European Journal of Clinical Microbiology & Infectious Diseases
https://doi.org/10.1007/s10096-024-05027-y

REVIEW

Artificial intelligence in antimicrobial stewardship: a systematic review and meta-analysis of predictive performance and diagnostic accuracy

Flavia Pennisi[1,2] · Antonio Pinto[2] · Giovanni Emanuele Ricciardi[1,2] · Carlo Signorelli[2] · Vincenza Gianfredi[3]

Received: 1 November 2024 / Accepted: 19 December 2024

Abstract

The increasing threat of antimicrobial resistance has prompted a need for more effective antimicrobial stewardship programs (AMS). Artificial intelligence (AI) and machine learning (ML) tools have emerged as potential solutions to enhance decision-making and improve patient outcomes in AMS. This systematic review and meta-analysis aims to evaluate the impact of AI in AMS and to assess its predictive performance and diagnostic accuracy. We conducted a comprehensive literature search across PubMed/MEDLINE, Scopus, EMBASE, and Web of Science to identify studies published up to July 2024. Studies included were observational, cohort, or retrospective, focusing on the application of AI/ML in AMS. The outcomes assessed were the area under the curve (AUC), accuracy, sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV). We calculated the mean pooled effect size (ES) and its 95% confidence interval (CI) using a random-effects model. The risk of bias was assessed using the QUADAS-AI tool, and the protocol was registered in PROSPERO. Out of 3,458 retrieved articles, 80 studies met the inclusion criteria. Our meta-analysis demonstrated that ML models exhibited strong predictive performance and diagnostic accuracy, with the following results: AUC [ES: 72.28 (70.42–74.14)], accuracy [ES: 74.97 (73.35–76.58)], sensitivity [ES: 76.89; (71.90–81.89)], specificity [ES: 73.77; (67.87–79.67)], NPV [ES:79.92 (76.54–83.31)], and PPV [ES: 69.41 (60.19–78.63)] across various AMS settings. AI and ML tools offer promising enhancements due to their strong predictive performance. The integration of AI into AMS could lead to more precise antimicrobial prescribing, reduced antimicrobial resistance, and better resource utilization.

Keywords Artificial intelligence · Antimicrobial stewardship · Machine learning · Systematic review · Meta-analysis

Introduction

Antimicrobial resistance (AMR) is one of the most pressing global health threats, contributing to increased morbidity, mortality, and healthcare costs worldwide, leading to over 1.5 million deaths worldwide in 2019 alone [1], with a broader implication of 4.95 million deaths when associated with resistant bacterial infections [2]. The emergence of resistant pathogens has significantly complicated clinical decision-making, particularly in the context of antimicrobial stewardship (AMS), where the goal is to use antimicrobials judiciously to reduce resistance development while ensuring effective treatment [3]. However, studies have shown that approximately 30–50% of antimicrobial use in hospitals is inappropriate, contributing to the acceleration of resistance [4]. In the context of AMS, the ability to predict resistance patterns using rapid diagnostics is crucial. However, traditional culture-based methods of identifying resistance patterns rely heavily on culture-based diagnostics, which can take up to 48–72 h, delaying optimal therapy. This uncertainty in clinical decision-making increases reliance on empirical, often broad-spectrum antibiotic use, exacerbating resistance problems and negatively impacting patient outcomes [5].

Artificial intelligence (AI) and machine learning (ML) technologies have emerged as promising tools to support AMS by enabling rapid, data-driven predictions of AMR and

✉ Vincenza Gianfredi
vincenza.gianfredi@unimi.it

1 PhD National Programme in One Health approaches to infectious diseases and life science research, Department of Public Health, Experimental and Forensic Medicine, University of Pavia, 27100 Pavia, Italy

2 Faculty of Medicine, University Vita-Salute San Raffaele, Milan, Italy

3 Department of Biomedical Sciences for Health, University of Milan, Via Pascal 36, 20133 Milan, Italy

Published online: 06 January 2025

guiding antimicrobial therapy decisions. ML models, particularly those utilizing ML algorithms, have shown potential in various domains of healthcare, including diagnostic accuracy, predictive analytics, and clinical decision support [5]. In the context of AMS, AI-driven approaches could offer more timely and accurate predictions of antimicrobial resistance, potentially guiding clinicians towards more targeted and appropriate therapies [6].

Despite these advancements, there remains a need to comprehensively evaluate the clinical utility of AI and ML models in AMS. Previous studies have reported various performance metrics, and some preliminary reviews have consolidated these findings [7]. However, the absence of any previous meta-analysis makes it impossible to statistically evaluate the overall predictive performance and diagnostic accuracy of AI and ML tools in this domain, which is essential for understanding their value in real-world clinical settings.

This systematic review and meta-analysis aims to evaluate the impact of AI and ML on antimicrobial stewardship, with a specific focus on their predictive performance and diagnostic accuracy. By analyzing key outcomes, including area under the curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), this study will provide insights into the potential role of AI in improving clinical outcomes in the fight against AMR. Through the analysis of the current literature, we aim to inform future research and the implementation of AI technologies in AMS programs.

Methods

Protocol and registration

This systematic review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO), registration number [PROSPERO ID CRD42024567640].

Literature search strategy

A comprehensive literature search was performed across PubMed/MEDLINE, Scopus, EMBASE, and Web of Science (WoS) in July 2024. This systematic review seeks to answer the following question: "What is the impact and effectiveness of AI in AMS aimed at reducing AMR?". The search strategy incorporated two key elements: artificial intelligence (and synonyms) and antimicrobial stewardship or antimicrobial resistance (and synonyms). Selected keywords, including MeSH terms and Title/Abstract terms, were combined using the Boolean operators "AND" and "OR". The search strategy was initially developed for PubMed/MEDLINE and subsequently adapted for Scopus, EMBASE, and WoS. The search strategies for each database are detailed in Table S1. Additionally, forward and backward citation searches were performed. No temporal filter was applied.

Eligibility criteria

Inclusion and exclusion criteria were defined based on the Population, Intervention, Comparison, Outcome, and Study design (PICOS) guidelines. Only studies involving patients (both inpatient or outpatient) requiring antimicrobial prescriptions in settings (population) and using ML models in AMS were included. We included studies that developed and, where applicable, validated ML models aimed at improving antimicrobial stewardship practices. Moreover, our primary outcomes included predictive performance metrics (AUC, c-statistic, accuracy, sensitivity, specificity, PPV, NPV, and F1-score). Secondary outcomes included clinical outcomes and improvement in patient health outcomes. Lastly, only observational studies published in English in international, peer-reviewed scientific journals, were considered eligible. A detailed description of the inclusion criteria, defined according to PICOS, is provided in Table 1.

Table 1 Inclusion criteria declined according to Population, Intervention, Comparison, Outcome, and Study design (PICOS) guidelines

	Inclusion criteria
Population (P)	Patients who require antimicrobial prescriptions in both inpatient or outpatient settings
Intervention (I)	Artificial intelligence or machine learning tools used in antimicrobial stewardship
Comparison (C)	Traditional clinical decision-making processes without the aid of artificial intelligence
Outcome (O)	Main outcome: measure of the effectiveness of machine learning algorithms to support antimicrobial stewardship programs: AUC, c-statistic, accuracy, sensitivity, specificity, PPV, NPV and F1-score Secondary outcome: clinical outcomes and improvement in patient health outcomes
Study design (S)	Observational, including cohort studies, prospective or retrospective
Time filter	No temporal filter

AUC Area under the curve; *PPV* positive predictive value; *NPV* negative predictive value

Study selection

Search results were imported into Mendeley for organization and de-duplication. Titles and abstracts were independently screened, in blind, by two reviewers using Rayyan software (https://rayyan.qcri.org). Full-text articles were obtained and assessed, in blind by two reviewers against the eligibility criteria. Discrepancies were resolved through discussion or by consulting a third reviewer. A PRISMA flowchart was used to document the study inclusion process, recording the number of articles retained at each screening stage and reasons for exclusion at the full-text screening stage.

Data extraction

Data extraction was performed in blind using a standardized template in Microsoft Excel (Microsoft Excel® for Microsoft 365 MSO, Redmond, WA, USA, 2019). The spreadsheet was initially tested on three randomly selected articles to enhance consistency and agreement among the authors. Extracted data included: study characteristics (author name, year of publication, country where the study took place, study design), population characteristics (demographic data, sample size, hospital setting, type of infection, clinical settings), ML model details (ML methods used, training data sets, number of features, data source i.e. clinical and/or laboratory data), and outcomes (predictors, performance validation, clinical outcomes). Discrepancies in data extraction were resolved through discussion or by consulting a third reviewer.

Risk of bias assessment

The risk of bias in the included studies was assessed independently in blind by two authors using the QUADAS-AI tool [8]. This tool allows for a detailed and structured assessment of various bias domains, including patient selection, index test, reference standard, and type of validation. Discrepancies in assessments will be resolved through discussion and, if necessary, by consulting a third senior reviewer to reach a consensus. This comprehensive assessment is crucial for understanding the reliability and validity of the study findings.

Data synthesis and meta-analysis

Data synthesis was organized into tables to provide an overview of study characteristics and findings. In the meta-analysis, only performance measures with data from at least 5 distinct studies were included. Quantitative analyses were conducted on performance measurements related to validation, where available. A quantitative synthesis of the data was planned if feasible, using random-effects models to assess the AUC, accuracy, sensitivity, specificity, PPV, and NPV of ML models. Heterogeneity among studies was assessed using I^2 statistics. Publication bias was assessed using funnel plots and Egger's test if at least ten studies were included in the meta-analysis.

ML models were grouped based on functional similarities, considering their approach to data processing, underlying mechanisms and typical applications. This grouping facilitated a more coherent comparison across studies by ensuring that models with comparable characteristics were analysed together. Specifically, the following groups were identified:

- Decision Trees, including Random Forest (RF), Decision Tree (DT), Classification and Regression Trees (CART), and J48 (C4.5).
- Boosted Models, including Extreme Gradient Boosting (XGB), Gradient Boosted Decision Trees (GBDT), Categorical Boosting (CatBoost), Adaptive Boosting (AdaBoost), Boosted Decision Trees (Boosted DT), Boosted Logistic Regression (Boosted LR), and Gradient Boosting Machine (GBM).
- Neural Networks were divided into two categories: generic networks, such as Artificial Neural Networks (ANN), Neural Networks (NN), Backpropagation Neural Networks (Backpropagation NN), and Neural Networks with SHapley Additive exPlanations (NN with SHAP); and recurrent networks, including Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bidirectional LSTM), Recurrent Neural Networks (RNN), and Gated Recurrent Unit (GRU).
- Support Vector Machines (SVMs), including standard Support Vector Machines (SVM), SVM with Radial Basis Function kernel (SVM with RBF kernel), Sequential Minimal Optimization (SMO), Linear Support Vector Machines (Linear SVM), Polynomial Support Vector Machines (Polynomial SVM), and SVM C-Support Vector.

Results

Literature search

A total of 3,458 records were identified through searches in PubMed/MEDLINE ($n = 992$), Scopus ($n = 1,134$), Embase (705), and Web of Science (627) databases. Additionally, two articles were included based on reference screening. Following the initial removal of duplicates ($n = 1,825$), a total of 1,633 records underwent screening based on title and abstract. Subsequently, 1,477 records were eliminated due to non-original content and focus on different topics, resulting in 156 records deemed eligible for inclusion. Three

articles [9–11] lacked full-text availability. Following the full texts assessment, 75 records were excluded, including 80 records [12–90]. The study selection process is visually represented in Fig. 1.

Descriptive characteristics of included studies

The systematic review comprised 80 studies, encompassing a total of over 1.3 million patients. All studies were published after the year 2000. The geographic distribution is notably concentrated (Figure S1), with the highest contribution from the USA ($n=20$ studies) [14, 19, 20, 22, 28, 30, 31, 36, 41, 43–45, 61, 62, 66, 67, 69, 72, 89, 90], Taiwan ($n=9$) [13, 32, 39, 50, 59, 60, 78, 83, 84], Israel [21, 35, 53, 56, 68, 73, 82, 91] and Spain [34, 49, 54, 57, 64, 65, 74, 80] with 8 studies each. Most studies (70%, $n=56$) [12, 15–19, 21–30, 32, 35, 38, 39, 46–50, 53–58, 60–62, 64, 65, 68, 70, 72–84, 86–89, 92] were monocentric, conducted at a single institution (Table 2). Genitourinary infections were the predominant area of research focus, comprising 38.8% ($n=31$) of total studies [12, 13, 16–20, 22, 23, 25, 33, 35, 41, 44, 53, 56–60, 62, 68–71, 75, 76, 83, 86, 91, 92]. Bloodstream infections accounted for 26.3% ($n=21$) of the studies [13, 16, 17, 24, 31, 43, 46, 49, 50, 59, 60, 62–65, 69, 70, 77, 78, 80, 83], followed by respiratory infections (21.3%, $n=17$) [13, 16, 19, 22, 26, 42, 51, 59, 60, 64, 67, 69–71, 75, 85, 92]. The majority of the studies (82.5%, $n=66$) [12–25, 28–30, 32, 34–37, 40, 42, 43, 45–62, 64–69, 71–74, 77, 78, 80–82, 85–92] report data on inpatient (I) populations, while only 2.5% ($n=2$) [27, 33] have outpatient data (O). A further 7.5% ($n=6$) [41, 44, 63, 75, 76, 83] include both inpatient and outpatient data (I/O), 5% ($n=4$) [26, 31, 38, 84] report data on emergency patients, and 2.5% ($n=2$) [39, 70] do not specify this information (NA).

Of the 80 included studies, only 7 [12, 29, 36, 44, 51, 59, 83] used external validation and one [71] employed both

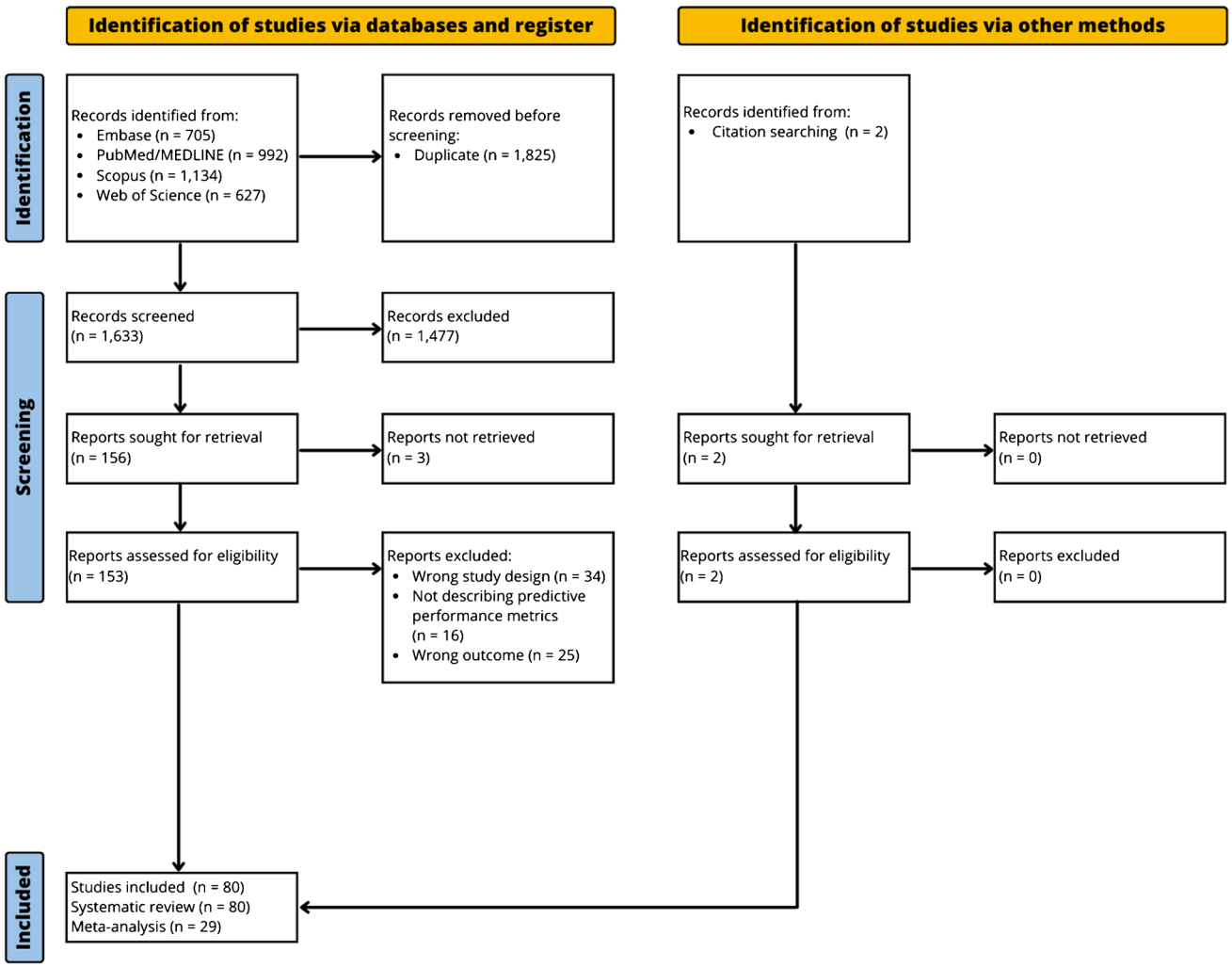

Fig. 1 Flow diagram depicting the selection process

Table 2 Descriptive characteristics of included studies

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Abu-Aqil G, 2022 [35]	RCS	Israel	1	NA	To prove the potential of FTIR spectroscopy combined with ML algorithms for determining the susceptibility of E. coli	Patients with UTI	Test 1798 distinct isolates of E. coli	I	UTI	Yes / Internal	No	NA
Abu-Aqil G, 2023 [91]	ERS	Israel	1	NA	To assess the effectiveness of IR with XGB ML for rapid E. coli identification and antibiotic susceptibility from urine samples, reducing diagnosis time	Patients with suspected UTIs from whom urine samples were collected to identify E. coli isolates	1765 E. coli isolate	I	Urinary	Yes / Internal	NA	16/398
Abu-Aqil G, 2023 [53]	OBS	Israel	1	NA	Evaluate using IR with ML for rapid detection of Proteus mirabilis and P. aeruginosa in urine samples and determining their antibiotic susceptibility	Urine samples from patients with UTI	3446 urine samples analyzed (360 Proteus mirabilis, 353 P. aeruginosa)	I	UTI	NA / NA	No	469

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Abu-Aqil G, 2024 [56]	RCS	Israel	1	NA	To assess the feasibility of IR-based ML for rapid, accurate K. pneumoniae identification and antibiotic susceptibility from urine samples	Patients with UTIs, specifically those infected with K. pneumoniae	2333 patients, 636 with K. pneumoniae infections, and the rest with other UTI-related bacterial infections	I	Urinary	Yes / Internal	No	50–400/469
Agbaria A-H, 2020 [82]	PCS	Israel	1	NA	To evaluate the potential of mid-IR combined with ML algorithms for rapid diagnosis of infection etiology (bacterial versus viral) in febrile pediatric oncology patients	Febrile pediatric oncology patients aged 0–18 years treated with chemotherapy	141 blood samples (50 viral, 71 bacterial, and 20 controls)	I	NA	Yes / Internal	NA	120–127
Al-khlifeh E-M, 2024 [75]	RCS	Jordan	1	Oct 2020-Dec 2022	To predict the AMR for various antibiotic categories using two ML models and assess the impact of demographic factors on AMR	The dataset included 2893 reports; 72.2% female, age range from less than one week to 102 years	2893	I/O	UTI, gastrointestinal, respiratory	Yes / Internal	No	NA

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Ananda-Rajah M-R, 2017 [85]	CCS	Australia	3	NA	Develop an expert system for IMD surveillance using NLP of CT reports, microbiology, and antifungal drug data to improve IMD prediction	Patients from hematology-oncology departments, focusing on high-risk cases for invasive mold disease	123; 64 with IMD and 59 controls	I	Respiratory	NA / NA	No	NA
Ayyildiz H, 2021 [58]	RCS	Turkey	1	2019–2020	To evaluate if E. coli antibiotic resistance can be predicted using ML models on routine lab data without conventional antibiogram tests	Patients with E. coli infections	103	I	UTI	Yes / Internal	No	12/12
Beaudoin M, 2016 [15]	PCS	Canada	1	Feb-Nov 2012; Nov-Dec 2013	To evaluate the ability of the algorithm to discover rules for identifying inappropriate prescriptions of TZP	Patients monitored by APSS who received at least one prescription of TZP at the Centre Hospitalier Universitaire de Sherbrooke	421	I	NA	Yes / Internal	No	NA

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Bhavani S, 2020 [43]	RCS	USA	2	2007–2018	To develop and validate ML models to predict blood culture results at the time of order using routine EHR data	Hospitalized adult patients who received a blood culture	Over 250000 blood culture days across both centers	I	Bloodstream	Yes / Internal	Yes	NA
Bolton W-J, 2022 [71]	RCS	UK	1	June 2017-Jan 2019	To estimate patients' ICU LOS and mortality outcomes for any given day, under the alternative scenarios of stopping versus continuing antibiotic treatment	Patients with UTI and wound infections from Maccabi Healthcare Services with a positive wound culture record	18988 patients, associated with 22845 unique ICU stays; train 15991 stays, validation 3427 stays, test 3427 stays	I	UTI	Yes / Internal	No	43
Bolton W-J, 2024 [62]	OBS	UK	Multiple	2008–2019	To develop and evaluate a ML model to predict the appropriateness of switching from IV to oral antibiotic therapy in ICU settings	ICU patients from MIMIC database and eICU Collaborative Research Database who received both IV and oral antibiotic treatments	8694 unique ICU stays from MIMIC, 1668 from eICU	I	UTI, Respiratory, Sepsis	Yes / Internal (MIMIC), External (eICU)	Yes	Short model: 5, long model: 37

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Brokowski T-J, 2022 [62]	RCS	USA	1	2001–2012	To develop and validate ML models predicting antibiotic susceptibility in S. infections, improving stewardship and optimizing antibiotic selection	Adult patients suspected of having S. infections	4734 unique ICU admissions, with 7655 antibiotic prescriptions	I	Bloodstream, urine, cerebral spinal fluid, pleural cavity, joints	Yes / Internal	No	87/4754
Brown D-G, 2023 [45]	PCS	USA	5	2017–2020	To derive and validate a clinical prediction rule for identifying USA international travelers at risk of acquiring ESBL-PE during travel	USA international travelers ESBL-negative pre-travel who provided stool samples before and after travel	528 travelers	I	Gastrointestinal	Yes / Internal	Yes	27
Bystritsky R-J, 2020 [17]	RCS	EUA	1	Dec 2015-Aug 2017	Predict whether antibiotic therapy requires intervention on a given day based on the note from the antimicrobial stewardship team in the patient's record	Hospitalized patients who received at least one antimicrobial tracked by the ASP at University of California, San Francisco MC	9651; split randomly into derivation (80%) and validation (20%) data sets	I	Bloodstream, UTI	Yes / Internal	Yes	56

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Çağlayan Ç, 2022 [28]	RCS	USA	1	2017–2018	To build a predictive framework for accurate MDRO colonization detection, ensuring high sensitivity and specificity to optimize resource use	Patients admitted to a surgical or medical ICU	3958 patients; 17.59% MDRO, 13.03% VRE, 1.45% CRE, 7.47% MRSA	I	NA	Yes / Internal	No	26
Cai T, 2023 [33]	RCS	Italy	2	Jan 2012-Dec 2020	Predicting the clinical efficacy of the empiric antimicrobial treatment in women with recurrent UTI	Patients who had undergone antimicrobial treatment for uncomplicated cystitis	1043 (725 train, 318 testing)	O	Urinary	Yes / Internal	Yes	NA
Candela A, 2022 [80]	OBS	Spain	1	2017–2019	To develop a MALDI-TOF-based classifying algorithm for discriminating between VRE and VSE faecium isolates	Clinical isolates of Enterococcus faecium from hospitalized patients	178 clinical isolates (92 vancomycin-susceptible, 31 VanA-resistant, and 55 VanB-resistant)	I	Bloodstream, gastrointestinal	Yes / Internal	No	18

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Cavallaro M, 2023 [55]	RCS	UK	1	Jan 2010-Oct 2016	To develop an explainable AI model predicting antimicrobial resistance in hospitals, reducing treatment mismatches, and improving antibiotic stewardship	Inpatients who had GNB isolates from blood and urine cultures, specifically focusing on E. coli, K. pneumoniae, and P. aeruginosa	Amoxicillin + Clavulanic Acid: 3984 patients; CIP: 5124 patients; Meropenem: 5190 patients; Tazobactam: 4402 patients	I	NA	Yes / Internal	No	21/125
Chan A-L, 2006 [32]	RCS	Taiwan	1	NA	To explore using simple, effective supervised learning techniques to predict vancomycin dosage	Patients who received a 5-days course of vancomycin	833	I	NA	Yes / Internal	No	8
Corbin C-K, 2020 [72]	RCS	USA	1	2008–2014	To develop ML models to predict bacterial susceptibility to antibiotics and create personalized antibiograms for patients, improving the empirical selection of antibiotics	Hospitalized patients with confirmed bacterial infections from microbial cultures	NA	I	NA	Yes / Internal	NA	4261

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Corbin C-K, 2022 [20]	RCS	USA	5	Jan 2009 – Dec 2019	To investigate the utility of ML-based clinical decision support for antibiotic prescribing stewardship	Patients who presented to Stanford ED, Massachusetts General Hospital, and Brigham and Women's Hospital in Boston	Stanford: 8342 infections from 6920 patients; Boston: 15806 uncomplicated UTI from 13862 patients. Training (5804 from 2009–2017), validation (1218 from 2018), and test (1320 from 2019)	I	Stanford: NA; Boston: UTI	Yes / Internal	Yes	Stanford 43220; Boston 788
de Vries S, 2022 [12]	RCS	Netherlands	1	Jan 2017-Dec 2018	To report design and evaluation of a CDSS to predict UTI before the urine culture results are available	Inpatients of the University MC Utrecht	906 cultures from 810 patients	I	UTI	Yes / External	No	36
Eickelberg G, 2020 [21]	RCS	Israel	1	2011–2012	To identify ICU patients with low risk of bacterial infection as candidates for earlier EAT discontinuation	ICU adults with suspected community-acquired bacterial infection	10290 (12232 ICU encounters); split into training and test set following a 70/30 split	I	NA	Yes / Internal	No	NA
Feretzakis G, 2020 [92]	RCS	Greece	1	Jan 2017-Dec 2018	To compare the performance of eight ML algorithms to assess antibiotic susceptibility predictions	ICU patients in a public tertiary hospital	345	I	Respiratory, urinary, wound	Yes / Internal	No	23067

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Feretzakis G, 2020 [87]	OBS	Greece	1	Jun 2018-Dec 2018	To evaluate the ability of different ML algorithms to predict antimicrobial resistance using patient demographic data, sample type, and Gram stain results	Patients hospitalized in two Internal Medicine Departments at a Greek public hospital	5590	I	NA	Yes / Internal	NA	5
Feretzakis G, 2021 [16]	RCS	Greece	1	2018	To assess the effectiveness of AutoML trained models to predict AMR	Patients admitted to the internal medicine wards of a public hospital	499 (11496 instances)	I	NA	Yes / Internal	No	6
Feretzakis G, 2021 [88]	OBS	Greece	1	2019	To evaluate the performance of different ML models in predicting AMR for acinetobacter baumannii, K. pneumoniae, and P. aeruginosa using microbiological data	Patients hospitalized in the ICU of a Greek public hospital with isolates of acinetobacter baumannii, K. pneumoniae, and P. aeruginosa	6086	I	Respiratory, bloodstream, urinary	NA / NA	NA	NA

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Garcia-Vidal C, 2021 [65]	RCS	Spain	1	Jan 2008-Dec 2017	Assess MDR-GNB infection risk factors via large data analysis and evaluate ML's ability to predict infection risk at febrile neutropenia onset	Hematological patients with febrile neutropenia	349 (3235 episodes of febrile neutropenia)	I	Bloodstream	Yes / Internal	No	28/43
Goodman K-E, 2022 [22]	RCS	USA	1	July 2017-Dec 2019	To identify patient and treatment characteristics linked to intervention likelihood in a PAF program and develop models to safely exclude certain antimicrobial orders from review	Patient with antimicrobial orders from University of Maryland MC	17503; testing set 3435	I	Bacteremia, bone/joint, central nervous system, cardiac/vascular, etc	Yes / Internal	No	33
Guerrero-López A, 2023 [34]	RCS	Spain	20	2014–2019	To evaluate ML application in MALDI-TOF mass spectrometry for predicting K. pneumoniae resistance	Patients with K. pneumoniae infection	402	I	NA	Yes / Internal	No	76

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Hassan N, 2023 [40]	RCS	UK	Multiple	2023	To develop and validate an AI model predicting post-operative infection in surgical patients	Surgical patients undergoing elective abdominal surgery	2716	I	Abdominal	Yes / Internal	No	74
Herman B, 2021 [42]	XS	Indonesia	Multiple	Model Building Jan 2015-Dec 2019; Testing Jan 2020-Oct 2020	To develop and evaluate CUHAS-ROBUST, an AI tool for screening RR-TB in resource-limited settings lacking rapid diagnostic tests	Patients with suspected RR-TB	487 for model building + 157 for testing (644 total)	I	Respiratory	Yes / Internal	No	19
Hernàndez-Carnerero À, 2023 [49]	RCS	Spain	1	Jul 2004 -May 2019	To develop a method using LSTM neural networks for predicting antibiotic resistance in pseudomonas aeruginosa nosocomial infections in the ICU	ICU patients from the University Hospital of Fuenlabrada with pseudomonas aeruginosa detected in cultures	Cultures: 755 AMG, 643 CAR, 749 Cephalosporins-4th generation, 749 Penicillins + β-lactamase inhibitors, 483 POLY, 708 QUI	I	NA	Yes / Internal	No	AMG: 36, CAR: 18, Cephalosporins-4th gen: 67, Penicillins + β-lactamase inhibitors: 24, POLY: 55, QUI: 49

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Hirano Y, 2021[61]	RCS	USA	1	2008–2019	To develop and validate an XGB ML model to predict MRSA in mechanically ventilated patients using MIMIC-IV data	Adult patients on mechanical ventilation screened for MRSA during their ICU stay	809 screened for MRSA	I	NA	Yes / Internal	No	28/28
Huang T-S, 2020 [39]	CCS	Taiwan	1	Jan 2016-Oct 2017	Detection of CRKP using MALDI	NA	K. Pneumoniae: 46 resistant, 49 susceptibile	NA	NA	Yes / Internal	No	NA
İlhanlı N, 2024 [76]	RCS	South Korea	1	Oct 2012-Oct 2022	To predict AMR in patients with a UTI using ML models and to interpret these models	Patients with UTI (19–100 years)	3865 (Cephalosporin 3535, TZP 737, TMP-SMX 708, Fluoroquinolone 1582, CAR 1365)	I/O	UTI	Yes / Internal	No	71
Jeon K, 2022 [70]	OBS	South Korea	1	2018	To evaluate the performance of a ML-based system for the presumptive identification of MRSA using MALDI-TOF and comparison of accuracy according to SCCmec types	S. aureus clinical isolates from various specimens (blood, sputum, ear swab, soft tissue, urine, tracheal aspiration, bronchoalveolar lavage, etc.)	194 clinical isolates (106 MRSA and 88 MSSA)	NA	Bloodstream, soft tissue, urinary, respiratory, gastrointestinal, etc	Yes / Internal	Yes	NA

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Jiménez F, 2020 [54]	ERS	Spain	1	Jan 2009-Jan 2018	To develop and evaluate ML models, focusing on multivariate time series forecasting with feature selection, to predict antibiotic resistance outbreaks	Patients at risk of or with antibiotic-resistant infections, particularly in hospitals	NA	I	NA	Yes / Internal	No	Initially: 508; SVM: 13, DT: 7, RF: 5
Kim C, 2023 [86]	CS	South Korea	1	2010–2021	To develop prediction models for antibiotic non-susceptibility to support the selection of empiric antibiotics for HA-UTIs	Patients aged (≥ 18 years) with suspected HA-UTIs	10474	I	UTI	Yes / Internal	Yes	Approximately 20000 predictors; final stage: 140
Kong P–H, 2022 [50]	RCS	Taiwan	1	Dec 2020-Jul 2021	To develop a ML model using MALDI-TOF data to discriminate MRSA from MSSA in patients with bacteremia	Patients with S. aureus bacteremia	366 samples (147 MRSA, 219 MSSA) for training; 182 samples (72 MRSA, 110 MSSA) for testing	I	Bloodstream	Yes / Internal	No	Initially: 508; Included: SVM: 13, DT: 7, RF: 5, PR: 15

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Lapp Z, 2021[69]	PCS	USA	21	NA	Determine the importance of patient characteristics and CRKP genetic background in predicting infection risk in CRKP-colonized patients	Patients with extraintestinal CRKP colonization	331 CRKP Sequence Type 258 isolates analyzed (355 isolates total, 24 were excluded for not being ST258)	I	Respiratory, urinary, bloodstream	Yes / Internal	No	141 (50 patient, 91 genomic)
Lee A-L–H, 2021 [37]	RCS	Hong Kong	3	Jan 2015-Dec 2019	To predict ESBL production in community-onset Enterobacteriaceae bacteraemia	Patients with community-onset bacteraemia	5625; split into a training set and a test set in a 4:1 ratio	I	NA	Yes / Internal	No	133
Lewin-Epstein O, 2021 [73]	CS	Israel	1	May 2013-Dec 2015	To predict the antibiotic resistance profiles of bacterial infections in hospitalized patients using ML algorithms applied to EHR	Hospitalized patients with culture-confirmed bacterial infections, covering various species and sites	More than 16000 AMR tests (samples: Ceftazidime: 2942, GEN: 4360, Imipenem: 2235, Ofloxacin: 3117, TMP-SMX: 3544)	I	NA	Yes / Internal	NA	NA

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Liang Q-Q, 2022 [51]	RCS	China	2	2015–2019	To develop and prospectively validate an ML model predicting CAR-GNB carriage within one week in ICU patients	ICU patients with potential CAR-GNB carriage	10247; 1385 with positive CAR-GNB cultures, 1535 with negative CAR-GNB cultures	I	Respiratory	Yes / External	No	16
Liang Q-Q, 2024 [77]	RCS	China	1	Jan 2015-Dec 2021	To predict whether the pathogen causing bloodstream infections in the ICU is CAR-GNB using ML algorithms	Critically ill patients admitted to the ICU with suspected bloodstream infections	952 positive blood cultures (CAR-GNB 418, non-CAR-GNB 534), 1422 negative blood; 70% training, 15% validation, 15% test	I	Bloodstream	Yes / Internal	NA	NA
Liu Q, 2024 [46]	RCS	China	1	2012–2022	To develop ML models predicting pathogens in secondary immunodeficiency patients, aiding antibiotic decisions before culture results	Patients with secondary immunodeficiency from conditions like hematological malignancies, autoimmune diseases, etc	443	I	Bloodstream, abdominal cavity, etc	Yes / Internal	No	88

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Mancini A, 2020 [18]	RCS	Italy	1	2012–2019	To design, develop, and evaluate, with a real AS dataset, a predictive model useful for predicting MDR UTI onset after patient hospitalization	Patients with nosocomial UTI from Principe di Piemonte Hospital in Senigallia	1486; randomly divided in training set (70%) and test set (30%)	I	UTI	Yes / Internal	Yes	6
Martínez-Agüero S, 2019 [64]	RCS	Spain	1	2003–2015	Develop ML models to predict antimicrobial resistance in P. aeruginosa, aiming for early detection of resistant bacteria to improve patient outcomes and limit the spread of infections	ICU patients	Total patients 2630: AMG: 2177; CAR: 1458; 4th-gen Cephalosporins: 1582; Broad-spectrum antibiotics: 2309; POLY: 570; QUI: 1952	I	Bloodstream	Yes / NA	NA	20–30/78–127

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Martínez-Agüero S, 2022 [57]	RCS	Spain	1	2004–2020	To develop interpretable models for early prediction of AMR in ICU patients using multivariate time-series data and intelligent feature selection	ICU patients with bacterial infections	3470 (627 identified with AMR)	I	Respiratory, bloodstream, urinary, surgical sites, etc	Yes / NA	NA	26/1252
McFadden B-R, 2023 [63]	RCS	Australia	Multiple	Jan 2018-May 2020	Establish an ML pipeline to predict blood culture outcomes from routine blood data, reducing unnecessary cultures and improving diagnostic stewardship	Adult patients whose blood samples were analyzed by Sysmex XN-2000 analyzers	Training: 10965 samples; internal validation: 318 samples; external validation: 1245 samples	I/O	Bloodstream	Yes / Internal	No	NA
McGuire R-J, 2021 [30]	CCS	USA	1	Jan 2012-Oct 2017	To develop a prediction model for CAR resistance	All patients 18 years of age and older admitted to Barnes-Jewish Hospital / Washington University in St. Louis	58752 patients with non-CAR resistant; 1088 patients with CAR resistants; randomly split into training (80%) and test (20%) data sets	I	NA	Yes / Internal	Yes	67

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Moehring R-W, 2021 [14]	RCS	USA	3	Oct 2015-Sep 2017	To evaluate whether variables derived from the EHR accurately identify inpatient antimicrobial use	Adult and pediatric inpatient from Duke University Health System	170294; 80% training and 20% testing sets randomly sampled	I	NA	Yes / Internal	Yes	204
Nigo M, 2024 [36]	RCS	USA	2	Jan 2018-Apr 2021	To make an accurate risk stratification of MRSA	All patients from Memorial Hermann Hospital, Houston	Test 30557; validation 393713	I	NA	Yes / External	No	NA
Oonsivilai M, 2018 [24]	RCS	Cambodia	1	Feb 2013-Jan 2016	To predict Gram stains and whether bacterial pathogens could be treated with standard empiric antibiotic regimens	Children with at least one positive blood culture from Angkor Hospital for Children	AMP + GEN: 243; Ceftriaxone: 68;	I	Bloodstream	Yes / Internal	Yes	35
Page P-S, 2022 [89]	RCS	USA	1	1995–2020	To develop a predictive model to evaluate the likelihood of failure of medical therapy in the setting of SEA	All patients presenting with SEA from primary spinal infections	159	I	SEA	No / NA	Yes	18

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Pascual-Sanchez L, 2021 [74]	RCS	Spain	1	2004–2020	To anticipate the antibiogram outcomes associated with MR bacteria in ICU patients by applying ML techniques	ICU patients with MR bacteria	3476 (628 with MR bacteria)	I	NA	Yes / Internal	NA	174 initially, reduced through feature selection methods
Ravkin H–D, 2024 [68]	RCS	Israel	1	Jan 2010-Dec 2020	Develop an ML tool to identify patients in ESBL-endemic areas who can be safely treated without ESBL-targeted therapy, reducing unnecessary broad-spectrum antibiotic use	Hospitalized patients with positive urine cultures for E. coli or K. pneumoniae	17913	I	UTI	Yes / Internal	Yes	7
Rhodes N-J, 2023 [67]	CCS	USA	2	Jan 2014-Mar 2018	To develop and validate ML models to predict the risk of MRSA in hospitalized patients with community-acquired pneumonia	Adult patients (≥ 18 years) hospitalized with community-acquired pneumonia	1823	I	Respiratory	Yes / Internal	No	NA

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Rich S–N, 2022 [44]	RCS	USA	47	2011–2019	To develop a prediction model using EHR data to identify high-risk patients for antibiotic-resistant UTIs and guide antimicrobial therapy decisions	Patients diagnosed with a UTI who had antibiotic susceptibility tests	TMP-SMX: 9072; NIT: 9762; CIP: 9688; any MDR: 10340	I/O	Urinary	Yes / External	Yes	NA
Schinkel M, 2023 [38]	PCS	Netherlands	1	Oct 2021-Sep 2022	To predict blood culture outcomes and enhance diagnostic stewardship	Adult patients	3035; testing 1606	E	NA	Yes / Internal	No	NA
Seheult J-N, 2023 [41]	RCS	USA	5	Jan 2017-Dec 2017	To develop and validate a ML DT algorithm using retrospective urinalysis and urine culture data	Adult patients from non-maternity inpatient and outpatient units	38361 paired urinalysis and urine culture cases	I/O	Urinary	Yes / Internal	No	NA
Shang J-S, 2000 [90]	RCS	USA	5	Mar 1996-Mar 1997	To investigate the potential of using NN and LR approach in diagnosing MRSA	S. aureus infected patients were obtained from five medical facilities in Pittsburgh	504	I	NA	Yes / Internal	No	8

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Shi Z-Y, 2022 [13]	RCS	Taiwan	25	May 2013-May 2014	To develop efficient ML models for auditing appropriate surgical antimicrobial prophylaxis	Patients with healthcare associated infections receiving at least one antimicrobial drug	7377	I	Bloodstream, urinary, respiratory and surgical site	Yes / Internal	No	26
Sophonsri A, 2023 [66]	RCS	USA	2	Mar 2014-Dec 2019	Identify risk factors for development and mortality of vHABP and evaluate empiric antibiotic management with an ML approach	Hospitalized adults (≥18 years) who developed nosocomial pneumonia within the specified time frame	457 (32.4% vHABP, 37.0% nvHABP, and 30.6% VABP)	I	NA	Yes / Internal	No	18/50
Stracy M, 2022 [19]	RCS	USA	1	2008–2019	To estimate patients' ICU LOS and mortality outcomes for any given day under the alternative scenarios of if they were to stop versus continue antibiotic treatment	Patients who received IV antibiotic treatment for a duration between 1–21 days during an ICU stay, at Beth Israel Deaconess MC, Boston	18988 (22845 unique stays)	I	Respiratory, UTI	Yes / NA	No	43

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Tacconelli E, 2020 [52]	PCS	Italy, Serbia, and Romania	3	Sep 2010-Jun 2013	To measure the impact of antibiotic exposure on ESBL-GNB colonization, accounting for confounders using ML methods	Hospitalized patients in medical and surgical wards	10034 (28322 rectal swab)	I	Gastrointestinal	Yes / NA	NA	39
Takács A-T, 2023 [48]	RCS	Hungary	1	Jan 2016-Dec 2017	To develop a DT model using ML to distinguish EBV and CMV infections from other tonsillopharyngitis causes in children, reducing unnecessary antibiotic use	Children (0–18 years) with tonsillopharyngitis	242	I	Pharynx	Yes / Internal	No	23
Tran-The T, 2024 [47]	RCS	Republic of Korea	1	Jan 2014-Dec 2019	To develop explainable ML models to assist ASP by prioritizing inpatients who would benefit most from interventions	Adult inpatients (18–90 years) receiving systemic antibiotics at the hospital	19468	I	NA	Yes / Internal	No	406

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Tsai W–C, 2023 [84]	RCS	Taiwan	1	Jan 2017-Dec 2020	To evaluate the ability of the RF algorithm to predict bacteremia in adult febrile patients in the ED	Febrile adults (≥20 years) with at least one blood culture taken in the ED	5647; divided into derivation (3369) and validation (2278)	E	NA	Yes / Internal	NA	21
Tsurumi A, 2023 [31]	RCS	USA	7	2003–2009	To enable precision medicine in pediatric burn care and develop predictive tools for early intervention in thermally injured children	Children with burns	82	E	Bloodstream	No / NA	No	NA
Tzelves L, 2022 [23]	RCS	Greece	1	2019	To evaluate the performance of ML techniques in predicting bacterial resistance in patients with urolithiasis	Patients with urolithiasis hospitalized in a urology department	239	I	UTI	Yes / Internal	NA	NA
Viswanathan V, 2024 [27]	PCS	India	1	NA	To evaluate the efficacy of autofluorescence imaging of diabetic foot ulcers using a novel artificial intelligence	Patient with diabetic foot ulcers	178 (203 tissue samples)	O	Foot	No / NA	No	NA

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Wang C-X, 2022 [60]	RCS	Taiwan	1	Jun 2013-Feb 2018	To develop and validate ML models for rapid detection of CIRKP using mass spectrometry data combined with patient demographics	Patients infected with K. pneumoniae	16697 samples initially, with 15782 samples selected after quality control (11354 CISKP and 4428 CIRKP)	I	Bloodstream, body fluid, wound, urinary, etc	Yes / Internal	No	102/480
Wang J, 2022 [81]	CS	China	1	Jan 2020-Dec 2021	To evaluate the ability of RF, SVM, and SVM with Kernel algorithms to distinguish CRKP from CSKP using MALDI-TOF data	Clinical isolates of K. pneumoniae (95 CRKP and 76 CSKP)	171 isolates (95 CRKP, 76 CSKP)	I	NA	Yes / Internal	No	211, with 47 housekeeping peaks
Wang Y, 2023 [29]	PCS	China	1	Aug 2021-Jan 2022; validation May 2022-Jul 2022	To develop a model for predicting MDRO infection in critically ill patients	Patients admitted to a ICU	688; divided randomly into training set (80%) and test set (20%); validation set 238	I	NA	Yes / External	Yes	9

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Wang Z, 2021 [59]	RCS	Taiwan	2	Discovery cohort 2013–2019; Replication cohort 2015–2017	To build a predictive model using mass spectrometry and demographic data to characterize methicillin resistance in S. aureus	Patients with S. aureus infections	Discovery population: 23099; replication population: 4962	I	Wound, respiratory, bloodstream, sterile body fluid, UTI	Yes / External	No	217/217
Wong J-G, 2020 [26]	PCS	Singapore	1	Jun 2016-Nov 2018	To develop prediction models based on local clinical and laboratory data to guide antibiotic prescribing for adult patients with uncomplicated upper respiratory tract infections	Patients with uncomplicated URTI at the ED at Tan Tock Seng Hospital	715; using 70% of the participants as training set	E	Respiratory	Yes / Internal	No	50

Table 2 (continued)

Author, year (Ref)	Study type	Country	Cen.	Data time frame	Study objective	Target population	Number of patients	HS	Infection site	TDS/Validation	Calib.	Features
Yu J, 2023 [78]	RCS	Taiwan	1	Oct 2022-Dec 2022	To evaluate the ability of a ML-based MALDI-TOF approach combined with the Rapid Sepsityper protocol to predict MRSA and CRKP from flagged blood cultures	Patients with monomicrobial bloodstream infections	461 (MRSA: 44 S. aureus isolates; CRKP: 126 K. pneumoniae isolates)	I	Bloodstream	Yes / Internal	NA	NA
Zhang Y-M, 2023 [2]	RCS	Taiwan	1	Apr 2018-Mar 2022	To establish an artificial NN model based on MALDI-TOF MS data to rapidly and accurately identify CRKP from CSKP	Clinical isolates of K. pneumoniae collected from various specimens (urine, sputum, blood)	2683 isolates (369 CRKP and 2314 CSKP)	I/O	Urinary, bloodstream, wound, vagina, etc	Yes / External	NA	4500

RCS Retrospective Cohort Study; *ERS* Evaluative Research Study; OBS: Observational Study; *PCS* Prospective Cohort Study; *CCS* Case–Control Study; *XS* Experimental Study; *AMR* Antimicrobial Resistance; *ICU* Intensive Care Unit; *EHR* Electronic Health Record; *MALDI-TOF* Matrix-Assisted Laser Desorption/Ionization Time-of-Flight; *ML* Machine Learning; *NLP* Natural Language Processing; *NA* Not Applicable; *MDRO* Multidrug-Resistant Organism; *CIRKP* Carbapenem-Resistant Klebsiella pneumoniae; *MRSA* Methicillin-Resistant Staphylococcus aureus; *CSKP* Carbapenem-Susceptible Klebsiella pneumoniae; *VRE* Vancomycin-Resistant Enterococcus; *VSE* Vancomycin-Susceptible Enterococcus; *TMP-SMX* Trimethoprim-Sulfamethoxazole; *CIP* Ciprofloxacin; *EAT* Empirical Antibiotic Therapy; *CCT* Clinical Decision Support Tool; *HA-UTI* Hospital-Acquired Urinary Tract Infection; *S.* Staphylococcus; *GNB* Gram-Negative Bacteria; *Bacteremia* Presence of bacteria in the blood; *LOS* Length of Stay; *SEA* Spinal Epidural Abscess; *RR-TB* Rifampicin-Resistant Tuberculosis; *vHABP* Ventilator-Associated Hospital-Acquired Bacterial Pneumonia; *CRKP* Carbapenem-Resistant Klebsiella pneumoniae; *NIT* Nitrofurantoin

internal and external. The number of features used in the studies varied greatly from a minimum of 5 [92] features to a maximum of 43,220 [20], with a mean of 1,375.3. Calibration was performed only in 16 models [14, 17, 18, 20, 24, 29, 30, 33, 43–45, 68, 70, 71, 86, 89].

Characteristics of the features of the included studies

Figure 2 illustrates the distribution of key features analysed in the included studies. The most common were microbiological/laboratory data (91.3%, $n=73$) [12, 14–26, 28–43, 46–60, 62–70, 72, 73, 75–90, 92], followed by clinical data (82.5%, $n=66$)[12–15, 17, 19–26, 28–31, 33, 36–38, 40–62, 64–69, 71–78, 82, 84–90]. Only 11.3% ($n=9$) of the studies included pediatric populations [12–14, 24, 31, 41, 48, 75, 82].

ML algorithms and performance

Table 3 summarizes the algorithms used in the included studies. The most used algorithm was RF ($n=39$) [12, 14, 16, 20–22, 24, 28, 35, 39, 42, 44–46, 50–52, 54, 60, 62–66, 69, 72, 74–77, 80–82, 84–87, 91, 92]. Boosted models were applied 34 times, with XGB used in 21 cases [12, 21, 28, 30, 35, 42, 47, 51, 53, 55, 56, 60, 61, 63, 65, 72, 74, 77, 82, 84, 86], along with other Boosting algorithms (Boosted DT $n=3$[17, 24, 32], Gradient Boosted DT $n=3$ [20, 68, 73], GBM $n=3$ [43, 46, 65], Boosted LR $n=2$ [44, 46], CatBoost $n=1$ [18], AdaBoost ($n=1$)[13]. SVM models were used 22 times (most frequently SVM $n=13$ [12, 18, 21, 34, 39, 50, 54, 58, 60, 77, 80, 82, 85]and SVM with RBF kernel $n=4$ [24, 34, 69, 81]). LASSO was used 8 times [20, 26, 31, 59, 72–74, 86], primarily for variable selection and dimensionality reduction. Neural Networks, including both general NN (Artificial NN $n=5$ [33, 42, 54, 58, 83], NN $n=4$ [18, 37, 73, 90], Backpropagation NN $n=1$ [29], NN with SHAP $n=1$ [71]) and more advanced variants like Recurrent NNs (LSTM $n=2$ [49, 57], Bidirectional LSTM $n=2$ [19, 57], Recurrent NN $n=2$ [25, 36], GRU $n=1$ [57]), were also widely employed, totalling 18 instances. DT models were also common, with DT used in 16 cases [13, 24, 28, 32, 41, 42, 44, 48, 50–52, 58, 64, 67, 74, 77], J48 (C4.5) in three [85, 87, 92], and CART in two [26, 75].

Moreover, Table 3 provides a comprehensive quantitative summary of the predictive performance of various ML models, as measured by AUC. The AUC values range from 41.6 [49] to 99.2 [13]. It was evaluated in three main contexts: predicting antibiotic resistance using ML models, assessing resistance to specific antibiotics, and predicting pathogen-specific resistance to various antibiotics.

Accuracy ranged from 41.6 [49] to 100.0 [75], the latter achieved by an RF model (Table S2). Sensitivity values varied from 0.0 [49], for predicting Quinolone resistance by an LSTM model to 100.0 [63, 82], attained by RF and XGB models (Table S3). Specificity ranged from 15.0 [63] to 100.0 [39, 49, 65], with RF, Naïve Bayes, and LSTM models reaching the upper limit (Table S4). PPV ranged from 6.3 [67]to 100.0 [65, 71, 82]; NPV spanned from 5.0 [71] to 99.6 [67] (Table S5).

Meta-analysis

In our meta-analysis, ML models demonstrated strong predictive performance and diagnostic accuracy across various AMS settings, despite the considerable heterogeneity observed among the included studies. For the AUC, 25 studies and 95 ML models were analyzed. The fixed

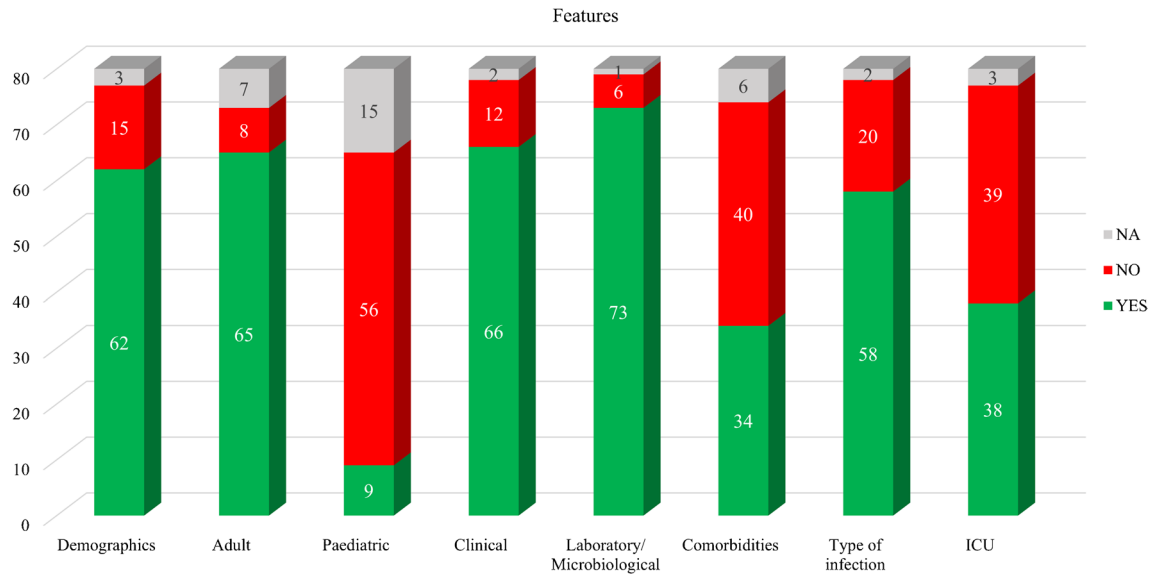

Fig. 2 Distribution of the main features included in the Machine Learning (ML) models. NA = not available; ICU = Intensive care Unit setting

Table 3 Machine Learning (ML) Algorithms and Area Under the Curve (AUC) of ML models from the included studies

Author, year	ML Algorithms	% Area Under the Curve
Abu-Aqil G, 2022	SVM, XGB, RF, k-NN	Amoxicillin: LDA 67.0, XGB 78.0, RF 76.0; Cefuroxime axetil: LDA 75.0, XGB 83.0, RF 82.0; GEN: LDA 73.0, XGB 83.0, RF 83.0; Piperacillin: LDA 67.0, XGB 75.0, RF 72.0; SMX: LDA 68.0, XGB 80.0, RF 79.0
Abu-Aqil G, 2023	MLP, DT, SimpleLogistic, and DT with SpreadSubsample, Bagging, SMOTE and AdaBoost	Ceftazidime: 86.0; Ceftriaxone: 85.0; Cefuroxime: 83.0; Cephalexin: 84.0; CIP: 89.0
Abu-Aqil G, 2023	A 2-stage RF	Ceftazidime: proteus: 80.0, P. aeruginosa: 78.0; CIP: proteus: 86.0, P. aeruginosa: 75.0; GEN: proteus: 82.0, P. aeruginosa: 85.0
Abu-Aqil G, 2024	Combined system, Learning modules	Taxonomic classification of K. pneumoniae versus other UTI bacteria: 99.0; Amoxicillin: 77.0; Ceftazidime: 78.0; Cephalexin: 78.0; CIP: 79.0; Piperacillin: 74.0
Agbaria A-H, 2020	Ensemble Model combining Algorithms, VotingEnsemble, SparseNormalizer with XGB and LightGBM	First Step (Control vs Infected): Naïve Bayes: 97.0, XGB: 98.0, RF: 97.0, SVM: 96.0; Second Step (Viral vs Bacterial): Naïve Bayes: 91.0, XGB: 93.0, RF: 93.0, SVM: 91.0
Al-khlifeh E-M, 2024	Boosted DT	NA
Ananda-Rajah M-R, 2017	Catboost, SVM, NN	Baseline Text Classifier: 73.9 [95% CI 67.1–80.6]; Naive Bayes: 92.8 [95% CI 88.0–97.5]; RF: 94.1 [95% CI 89.8–98.3]; SMO: 92.4 [95% CI 87.5–97.3]; J48(C4.5): 87.2 [95% CI 80.6–93.7]
Ayyildiz H, 2021	Bidirectional LSTM (create with PyTorch)	NA
Beaudoin M, 2016	RF, Gradient Boosted DT, LASSO, Ridge Regression	NA
Bhavani S, 2020	RF, XGB, MLP, SVM, k-NN, VotingEnsemble	Bacteremia Prediction: 78.0 [95% CI 77.0–78.0]
Bolton W-J, 2022	RF	77.0 [95% CI 73.0–80.0]
Bolton W-J, 2024	SVM C-support Vector; SMO; k-NN; J48(C4.5); RF; RIPPER; MLP	MIMIC—Short model: 1st threshold: 78.0 [sd ± 2.0], 2nd threshold: 69.0 [sd ± 3.0]; MIMIC—Long model: 1st threshold: 80.0 [sd ± 1.0], 2nd threshold: 75.0 [sd ± 2.0]; MIMIC Incomplete absorption—Short model: 1st threshold: 73.0 [sd ± 3.0], 2nd threshold: 67.0 [sd ± 5.0]; MIMIC Incomplete absorption—Long model: 1st threshold: 77.0 [sd ± 2.0], 2nd threshold: 73.0 [sd ± 3.0]; MIMIC Malabsorption—Short model: 1st threshold: 76.0 [sd ± 10.0], 2nd threshold: 75.0 [sd ± 11.0]; MIMIC Malabsorption—Long model: 1st threshold: 75.0 [sd ± 7.0], 2nd threshold: 71.0 [sd ± 16.0]; MIMIC UTI—Short model: 1st threshold: 77.0 [sd ± 3.0], 2nd threshold: 74.0 [sd ± 4.0]; MIMIC UTI—Long model: 1st threshold: 78.0 [sd ± 2.0], 2nd threshold: 77.0 [sd ± 4.0]; MIMIC Pneumonia—Short model: 1st threshold: 76.0 [sd ± 3.0], 2nd threshold: 76.0 [sd ± 3.0]; MIMIC Pneumonia—Long model: 1st threshold: 77.0 [sd ± 2.0], 2nd threshold: 74.0 [sd ± 4.0]; MIMIC Sepsis—Short model: 1st threshold: 82.0 [sd ± 5.0], 2nd threshold: 79.0 [sd ± 12.0]; MIMIC Sepsis—Long model: 1st threshold: 77.0 [sd ± 7.0], 2nd threshold: 76.0 [sd ± 18.0]; eICU—Short model: 1st threshold: 72.0 [sd ± 2.0], 2nd threshold: 65.0 [sd ± 5.0]; eICU—Long model: 1st threshold: 72.0 [sd ± 2.0], 2nd threshold: 64.0 [sd ± 6.0]; BEST MODEL 80.0 [sd ± 1.0]
Brokowski T-J, 2022	RF, DT, Boosted DT, Linear SVM, Polynomial SVM, SVM with RBF kernel, kNN	LEV 90.0; Clindamycin 88.0; Vancomycin 87.0; Linezolid 93.0; Oxacillin 89.0; Gentamicin 91.0; Erythromycin 88.0; Daptomycin 92.0. Model not specified
Brown D-G, 2023	Recurrent NN autoencoder	10-Features RF Model: 67.0 [95% CI 66.0–68.0]; 4-Features RF Model: 63.0 [95% CI 62.0–64.0]

Table 3 (continued)

Author, year	ML Algorithms	% Area Under the Curve
Bystritsky R-J, 2020	LASSO, CART	75.0 [95% CI 72.0–79.0]
Çağlayan Ç, 2022	Illuminate® (Adiuvo Diagnostics Private Limited, Chennai, India). Algorithm not defined	RF VRE 77.0; XGB VRE 77.0; RF MRSA 70.0; XGB MRSA 66.0; RF CRE 72.0; XGB CRE 72.0; RF MDRO 89.0; XGB MDRO 87.0; DT MDRO 81.0
Cai T, 2023	RF, XGB, DT	86.7
Candela A, 2022	Backpropagation NN	NA
Cavallaro M, 2023	XGB	Amoxicillin + Clavulanic Acid: 78.0 [95% CI 76.0–81.0]; CIP: 87.0 [95% CI 85.0–89.0]; Meropenem: 99.0 [95% CI 98.0–99.0]; Tazobactam: 79.0 [95% CI 75.0–81.0]
Chan A-L, 2006	LASSO	NA
Corbin C-K, 2020	DT, Boosted DT	E. coli [*N* = 2,424]: AMP: best model: RF 60.0 [95% CI 54.0–65.0]; Cefepime: best model: XGB 73.0 [95% CI 65.0–78.0]; Ceftriaxone: best model: RF 63.0 [95% CI 56.0–69.0]; Cefazolin: best model: RF 60.0 [95% CI 54.0–67.0]; CIP: best model: RF 67.0 [95% CI 63.0–72.0]; LEV: best model: RF 67.0 [95% CI 64.0–72.0]. K. pneumoniae [*N* = 671]: Cefepime: best model: RF 74.0 [95% CI 62.0–84.0]; Ceftriaxone: best model: RF 70.0 [95% CI 61.0–78.0]; Cefazolin: best model: RF 61.0 [95% CI 52.0–70.0]; CIP: best model: LASSO 56.0 [95% CI 44.0–67.0]; LEV: best model: XGB 56.0 [95% CI 44.0–67.0]. P. aeruginosa [*N* = 693]: Cefepime: best model: RF 66.0 [95% CI 50.0–76.0]; CIP: best model: XGB 66.0 [95% CI 53.0–75.0]
Corbin C-K, 2022	Artificial NN	STANFORD DATASET Vancomycin Gradient Boosted DT 72.0 [95% CI 68.0–75.0]; AMP Gradient Boosted DT 62.0 [95% CI 59.0–65.0]; Cefazolin Gradient Boosted DT 67.0 [95% CI 64.0–70.0]; CIP RF 61.0 [95% CI 58.0–64.0]; Ceftriaxone Gradient Boosted DT 69.0 [95% CI 66.0–72.0]; Cefepime RF 65.0 [95% CI 61.0–69.0]; Vancomycin + Ceftriaxone Gradient Boosted DT 67.0 [95% CI 63.0–71.0]; Meropenem Gradient Boosted DT 69.0 [95% CI 65.0–72.0]; TZP RF 64.0 [95% CI 59.0–69.0]; Vancomycin + TZP RF 70.0 [95% CI 62.0–77.0]; Vancomycin + Cefepime RF 70.0 [95% CI 62.0–78.0]; Vancomycin + Meropenem Gradient Boosted DT 73.0 [95% CI 65.0–81.0] BOSTON DATASET TMP-SMX Gradient Boosted DT 60.0 [95% CI 58.0–62.0]; NIT Gradient Boosted DT 57.0 [95% CI 54.0–61.0]; CIP LASSO 64.0 [95% CI 60.0–68.0]; LEV LASSO 64.0 [95% CI 60.0–67.0]
de Vries S, 2022	Gaussian Process with PIKE, KSSHIBA with RBF kernel, SVM with RBF kernel	SVM: 78.2 [sd ± 0.9]; XGB: 80.1 [sd ± 1.1]; RF 80.0 [sd ± 1.0]; k-NN: 78.6 [sd ± 1.0]
Eickelberg G, 2020	LDA, XGB, RF	72 h: RF: 79.3; XGB: 79.5; MLP: 77.9; SVM: 77.8; k-NN: 73.4; VotingEnsemble: 79.3. 48 h: RF: 78.8; XGB: 76.9; MLP: 77.1; SVM: 77.3; kNN: 73.3; VotingEnsemble: 78.8. 24 h: RF: 77.4; XGB: 77.6; MLP: 76.4; SVM: 76.3; k-NN: 71.4; VotingEnsemble: 77.0
Feretzakis G, 2020	Recurrent NN	LibrarySVM C-Support Vector 66.0; SMO 65.9; 1-Nearest Neighbors 68.2; 5-Nearest Neighbors 71.1; J48(C4.5) 72.4; RF 70.3; RIPPER 69.9; MLP 72.6
Feretzakis G, 2020	NN	Multinomial LR: 75.8; MLP: 74.6; k-NN: 72.7; J48 (C4.5): 72.2; RF: 72.1

Table 3 (continued)

Author, year	ML Algorithms	% Area Under the Curve
Feretzakis G, 2021	Blood culture prediction tool. Algorithm not specified	StackEnsemble 85.0; VotingEnsemble 85.0; SparseNormalizer, XGB 84.2; SparseNormalizer, LightGBM 83.7
Feretzakis G, 2021	RF, Naïve Bayes, k-NN, SVM	Original Data: JRip: 87.0, RF: 86.9, MLP: 86.5, Classification via Regression: 93.3, REPT: 87.3; Balanced Data: JRip: 85.7, RF: 87.2, MLP: 86.3, Classification via Regression: 91.8, REP: 87.1
Garcia-Vidal C, 2021	Ensemble Model combining Algorithms: using GBM and other ML algorithms	RF: 78.9; GBM: 78.7; XGB: 79.4
Goodman K-E, 2022	Pittsburgh Urinary Tract (PittUDT) DT	76.0 [95% CI 75.0–77.0]
Guerrero-López A, 2023	Artificial NN, DT, RF, XGB	K. pneumoniae wild type 75.0 [sd ± 11.0]; K. pneumoniae ESBL 72.0 [sd ± 14.0]; K. pneumoniae ESBL + CAR-producers 85.0 [sd ± 14.0]
Hassan N, 2023	GBM	88.6 in training; 85.5 in validation
Herman B, 2021	Boosted LR, RF, DT	Artificial NN 96.0 in training data
Hernàndez-Carnerero À, 2023	RF	AMG: without feature selection: 76.0 [sd ± 5.7], with feature selection: 76.8 [sd ± 5.6]; CAR: without feature selection: 65.8 [sd ± 3.5], with feature selection: 66.5 [sd ± 3.3]; Cephalosporins—4th gen: without feature selection: 63.0 [sd ± 2.9], with feature selection: 63.5 [sd ± 2.6]; Penicillins + β-lactamase inhibitors: without feature selection: 71.1 [sd ± 3.9], with feature selection: 71.6 [sd ± 3.5]; POLY: without feature selection: 41.6 [sd ± 6.1], with feature selection: 42.3 [sd ± 6.7]; QUI: without feature selection: 88.3 [sd ± 3.7], with feature selection: 89.4 [sd ± 3.3]
Hirano Y, 2021	k-NN, Boosted LR, RF, GBM	89.0 [95% CI 83.0–95.0]
Huang T-S, 2020	XGB, LightGBM	NA
İlhanlı N, 2024	DT	RF—Training: Cephalosporin: 77.7 [95% CI 77.5–77.9]; TZP: 86.4 [95% CI 86.2–86.7]; CAR: 87.7 [95% CI 87.4–88.0]; TMP-SMX: 88.1 [95% CI 87.9–88.2]; Fluoroquinolone: 88.4 [95% CI 88.4–88.5]. RF—Test: Cephalosporin: 63.8 [95% CI 63.5–64.2]; TZP: 63.0 [95% CI 62.6–63.4]; CAR: 66.5 [95% CI 65.9–67.1]; TMP-SMX: 67.0 [95% CI 66.6–67.3]; Fluoroquinolone: 72.1 [95% CI 71.8–72.4]
Jeon K, 2022	LSTM	total 87.6; blood isolates 90.6; non-blood isolates 86.3
Jiménez F, 2020	SVM, DT, RF, PR	SVM: 86.6; RF: 82.0; Artificial NN: 80.0; Gaussian Processes: 81.0

Table 3 (continued)

Author, year	ML Algorithms	% Area Under the Curve
Kim C, 2023	DT, RF, XGB	Full Model: AMP: LASSO 73.4, GBM: 72.4, RF: 72.3, stacked ensemble: 68.8; AMP/Sulbactam: LASSO: 71.0, GBM: 74.8, RF: 72.2, stacked ensemble: 76.9; Cefepime: LASSO: 66.0, GBM: 70.7, RF: 67.3, stacked ensemble: 76.1; CIP: LASSO: 71.7, GBM: 70.4, RF: 70.7, stacked ensemble: 77.0; GEN: LASSO: 62.0, GBM: 66.9, RF: 65.9, stacked ensemble: 64.0; Imipenem: LASSO: 61.8, GBM: 73.0, RF: 75.2, stacked ensemble: 80.6; TZP: LASSO: 69.5, GBM: 65.0, RF: 68.8, stacked ensemble: 76.1; TMP/SMX: LASSO: 73.9, GBM: 72.1, RF: 75.7, stacked ensemble: 76.5. Parsimonious Model: AMP: LASSO: 70.6, GBM: 70.2, RF: 69.1, stacked ensemble: 63.7; AMP/Sulbactam: LASSO: 71.8, GBM: 68.6, RF: 67.0, stacked ensemble: 55.9; Cefepime: LASSO: 73.0, GBM: 70.4, RF: 70.9, stacked ensemble: 57.4; CIP: LASSO: 70.0, GBM: 70.3, RF: 69.8, stacked ensemble: 62.5; GEN: LASSO: 65.9, GBM: 62.3, RF: 59.4, stacked ensemble: 54.0; Imipenem: LASSO: 70.8, GBM: 70.7, RF: 68.3, stacked ensemble: 71.3; TZP: LASSO: 64.4, GBM: 67.2, RF: 63.9, stacked ensemble: 56.0; TMP/SMX: LASSO: 73.0, GBM: 71.0, RF: 68.5, stacked ensemble: 54.5
Kong P–H, 2022	RF, DT	SVM: 86.6; DT: 79.6; RF: 82.4; PR: 82.4
Lapp Z, 2021	XGB	Median range = 55.0–68.0 across different feature sets and models
Lee A-L–H, 2021	Gaussian Processes, Artificial NN, RF, SVM	76.1 [95% CI 72.5–79.7]
Lewin-Epstein O, 2021	XGB	Ensemble models: Bacterial species excluded: Imipenem 79.2, Ofloxacin 76.3, Ceftazidime 76.0, SMX-TMP 74.3, Combined 73.9, GEN 72.7; Bacterial species included: Imipenem 88.1, Ofloxacin 80.3, Ceftazidime 82.0, SMX-TMP 79.9, Combined 80.8, GEN 80.4. Ensemble antibiotics performance: Bacterial species excluded: LASSO 75.3, XGB 74.8, NN 74.3, Ensemble models 75.7; Bacterial species included: LASSO 82.0, XGB 82.0, NN 80.3, Ensemble models 82.1
Liang Q-Q, 2022	XGB	Validation set: RF 91.0, XGB 91.0, DT 90.0; Test set: RF 90.0, XGB 89.0, DT 89.0
Liang Q-Q, 2024	MLP, GRU, LSTM, Bidirectional LSTM (with Full Feature, Balanced Binary Cross-Entropy, Masking)	Bloodstream Infection Model: DT: 77.0, RF: 86.0, SVM: 83.0, XGB: 85.0; CR-GNB Bacteremia Model: DT: 69.0, RF: 87.0, SVM: 88.0, XGB: 80.0
Liu Q, 2024	k-NN, Artificial NN, SVM, DT	k-NN: Micro-average: 87.0, Macro-average: 91.0; Boosted LR: Micro-average: 93.0, Macro-average: 92.0; RF: Micro-average: 98.0, Macro-average: 98.0; GBM: Micro-average: 98.0, Macro-average: 97.0
Mancini A, 2020	LASSO	Catboost 73.9 [sd ± 2.1]; SVM 62.8 [sd ± 2.5]; NN 65.2 [sd ± 2.3]
Martínez-Agüero S, 2019	SVM, XGB, RF	NA
Martínez-Agüero S, 2022	XGB	GRU: 66.0 [sd ± 1.3]; LSTM: 66.7 [sd ± 1.8]; Bidirectional LSTM: 64.4 [sd ± 0.8]
McFadden B-R, 2023	LightGBM, RF	Internal validation set: XGB 76.0, RF 82.0; External validation set: 76.0
McGuire R-J, 2021	XGB, RF	Detection CAR resistant 84.6
Moehring R-W, 2021	k-NN, DT, RF, MLP	Adult all antibacterials 87.4 [95% CI 87.0–87.8]; Adult C. difficile agents 86.8 [95% CI 86.4–87.3]; Adult resistant Gram + 91.0 [95% CI 90.6–91.3]; Adult narrow spectrum beta-lactam 88.7 [95% CI 88.3–89.2]; Pediatrical antibacterials 93.3 [95% CI 92.6–93.9]; Pediatric azithromycin 94.4 [95% CI 92.6–96.3]; Pediatric C. difficile agents 94.2 [95% CI 93.4–94.9]; Pediatric Gram + 94.6 [95% CI 93.8–95.5]

Table 3 (continued)

Author, year	ML Algorithms	% Area Under the Curve
Nigo M, 2024	RF, GBM, XGB	91.1 [95% CI 90.0–91.6]
Oonsivilai M, 2018	RF	Resistant to ceftriaxone: RF: 80.0 [95% CI 66.0–94.0]; Boosted DT: 88.0 [95% CI 77.0–100.0]; DT: 79.0 [95% CI 65.0–92.0]; Regularized LR: 75.0 [95% CI 58.0–91.0]; k-NN: 59.0 [95% CI 42.0–77.0]; SVM Radial Kernel: 80.0 [95% CI 64.0–96.0]; SVM Linear Kernel: 81.0 [95% CI 68.0–95.0]; SVM Polynomial Kernel: 83.0 [95% CI 71.0–95.0] Resistant to AMP + GEN: RF: 74.0 [95% CI 59.0–89.0]; Boosted DT: 61.0 [95% CI 43.0–78.0]; DT: 54.0 [95% CI 46.0–63.0]; Regularized LR: 59.0 [95% CI 43.0–75.0]; k-NN: 60.0 [95% CI 43.0–76.0]; SVM Radial Kernel: 56.0 [95% CI 38.0–73.0]; SVM Linear Kernel: 58.0 [95% CI 41.0–75.0]; SVM Polynomial Kernel: 63.0 [95% CI 46.0–80.0] Gram stain: RF: 71.0 [95% CI 57.0–86.0]; Boosted DT: 62.0 [95% CI 46.0–79.0]; DT: 63.0 [95% CI 48.0–77.0]; Regularized LR: 66.0 [95% CI 51.0–82.0]; k-NN: 61.0 [95% CI 45.0–77.0]; SVM Radial Kernel: 68.0 [95% CI 53.0–84.0]; SVM Linear Kernel: 75.0 [95% CI 60.0–89.0]; SVM Polynomial Kernel: 75.0 [95% CI 62.0–89.0]
Page P-S, 2022	DT(Classification Tree Analysis)	80.4
Pascual-Sanchez L, 2021	Gradient Boosted DT	Experiment 1 (Statistics of all features): highest: 76.0 with XGB and Chi-Square feature selection; Experiment 4 (Raw data): highest: 85.0 with XGB and LASSO feature selection; Experiment 2 (Patient features only): highest: 71.0 with XGB and Chi-Square; Experiment 3 (ICU features only): highest: 72.0 with RF and Chi-Square
Ravkin H–D, 2024	Ridge regression, Elastic Net, RF, SVM with RBF kernel	72.0
Rhodes N-J, 2023	Enhanced RF	77.5
Rich S–N, 2022	NN with SHAP	Resistant UTI: TMP-SMX: Boosted LR: 58.0 [SD 1.0], DT: 57.0 [SD 1.0], RF: 57.0 [SD 1.0]; NIT: Boosted LR: 62.0 [SD 1.0], DT: 60.0 [SD 2.0], RF: 57.0 [SD 1.0]; CIP: Boosted LR: 64.0 [SD 1.0], DT: 62.0 [SD 1.0], RF: 63.0 [SD 1.0]. MDR UTI: Boosted LR: 66.0 [SD 2.0], DT: 63.0 [SD 2.0], RF: 62.0 [SD 3.0]
Schinkel M, 2023	LASSO, RF, XGB	78.0 [sd ± 2.0]
Seheult J-N, 2023	LASSO, Gradient Boosted DT, NN, Ensemble Model combining Algorithms	Urine WBC: 79.0; Leukocyte Esterase: 78.0; Bacteria: 77.0
Shang J-S, 2000	RF, LASSO, XGB, DT, MLP	NN 92.8 [sd ± 1.6]
Shi Z-Y, 2022	RF, CART	DT with SpreadSubsample: 97.8; DT: 98.5; Bagging with DT: 99.1; Adaboost with DT: 98.3; DT with SMOTE: 98.5; SimpleLogistic: 98.7; MLP: 99.2
Sophonsri A, 2023	RF	vHABP: 78.0 [95% CI 70.0–85.0]; nvHABP: 80.0 [95% CI 71.0–89.0]; VABP: 83.0 [95% CI 75.0–90.0]
Stracy M, 2022	RF, SVM, DT, XGB	UTI: TMP-SMX 66.0; CIP 74.0; Amoxicillin/Clavulanic Acid 62.0; Cefuroxime—Axetil 67.0; Cephalexin 70.0; NIT 89.0; Fosfomycin 74.0. Wound Infections: TMP-SMX 63.0; CIP 60.0; Amoxicillin/CA 96.0; Cefuroxime—Axetil 58.0; Cephalexin 68.0
Tacconelli E, 2020	LightGBM	Validation set: RF 91.8, XGB 91.2, DT 90.7; Test set: RF 90.8, XGB 89.4, DT 89.8

Table 3 (continued)

Author, year	ML Algorithms	% Area Under the Curve
Takács A-T, 2023	RF	GPT as a standalone marker: 86.0
Tran-The T, 2024	SVM, RF, Partial Least Squares-Discriminant Analysis	Discontinuation: XGB: 80.0, LightGBM: 80.0; IV to Oral: XGB: 80.0, LightGBM: 81.0; Early De-escalation: XGB: 78.0, LightGBM: 78.0; Late De-escalation: XGB: 69.0, LightGBM: 72.0
Tsai W–C, 2023	RF, SVM, SVM with RBF kernel	RF: 76.1; MLP: 74.5; XGB: 74.0; LightGBM: 73.1; qSOFA score: 56.0; RF (validation): 70.9
Tsurumi A, 2023	XGB, RF, Gaussian Naïve Bayes, SVM	Multibiomarker panel: 93.8 [95% CI 88.1–98.1]; Multibiomarker panel + TBSA: 96.5 [95% CI 93.9–99.9]; Multibiomarker panel + TBSA + inhalation injury status: 97.8 [95% CI 94.1–100.0]
Tzelves L, 2022	Artificial NN	Bagging classifier: 87.4; Multinomial LR 76.8
Viswanathan V, 2024	RF, MLP, XGB, LightGBM	NA
Wang C-X, 2022	Naive Bayes, RF, SVM, J48(C4.5), SMO	SVM: 89.0; XGB: 89.0; RF: 85.0
Wang J, 2022	LASSO, XGB, RF, Ensemble Model combining Algorithms	RF: 90.1; SVM: 92.9; SVM with RBF kernel: 93.6
Wang Y, 2023	J48 (C4.5), RF, Multinomial LR, k-NN, MLP	81.1 [95% CI 73.1–89.1]
Wang Z, 2021	Multinomial LR; bagging classifier	Discovery Population: 89.0 [95% CI 88.5–89.3]; Replication Population: 85.0 [95% CI 84.1–86.1]
Wong J-G, 2020	RIPPER (JRip), RF, MLP, Classification via Regression, REPT	LASSO 70.0 [95% CI 62.0–77.0]; CART 67.0 [95% CI 59.0–74.0]
Yu J, 2023	Risk-Calibrated Supersparse Linear Integer Model	S. aureus: training: 85.0, testing: 87.0; K. pneumoniae: training: 82.0, testing: 82.0
Zhang Y-M, 2023	NN	91.0

AMP Ampicillin, *AI* Artificial Intelligence, *CAR* Carbapenems, *CART* Classification and Regression Trees, *C.* Clostridioides, *CIP* Ciprofloxacin, *CRE* Carbapenem-Resistant Enterobacteriaceae, *CR* Carbapenem-Resistant, *DT* Decision Tree, *E.* Escherichia, *eICU* eICU Collaborative Research Database, *ESBL* Extended-Spectrum Beta-Lactamase, *GBM* Gradient Boosting Machine, *GEN* Gentamicin, *GPT* Glutamate Pyruvate Transaminase, *GRU* Gated Recurrent Unit, *GNB* Gram-Negative Bacteria, *K.* Klebsiella, *KSSHIBA* Kernelized Sparse Semi-Supervised Inter-Battery Bayesian Analysis, *LASSO* Least Absolute Shrinkage and Selection Operator, *LDA* Linear Discriminant Analysis, *LEV* Levofloxacin, *LR* Logistic Regression, *LSTM* Long Short-Term Memory, *MDRO* Multidrug-Resistant Organism, *MIMIC* Medical Information Mart for Intensive Care, *ML* Machine Learning, *MLP* Multilayer Perceptron, *MRSA* Methicillin-Resistant Staphylococcus Aureus, *NIT* Nitrofurantoin, *NN* Neural Network, *nvHABP* Non-Ventilator-associated Hospital-Acquired Bacterial Pneumonia, *P.* Pseudomonas, *PIKE* Peak Information Kernel, *POLY* Polymyxins, *PR* Polynomial Regression, *QUI* Quinolones, *RBF* Radial Basis Function, *REPT* Reduced Error Pruning Tree, *RF* Random Forest, *RIPPER* Repeated Incremental Pruning to Produce Error Reduction, *SHAP* SHapley Additive exPlanations, *SMO* Sequential Minimal Optimization, *SMOTE* Synthetic Minority Over-sampling Technique, *SMX* Sulfamethoxazole, *SVM* Support Vector Machine, *TBSA* Total Burn Surface Area, *TMP* Trimethoprim, *TZP* Piperacillin/tazobactam, *UTI* Urinary Tract Infection, *VABP* Ventilator-Associated Bacterial Pneumonia, *vHABP* Ventilator-associated Hospital-Acquired Bacterial Pneumonia, *VRE* Vancomycin-Resistant Enterococci, *WBC* White Blood Cells, *XGB* Extreme Gradient Boosting

effects model (FEM) revealed an ES of 83.40 (95% CI: 83.40–83.40, $p < 0.001$) based on a total of 1,075,058 participants, though significant heterogeneity was observed ($I^2 = 100\%$, $p < 0.001$). Applying a random effects model (REM), the ES decreased to 72.28 (95% CI: 70.42–74.14, $p < 0.001$), reflecting lower predictive performance when accounting for study variability. Publication bias was detected through visual inspection of the funnel plot and confirmed by Egger's regression test (intercept: −484.04, $p = 0.020$). These results are presented in Fig. 3 (a: Forest plot, b: Funnel plot).

Regarding accuracy, we evaluated 8 studies comprising 48 ML models. The FEM showed an ES of 76.15 (95% CI: 76.14–76.16, $p < 0.001$), again with high heterogeneity ($I^2 = 100\%$) and a total sample of 95,626 patients. The REM yielded a comparable ES of 74.97 (95% CI: 73.35–76.58, $p < 0.001$). No evidence of publication bias was found in this case, as indicated by the funnel plot and Egger's test (intercept: 31.69, $p = 0.368$). Results are illustrated in Fig. 4 (a: Forest plot, b: Funnel plot).

For sensitivity, 10 studies with 51 models were included. The FEM showed an ES of 72.68 (95% CI: 72.67–72.70, $p < 0.001$), alongside substantial heterogeneity ($I^2 = 100\%$, $p < 0.001$). The REM, however, indicated an improved ES of 76.89 (95% CI: 71.90–81.89, $p < 0.001$), based on 69,982 patients. No publication bias was detected, as confirmed by Egger's test (intercept: 117.04, $p = 0.113$). Corresponding plots are shown in Fig. 5.

When assessing specificity, 13 studies encompassing 59 models were evaluated. The FEM provided an ES of 99.84 (95% CI: 99.84–99.84, $p < 0.001$), although heterogeneity remained high ($I^2 = 100\%$, $p < 0.001$). The REM showed a significantly lower ES of 73.77 (95% CI: 67.87–79.67, $p < 0.001$), based on a total of 94,113 participants. Visual analysis of the funnel plot and Egger's test (intercept: 401.75, $p < 0.001$) confirmed potential publication bias. These findings are presented in Fig. 6.

For NPV, 6 studies with 12 models were included. The FEM reported an ES of 81.16 (95% CI: 81.14–81.18, $p < 0.001$), with high heterogeneity ($I^2 = 100\%$, $p < 0.001$), and a sample size of 21,338. The REM demonstrated a slightly lower ES of 79.92 (95% CI: 76.54–83.31, $p < 0.001$). No publication bias was identified (Egger's test: intercept 70.68, $p = 0.301$). Results are shown in Fig. 7.

Lastly, for PPV, 5 studies with 11 models were analyzed. The FEM yielded an ES of 56.83 (95% CI: 56.81–56.85, $p < 0.001$), with high heterogeneity ($I^2 = 100\%$, $p < 0.001$) based on 10,976 participants. The REM showed a higher ES of 69.41 (95% CI: 60.19–78.63, $p < 0.001$). No publication bias was detected (Egger's test: intercept 298.55, $p = 0.533$). These results are shown in Fig. 8.

Sensitivity analysis

Sensitivity analysis by study design was not conducted, as all included studies were cross-sectional. Instead, we performed sensitivity analysis based on the type of ML models used. For AUC, when restricted to studies using DT models (CART, RF, J48 [C4.5], DT), 11 studies and 36 ML models were analyzed, covering a total of 291,984 patients. The FEM demonstrated an ES of 79.44 (95% CI: 79.44–79.44, $p < 0.001$), although there was considerable heterogeneity ($I^2 = 100\%$, $p < 0.001$). The REM showed a lower ES of 68.18 (95% CI: 64.10–72.25, $p < 0.001$) (Fig. 9a and S2). Conversely, studies employing boosted models (XGB, GBM, CatBoost, Boosted LR, Boosted DT), comprising 9 studies and 18 models with a sample of 324,381 patients, showed an ES of 78.26 (95% CI: 78.26–78.26, $p < 0.001$) under the FEM. The REM showed an ES of 74.05 (95% CI: 68.02–80.09, $p < 0.001$) (Fig. 9b and S3).

For sensitivity, studies using decision tree-based ML models (5 studies, 18 models) with a sample of 23,155 patients, showed an ES of 78.46 (95% CI: 78.43–79.49, $p < 0.001$) in the FEM, with high heterogeneity ($I^2 = 99.99\%$, $p < 0.001$). The REM produced an ES of 74.97 (95% CI: 71.20–78.74, $p < 0.001$) (Fig. S4). In contrast, 5 studies using general neural networks (Artificial NN, NN, MLP), with 10 models and a sample size of 18,307 patients, showed an ES of 60.98 (95% CI: 60.95–61.00, $p < 0.001$) for the FEM, again with substantial heterogeneity ($I^2 = 99.99\%$, $p < 0.001$). The REM indicated an improved ES of 75.76 (95% CI: 61.25–90.27, $p < 0.001$) (Figure S5).

For specificity, 6 studies using DT models (RF, CART, DT, J48) were included, covering 20 models and a sample of 24,211 patients. The FEM showed an ES of 51.89 (95% CI: 51.86–51.91, $p < 0.001$), with substantial heterogeneity ($I^2 = 100\%$, $p < 0.001$). The REM showed a higher ES of 71.83 (95% CI: 59.13–84.53, $p < 0.001$) (Figure S6).

The high heterogeneity observed ($I^2 = 100\%$) can be attributed to differences in study designs, patient populations (e.g., ICU vs. general wards, adult vs. pediatric), and infection types (e.g., bloodstream vs. urinary infections). Additionally, variations in data sources (clinical vs. microbiological) contributed to the variability in outcomes.

Quality Assessment of included studies: QUADAS-AI

A detailed analysis of the risk of bias assessment and concerns regarding applicability was performed for each study and summarized in Figs. 10a and b. Risk of bias: 19 studies (24%) [13, 17, 26–28, 30, 32, 39–41, 53, 54, 56, 64, 67–70, 85] had a high risk of bias in patient selection mostly due to the lack of a clear rationale for its sample size and the unspecified data source. Most of the studies had a high risk of bias for index test (55%), which was most often due to the

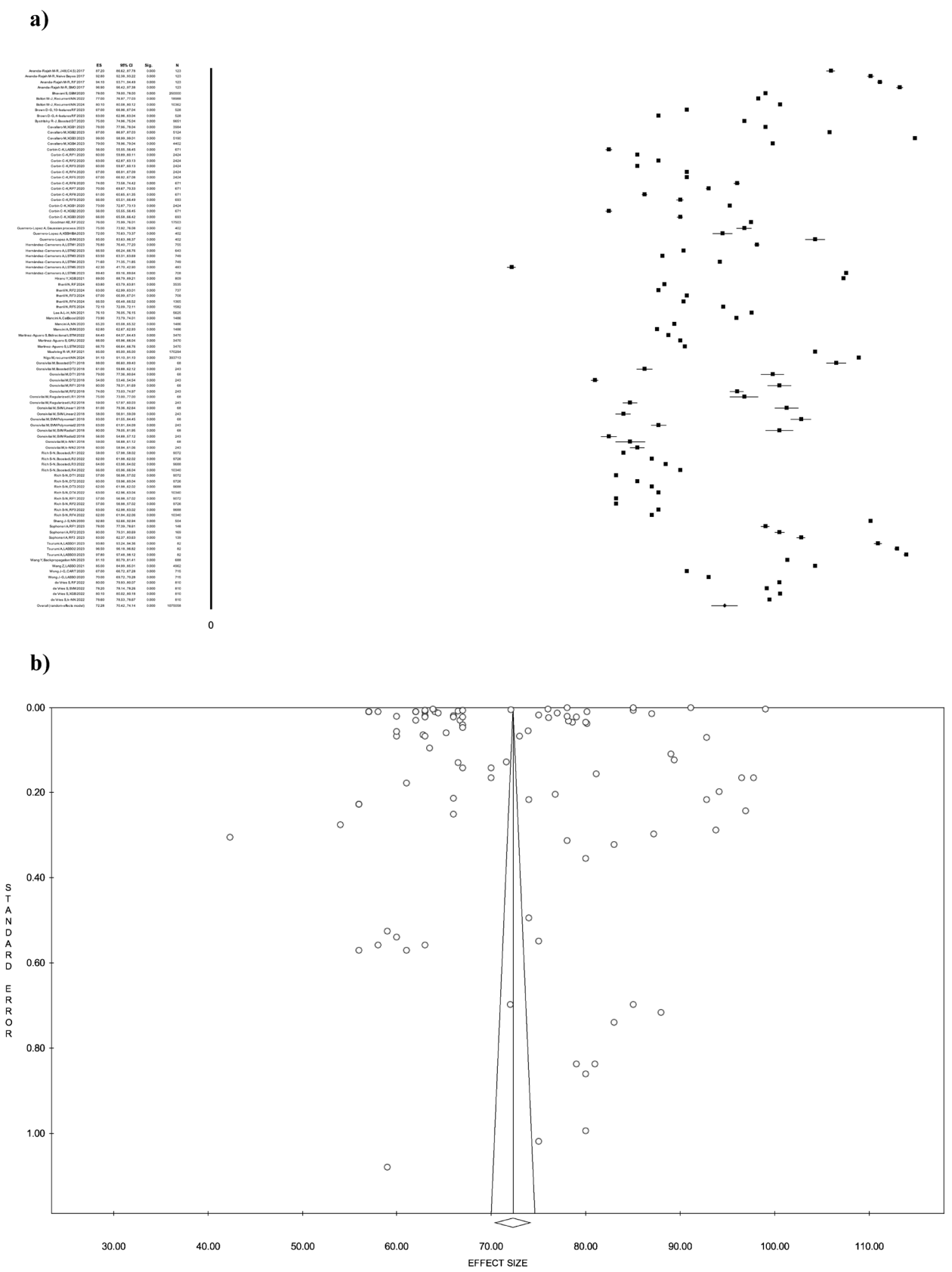

Fig. 3 **a** Forest plot and **b** funnel plot of the random effect model assessing the AUC

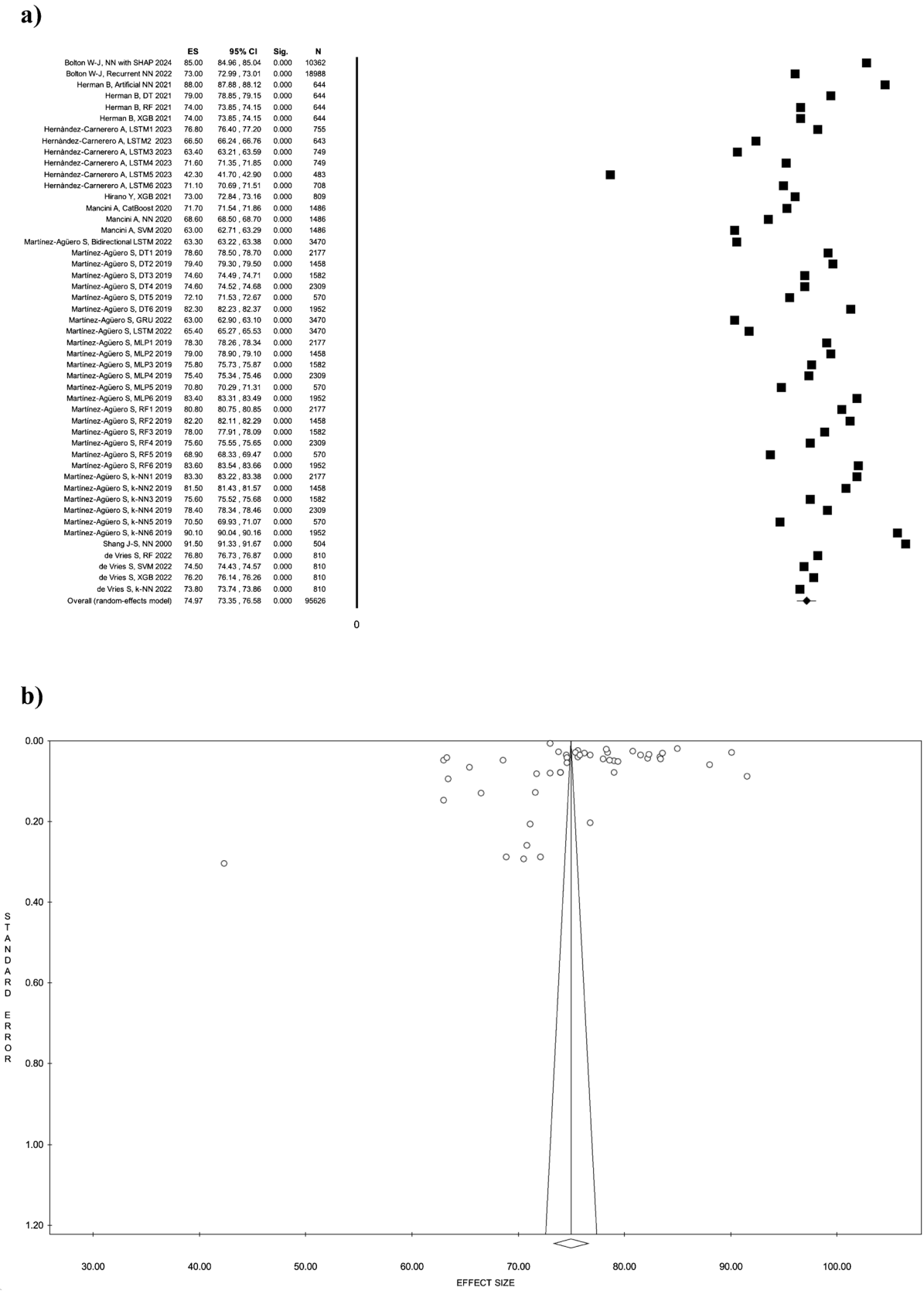

Fig. 4 **a** Forest plot and **b** funnel plot of the random effect model assessing the Accuracy

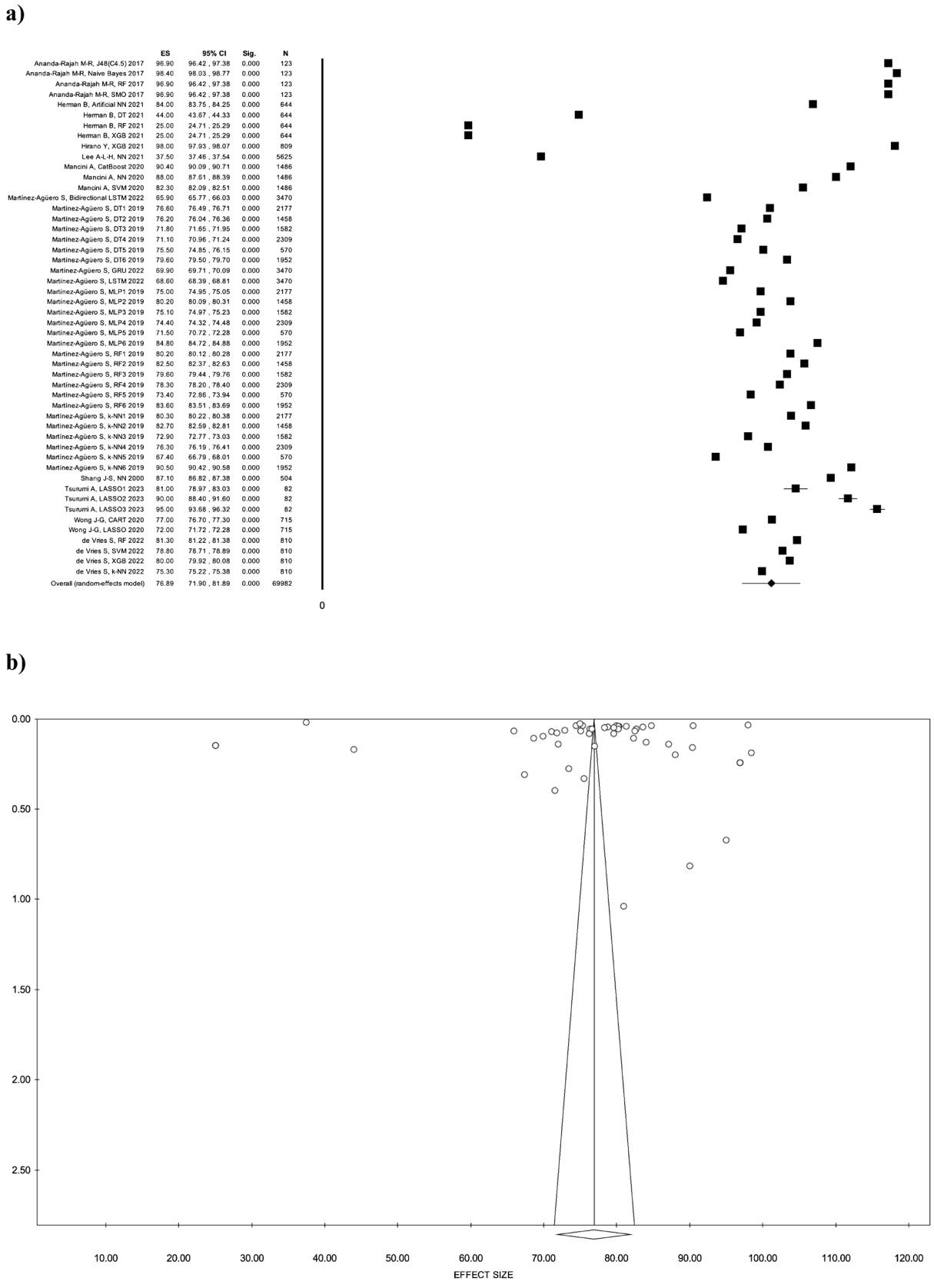

Fig. 5 **a** Forest plot and **b** funnel plot of the random effect model assessing the Sensitivity

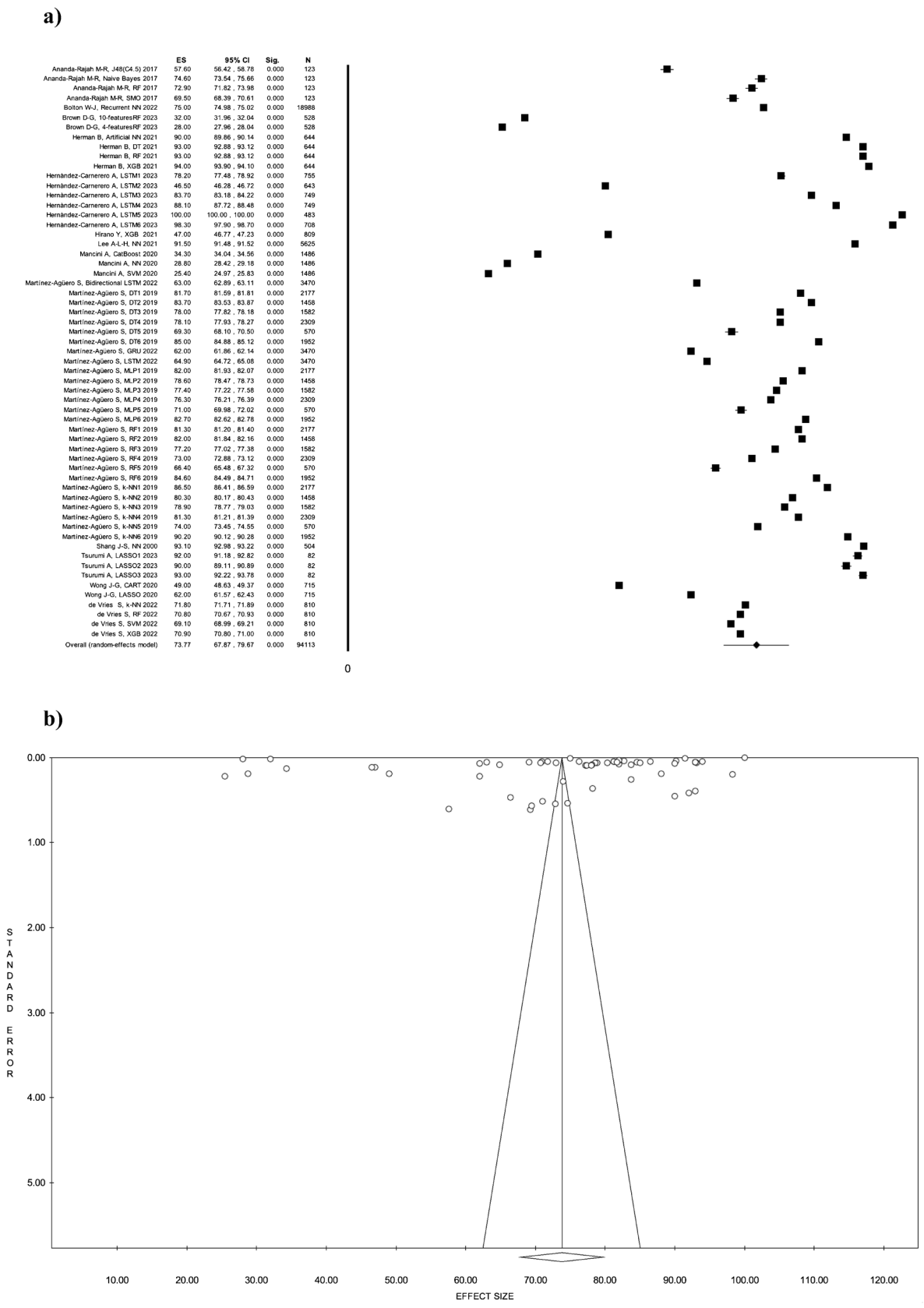

Fig. 6 **a** Forest plot and **b** funnel plot of the random effect model assessing the Specificity

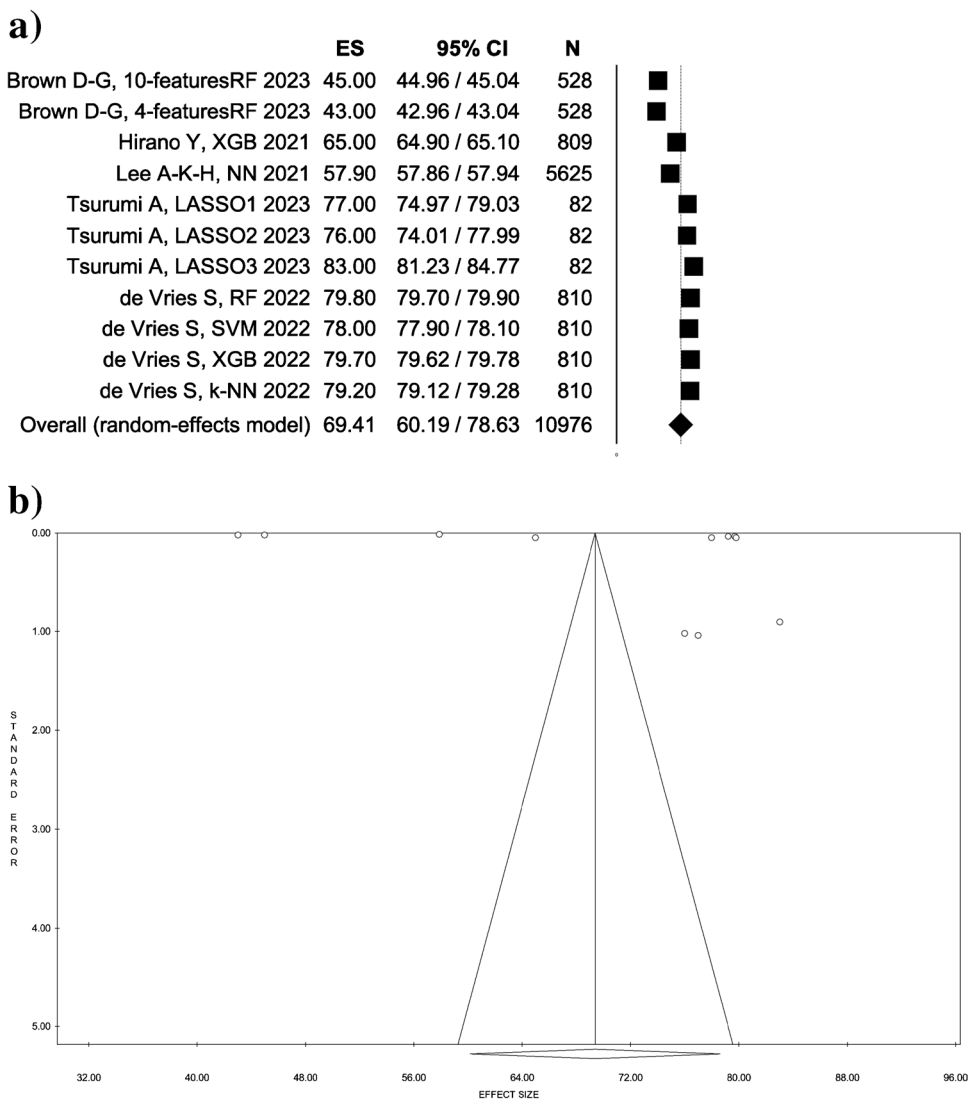

Fig. 7 **a** Forest plot and **b** funnel plot of the random effect model assessing the PPV

lack of validation or testing on external data. In the reference standard domain, which assesses whether the method used as the gold standard for diagnosis or outcome measurement is reliable and applied appropriately, 10% of articles [26, 31, 39, 53, 68, 74, 75, 89] were at a high risk of bias. Finally, the risk of bias in the flow and timing domain was high in 18% of studies [12, 13, 17, 23, 32, 34, 38, 39, 53, 54, 66, 89, 90, 92]. Applicability concerns: In the patient selection domain, concerns about applicability were low only in 28% of the included studies [12–14, 16, 23, 31, 33, 34, 36, 37, 59, 63, 66, 67, 69, 71, 72, 85, 86, 90, 92, 93]. In the index test domain, concerns about applicability were high in 50% of the studies due to the lack of detail on the construct or architecture of the algorithm. Finally, in the reference standard domain, concerns about applicability were high in 14% of studies [26, 27, 31, 41, 53, 54, 75, 78, 80, 89, 94]. The main reasons for lower applicability included using surrogate or proxy measures instead of gold-standard diagnostic tests, reliance on internal validation only and patient populations that may not be representative of the broader clinical setting. Additionally, issues like improper patient flow, where inclusion or exclusion criteria are unclear, or where the patient population does not mirror real-world clinical practice, could introduce bias and further reduce applicability.

Discussion

Interpretation of the results

The current systematic review of the literature assessed a total of more than 3 thousand studies, ultimately including 80 after the selection process. This data allows us to understand that, despite the topic being relatively new, the first study dates back to 2000, it is nonetheless gaining significant interest from the scientific community.

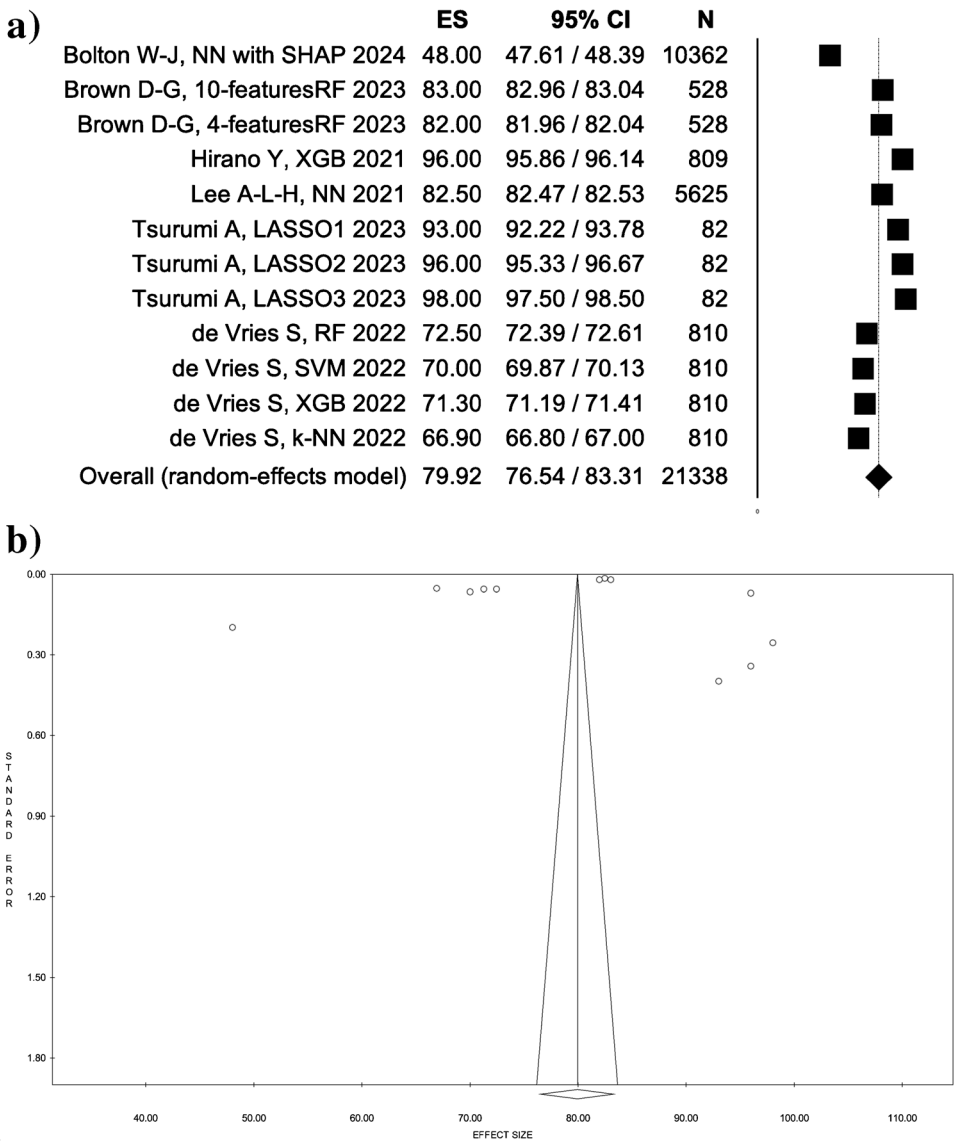

Fig. 8 **a** Forest plot and **b** funnel plot of the random effect model assessing the NPV

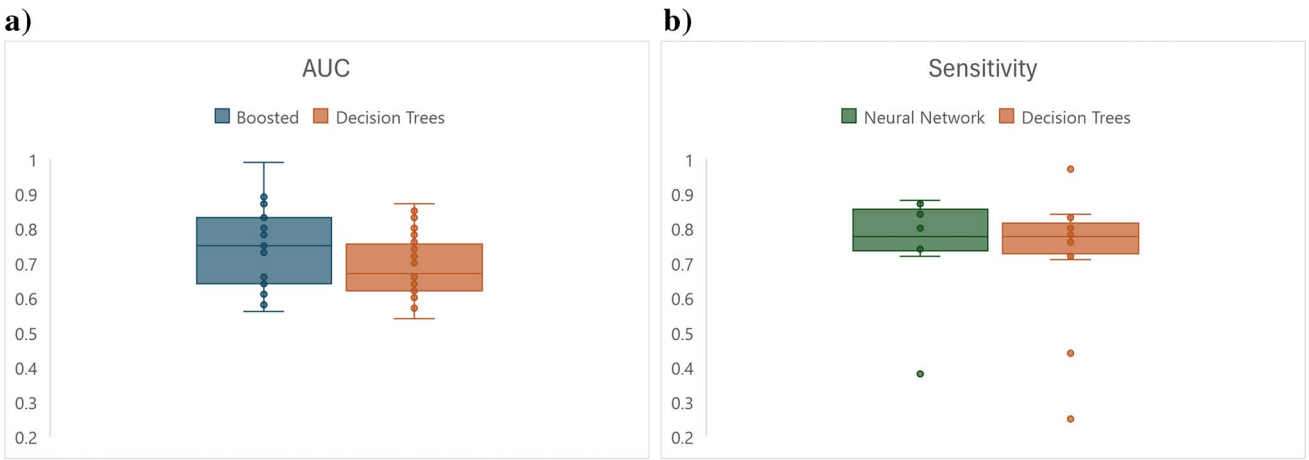

Fig. 9 **a** Sensitivity analysis for the area under the curve (AUC); and **b** for the sensitivity

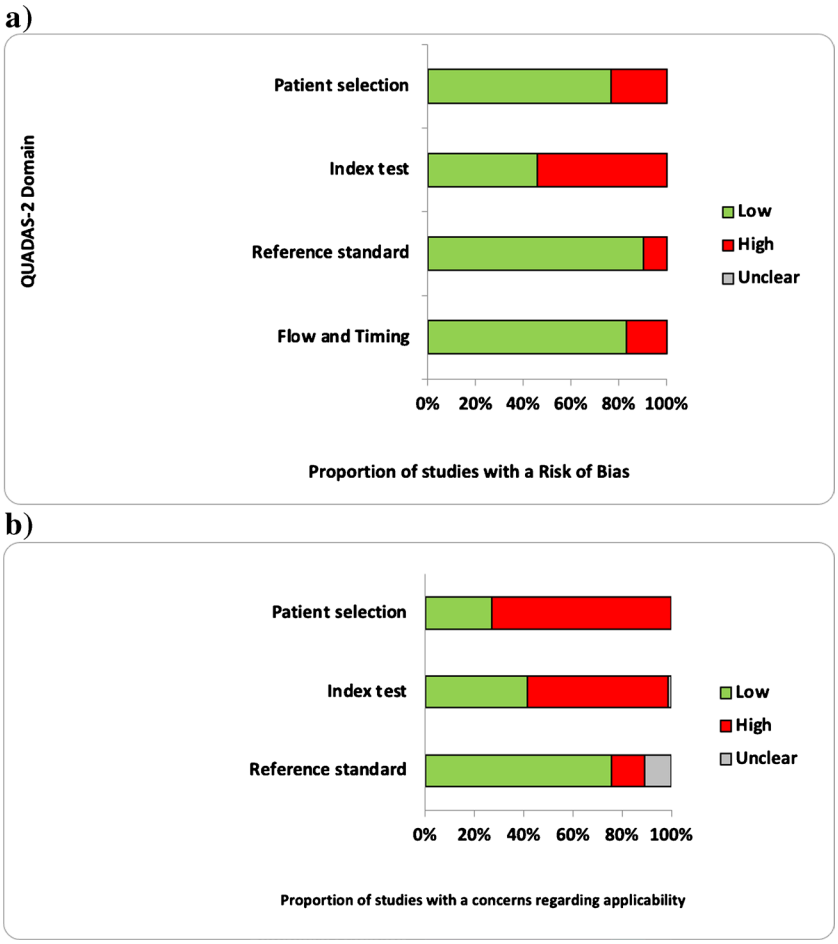

Fig. 10 **a** Proportion of studies with a Risk of bias; and **b** with concerns regarding applicability

However, at the same time, the total number of studies that could be included in the meta-analytical evaluation was 29, highlighting the considerable heterogeneity in how data was collected and analyzed in each individual original study. This meta-analysis and systematic review demonstrated that ML models consistently show strong predictive performance in AMR prediction across different settings, as evidenced by the high AUC scores and diagnostic accuracy metrics. However, there was significant heterogeneity across the included studies, with variability in model performance depending on the algorithm used, the clinical setting, and the infection type. The REM consistently showed lower ES for AUC, accuracy, sensitivity, specificity, and predictive values, emphasizing the need to account for study variability in pooled analyses. The sensitivity analysis, which stratified results by the type of ML model employed, provided deeper insight into the differences between models.

The analysis showed that DT models, such as RF, CART, J48, and traditional DT, generally performed well, achieving consistent ES across metrics. For example, DT models showed a strong AUC in both FEMs (ES = 79.44) and REMs (ES = 68.18). DT models tend to perform well in clinical applications because of their relative simplicity and interpretability [95]. In the context of AMR prediction, decision trees may outperform other models because their structure mirrors clinical decision-making processes, making them easy to validate and implement in real-world settings [96]. This could explain why DT models often exhibit high AUC and specificity in predicting AMR. One of the key determinants of model performance is the type and number of features included [97]. The studies included in this review had a wide range of feature numbers, from as few as five features to over 43,000. Models that can effectively select and prioritize important features, such as LASSO for dimensionality reduction or RF models that rank feature importance, are more likely to generalize well [98]. RF models, the most frequently used in the studies, inherently manage feature importance and handle large feature sets effectively. These capabilities contribute to their strong performance in predicting AMR across various datasets.

Boosted models similarly excel when feature selection is well-curated, as they iteratively refine the contribution of each feature to the prediction. Boosted models demonstrated superior performance compared to decision tree models in certain metrics, particularly AUC (ES = 74.05 under REMs),

indicating that these models might be better suited for AMR prediction in larger, more complex datasets. Despite their high predictive power, these models require vast amounts of high-quality data [99, 100]. Obtaining such data in healthcare is challenging due to privacy concerns, limited access to comprehensive datasets, and inconsistent record-keeping, particularly in underdeveloped regions [101–104]. Additionally, the temporal nature of medical data, such as evolving disease patterns, poses a challenge for models that rely on static data representations. While advanced architectures like bidirectional LSTMs attempt to address this, temporal dependencies in healthcare data remain complex [102]. These factors contribute to the mixed results seen in the application of neural networks despite their theoretical potential.

NN models, such as general ANNs, RNNs, and LSTM models, demonstrated inconsistent performance compared to decision tree and boosted models. Although neural networks are highly flexible and capable of handling large datasets, they often require substantial data preprocessing and hyperparameter tuning [105]. This might explain why, despite their theoretical potential, neural networks achieved lower ES values in the FEM for sensitivity (ES = 60.98) but showed an improved ES under REMs (ES = 75.76), particularly in studies that used more advanced recurrent architectures such as bidirectional LSTMs.

The predominance of internal validation suggests that some models may have been overfitted to specific datasets. Internal validation, while useful, can inflate performance metrics such as AUC or accuracy by tailoring the model too closely to the training data. In contrast, external validation provides a more rigorous test of model generalizability. This could explain why models that underwent external validation, such as those using decision trees and RF, showed lower but more reliable performance metrics in the REMs. The limited use of external validation in healthcare-related ML models points to a relevant challenge in evaluating true model performance.

The sensitivity analysis confirmed high heterogeneity ($I^2 = 100\%$ across nearly all models and metrics). This likely stems from differences in study design, patient populations, infection types, and the algorithms used. For example, studies focused on high-risk populations (e.g., ICU patients) or specific infections (e.g., genitourinary infections) may produce different results compared to those with more general patient populations. Additionally, low standardization in ML methods might also contributed in high heterogeneity. Publication bias, especially noted in specificity measures, suggests that studies reporting poor performance may have been underreported. This may skew the apparent superiority of certain models, like RF, in specific performance metrics, even if their generalizability to diverse clinical settings is less clear. This bias should be considered when interpreting these findings. Overall, the results confirm that ML models, particularly decision trees and boosted models, perform well in predicting AMR, but their utility is influenced by the specific clinical context, model selection, and dataset characteristics.

Comparison with existing literature

The findings of this review are consistent with existing literature on the use of ML in healthcare, particularly in infectious disease management. Previous studies have shown the utility of ML algorithms, such as RF, SVM, and NN in improving the diagnostic capabilities of predictive models in a variety of clinical settings. Similar to prior reviews [7, 106], this study found that RF and boosted models, such as XGBoost, performed particularly well in terms of AUC, accuracy, and specificity. The widespread application of ML in AMR prediction aligns with earlier findings that ML can assist in stratifying patients and guiding therapeutic decisions in complex cases, such as bloodstream infections and respiratory infections. However, the heterogeneity observed in this review mirrors challenges noted in the broader literature, where differences in data quality, feature selection, and study designs lead to variable model performance. This underscores the need for standardized reporting and model validation practices to improve the comparability of ML models across studies.

Public health implications

This systematic review emphasizes the significant role ML models could play in AMS programs to combat AMR, offering faster, accurate predictions that optimize antibiotic use and minimize the spread of resistance. Integrating ML into clinical settings can reduce reliance on traditional culture tests, which take days to yield results, allowing clinicians to make informed decisions within hours and avoid premature antibiotic use. This aligns with the One Health approach, as reducing AMR in humans also benefits animal and environmental health. The review shows ML's potential in precision medicine and infection control, especially in high-risk environments like ICUs, with 65% of studies focusing on hospital settings. Additionally, ML can help identify facility-level antibiotic usage predictors, aiding in targeted infection control measures. However, due to publication bias in some studies, further independent validation of these ML models is necessary before broader implementation.

Future directions

Future research should focus on addressing the limitations identified in this review, including the need for more external validation of ML models. Future studies should

aim to validate ML models in diverse clinical settings, particularly in ambulatory and emergency care environments, which were underrepresented in the current review. Additionally, there is a need for more research focused on pediatric populations, as only 11.3% of the studies included children. Another priority for future research should be the development of models that integrate data from multiple institutions, enhancing their generalizability. Finally, the detection of publication bias in some areas suggests the need for more transparency in reporting, and future research should prioritize open data sharing and the use of standardized metrics for evaluating ML model performance.

Strengths and limitations

The main strength of this review is its comprehensive analysis of ML models for AMR prediction across diverse regions and settings, covering over 1.3 million patients to enhance generalizability. The systematic review with meta-analysis rigorously evaluates AUC, accuracy, sensitivity, specificity, and predictive values, highlighting the strengths and weaknesses of different models. Conducted per international guidelines, this work ensures rigor and transparency. However, high study heterogeneity and publication bias, especially in specificity, limit the ability to draw definitive conclusions.

Conclusion

This systematic review and meta-analysis demonstrate the significant potential of ML models to enhance AMS by improving diagnostic accuracy and predicting AMR. ML models, particularly DT, RF, and boosted algorithms, showed strong predictive performance and could serve as valuable tools in assisting AMS teams, and helping to control the global rise of AMR. However, several limitations emerged from this review. The generalizability of these models remains a concern, as many studies lacked standardization in methodologies and validation approaches, potentially limiting their applicability in diverse clinical settings. Additionally, the complexity and interpretability of some ML models pose challenges to their adoption in practice, as they may not be easily understood or trusted by clinicians. In conclusion, while ML models hold great promise for improving AMS, further research is needed to ensure their effective implementation. Standardized guidelines should be introduced to ensure consistency across future studies, and these models should be tested in randomized controlled trials or real-world settings to validate their practical utility in diverse healthcare environments.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10096-024-05027-y.

Author contributions F. P. and V. G. contributed to the study conception and design. Material preparation, the literature search and data analysis were performed by all the authors. The first draft of the manuscript was written by all the authors. F. P. and V. G. critically revised the work. All authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability Data are available from the corresponding author upon reasonable request (risk of bias assessment data).

Declarations

Competing interests The authors declare no competing interests.

References

1. Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A et al (2022) Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 399:629–655
2. Zhang C, Fu X, Liu Y, Zhao H, Wang G (2024) Burden of infectious diseases and bacterial antimicrobial resistance in China: a systematic analysis for the global burden of disease study 2019. Lancet Reg Health West Pac 43:1–12
3. Majumder MAA, Rahman S, Cohall D, Bharatha A, Singh K, Haque M et al (2020) Antimicrobial stewardship: Fighting antimicrobial resistance and protecting global public health. Infect Drug Resist 13:4713–4738
4. Aiesh BM, Nazzal MA, Abdelhaq AI, Abutaha SA, Zyoud SH, Sabateen A (2023) Impact of an antibiotic stewardship program on antibiotic utilization, bacterial susceptibilities, and cost of antibiotics. Sci Rep. 13:1–9. https://doi.org/10.1038/s41598-023-32329-6
5. Fanelli U, Chiné V, Pappalardo M, Gismondi P, Esposito S (2020) Improving the quality of hospital antibiotic use: impact on multidrug-resistant bacterial infections in children. Front Pharmacol 11:2019–2021
6. Ali T, Ahmed S, Aslam M (2023) Artificial intelligence for antimicrobial resistance prediction: challenges and opportunities towards practical implementation. Antibiotics 12:523
7. Pinto-de-Sá R, Sousa-Pinto B, Costa-de-Oliveira S (2024) Brave new world of artificial intelligence: its use in antimicrobial stewardship—a systematic review. Antibiotics 13:307
8. Guni A, Sounderajah V, Whiting P, Bossuyt P, Darzi A, Ashrafian H (2024) A revised tool for the quality assessment of diagnostic accuracy studies utilising AI: protocol for QUADAS-AI. JMIR Res Protoc 13:e58202
9. Yarahuan JKW, Kisvarday S, Kim E, Yan AP, Nakamura MM, Jones SB et al (2024) An Algorithm to Assess Guideline Concordance of Antibiotic Choice in Community-Acquired Pneumonia. Hosp Pediatr 14:137–145
10. Agbaria AH, Salman A, Beck G, Lapidot I, Rich DH, Kapelushnik J et al (2019) Potential of bacterial infection diagnosis using infrared spectroscopy of WBC and machine learning algorithms. Optics InfoBase Conference Papers. Part F142-:2024
11. Imamović E, Deumić A, Khouly A, Pisil KT, Avdić E, Hukić M et al (2021) Prediction of multi-drug resistance in escherichia coli using machine learning algorithms. IFMBE Proc 84:155–163
12. de Vries S, ten Doesschate T, Totté JEE, Heutz JW, Loeffen YGT, Oosterheert JJ et al (2022) A semi-supervised decision support

system to facilitate antibiotic stewardship for urinary tract infections. Comput Biol Med 146:105621
13. Shi ZY, Hon JS, Cheng CY, Chiang HT, Huang HM (2022) Applying machine learning techniques to the audit of antimicrobial prophylaxis. Appl Sci (Switzerland) 12:2586
14. Moehring RW, Phelan M, Lofgren E, Nelson A, Dodds Ashley E, Anderson DJ et al (2021) Development of a machine learning model using electronic health record data to identify antibiotic use among hospitalized patients. JAMA Netw Open 4:1–12
15. Beaudoin M, Kabanza F, Nault V, Valiquette L (2016) Evaluation of a machine learning capability for a clinical decision support system to enhance antimicrobial stewardship programs. Artif Intell Med 68:29–36. https://doi.org/10.1016/j.artmed.2016.02.001
16. Feretzakis G, Sakagianni A, Loupelis E, Kalles D, Skarmoutsou N, Martsoukou M et al (2021) Machine learning for antibiotic resistance prediction: A prototype using off-the-shelf techniques and entry-level data to guide empiric antimicrobial therapy. Healthc Inform Res 27:214–221
17. Bystritsky RJ, Beltran A, Young AT, Wong A, Hu X, Doernberg SB (2020) Machine learning for the prediction of antimicrobial stewardship intervention in hospitalized patients receiving broad-spectrum agents. Infect Control Hosp Epidemiol 41:1022–1027
18. Mancini A, Vito L, Marcelli E, Piangerelli M, De Leone R, Pucciarelli S et al (2020) Machine learning models predicting multidrug resistant urinary tract infections using "dsaaS." BMC Bioinforma 21:1–12. https://doi.org/10.1186/s12859-020-03566-7
19. Stracy M, Snitser O, Yelin I, Amer Y, Parizade M, Katz R et al (1979) Minimizing treatment-induced emergence of antibiotic resistance in bacterial infections. Science 2022(375):889–894
20. Corbin CK, Sung L, Chattopadhyay A, Noshad M, Chang A, Deresinksi S et al (2022) Personalized antibiograms for machine learning driven antibiotic selection. Commun Med 2:38
21. Eickelberg G, Sanchez-Pinto LN, Luo Y (2020) Predictive modeling of bacterial infections and antibiotic therapy needs in critically ill adults. J Biomed Inform 109:103540. https://doi.org/10.1016/j.jbi.2020.103540
22. Goodman KE, Heil EL, Claeys KC, Banoub M, Bork JT (2022) Real-world antimicrobial stewardship experience in a large academic medical center: using statistical and machine learning approaches to identify intervention "Hotspots" in an antibiotic audit and feedback program. Open Forum Infect Dis 9:ofac289
23. Tzelves L, Lazarou L, Feretzakis G, Kalles D, Mourmouris P, Loupelis E et al (2022) Using machine learning techniques to predict antimicrobial resistance in stone disease patients. World J Urol. 40:1731–6. https://doi.org/10.1007/s00345-022-04043-x
24. Oonsivilai M, Mo Y, Luangasanatip N, Lubell Y, Miliya T, Tan P et al (2018) Using machine learning to guide targeted and locally-tailored empiric antibiotic prescribing in a children's hospital in Cambodia. Wellcome Open Res 3:1–18
25. Bolton WJ, Rawson TM, Hernandez B, Wilson R, Antcliffe D, Georgiou P et al (2022) Machine learning and synthetic outcome estimation for individualised antimicrobial cessation. Front Digit Health 4:1–12
26. Wong JG, Aung AH, Lian W, Lye DC, Ooi CK, Chow A (2020) Risk prediction models to guide antibiotic prescribing: a study on adult patients with uncomplicated upper respiratory tract infections in an emergency department. Antimicrob Resist Infect Control 9:1–11
27. Viswanathan V, Govindan S, Selvaraj B, Rupert S, Kumar R (2024) A clinical study to evaluate autofluorescence imaging of diabetic foot ulcers using a novel artificial intelligence enabled noninvasive device. Int J Low Extrem Wounds 23:169–176
28. Cağlayan Ç, Barnes SL, Pineles LL, Harris AD, Klein EY (2022) A data-driven framework for identifying intensive care unit admissions colonized with multidrug-resistant organisms. Front Public Health 10:1–17
29. Wang Y, Wang G, Zhao Y, Wang C, Chen C, Ding Y et al (2023) A deep learning model for predicting multidrug-resistant organism infection in critically ill patients. J Intensive Care. 11:1–11. https://doi.org/10.1186/s40560-023-00695-y
30. McGuire RJ, Yu SC, Payne PRO, Lai AM, Vazquez-Guillamet MC, Kollef MH et al (2021) A pragmatic machine learning model to predict carbapenem resistance. Antimicrob Agents Chemother 65:1–10
31. Tsurumi A, Flaherty PJ, Que YA, Ryan CM, Banerjee A, Chakraborty A et al (2023) A preventive tool for predicting bloodstream infections in children with burns. Shock 59:393–399
32. Chan ALF, Chen JX, Wang HY (2006) Application of data mining to predict the dosage of vancomycin as an outcome variable in a teaching hospital population. Int J Clin Pharmacol Ther. 44:533–8
33. Cai T, Anceschi U, Prata F, Collini L, Brugnolli A, Migno S et al (2023) Artificial intelligence can guide antibiotic choice in recurrent UTIs and become an important aid to improve antimicrobial stewardship. Antibiotics. 12:375
34. Guerrero-López A, Sevilla-Salcedo C, Candela A, Hernández-García M, Cercenado E, Olmos PM et al (2023) Automatic antibiotic resistance prediction in Klebsiella pneumoniae based on MALDI-TOF mass spectra. Eng Appl Artif Intell 118:105644. https://doi.org/10.1016/j.engappai.2022.105644
35. Abu-Aqil G, Sharaha U, Suleiman M, Riesenberg K, Lapidot I, Salman A et al (2022) Culture-independent susceptibility determination of E. coli isolated directly from patients' urine using FTIR and machine-learning. Analyst. 147:4815–23
36. Nigo M, Rasmy L, Mao B, Kannadath BS, Xie Z, Zhi D (2024) Deep learning model for personalized prediction of positive MRSA culture using time-series electronic health records. Nat Commun 15:1–11
37. Lee ALH, To CCK, Lee ALS, Chan RCK, Wong JSH, Wong CW et al (2021) Deep learning model for prediction of extended-spectrum beta-lactamase (ESBL) production in community-onset Enterobacteriaceae bacteraemia from a high ESBL prevalence multi-centre cohort. Eur J Clin Microbiol Infect Dis 40:1049–1061
38. Schinkel M, Boerman AW, Paranjape K, Wiersinga WJ, Nanayakkara PWB (2023) Detecting changes in the performance of a clinical machine learning tool over time. EBioMedicine 97:104823. https://doi.org/10.1016/j.ebiom.2023.104823
39. Huang TS, Lee SSJ, Lee CC, Chang FC (2020) Detection of carbapenem-resistant Klebsiella pneumoniae on the basis of matrix-assisted laser desorption ionization time-of-flight mass spectrometry by using supervised machine learning approach. PLoS One 15:1–13
40. Hassan N, Slight R, Weiand D, Morgan G, Vellinga A, Mamdouh H et al (2023) Developing an AI-predictive model for predicting the likelihood of post-operative infection in surgical patients. Int J Pharm Pract 31:ii22-3
41. Seheult JN, Stram MN, Contis L, Pontzer RE, Hardy S, Wertz W et al (2023) Development, evaluation, and multisite deployment of a machine learning decision tree algorithm to optimize urinalysis parameters for predicting urine culture positivity. J Clin Microbiol 61:e0029123

42. Herman B, Sirichokchatchawan W, Pongpanich S, Nantasenamat C (2021) Development and performance of CUHASROBUST application for pulmonary rifampicin-resistance tuberculosis screening in Indonesia. PLoS One. 16:1–19. https://doi.org/10.1371/journal.pone.0249243
43. Bhavani S, Lonjers Z, Carey K, Afshar M, Gilbert E, Shah N et al (2024) The development and validation of a machine learning model to predict bacteremia and fungemia in hospitalized patients using electronic health record data. Crit Care Med 48:1–16
44. Rich SN, Jun I, Bian J, Boucher C, Cherabuddi K, Morris JG et al (2022) Development of a prediction model for antibiotic-resistant urinary tract infections using integrated electronic health records from multiple clinics in North-Central Florida. Infect Dis Ther. 11:1869–82. https://doi.org/10.1007/s40121-022-00677-x
45. Brown DG, Worby CJ, Pender MA, Brintz BJ, Ryan ET, Sridhar S et al (2023) Development of a prediction model for the acquisition of extended spectrum beta-lactam-resistant organisms in U.S. international travellers. J Travel Med 30:1–9
46. Liu Q, Chen Y, Xie P, Luo Y, Wang B, Meng Y et al (2024) Development of a predictive machine learning model for pathogen profiles in patients with secondary immunodeficiency. BMC Med Inform Decis Mak. 24:1–11. https://doi.org/10.1186/s12911-024-02447-w
47. Tran-The T, Heo E, Lim S, Suh Y, Heo KN, Lee EE et al (2024) Development of machine learning algorithms for scaling-up antibiotic stewardship. Int J Med Inform. 181:105300. https://doi.org/10.1016/j.ijmedinf.2023.105300
48. Takács AT, Bukva M, Bereczki C, Burián K, Terhes G (2023) Diagnosis of Epstein-Barr and cytomegalovirus infections using decision trees: an effective way to avoid antibiotic overuse in paediatric tonsillopharyngitis. BMC Pediatr 23:1–9
49. Hernàndez-Carnerero À, Sànchez-Marrè M, Mora-Jiménez I, Soguero-Ruiz C, Martínez-Agüero S, Álvarez-Rodríguez J (2023) Dimensionality reduction and ensemble of LSTMs for antimicrobial resistance prediction. Artif Intell Med. 138:102508. https://doi.org/10.1016/j.artmed.2023.102508
50. Kong PH, Chiang CH, Lin TC, Kuo SC, Li CF, Hsiung CA et al (2022) Discrimination of Methicillin-Resistant Staphylococcus aureus by MALDI-TOF mass spectrometry with machine learning techniques in patients with Staphylococcus aureus Bacteremia. Pathogens 11:586
51. Liang Q, Zhao Q, Xu X, Zhou Y, Huang M (2022) Early prediction of carbapenem-resistant Gram-negative bacterial carriage in intensive care units using machine learning. J Glob Antimicrob Resist. 29:225–31. https://doi.org/10.1016/j.jgar.2022.03.019
52. Tacconelli E, Górska A, De Angelis G, Lammens C, Restuccia G, Schrenzel J et al (2020) Estimating the association between antibiotic exposure and colonization with extended-spectrum β-lactamase-producing Gram-negative bacteria using machine learning methods: a multicentre, prospective cohort study. Clin Microbiol Infect 26:87–94
53. Abu-Aqil G, Lapidot I, Salman A, Huleihel M (2023) Quick detection of proteus and pseudomonas in patients' urine and assessing their antibiotic susceptibility using infrared spectroscopy and machine learning. Sensors. 23:8132
54. Jiménez F, Palma J, Sánchez G, Marín D, Francisco Palacios MD, Lucía López MD (2020) Feature selection based multivariate time series forecasting: An application to antibiotic resistance outbreaks prediction. Artif Intell Med. 104:101818. https://doi.org/10.1016/j.artmed.2020.101818
55. Cavallaro M, Moran E, Collyer B, McCarthy ND, Green C, Keeling MJ (2023) Informing antimicrobial stewardship with explainable AI. PLOS Digital Health. 2:1–20. https://doi.org/10.1371/journal.pdig.0000162
56. Abu-Aqil G, Suleiman M, Lapidot I, Huleihel M, Salman A (2024) Infrared spectroscopy-based machine learning algorithms for rapid detection of Klebsiella pneumoniae isolated directly from patients' urine and determining its susceptibility to antibiotics. Spectrochim Acta A Mol Biomol Spectrosc. 314:124141. https://doi.org/10.1016/j.saa.2024.124141
57. Martínez-Agüero S, Soguero-Ruiz C, Alonso-Moral JM, Mora-Jiménez I, Álvarez-Rodríguez J, Marques AG (2022) Interpretable clinical time-series modeling with intelligent feature selection for early prediction of antimicrobial multidrug resistance. Futur Gener Comput Syst 133:68–83. https://doi.org/10.1016/j.future.2022.02.021
58. Ayyıldız H, Tuncer SA (2021) Is it possible to determine antibiotic resistance of E. coli by analyzing laboratory data with machine learning? Turk J Biochem 46:623–30
59. Wang Z, Wang HY, Chung CR, Horng JT, Lu JJ, Lee TY (2021) Large-scale mass spectrometry data combined with demographics analysis rapidly predicts methicillin resistance in Staphylococcus aureus. Brief Bioinform 22:1–12
60. Wang C, Wang Z, Wang HY, Chung CR, Horng JT, Lu JJ et al (2022) Large-scale samples based rapid detection of ciprofloxacin resistance in Klebsiella pneumoniae using machine learning methods. Front Microbiol 13:1–14
61. Hirano Y, Shinmoto K, Okada Y, Suga K, Bombard J, Murahata S et al (2021) Machine learning approach to predict positive screening of methicillin-resistant staphylococcus aureus during mechanical ventilation using synthetic dataset from MIMIC-IV database. Front Med (Lausanne) 8:1–9
62. Brokowski TJ, Chiang JN (2022) Machine learning for antibiotic stewardship in the treatment of stapholycoccus bacterial infections. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2021629655&from=export. https://doi.org/10.1101/2022.11.28.22282797. Accessed 30 Oct 2024
63. McFadden BR, Inglis TJJ, Reynolds M (2023) Machine learning pipeline for blood culture outcome prediction using Sysmex XN-2000 blood sample results in Western Australia. BMC Infect Dis 23:1–10. https://doi.org/10.1186/s12879-023-08535-y
64. Martínez-Agüero S, Mora-Jiménez I, Lérida-García J, Álvarez-Rodríguez J, Soguero-Ruiz C (2019) Machine learning techniques to identify antimicrobial resistance in the intensive care unit. Entropy 21:1–24
65. Garcia-Vidal C, Puerta-Alcalde P, Cardozo C, Orellana MA, Besanson G, Lagunas J et al (2021) Machine learning to assess the risk of multidrug-resistant gram-negative bacilli infections in febrile neutropenic hematological patients. Infect Dis Ther 10:971–983
66. Sophonsri A, Lou M, Ny P, Minejima E, Nieberg P, Wong-Beringer A (2023) Machine learning to identify risk factors associated with the development of ventilated hospital-acquired pneumonia and mortality: implications for antibiotic therapy selection. Front Med (Lausanne) 10:1–11
67. Rhodes NJ, Rohani R, Yarnold PR, Pawlowski AE, Malczynski M, Qi C et al (2023) Machine learning to stratify methicillin-resistant staphylococcus aureus risk among hospitalized patients with community-acquired pneumonia. Antimicrob Agents Chemother 67:1–10

68. Ravkin HD, Ravkin RM, Rubin E, Nesher L (2024) Machine-learning-based risk assessment tool to rule out empirical use of ESBL-targeted therapy in endemic areas. J Hosp Infect. 149:90–7. https://doi.org/10.1016/j.jhin.2024.04.005
69. Lapp Z, Wiens J (2021) Patient and microbial genomic factors associated with. Am Soc Microbiol 6:1–12
70. Jeon K, Kim JM, Rho K, Jung SH, Park HS, Kim JS (2022) Performance of a machine learning-based methicillin resistance of staphylococcus aureus identification system using MALDI-TOF MS and comparison of the accuracy according to SCCmec types. Microorganisms 10:1903
71. Bolton WJ, Wilson R, Gilchrist M, Georgiou P, Holmes A, Rawson TM (2024) Personalising intravenous to oral antibiotic switch decision making through fair interpretable machine learning. Nat Commun 15:1–13
72. Corbin CK, Medford RJ, Osei K, Chen JH (2020) Personalized Antibiograms: Machine Learning for Precision Selection of Empiric Antibiotics. AMIA Jt Summits Transl Sci Proc. 2020:108–15. Available from: https://www.ncbi.nlm.nih.gov/pubmed/32477629%0A, https://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7233062. Accessed 30 Oct 2024
73. Lewin-Epstein O, Baruch S, Hadany L, Stein GY, Obolski U (2021) Predicting antibiotic resistance in hospitalized patients by applying machine learning to electronic medical records. Clin Infect Dis 72:E848–E855
74. Pascual-Sanchez L, Mora-Jimenez I, Martinez-Aguero S, Alvarez-Rodriguez J, Soguero-Ruiz C (2021) Predicting Multidrug Resistance Using Temporal Clinical Data and Machine Learning Methods. Proceedings - 2021 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021. 2826–33
75. Al-khlifeh EM, Hassanat AB (2024) Predicting the distribution patterns of antibiotic-resistant microorganisms in the context of Jordanian cases using machine learning techniques. J Appl Pharm Sci 14:174–183
76. İlhanlı N, Park SY, Kim J, Ryu JA, Yardımcı A, Yoon D (2024) Prediction of antibiotic resistance in patients with a urinary tract infection: algorithm development and validation. JMIR Med Inform 12:1–11
77. Liang Q, Ding S, Chen J, Chen X, Xu Y, Xu Z et al (2024) Prediction of carbapenem-resistant gram-negative bacterial bloodstream infection in intensive care unit based on machine learning. BMC Med Inform Decis Mak. 24:1–11. https://doi.org/10.1186/s12911-024-02504-4
78. Yu J, Lin HH, Tseng KH, Lin YT, Chen WC, Tien N et al (2023) Prediction of methicillin-resistant Staphylococcus aureus and carbapenem-resistant Klebsiella pneumoniae from flagged blood cultures by combining rapid Sepsityper MALDI-TOF mass spectrometry with machine learning. Int J Antimicrob Agents 62:106994. https://doi.org/10.1016/j.ijantimicag.2023.106994
79. Abu-Aqil G, Suleiman M, Sharaha U, Riesenberg K, Lapidot I, Huleihel M et al (2023) Fast identification and susceptibility determination of E. coli isolated directly from patients' urine using infrared-spectroscopy and machine learning. Spectrochim Acta A Mol Biomol Spectrosc 285:121909. https://doi.org/10.1016/j.saa.2022.121909
80. Candela A, Arroyo MJ, Sánchez-molleda Á, Méndez G, Quiroga L, Ruiz A et al (2022) Rapid and reproducible MALDI-TOF-based method for the detection of vancomycin-resistant enterococcus faecium using classifying algorithms. Diagnostics 12:1–13
81. Wang J, Xia C, Wu Y, Tian X, Zhang K, Wang Z (2022) Rapid detection of carbapenem-resistant Klebsiella pneumoniae using machine learning and MALDI-TOF MS platform. Infect Drug Resist 15:3703–3710
82. Agbaria AH, Rosen GB, Lapidot I, Rich DH, Mordechai S, Kapelushnik J et al (2020) Rapid diagnosis of infection etiology in febrile pediatric oncology patients using infrared spectroscopy of leukocytes. J Biophotonics 13:1–12
83. Zhang YM, Tsao MF, Chang CY, Lin KT, Keller JJ, Lin HC (2023) Rapid identification of carbapenem-resistant Klebsiella pneumoniae based on matrix-assisted laser desorption ionization time-of-flight mass spectrometry and an artificial neural network model. J Biomed Sci. 30:1–10. https://doi.org/10.1186/s12929-023-00918-2
84. Tsai WC, Liu CF, Ma YS, Chen CJ, Lin HJ, Hsu CC et al (2023) Real-time artificial intelligence system for bacteremia prediction in adult febrile emergency department patients. Int J Med Inform 178:105176. https://doi.org/10.1016/j.ijmedinf.2023.105176
85. Ananda-Rajah MR, Bergmeir C, Petitjean F, Slavin MA, Thursky KA, Webb GI (2017) Toward electronic surveillance of invasive mold diseases in hematology-oncology patients: an expert system combining natural language processing of chest computed tomography reports, microbiology, and antifungal drug data. JCO Clin Cancer Inform. 1:1–10
86. Kim C, Choi YH, Choi JY, Choi HJ, Park RW, Rhie SJ (2023) Translation of machine learning-based prediction algorithms to personalised empiric antibiotic selection: a population-based cohort study. Int J Antimicrob Agents. 62:106966. https://doi.org/10.1016/j.ijantimicag.2023.106966
87. Feretzakis G, Loupelis E, Sakagianni A, Kalles D, Lada M, Christopoulos C et al (2020) Using machine learning algorithms to predict antimicrobial resistance and assist empirical treatment. Stud Health Technol Inform. 272:75–8
88. Feretzakis G, Sakagianni A, Loupelis E, Kalles D, Martsoukou M, Skarmoutsou N et al (2021) Using machine learning to predict antimicrobial resistance of acinetobacter baumannii, klebsiella pneumoniae and pseudomonas aeruginosa strains. Public Health Inform Proc MIE 2021:43–47
89. Page PS, Greeneway GP, Ammanuel SG, Brooks NP (2022) Development and validation of a predictive model for failure of medical management in spinal epidural abscesses. Neurosurgery 91:422–426
90. Shang JS, Lin YSE, Goetz AM (2000) Diagnosis of MRSA with neural networks and logistic regression approach. Health Care Manag Sci 3:287–297
91. Abu-Aqil G, Suleiman M, Sharaha U, Lapidot I, Huleihel M, Salman A (2023) Instant detection of extended-spectrum β-lactamase-producing bacteria from the urine of patients using infrared spectroscopy combined with machine learning. Analyst 148:1130–1140
92. Feretzakis G, Loupelis E, Sakagianni A, Kalles D, Martsoukou M, Lada M et al (2020) Using machine learning techniques to aid empirical antibiotic therapy decisions in the intensive care unit of a general hospital in Greece. Antibiotics 9:50
93. Kanjilal S, Oberst M, Boominathan S, Zhou H, Hooper DC, Sontag D (2020) A decision algorithm to promote outpatient antimicrobial stewardship for uncomplicated urinary tract infection. Sci Transl Med 12:568
94. Chan E, Leroi M (2021) Evaluation of the VITEK 2 Advanced Expert System performance for predicting resistance mechanisms in Enterobacterales acquired from a hospital-based screening program. Pathology 53:763–767
95. Azar AT, El-Metwally SM (2013) Decision tree classifiers for automated medical diagnosis. Neural Comput Appl 23:2387–2403
96. Dos Santos CFG, Papa JP (2022) Avoiding overfitting: a survey on regularization methods for convolutional neural networks. ACM Comput Surv. 54:1–25
97. Bailly A, Blanc C, Francis É, Guillotin T, Jamal F, Wakim B et al (2022) Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models. Comput Methods Programs Biomed. 213:106504. https://doi.org/10.1016/j.cmpb.2021.106504
98. Bailey J, Oliveri A, Levin E (2013) A review of feature reduction techniques in neuroimagin. Bone 23:1–7

99. Ahmed SF, Alam MS Bin, Hassan M, Rozbu MR, Ishtiak T, Rafa N et al (2023) Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artif Intell Rev. Springer Netherlands. https://doi.org/10.1007/s10462-023-10466-8
100. Szmigiel A, Apel DB, Pu Y, Pourrahimian Y, Dehghanpour H (2024) Exploring machine learning techniques for open stope stability prediction: a comparative study and feature importance analysis. Rock Mech Bull. 100146. https://doi.org/10.1016/j.rockmb.2024.100146
101. Zion I, Ozuomba S, Asuquo P (2020) An Overview of Neural Network Architectures for Healthcare. 2020 International Conference in Mathematics, Computer Engineering and Computer Science, ICMCECS 2020. 10
102. Pandey SK, Janghel RR (2019) Recent deep learning techniques, challenges and its applications for medical healthcare system: a review. Neural Process Lett. 50:1907–35. https://doi.org/10.1007/s11063-018-09976-2
103. Perez H, Tah JHM (2020) Improving the accuracy of convolutional neural networks by identifying and removing outlier images in datasets using t-SNE. Mathematics. 8:662
104. Chen RC, Dewi C, Huang SW, Caraka RE (2020) Selecting critical features for data classification based on machine learning methods. J Big Data. 7. https://doi.org/10.1186/s40537-020-00327-4
105. Munappy AR, Bosch J, Olsson HH, Arpteg A, Brinne B (2022) Data management for production quality deep learning models: challenges and solutions. J Syst Softw. 191:111359. https://doi.org/10.1016/j.jss.2022.111359
106. Tang R, Luo R, Tang S, Song H, Chen X (2022) Machine learning in predicting antimicrobial resistance: a systematic review and meta-analysis. Int J Antimicrob Agents. 60:106684. https://doi.org/10.1016/j.ijantimicag.2022.106684