memory

ML: machine learning

NB: native Bayes

NHSN: National Healthcare Safety Network

PFI: permutation feature importance

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

PROBAST+AI: Prediction Model Risk of Bias Assessment Tool for Prediction Models Using Regression or Artificial Intelligence

ROB: risk of bias

SHAP: Shapley Additive Explanations

SSI: surgical site infection

WHO: World Health Organization

XGBoost: extreme gradient boosting

Edited by Stefano Brini; peer-reviewed by Batuhan Kilic, Javed Al Faysal; submitted 08.Mar.2026; final revised version received 12.Aug.2026; accepted 20.Aug.2026; published 30.Sep.2026

Please cite as:

Sun R, Liu Y, Lu S, Tartari E, Birgand G, Yang L, Zhou L

Explainable Machine Learning Predictive Models for Surgical Site Infections: Scoping Review

J Med Internet Res 2026;28:e94617

URL: <https://www.jmir.org/2026/1/e94617>

doi: [10.2196/94617](https://doi.org/10.2196/94617)

© Rui Sun, Yao Liu, Shuya Lu, Ermira Tartari, Gabriel Birgand, Lin Yang, Lei Zhou. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 30.Sep.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 Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.