

1 Recent Advancements in Machine Learning-Based 2 Bloodstream Infection Prediction: A Systematic Review 3 and Meta-analysis of Diagnostic Test Accuracy

4
5 Rajeev Bopche, MSc ^{1*} and Jan Kristian Damås, MD, PhD ^{2,3}

6 ¹ *Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway.*

7 ² *Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway*

8 ³ *Department of Infectious Diseases, Clinic of Medicine, St Olavs Hospital, Trondheim, Norway*

9 **Abstract.**

- 10 • **Purpose:** Bloodstream infections (BSIs) present significant public health challenges. With the advent of machine learning (ML),
11 promising predictive models have been developed. This study evaluates their performance through a systematic review and meta-
12 analysis.
- 13 • **Methods:** We performed a comprehensive systematic review across multiple databases, including PubMed, IEEE Xplore,
14 ScienceDirect, ACM Digital Library, SpringerLink, Web of Science, Scopus, and Google Scholar. Eligible studies focused on
15 BSIs within any hospital setting, employing ML models as the diagnostic test. We evaluated the risk of bias with the Quality
16 Assessment of Diagnostic Accuracy Studies (QUADAS-2) checklist and assessed the quality of evidence using the Grading of
17 Recommendations, Assessment, Development, and Evaluation (GRADE) approach. Models reporting the area under the receiver
18 operating characteristic curve (AUROC) were included in the meta-analysis to identify key performance drivers.
- 19 • **Results:** After screening, a total of 30 studies were eligible for synthesis, from which 41 models and 8 data types were extracted.
20 Most of the studies were carried out in the inpatient settings (n=17; 56%), followed by the emergency department (ED) settings
21 (n=7; 23%), and followed by the ICU settings (n=6; 20%). The reported AUROCs in the hospital inpatients settings, ranged from
22 0.51-0.866, in the ICU settings AUROCs ranged from 0.668-0.970, and in the emergency department (ED) settings the AUROCs
23 of the models ranged from 0.728-0.844. One study reported prospective cohort study, while two prospectively validated their
24 models. In the meta-analysis, laboratory tests, Complete Blood Count/Differential Count (CBC/DC), and ML model type
25 contributed the most to model performance.
- 26 • **Conclusion:** This systematic review and meta-analysis show that on retrospective data, individual ML models can accurately
27 predict BSIs at different stages of patient trajectory. Although they enable early prediction of BSI, a comprehensive approach to

^{1*} Corresponding Author: Rajeev Bopche, rajeev.bopche@ntnu.no

28 integrate data types and models is necessary. Systematic reporting, externally validated, and clinical implementation studies are
29 needed to establish clinical confidence.

30 **Keywords.** Bloodstream infection, Machine Learning, Bacteremia, Artificial intelligence, Predictive modelling,
31 Sepsis

32 1. INTRODUCTION

33 Bloodstream infections (BSIs) pose a severe public health threat, often escalating to critical conditions such
34 as sepsis and septic shock, particularly if not promptly recognized and treated. The rapid progression of these
35 infections contributes to high morbidity, mortality, and significant healthcare costs, making BSIs a pivotal
36 challenge in clinical care [1, 2]. Currently available clinical decision tools for BSIs and sepsis are primarily
37 based on changes in vital signs and abnormal blood test results, which lack sufficient accuracy [3, 4]. Enhancing
38 prediction capabilities could lead to more efficient resource allocation and cost reduction. Accurate initial
39 stratification helps direct resources to high-risk BSI patients and minimizes unnecessary tests and treatments for
40 those at low risk. Moreover, ambiguous results from blood cultures with contaminants can prolong hospital
41 stays and lead to unnecessary antibiotic administration [5, 6]. Therefore, improving the predictive value of these
42 tests and reducing their unnecessary use are critical for optimizing healthcare resource allocation and
43 minimizing costs.

44
45 Despite the extensive application of machine learning (ML) across various healthcare diagnostics, its
46 potential in BSI prediction has not been fully exploited, particularly when compared to its use in predicting
47 sepsis, where it has been more thoroughly researched [7-14]. This gap underscores a critical oversight; early
48 detection of BSIs can significantly limit their progression to sepsis, thus reducing patient suffering and
49 healthcare expenses [16]. This systematic literature review (SLR) and meta-analysis (MA) seek to bridge this
50 gap by highlighting recent advancements in ML models for BSI prediction. Our analysis spans a range of
51 algorithms and examines the influence of different data types used within these models, providing a
52 comprehensive overview of the current landscape and suggesting directions for future research.

53 **1.1. Related works**

54 Recent systematic reviews on ML for BSI prediction have illuminated the capabilities of validated models
55 in clinical environments. For example, Eliakim-Raz et al. (2015) emphasize the variability in study populations,
56 model parameters, and validation processes, which highlights the complexity in creating universally effective
57 ML models for BSI prediction [16]. In contrast, the review by Coburn et al. (2012) focuses on the critical
58 clinical and laboratory markers needed to determine the necessity for blood cultures in suspected bacteremia
59 cases, further supporting the role of ML in enhancing diagnostic accuracy and reducing false positives [2].
60 These studies demonstrate the untapped potential of ML to streamline diagnostic processes and improve patient
61 care in BSI contexts, advocating for its increased integration into BSI prediction strategies.

62 **2. METHODS**

63 The guidelines provided by the Preferred Reporting Items for Systematic Reviews and Meta-Analysis of
64 Diagnostic Test Accuracy studies (PRISMA-DTA) statement [15] was followed for the conduct of this literature
65 review and meta-analysis.

66 **2.1. Search Strategy**

67 The literature search spanned nine databases to cover a wide array of relevant literature. The databases
68 included: PubMed, IEEE Xplore Digital Library, ScienceDirect, ACM Digital Library, SpringerLink, Web of
69 Science, Scopus, and Google Scholar. These databases were selected to encompass a vast range of medical and
70 computer science literature, including journal articles and conference proceedings. To ensure a comprehensive
71 and relevant literature retrieval, we first identified key terms and their possible alternatives. This approach
72 ensures coverage of a broad spectrum of pertinent studies. The core search terms, along with their alternatives,
73 were: Bloodstream Infection (BSI) – Blood culture, Bacteremia, Blood culture test; Prediction – Diagnosis,
74 Identification, Detection; Machine Learning – Artificial Intelligence, Deep Learning, Supervised Learning,
75 Computational Intelligence; Clinical Data – Electronic Health Records (EHR), Electronic Patient Records (EPR),
76 Electronic Medical Records (EMR). Adjustments were made to the query to accommodate the syntax of
77 different database search engines. The search query was constructed by combining these terms using Boolean
78 operators:

79

80 – (Bloodstream Infection OR Blood culture OR Bacteremia) AND (Prediction OR Diagnosis OR
81 Identification OR Detection) AND (Machine Learning OR Artificial Intelligence OR Deep Learning
82 OR Supervised Learning OR Computational Intelligence) AND (Clinical Data OR Electronic Health
83 Records OR Electronic Patient Records OR Electronic Medical Records)

84

85 The selection of relevant studies involved a multi-step process. First, The execution of the search query in each
86 database. Then, Initial screening based on titles and abstracts to exclude non-relevant records. In the third step,
87 detailed screening involving full-text reviews was carried out. The reasons for exclusion were documented.
88 Following this, references in selected articles were examined to identify additional relevant studies,
89 implementing snowballing technique for accessing new literature. This structured approach ensured
90 comprehensive coverage of the topic, allowing the retrieval of documents that met the inclusion criteria for data
91 extraction and analysis to address the research questions of this SLR.

92 2.2. Study Selection

93 2.2.1. Inclusion Criteria

- 94 - **Study Focus:** This review specifically targeted studies that utilized machine learning models for the
95 prediction of bacteremia, BSI, or positive blood culture results. The primary objective was the
96 development or validation of these predictive models.
- 97 - **Types of Publications:** The review was confined to peer-reviewed journal articles and conference
98 papers.
- 99 - **Target Population:** The studies included focused on adult populations.
- 100 - **Model Specification:** Included studies clearly described the ML model(s) used, including the type of
101 algorithm, data inputs, and validation methods.
- 102 - **Outcome Measures:** Studies reported specific outcomes related to the accuracy, sensitivity, specificity,
103 or predictive value of the ML models for BSI prediction.

104 2.2.2. Exclusion Criteria

- 105 - **Focus on Sepsis and Severe Sepsis:** Studies aimed primarily at predicting the onset of sepsis or severe
106 sepsis, rather than BSI, were excluded. This was to concentrate on predictive research for BSI, an
107 antecedent condition, rather than sepsis, a consequent condition.

- 108 - **Pediatric Studies:** Research focusing on neonatal, pediatric, or infant populations was excluded, as the
109 review was limited to adult populations.
- 110 - **Post-Diagnosis Prediction:** Studies focusing on predicting outcomes or complications for patients
111 already diagnosed with BSI were excluded. The emphasis was on the initial prediction of BSI.
- 112 - **Non-Machine Learning Models:** Research employing statistical or other predictive methodologies
113 without using ML algorithms was not included.
- 114 - **Non-Peer Reviewed Literature:** Studies that had not undergone a peer review process, including grey
115 literature, technical reports, and unpublished manuscripts, were excluded.
- 116 - **Opinion Pieces and Theoretical Works:** Commentaries, opinion pieces, and purely theoretical works
117 lacking empirical data or validation were omitted.

118 2.2.3. Timeframe and Language Restrictions

119

- 120 - **Language Restrictions:** The review included only studies published in English. This criterion was
121 applied to ensure the feasibility of thorough analysis and comprehension of the research findings.
- 122 - **Timeframe of Publication:** The review focused on studies published within the last five years,
123 capturing the most recent advancements in machine learning applications for bloodstream infection
124 prediction.

125

126 2.3. Data Extraction

127 A structured approach was used to extract relevant data from identified studies. The data extraction process
128 involved the following key elements:

- 129 - **Study Identification:** Details including authors, year of publication, and title of the study.
- 130 - **Population Characteristics:** Information on the patient demographics, clinical settings, and specific
131 population subgroups studied.
- 132 - **Dataset Description:** Description of the dataset used in the study, including size, source, prevalence
133 rates, and period of data collection.
- 134 - **Algorithms Employed:** Identification of the specific ML algorithms used in the study.
- 135 - **Reported Metrics:** Reporting on the performance metrics used in the study.

136 – **Key Findings and Predictors:** Summary of the main findings, including key predictors identified for
137 BSI prediction.

138 **2.4. Quality Assessment**

139 To evaluate risk of bias, we used the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2)
140 criteria [17]. Domains included patient selection, index test, reference standard, and flow and timing. In line
141 with the recommendations from the QUADAS-2 guidelines, questions per domain were tailored for this paper
142 and can be found in (Supplementary Information). If one of the questions was scored at risk of bias, the domain
143 was scored as high risk of bias. At least one domain at high risk of bias resulted in an overall score of high risk
144 of bias, only one domain scored as unclear risk of bias resulted in an overall score of unclear risk of bias for that
145 paper. We used the Grading of Recommendations Assessment, Development and Evaluation (GRADE)
146 methodology to assess the quality of evidence per hospital setting for all studies reporting the area under the
147 curve of the receiver operating characteristic (AUROC) as their performance metric [18]. In line with the
148 GRADE guidelines for diagnostic test accuracy, we included the study design, domains risk of bias (limitations),
149 and inconsistency of the results. One level of evidence was deducted for each domain with serious concerns or
150 high risk of bias, no factors increased the level of evidence (Supplementary Information). Overall level of
151 evidence is expressed in four categories (high, moderate, low, very low).

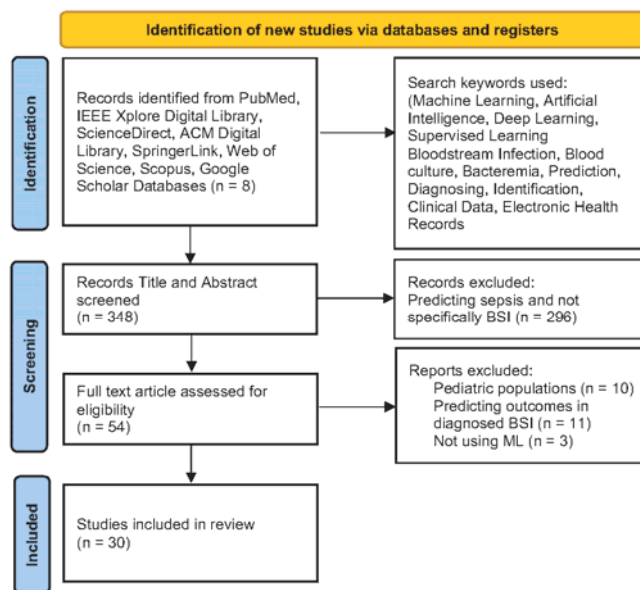
152 **2.5. Meta-analysis**

153 For meta-analysis we conducted univariate and multivariate random effects model analysis fitted using
154 restricted maximum likelihood estimation (REML) for the contribution of covariates towards model
155 performance [19]. We grouped medical data types utilized to train the models such as (demographics as features,
156 vital signs as features, laboratory tests as features, use of textual data, use of time series data, etc.) and included
157 them as covariates. We performed pooled regression analysis to study the most significant covariates per study
158 settings.

159 3. RESULTS

160 3.1 Study selection

161 A comprehensive literature search was carried out from November 2023 to March 2024, yielding 348 articles
162 post-duplication removal. Following an in-depth assessment for eligibility, 30 studies conformed to our
163 inclusion criteria, references [20-49]. The selection process, including reasons for exclusion at each stage, is
164 outlined in the PRISMA flow diagram (Figure 1). Notably, the majority (29 out of 30) of the included articles
165 employed retrospective study designs, with only two conducting prospective validations of their models [40, 41].
166 A single study presented a prospective cohort design [29].



167
168 Fig. 1 PRISMA flow diagram illustrating the systematic review methodology, depicting the screening, eligibility assessment, and inclusion
169 of studies, alongside exclusion justifications at each phase

170 3.2 Study characteristics

171 The studies predominantly took place in inpatient settings (56%, n=17), with the emergency department
172 (ED) (23%, n=7) and intensive care units (ICU) (20%, n=6) also represented. Within inpatient settings, nine
173 studies examined general populations [20-28], and others focused on specific patient cohorts including those
174 with central venous catheters (CVC) [30, 31], systemic inflammatory response syndrome (SIRS) [29], and
175 various other conditions. In ED settings, seven studies were identified [37-43], two of which addressed specific

176 patient groups. ICU-focused research numbered six studies [44-49], again with two concentrating on specific
 177 patient subsets. Bacteremia was the primary condition investigated in 24 studies, alongside candidemia and
 178 central line-associated bloodstream infection (CLABSI) in others. Contaminants were generally considered
 179 negative results, except in two studies [23, 45]. Table 1 presents a synthesis of the study characteristics,
 180 including target conditions, patient/sample sizes, data sources, and prevalence, categorized by setting.
 181

182 **Table 1.** Key characteristics and data sources of selected studies

183

Setting	First author, year	Target condition	No. Patients	Data source	Prevalence
Inpatients	Bhavani et al. (2020)	Bacteremia and Fungemia	76688	EHRs, University of Chicago Hospital (2008-2018) and Loyola University Medical Center, USA (2007-2017)	Bacteremia 7.7%, Fungemia 0.7%
	Lee et al. (2019)	Bacteremia	13402	EHRs, Gangnam Severance Hospital, Seoul, Republic of Korea, (2008-2012)	7.9%
	Lee et al. (2022)	Bacteremia	622771*	EHRs, Sinchon and Gangnam Severance Hospitals, Republic of Korea, (2007-2018)	6.2%
	Cheng et al. (2020)	Bacteremia	28043	EHRs, Zhengzhou University Hospital, China, (2017-2018)	10%
	McFadden et al. (2023)	Bacteremia	10965*	CBC/DC, CPD, Sir Charles Gairdner Hospital, Western Australia (2018-2020)	7.58%
	Lien et al. (2022)	Bacteremia	366586*	EHRs, CBC/DC, Linkou Chang Gung Memorial Hospital (CGMH) in Taiwan, (2014-2019)	8.2%
	Mahmoud et al. (2021)	Bacteremia	7157	EHRs, King Abdulaziz Medical City, Riyadh, Saudi Arabia (2017-2019)	11.4%
	Garnica et al. (2021)	Bacteremia	4357	EHRs, Microbiological data, Hospital Universitario de Fuenlabrada, Madrid, Spain, (2005-2015)	51.3%
	Murri et al. (2024)	HA-BSI	5660*	Generator Center at the Fondazione Policlinico Universitario A. Gemelli IRCCS (FPG), Rome, Italy (2016-2019)	33.6%
Inpatients	Ratzinger et al.	Bacteremia	466	EHRs Vienna General Hospital, Austria, (2011-	28.8%

with SIRS	(2018)			2012)	
Inpatients with CVC	Rahmani et al. (2022)	CLABSIs	27619	EHRs, a proprietary national longitudinal EHR repository, Houston, Texas, USA (2015-2020)	1%
	Beeler et al. (2018)	CLABSIs	70218 *	EHRs, Indiana University Health Academic Health Center, USA, (2013-2016)	0.6%
HD patients	Zhou et al. (2023)	Bacteremia	391	EHRs, Department of Nephrology, Affiliated Hospital of North Sichuan Medical College, Sichuan Province, China, (2018-2022)	18.9%
Cancer patients	Yoo et al. (2021)	Candidemia	34574	EHRs, academic single hospital in Seoul, Republic of Korea, (2010-2018)	0.6%
Maternity patients	Mooney et al. (2020)	Bacteremia	129	CBC parameters, Rotunda Hospital, Ireland (2019)	3%
Patients with PCT ≤2.0 ng/ml	Su et al. (2021)	Bacteremia	931	EHRs, Mindong Hospital Affiliated to Fujian Medical University, China, (2014-2020)	47%
HIV patients	Wu et al. (2023)	Bacteremia	498	EHRs, Wenzhou Central Hospital, China, (2014-2021)	34.3%
ED	Choi et al. (2022)	Bacteremia	24768	EHRs, An urban tertiary referral hospital, Republic of Korea, (2016-2018)	12%
	Choi et al. (2023)	Bacteremia	15362	EHRs, Seoul National University Hospital, Seoul National University Bundang Hospital, Republic of Korea, (2016-2018)	10.9%
	Boerman et al. (2022)	Bacteremia	4885	EHRs, Amsterdam UMC, location VU University Medical Center, NL, (2018-2020)	12.2%
	Chang et al. (2023)	Bacteremia	20636	EHRs, CPD, CBC/DC, China Medical University Hospital, Taiwan, (2021-2022)	10.4%
	Schinkel et al. (2022)	Bacteremia	6421	EHRs, Amsterdam UMC, (VUMC, AMC, ZMC, and BIDMC), NL, (2016-2021)	5.4% - 12.3%
ED patients with SIRS	Goh et al. (2022)	Bacteremia	40395	EHRs, National Cheng Kung University Hospital, Taiwan, (2015-2019)	10%
ED patients with fever	Tsai et al. (2023)	Bacteremia	3669	EHRs, Chi Mei Medical Center, Taiwan, (2017-2020)	13.8%
ICU patients	Roimi et al. (2020)	Bacteremia	3372	EHRs, BIDMC, Boston, Massachusetts, USA, (2008-2012), ICU of Rambam Healthcare Campus (RHCC), Israel, (2013-2017)	ICU acquired: 6.4% (BIDMC), 15.9% (RHCC)

	Van Steenkiste et al. (2019)	Bacteremia	2177	EHRs, ICU, Ghent University Hospital, Belgium, (2013-2015)	10.5%
	Boner et al. (2022)	Bacteremia	6557	EHRs, ICU, University of Virginia, USA, (2011-2015)	13.3%
	Pai et al. (2021)	Bacteremia	4275	EHRs, Taichung Veterans General Hospital ICU, Taiwan, (2015-2019)	13.8%
ICU patients with CVC	Parreco et al. (2018)	CLABSIs	57786 *	MIMIC-III database, USA, (2001-2012)	1.5%
ICU patients with new-onset SIRS	Yuan et al. (2021)	Candidemia	7932	EHRs, Peking Union Medical College Hospital, The Affiliated Hospital of Qingdao University, The First Affiliated Hospital of Fujian Medical University, China, (2013-2017)	1%

184

185 Summarizes the salient features and data sources of the selected studies, grouped by the hospital settings

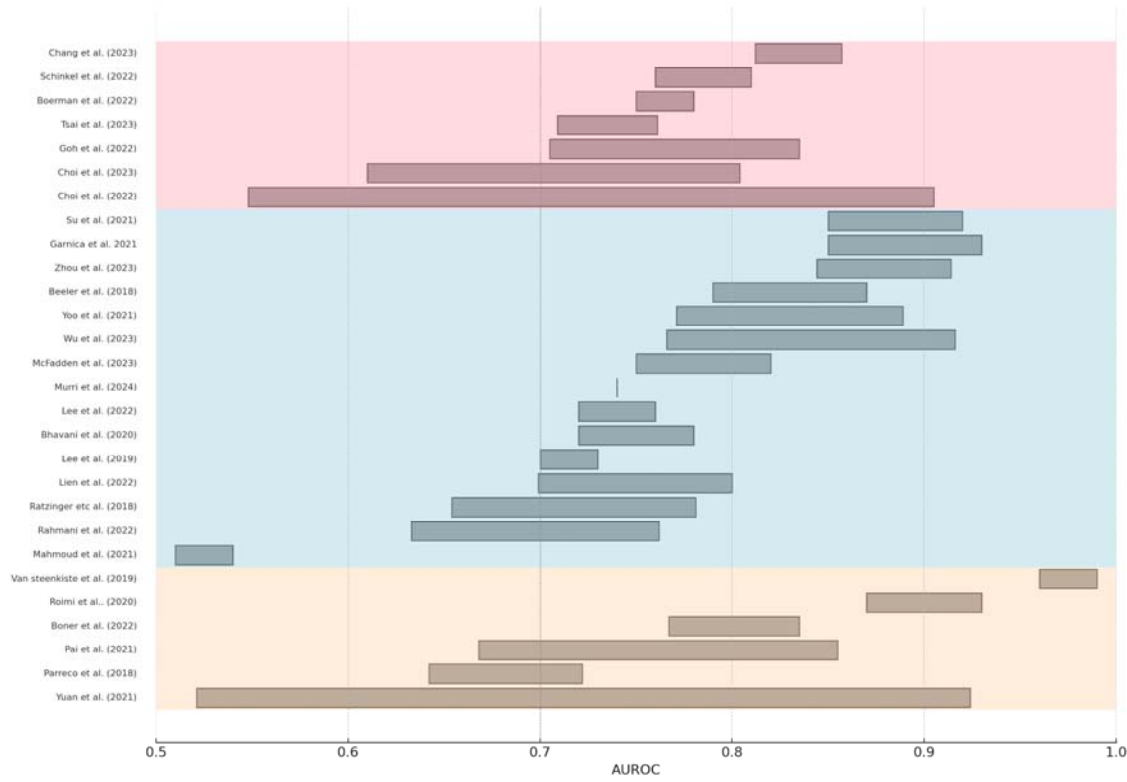
186 Note: * The given number represents the total number of BSI episodes included in the analysis as these studies did not explicitly mention
187 the number of patients.

188 Most of the reviewed studies were conducted within a single institution; however, three studies utilized
189 datasets encompassing two hospital systems [20, 22, 44], and two studies expanded their analysis to incorporate
190 multi-center data [41, 49]. External validation, which is critical for the generalizability of findings, was
191 performed in four studies [38, 40, 41, 44]. Compliance with the Transparent Reporting of a multivariable
192 prediction model for individual prognosis or diagnosis (TRIPOD) guidelines, which enhance the reliability of
193 predictive modeling, was confirmed in five studies [20, 28, 37, 39, 41, 50]. Data diversity is key in model
194 training and validation. Within this context, three studies utilized the publicly available Medical Information
195 Mart for Intensive Care (MIMIC) database [41, 44, 48], while four studies included Complete Blood
196 Count/Differential Count (CBC/DC) data [24, 25, 34, 40]. Furthermore, two of these studies also incorporated
197 Cell Population Data (CPD) [24, 40], highlighting the integration of detailed hematologic parameters.
198 Unstructured data utilization was observed in one study that utilized textual data [23]. In terms of data
199 accessibility, most studies employed proprietary hospital data. Data sharing policies varied: one study explicitly
200 stated that their data would not be shared [47], one study offered openly available data [34], and ten studies
201 indicated that deidentified data could be provided upon reasonable request [24, 25, 27, 28, 29, 38, 39, 41, 42,
202 49], thus contributing to transparency and reproducibility. Notably, six articles did not specify the number of
203 patients included in their analysis [22, 24, 25, 28, 31, 48].

204 3.3. Models and data types

205 The reported AUROCs in the hospital inpatients settings ranged from 0.51-0.866, while models on select
206 group of participants like HD patients [32], patients with low Procalcitonin (PCT) [35], and HIV patients [36]
207 achieved high AUROCs of 0.91, 0.92, and 0.91 respectively. For the emergency department (ED) settings the
208 AUROCs of the models ranged from 0.728-0.844. For the ICU settings AUROCs ranged from 0.668-0.970. All
209 articles reported high prediction performance (AUROC > 0.7) except for one article [26]. All except two articles
210 [23, 34] reported AUROC as their performance metric. Figure 2 presents a horizontal bar chart that delineates
211 the distribution of AUROC values across the various studies reporting AUROC, with the bar length indicating
212 the range of AUROC values reported across various ML models utilized in the study. The gradient color scheme
213 distinguishes the studies based on the clinical settings, ED (red), inpatient (blue), and ICU (yellow). On the right
214 axis the number of patients, or the number of BC episodes/samples analyzed in the study. Figure 3 depicts the
215 utilization frequency of different ML algorithms categorized into tree-based models, traditional ML algorithms,
216 and neural networks. Notably, tree-based models such as RF and XGB displayed a higher incidence of
217 application. LR was predominantly favored among traditional ML algorithms and within neural networks, ANN
218 and MLP were the commonly employed architectures. Figure 4 presents a horizontal bar graph enumerating the
219 occurrence of various performance metrics in the selected studies. The AUROC was the most reported metric.
220 Figure 5 illustrates the types and diversity of data inputs used with the ML models across the selected studies via
221 a stacked bar chart. The derived risk factors represent the derived clinical features included in the models. The
222 ML algorithms and top predictors for each study grouped by settings are given in Table 2.

223



224

225

226 **Fig. 2** Horizontal bar chart depicting the spread of AUROC values reported in various studies. The studies are listed on the vertical axis, and

227 each bar's horizontal extension represents the range of minimum and maximum AUROC value achieved among the various models

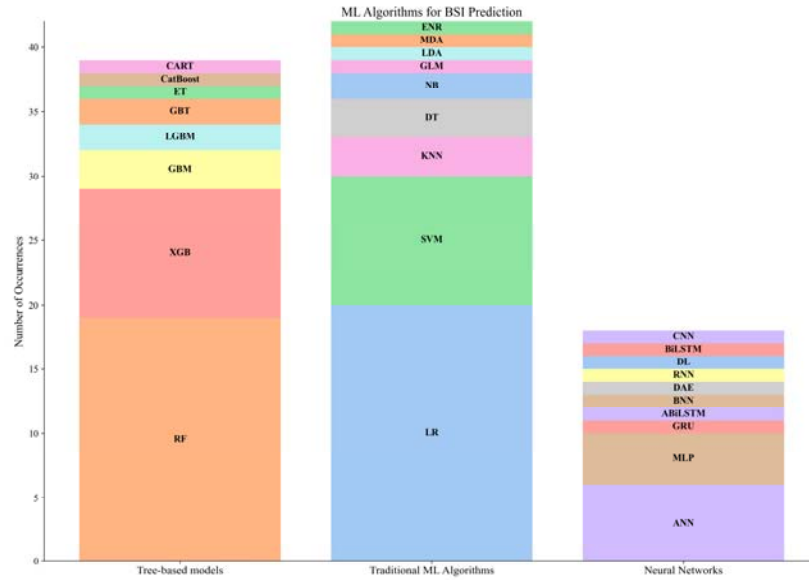
228 employed in the study. The color gradient red, indicates studies in ED, blue, indicating hospital inpatient settings, and the yellow area covers

229 the studies in ICU settings.

230

231

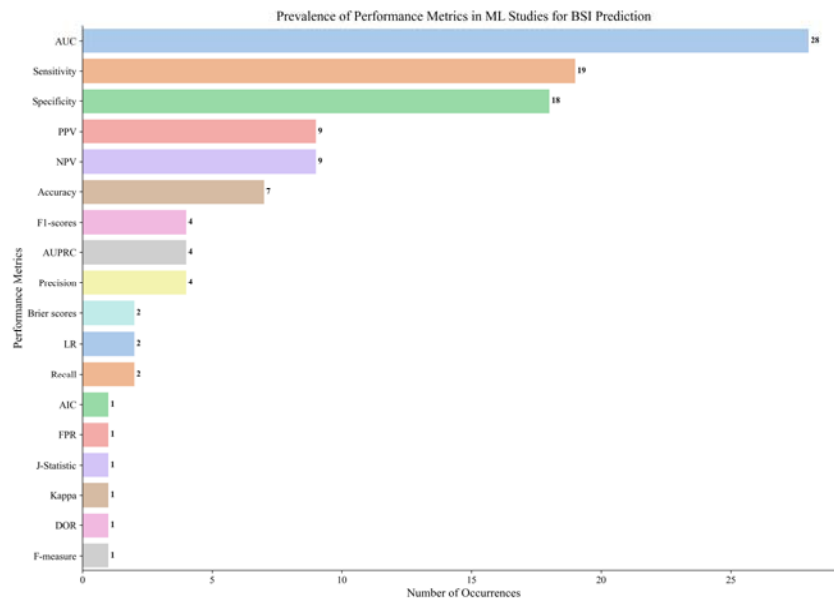
232



233

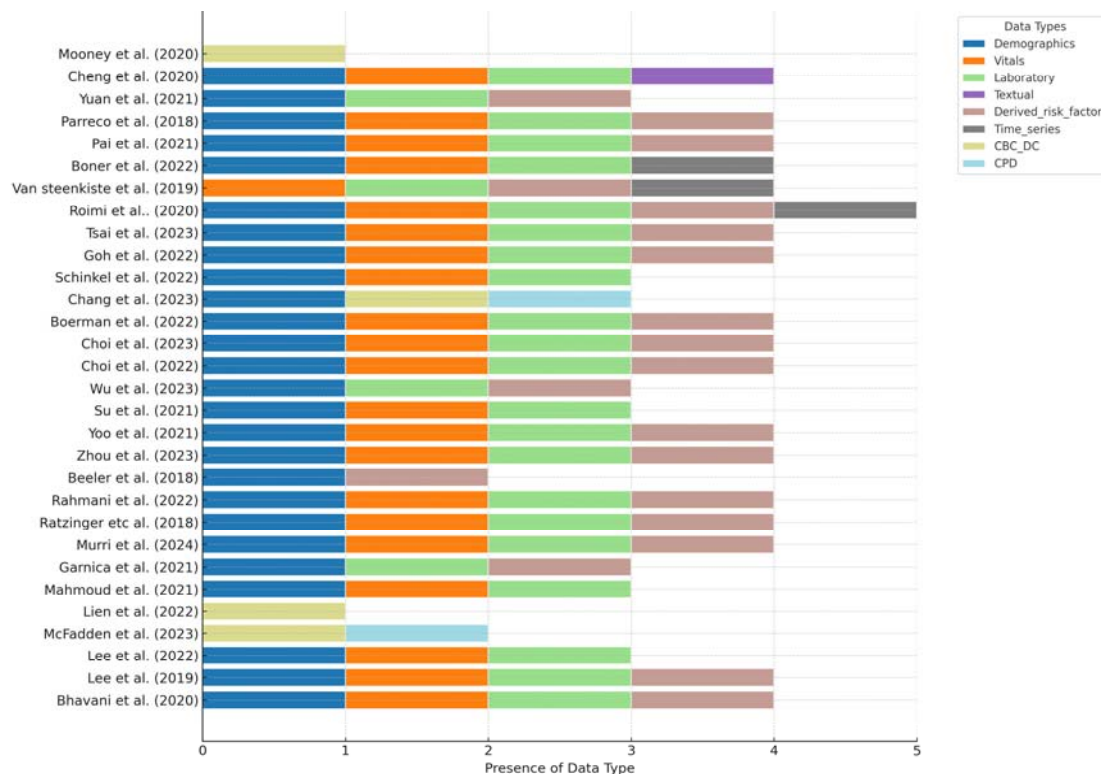
234 Fig. 3 Stacked bar chart displaying the occurrence of ML algorithms in BSI prediction studies across three categories: Tree-based models,
 235 Traditional ML Algorithms, and Neural Networks. For Tree-based models: Random Forest (RF), Extreme Gradient Boosting (XGB),
 236 Gradient Boosting Machine (GBM), Light Gradient Boosting Machine (LGBM), Extra Trees (ET), Gradient Boosted Trees (GBT),
 237 Classification and Regression Trees (CART), and Categorical Boosting (CatBoost). Traditional ML Algorithms include Logistic Regression
 238 (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), Naive Bayes (NB), Generalized Linear Model
 239 (GLM), Linear Discriminant Analysis (LDA), Multiple Discriminant Analysis (MDA), and Elastic Net Regression (ENR). Neural Networks
 240 are represented by Artificial Neural Network (ANN), Multilayer Perceptron (MLP), Bayesian Neural Network (BNN), Deep Learning (DL),
 241 Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), Bidirectional Long Short-Term
 242 Memory (BiLSTM), Attention-based Bidirectional Long Short-Term Memory (ABiLSTM), and Denoising Autoencoder (DAE)

243



244

245 Fig. 4 Horizontal bar graph summarizing the frequency of performance metrics reported in ML studies for BSI prediction. The metrics are
 246 displayed in descending order of occurrence, starting with Area Under the Receiver Operating Characteristic curve (AUC), followed by
 247 Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), Accuracy, F1-scores, Area Under the Precision-
 248 Recall Curve (AUPRC), Precision, Brier scores, Likelihood Ratio (LR), Recall, Akaike Information Criterion (AIC), False Positive Rate
 249 (FPR), J-Statistic, Cohen’s Kappa, F-measure, and Diagnostic Odds Ratio (DOR)



250

251

252

253 Fig. 5 Stacked bar chart representing the types of data used in machine learning models for predicting bloodstream infections, across various
 254 studies listed on the vertical axis. Data types include demographics, vital signs, laboratory tests, textual data, derived risk factors, time-series
 255 data, Complete Blood Count/Differential Count (CBC/DC), and Cell Population Data (CPD). Each bar's length indicates the number of data
 256 types used in each study, providing a comparison of data diversity across the research

257

258 **Table 2.** ML algorithms and key predictors in the selected studies

259

Setting	First author, year	ML models	Key predictors
Inpatients	Bhavani et al.	LR, GBM	Time from admission to BC, Temperature, Age, HR, Prior Bacteremia/Fungemia,

	(2020) [20]		WBC, BUN, Glucose, DBP, SBP, PPI, RR
	Lee et al. (2019) [21]	MLP, SVM, RF	ALP, PLT, Temperature, SBP, WBC, ICU stay, CRP, CVC, Age, PT, Hospital days to BC, HR, Gender, Antibiotics, RR, Creatinine
	Lee et al. (2022) [22]	MLP, RF, XGB	PLT, Monocyte, Neutrophil, Bilirubin, Albumin, and Hospital stay, BUN, ALP, RR, PR, DBP, TP, WBC, PT, Hb, CRP, Creatinine, ALT, AST, Sodium, Chloride, ESR
	Cheng et al. (2020) [23]	LR, NB, SVM, ADT, CNN, BiLSTM, ABiLSTM+ DAE	Textual chief complaints, Admission records, and Laboratory biochemical indicators.
	McFadden et al. (2023) [24]	RF, XGB	CBC, DIFF, and CPD
	Lien et al. (2022) [25]	RF, LR	CBC/DC, CRP, and PCT
	Mahmoud et al. (2021) [26]	NN, RF, LR, DT, NB, SVM	Age, Antibiotics use, Surgery within 14 days, CVC, length of hospitalization before BC, RR, SBP, Temperature, DBP, HR, WBC, Sodium, PLT, Albumin, Creatinine, Lactic acid level.
	Garnica et al. (2021) [27]	SVM, RF, KNN	The number of days in ICU before BC extraction, presence of Catheters, Chronic Respiratory disease, Fever, Age, CRP, PLT.
	Murri et al. (2024) [28]	LR	Time BSI > 12 days, Procalcitonin > 1 ng/mL, Presence of a CVC, PLT, Hypotension, BUN, Presence of urinary catheter, Fever, Tachycardia, Altered mental status, Age, Bilirubin, Creatinine
Inpatients with SIRS	Ratzinger et al. (2018) [29]	RF, ANN, ENR	PCT, LBP, Albumin, Bilirubin
Inpatients with CVC	Rahmani et al. (2022) [30]	XGB, DT, LR	Temperature, HGB, comorbidities, Age, WBC, Race, Neutrophil.
	Beeler et al. (2018) [31]	RF, LR	Age, Gender, history of CLABSI, CHG (Chlorhexidine Gluconate) Bathing Non-compliant Days, Line days.
HD patients	Zhou et al. (2023) [32]	LR, SVM, DT, RF, XGB	PCT, Temperature, Non-arteriovenous fistula dialysis access, NLR, Leukocyte, dialysis duration, LMR, Albumin, Neutrophil, PLT, Age, DBP, CRP, PLR, ALP, SBP, HR, history of BSI,
Cancer patients	Yoo et al. (2021) [33]	LR, ANN, RF, GBM, AML	Variables reflecting the dynamic status of patients with cancer, including blood urea nitrogen level, 7-day variance of RR, Total bilirubin level, 7-day variance of SBP, Body weight.
Maternity patients	Mooney et al. (2020) [34]	CART, LDA, KNN, SVM, RF	NLR, CBC parameters.
Patients with PCT ≤ 2.0 ng/ml	Su et al. (2021) [35]	ANN, KNN, LR, RF, SVM, and NB.	Interleukin-6, PCT, D-dimer, Lactic acid, Leukocytes, Neutrophil, and PLT.

HIV patients	Wu et al. (2023) [36]	SVM, ANN, GBM, GLM, MDA, PLR, NB, RF	Low Hb, CD4+T cell, PLT, LDH, BUN, splenomegaly, absence of ART treatment, Strip shadow, Nodular shadow, and Shock.
ED	Choi et al. (2022) [37]	XGB, RF, LR	Chief complaint, Age, Temperature, HR, and DBP at triage stage. Neutrophils, PLT, CRP, Chief complaints, and Creatinine at disposition stage.
	Choi et al. (2023) [38]	BNN	Age, HR, Temperature, DBP, History of chills, Ambulance use
	Boerman et al. (2022) [39]	GBT, LR	Bilirubin, Urea, lymphocyte, Pulse rate, CRP, Neutrophil, age, Temperature, DBP, Potassium, Glucose, Thrombocytes, Creatinine, ALP, SBP, Organ damage
	Chang et al. (2023) [40]	CatBoost, LGBM, XGB, RF, LR	Demographics, CPD, CBC/DC
	Schinkel et al. (2022) [41]	XGB, LR	Temperature, Creatinine, CRP, Lymphocytes, DBP, Bilirubin, Thrombocytes, Neutrophils, ALP, HR, SBP, Leukocytes, Glucose, Age, Potassium, BUN, Sodium, monocytes
ED patients with SIRS	Goh et al. (2022) [42]	LR, SVM, RF	Age, Gender, COPD, Uncomplicated DM, Hemato-oncology, WBC, Band cell, Platelet, Temperature, HR, mild liver disease, Mean arterial pressure, RR, GSC
ED patients with fever	Tsai et al. (2023) [43]	RF, LR, MLP, XGB, LGBM	Hypertension, Gender, Temperature, DM, Age, CRP, PLT, WBC, Malignancy, Eosinophil, HR, BMI, Hb, RR, SBP, DBP, Band, CKD, Liver Cirrhosis, COPD, GCS
ICU patients	Roimi et al. (2020) [44]	RF, XGB	Time duration (days) between sampling time and last defecation, Time duration (hours) between sampling time and the maximum BUN (mg/dL) value measured during the 5 days prior to sampling, Length of stay (days) between sampling time and ICU admission, The minimal weight (kg) during the 5 days prior to sampling, The time duration between sampling time and the maximum MCHC (g/dL) during the 5 days prior to sampling
	Van Steenkiste et al. (2019) [45]	BiLSTM, ANN, SVM, KNN, LR	Temperature, Thrombocytes, Leukocytes, CRP, sepsis-related organ failure assessment, HR, RR, PT, and mean systemic arterial pressure.
	Boner et al. (2022) [46]	FNN, GRU, CNN, LR	Temperature, BUN, BP, HR, Albumin, PLT, Chloride, Creatinine, Chloride, and Phosphorus.
	Pai et al. (2021) [47]	LR, SVM, MLP, RF, XGB	ALP, CVC period, prothrombin time, PLT, Albumin, Apache II score, Age, foley
ICU patients with CVC	Parreco et al. (2018) [48]	LR, GBT, DL	Severity of illness scores (like SAPS II, APS III, and OASIS) and comorbidities.
ICU patients with new-onset SIRS	Yuan et al. (2021) [49]	XGB, SVM, RF, ET, LR	Colonization, Diabetes, AKI, total number of parenteral nutrition days, history of fungal infection, CRRT days, Abdominal surgery, BDG, days of mechanical ventilation, Length of hospital and ICU stay, days of CVC

260

261 This table summarizes the ML algorithms and key predictors from the selected studies

262 **3.4. Data challenges and strategies**

263 BSIs may be relatively rare compared to the number of non-infection cases in a dataset. This imbalance can
264 lead models to become biased towards predicting the majority class, reducing their effectiveness in identifying
265 true infection cases. Most studies reported imbalanced dataset with prevalence rates of BSI as given in the Table
266 1. To overcome challenges with data imbalance, Synthetic Minority Over-sampling Technique (SMOTE) was
267 widely used for augmenting the minority class in the dataset by generating synthetic samples [21, 30, 49, 51].
268 Goh et al. (2022) employed oversampling, undersampling, and random oversampling (ROSE) methods for
269 model development [42]. Lien et al. (2022) and Van Steenkiste et al. (2019) employed Precision-Recall Area
270 Under Curve (PRAUC) metric for a more accurate assessment of model performance in imbalanced datasets [25,
271 45].. The study by Garnica et al. (2021) encountered significant issues with missing data across the patient
272 records used [27]. The types of missing data were classified into three categories: Missing Completely At
273 Random (MCAR), Missing At Random (MAR), and Missing Not At Random (MNAR) [52]. They employed
274 separate class method to represent the missing data, ensuring the ML models could handle these cases without
275 dropping significant amounts of data. Using patient data for training ML models can raise concerns about
276 privacy and data security. Ensuring patient anonymity and complying with regulations can limit the accessibility
277 and use of certain data. Boerman et al. (2022) faced difficulty with the limitation of not being able to use free-
278 text data such as physician and nurse reports due to privacy concerns [39]. To ensure patient privacy and
279 compliance with data protection regulations researchers can implement effective deidentification of patient
280 records, involving elimination or alteration of direct identifiers, such as names, age, gender, or location, which
281 could be combined to identify an individual.

282 **3.5. Quality of evidence and risk of bias**

283 The risk of bias for the retrospective diagnostic test accuracy studies was assessed following the QUADAS-
284 2 guidelines, and the results are shown in Table 3. For patient selection, six articles (20.6%) presented high risk
285 of bias and four articles (13.7%) presented unclear risk of bias for failing to describe the study population and
286 patient selection. For index test 12 articles scored a high risk of bias (41.3%) and five article (17.2%) scored
287 unclear risk of bias, due to non-reporting of data splits and cross-validation strategies. Three articles (10.3%)
288 scored high risk of bias and one score unclear risk of bias for not specifying the reference standards. One article
289 presented unclear risk of bias in the flow and timing.

290

291

Table 3. Risk of Bias Assessment Results According to QUADAS-2 Criteria

Setting	Study	Risk of bias			
		Patient selection	Index test	Reference standards	Flow and Timing
Inpatients	Bhavani et al. (2020) [20]	Low	Low	Low	Low
	Lee et al. (2019) [21]	Unclear	High	Low	Low
	Lee et al. (2022) [22]	Low	High	Low	Low
	Cheng et al. (2020) [23]	High	High	High	Low
	McFadden et al. (2023) [24]	Low	Low	Low	Low
	Lien et al. (2022) [25]	Low	Unclear	Low	Low
	Mahmoud et al. (2021) [26]	Unclear	High	Low	Low
	Garnica et al. (2021) [27]	Unclear	Low	Low	Low
	Murri et al. (2024) [28]	Low	Low	Low	Low
Inpatients with CVC	Rahmani et al. (2022) [30]	Low	High	Low	Low
	Beeler et al. (2018) [31]	Low	High	Low	Low
Hemodialysis patients	Zhou et al. (2023) [32]	Unclear	High	Low	Low
Cancer patients	Yoo et al. (2021) [33]	Low	Low	Low	Low
Maternity patients	Mooney et al. (2020) [34]	Low	Unclear	Low	Low
Patients with procalcitonin ≤ 2.0 ng/mL	Su et al. (2021) [35]	High	High	Unclear	Low
	Wu et al. (2023) [36]	Low	High	Low	Low
ED patients	Choi et al. (2022) [37]	Low	Low	Low	Low
	Choi et al. (2023) [38]	Low	Low	Low	Low
	Boerman et al. (2022) [39]	Low	Low	Low	Low
	Chang et al. (2023) [40]	Low	Unclear	Low	Low
	Schinkel et al. (2022) [41]	Low	Low	Low	Low
ED patients	Goh et al. (2022) [42]	Low	Unclear	Low	Low

with SIRS					
ED patients	Tsai et al. (2023) [43]	Low	High	Low	Low
with Fever					
ICU patients	Roimi et al. (2020) [44]	Low	Low	Low	Low
	Van Steenkiste et al. (2019) [45]	High	High	High	Unclear
	Boner et al. (2022) [46]	High	Low	High	Low
	Pai et al. (2021) [47]	Low	Unclear	Low	Low
	Parreco et al. (2018) [48]	High	High	Low	Low
ICU patients with new-onset SIRS	Yuan et al. (2021) [49]	High	High	Low	Low

292 Assessment of the risk of bias for included studies based on QUADAS-2 criteria, presented by healthcare setting (Inpatients, ED, ICU).

293 Columns display the evaluation for patient selection, index test, reference standards, and flow and timing.

294

295 The GRADE evidence profile was estimated by pooling studies based on settings. In this analysis we only
 296 considered studies focusing on general adult patient population and did not consider studies reporting specific
 297 study population. The GRADE evidence profile is given as Table 4. The studies in ICU and Inpatients settings
 298 aggregates were considered at high risk of bias and studies in ED settings aggregates were considered unclear
 299 risk of bias. One level of evidence was deducted for observational study design and Inconsistencies due to
 300 heterogeneity. Consequently, the quality of evidence for each of the settings was scored as low. Additionally,
 301 the outcome column distinguishes AUROC values for high and unclear risk of bias studies. High risk of bias
 302 studies reported lowest AUROC values in the inpatient settings and reported highest AUROC values in the ED
 303 and ICU settings.

304 **Table 4.** GRADE Evidence Profile and Quality Assessment of Studies by Setting

Settings	Studies Characteristics		Quality Assessment		Outcome	
	Number of studies	Design	Limitations (Unclear risk of bias studies/total)	Inconsistency	AUROC high risk of bias/unclear risk of bias	Quality of evidence
Inpatients	9 studies [20-28] (1,195,629) ^a	Observational cohort studies	High risk of bias (5/8)	High heterogeneity	0.51-0.762/0.74-0.86	Low

ED	5 studies (72,072)	Observational cohort studies	Unclear risk of bias (1/5)	Moderate heterogeneity	0.834-0.844/ 0.728-0.782	Low
	[37-41]					
ICU	4 studies (16,381)	Observational cohort studies	High risk of bias (2/4)	Considerable heterogeneity	0.835-0.97/ 0.668-0.851	Low
	[44-47]					

305 This table shows the GRADE evidence profile and overall quality of studies segregated by hospital setting (Inpatient, ED, ICU).

306 Note: ^a 4 studies in this pool did not report number of patients and so the reported number of BSI episodes were included to calculate the
307 sum.

308 3.6. Meta-analysis

309 In the meta-analysis, only models presented in articles with target condition bacteremia and general patients
310 population across the three settings were considered. Models for specific disease population, models not
311 reporting AUROC metric, and models predicting CLABSIs were not considered in the meta-analysis. A total of
312 41 models and 9 covariates were included in the meta-analysis. The univariate and multivariate random effect
313 model significant (p-value>0.05) results are shown in Table 5. The random effect model analysis shows that
314 laboratory tests, use of time-series data, use of CBC/DC data, BiLSTM model, XGB model, and RF model
315 positively contributed to the AUROC. We performed a pooled analysis by setting, to identify the best covariates
316 for each. The results of the regression analysis are given as Table 6. For Inpatient setting, demographics
317 (includes age), laboratory tests, and use of CBC/DC data positively contributed to the model performance and
318 use of vital signs negatively impacted the model performance. In contrast in the ICU setting, vital signs,
319 laboratory tests, derived risk factors, use of time-series data, and BiLSTM model employed for sequential data
320 positively impacted the performance metrics.

321 **Table 5.** Univariate and Multivariate Analysis of Predictive Performance Covariates

Variables	Univariate analysis			Multivariate analysis		
	Coeff	SE	p-value	Coeff	SE	p-value
Vital signs as features	-0.044	0.020	0.034	-0.068	0.012	0.001
Laboratory tests as features	0.211	0.035	0.000			
Time-series data	0.059	0.022	0.012	0.057	0.019	0.001
CBC/DC data	0.195	0.033	0.028			
BiLSTM as model	0.187	0.054	0.001	0.154	0.065	0.000

XGB as model	0.062	0.025	0.024
RF as model	0.052	0.025	0.050

322 This table presents the results of the univariate and multivariate regression analyses identifying the impact of various covariates on the
 323 predictive performance of ML models

324

325 **Table 6.** Pooled Regression Analysis Results by Study Setting

Variables	Regression analysis		
	Coeff	SE	p-value
Inpatients settings:			
Demographics	0.184	0.010	0.000
Vitals as features	-0.138	0.010	0.001
Laboratory features	0.184	0.010	0.000
CBC/DC data	0.282	0.006	0.000
ED settings:			
None with significant p-value			
ICU settings:			
Vitals as features	0.131	0.009	0.005
Laboratory features	0.131	0.009	0.005
Derived risk factors as features	0.184	0.026	0.020
Time-series data	0.100	0.024	0.050
BiLSTM as model	0.250	0.033	0.017

326 This table presents the coefficients, standard errors, and p-values from the regression analysis of significant covariates affecting model
 327 performance in different hospital settings (Inpatient, ED, ICU).

328 4. DISCUSSIONS

329 This is the first systematic review and meta-analysis that has integrated findings from diverse clinical
 330 settings to assess the performance of ML models in predicting BSIs. This work corroborates the growing
 331 evidence that ML models are viable tools for enhancing diagnostic accuracy and potentially reducing reliance on

332 traditional blood cultures, a point underscored by contemporary studies [3, 6]. Significantly, our meta-analysis
333 revealed that models performed variably across different clinical environments, with AUROC scores ranging
334 widely. In the inpatient settings, AUROCs ranged from 0.51 to 0.866, demonstrating moderate to high
335 diagnostic accuracy. This performance variance was even more pronounced in the ICU settings, where
336 AUROCs spanned from 0.668 to 0.970, suggesting that certain models are highly effective while others may
337 require further refinement. The variability in model performance aligns with the findings from previous studies
338 which have highlighted the complexity involved in creating universally effective ML models for BSI prediction
339 [16]. This analysis not only validates the potential of ML in clinical diagnostics but also highlights the critical
340 need for tailored approaches depending on specific hospital settings and patient populations. Such
341 differentiation in model efficacy emphasizes the importance of context-specific model development and
342 deployment, which should be informed by the distinct dynamics and needs of each clinical environment.
343 Furthermore, the integration of various data types, such as demographic features, laboratory tests, and derived
344 clinical features, has been shown to significantly impact model performance, echoing the findings of recent
345 meta-analyses which advocate for a multi-faceted approach to data integration within ML models to enhance
346 diagnostic precision and reliability [53].

347 **4.1. Clinical relevance and model performance**

348 The findings from our study underscore the potential of ML models to enhance the diagnostic accuracy of
349 bloodstream infections BSIs significantly. This enhancement is crucial, considering the current reliance on
350 traditional blood cultures, which, while standard, are not without their limitations such as delays in results and
351 the potential for contamination [54]. ML models, by leveraging a variety of clinical data, including patient
352 demographics, prior medical history, and laboratory results, can predict BSIs with notable accuracy. This
353 predictive capability is especially valuable in clinical settings where rapid decision-making is critical [55]. By
354 predicting BSIs accurately, ML models can facilitate earlier intervention strategies, potentially leading to
355 improved patient outcomes and reduced mortality rates [56]. Furthermore, the integration of ML into hospital
356 systems offers a pathway to more streamlined resource allocation [57-59]. By accurately identifying patients at
357 high risk of BSIs, hospitals can optimize the use of tests and allocate personnel and medical resources more
358 efficiently [60]. This not only helps in managing hospital resources but also in reducing unnecessary antibiotic
359 use, which is often a knee-jerk response to suspected infections [61].

360 **4.2. Future directions and academic contribution**

361 Our study have laid the groundwork for several promising directions that future investigations could take to
362 advance this critical area of medical informatics. While retrospective studies form the bulk of current research
363 on ML models for BSI prediction, there is a pronounced need for prospective studies. Such studies will allow
364 for real-time data collection and model validation, offering insights that are often obscured in retrospective
365 analyses [12]. This shift could also facilitate the adjustment of models in accordance with dynamic clinical
366 environments, ensuring that they remain robust and reliable over time. The potential of ML models to be
367 integrated into real-time clinical decision support systems represents a significant leap towards operationalizing
368 AI in everyday clinical practice [62]. However, achieving this requires rigorous testing and validation of these
369 models within clinical settings to ensure they perform reliably when interfaced with live data streams [63, 64].
370 There is substantial scope for exploring new predictive variables that could enhance the predictive accuracy of
371 ML models. The incorporation of genomic data, patient mobility patterns, and real-time monitoring data could
372 provide new insights into infection risk factors, potentially leading to more sophisticated and accurate prediction
373 models [65, 66]. To build trust and validate the efficacy of ML models in clinical settings, systematic reporting
374 and external validation are essential [67, 68]. Models need to be tested across diverse demographics and varied
375 clinical environments to assess their universality and reliability [69-72]. The work of Fleuren et al. (2020) and
376 Moor et al. (2021) highlights the importance of such validation in confirming the utility and accuracy of
377 predictive models for sepsis, which can be paralleled in BSI prediction [12, 13]. As ML applications in
378 healthcare continue to expand, it is crucial to consider the regulatory and ethical implications of their use [73,
379 74]. Ensuring patient privacy, securing data, and maintaining transparency in AI decision-making processes are
380 paramount [75, 76]. Future research must also address these aspects to foster a safe and trustful adoption of AI
381 technologies in healthcare.

382 **4.2. Strengths and Limitations**

383 Our systematic review and meta-analysis are grounded in a comprehensive, methodologically robust
384 approach that integrates a variety of data sources and analytical techniques. The extensive database search
385 across multiple platforms including PubMed, IEEE Xplore, and Scopus ensures a broad capture of relevant
386 studies, reducing the risk of publication bias. This wide-ranging data acquisition is supplemented by our
387 application of the QUADAS-2 framework and GRADE methodology, which enhances the credibility of our

388 findings by systematically assessing the risk of bias and the quality of evidence across studies. Furthermore, the
389 synthesis of data into figures and comprehensive tables enables clear visual representation and understanding of
390 the applications of ML models across varied clinical settings. Our findings are supported by rigorous statistical
391 analysis, including univariate and multivariate models, which reveal key performance drivers and validate the
392 predictive power of ML models for bloodstream infections. Despite these strengths, several limitations must be
393 acknowledged. A significant proportion of the included studies utilized retrospective designs. While
394 retrospective studies provide valuable historical insights, they are inherently limited by the data available, often
395 lacking the prospective validation needed to confirm the efficacy of predictive models under current clinical
396 conditions. This design limitation impacts the generalizability of our findings, as the models might perform
397 differently when deployed in real-time environments. To enhance the generalizability and applicability of future
398 research, several strategies can be adopted. Prospective study designs should be prioritized, as they allow for the
399 real-time evaluation and adjustment of ML models, ensuring that the models are tested and validated under
400 varied clinical conditions. This approach not only tests the robustness of the models but also helps in fine-tuning
401 them for practical deployment. Furthermore, multi-center studies involving diverse populations and settings
402 should be encouraged to test the efficacy of these predictive models across different demographics and
403 operational conditions. Such studies can help identify and mitigate any population-specific biases, thereby
404 enhancing the models' applicability and reliability. Lastly, ongoing validation and systematic reporting should
405 be integral to future research efforts. By continuously assessing and reporting the performance of ML models,
406 researchers can ensure that these tools remain effective and relevant in the ever-evolving clinical landscape. By
407 addressing these limitations and building on the strengths of our current approach, future research can
408 significantly advance the field of ML in medical diagnostics, particularly in the prediction and management of
409 bloodstream infections.

410 **5. CONCLUSIONS**

411 Our systematic review and meta-analysis have critically evaluated the efficacy of ML models in predicting
412 BSIs, a crucial domain where timely and accurate diagnosis can significantly influence clinical outcomes. ML
413 models, especially those incorporating diverse data types such as laboratory results and demographic
414 information, demonstrated a capacity to predict BSIs with a high degree of accuracy. The integration of these
415 models into clinical settings can potentially reduce the reliance on traditional blood cultures, which, while

416 foundational, are hampered by delays and susceptibility to contamination. This shift could streamline diagnostic
417 workflows and enhance the speed and precision of infection response interventions, thus improving patient care
418 and outcomes. To advance the use of ML in predicting BSIs, future research should focus on prospective studies
419 and the development of real-time, adaptive ML systems that can be integrated seamlessly into clinical decision-
420 support frameworks. Furthermore, external validation of these models across diverse populations and settings is
421 essential to bolster clinical confidence and foster wider adoption. In conclusion, while ML models hold
422 significant promise for transforming BSI diagnosis, their successful implementation will depend on meticulous
423 model development, validation, and customization tailored to the nuanced demands of different healthcare
424 settings. This holistic approach will be crucial in overcoming the current challenges and fully realizing the
425 potential of ML in clinical diagnostics.

426 **Acknowledgements**

427 We would like to thank the researchers at Mid-Norway Centre for Sepsis Research for valuable discussions and
428 feedback.

429

430 **Author's contributions**

431

432 RB conceptualized and designed the study with inputs from JKD and ØN. RB performed database search and
433 literature analysis. RB wrote the initial and subsequent drafts of the manuscript, which other authors reviewed
434 and approved.

435

436 **Funding**

437 Financial support for this study was provided by the Computational Sepsis Mining and Modelling project
438 through the Norwegian University of Science and Technology Health Strategic Area.

439

440 **Appendix: List of Medical Abbreviations**

- 441 • ALP: Alkaline Phosphatase
- 442 • ALT: Alanine Aminotransferase

- 443 • APS III: Acute Physiology Score III
- 444 • ART: Antiretroviral Therapy
- 445 • AST: Aspartate Aminotransferase
- 446 • AKI: Acute Kidney Injury
- 447 • BC: Blood Cultures
- 448 • BDG: Beta-D-Glucan
- 449 • BUN: Blood Urea Nitrogen
- 450 • CBC: Complete Blood Count
- 451 • CHG: Chlorhexidine Gluconate
- 452 • CKD: Chronic Kidney Disease
- 453 • CLABSI: Central Line-Associated Bloodstream Infection
- 454 • COPD: Chronic Obstructive Pulmonary Disease
- 455 • CPD: Cephalopelvic Disproportion
- 456 • CRP: C-reactive Protein
- 457 • CRRT: Continuous Renal Replacement Therapy
- 458 • CVC: Central Venous Catheter
- 459 • DBP: Diastolic Blood Pressure
- 460 • DIFF: Differential Count
- 461 • DM: Diabetes Mellitus
- 462 • ESR: Erythrocyte Sedimentation Rate
- 463 • GCS: Glasgow Coma Scale
- 464 • Hb: Hemoglobin
- 465 • HR: Heart Rate
- 466 • ICU: Intensive Care Unit
- 467 • LMR: Lymphocyte to Monocyte Ratio
- 468 • NLR: Neutrophil to Lymphocyte Ratio
- 469 • OASIS: Oxford Acute Severity of Illness Score
- 470 • PCT: Procalcitonin
- 471 • PLR: Platelet to Lymphocyte Ratio

- 472 • PLT: Platelet Count
- 473 • PPI: Proton Pump Inhibitor
- 474 • PT: Prothrombin Time
- 475 • RR: Respiratory Rate
- 476 • SAPS II: Simplified Acute Physiology Score II
- 477 • SBP: Systolic Blood Pressure
- 478 • WBC: White Blood Cell Count

479

480 **References**

481

- 482 [1] Goto, M., & Al-Hasan, M. N. (2013). Overall burden of bloodstream infection and nosocomial bloodstream infection in North America
483 and Europe. *Clinical Microbiology and Infection*, 19(6), 501-509. <https://doi.org/10.1111/1469-0691.12195>
- 484 [2] Coburn, B., Morris, A. M., Tomlinson, G., & Detsky, A. S. (2012). Does this adult patient with suspected bacteremia require blood
485 cultures?. *JAMA*, 308(5), 502–511. <https://doi.org/10.1001/jama.2012.8262>
- 486 [3] Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized
487 Patients. *JAMA Intern Med*. 2021;181(8):1065–1070. doi:10.1001/jamainternmed.2021.2626
- 488 [4] Schootman, M., Wiskow, C., Loux, T., Meyer, L., Powell, S., Gandhi, A., & Lacasse, A. (2022). Evaluation of the effectiveness of an
489 automated sepsis predictive tool on patient outcomes. *Journal of critical care*, 71, 154061. <https://doi.org/10.1016/j.jcrc.2022.154061>
- 490 [5] Bates, D. W., Cook, E. F., Goldman, L., & Lee, T. H. (1990). Predicting bacteremia in hospitalized patients. A prospectively validated
491 model. *Annals of internal medicine*, 113(7), 495–500. <https://doi.org/10.7326/0003-4819-113-7-495>
- 492 [6] Sakarikou, C., Altieri, A., Bossa, M. C., Minelli, S., Dolfi, C., Piperno, M., & Favalli, C. (2018). Rapid and cost-effective identification
493 and antimicrobial susceptibility testing in patients with Gram-negative bacteremia directly from blood-culture fluid. *Journal of*
494 *microbiological methods*, 146, 7–12. <https://doi.org/10.1016/j.mimet.2018.01.012>
- 495 [7] Buch, V. H., Ahmed, I. & Maruthappu, M. Artificial intelligence in medicine: Current trends and future possibilities. *Br. J. Gen. Pract.* 68, 143–
496 144. <https://doi.org/10.3399/bjgp18X695213> (2018).
- 497 [8] Rajpurkar, P., Chen, E., Banerjee, O. et al. AI in health and medicine. *Nat Med* 28, 31–38 (2022). [https://doi.org/10.1038/s41591-021-](https://doi.org/10.1038/s41591-021-01614-0)
498 [01614-0](https://doi.org/10.1038/s41591-021-01614-0)
- 499 [9] A. Baldominos, A. Puello, H. Oğul, T. Aşuroğlu and R. Colomo-Palacios, "Predicting Infections Using Computational Intelligence – A
500 Systematic Review," in *IEEE Access*, vol. 8, pp. 31083–31102, 2020, doi: 10.1109/ACCESS.2020.2973006.
- 501 [10] Hassan, N., Slight, R., Weiand, D., Vellinga, A., Morgan, G., Aboushareb, F., & Slight, S. P. (2021). Preventing sepsis; how can
502 artificial intelligence inform the clinical decision-making process? A systematic review. *International journal of medical informatics*,
503 *150*, 104457. <https://doi.org/10.1016/j.ijmedinf.2021.104457>
- 504 [11] Islam, K. R., Prithula, J., Kumar, J., Tan, T. L., Reaz, M. B. I., Sumon, M. S. I., & Chowdhury, M. E. H. (2023). Machine Learning-
505 Based Early Prediction of Sepsis Using Electronic Health Records: A Systematic Review. *Journal of clinical medicine*, 12(17), 5658.
506 <https://doi.org/10.3390/jcm12175658>

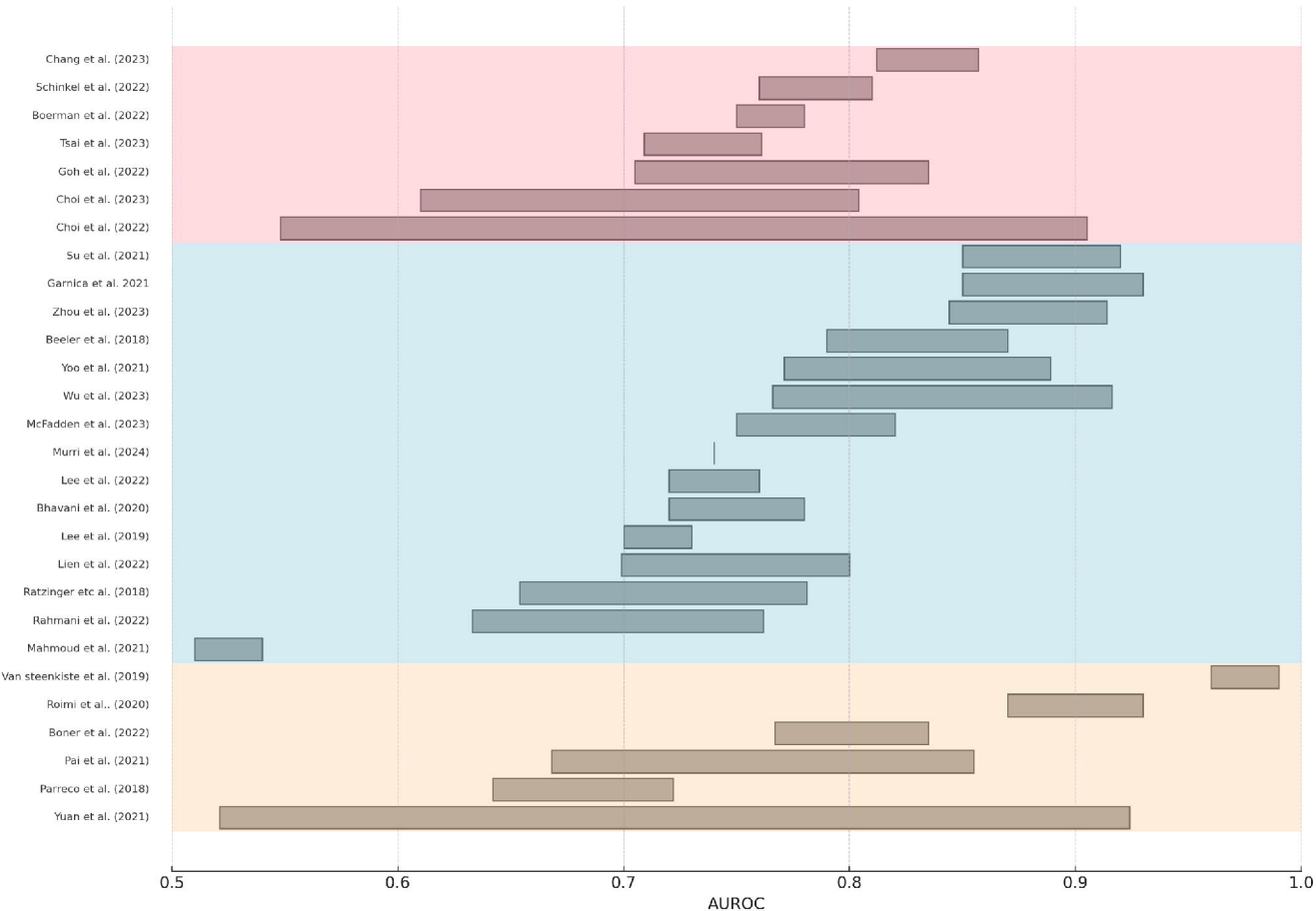
- 507 [12] Fleuren, L. M., Klausch, T. L. T., Zwager, C. L., Schoonmade, L. J., Guo, T., Roggeveen, L. F., Swart, E. L., Girbes, A. R. J., Thoralf,
508 P., Ercole, A., Hoogendoorn, M., & Elbers, P. W. G. (2020). Machine learning for the prediction of sepsis: a systematic review and
509 meta-analysis of diagnostic test accuracy. *Intensive care medicine*, 46(3), 383–400. <https://doi.org/10.1007/s00134-019-05872-y>
- 510 [13] Moor, M., Rieck, B., Horn, M., Jutzeler, C. R., & Borgwardt, K. (2021). Early Prediction of Sepsis in the ICU Using Machine
511 Learning: A Systematic Review. *Frontiers in medicine*, 8, 607952. <https://doi.org/10.3389/fmed.2021.607952>
- 512 [14] Deng, H. F., Sun, M. W., Wang, Y., Zeng, J., Yuan, T., Li, T., Li, D. H., Chen, W., Zhou, P., Wang, Q., & Jiang, H. (2021).
513 Evaluating machine learning models for sepsis
- 514 [15] McInnes, M. D. F., Moher, D., Thombs, B. D., McGrath, T. A., Bossuyt, P. M., and the PRISMA-DTA Group, Clifford, T., Cohen, J.
515 F., Deeks, J. J., Gatsonis, C., Hooft, L., Hunt, H. A., Hyde, C. J., Korevaar, D. A., Leeflang, M. M. G., Macaskill, P., Reitsma, J. B.,
516 Rodin, R., Rutjes, A. W. S., Salameh, J. P., ... Willis, B. H. (2018). Preferred Reporting Items for a Systematic Review and Meta-
517 analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA*, 319(4), 388–396.
518 <https://doi.org/10.1001/jama.2017.19163>
- 519 [16] Eliakim-Raz, N., Bates, D. W., & Leibovici, L. (2015). Predicting bacteraemia in validated models—a systematic review. *Clinical*
520 *microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 21(4),
521 295–301. <https://doi.org/10.1016/j.cmi.2015.01.023>
- 522 [17] Whiting PF, Rutjes AWS, Westwood ME et al (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy
523 studies. *Ann Intern Med* 155:529. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>
- 524 [18] Schünemann, H. J., Oxman, A. D., Brozek, J., Glasziou, P., Jaeschke, R., Vist, G. E., Williams, J. W., Jr, Kunz, R., Craig, J., Montori,
525 V. M., Bossuyt, P., Guyatt, G. H., & GRADE Working Group (2008). Grading quality of evidence and strength of recommendations
526 for diagnostic tests and strategies. *BMJ (Clinical research ed.)*, 336(7653), 1106–1110. <https://doi.org/10.1136/bmj.39500.677199.AE>
- 527 [19] Higgins, J. P., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal*
528 *Statistical Society. Series A, (Statistics in Society)*, 172(1), 137–159. <https://doi.org/10.1111/j.1467-985X.2008.00552.x>
- 529 [20] Bhavani, S. V., Lonjers, Z., Carey, K. A., Afshar, M., Gilbert, E. R., Shah, N. S., Huang, E. S., & Churpek, M. M. (2020). The
530 Development and Validation of a Machine Learning Model to Predict Bacteremia and Fungemia in Hospitalized Patients Using
531 Electronic Health Record Data. *Critical Care Medicine*, 48(11), E1020–E1028. <https://doi.org/10.1097/CCM.0000000000004556>
- 532 [21] Lee, K. H., Dong, J. J., Jeong, S. J., Chae, M.-H., Lee, B. S., Kim, H. J., Ko, S. H., & Song, Y. G. (2019). Early detection of
533 bacteraemia using ten clinical variables with an artificial neural network approach. *Journal of Clinical Medicine*, 8(10).
534 <https://doi.org/10.3390/jcm8101592>
- 535 [22] Lee, K. H., Dong, J. J., Kim, S., Kim, D., Hyun, J. H., Chae, M.-H., Lee, B. S., & Song, Y. G. (2022). Prediction of Bacteremia Based
536 on 12-Year Medical Data Using a Machine Learning Approach: Effect of Medical Data by Extraction Time. *Diagnostics*, 12(1).
537 <https://doi.org/10.3390/diagnostics12010102>
- 538 [23] Cheng, M., Zhao, X., Ding, X., Gao, J., Xiong, S., & Ren, Y. (2020). Prediction of blood culture outcome using hybrid neural network
539 model based on electronic health records. *BMC medical informatics and decision making*, 20(Suppl 3), 121.
540 <https://doi.org/10.1186/s12911-020-1113-4>
- 541 [24] McFadden, B. R., Inglis, T. J. J., & Reynolds, M. (2023). Machine learning pipeline for blood culture outcome prediction using
542 Sysmex XN-2000 blood sample results in Western Australia. *BMC Infectious Diseases*, 23(1). [https://doi.org/10.1186/s12879-023-](https://doi.org/10.1186/s12879-023-08535-y)
543 [08535-y](https://doi.org/10.1186/s12879-023-08535-y)

- 544 [25] Lien, F., Lin, H.-S., Wu, Y.-T., & Chiueh, T.-S. (2022). Bacteremia detection from complete blood count and differential leukocyte
545 count with machine learning: complementary and competitive with C-reactive protein and procalcitonin tests. *BMC Infectious*
546 *Diseases*, 22(1). <https://doi.org/10.1186/s12879-022-07223-7>
- 547 [26] Mahmoud, E., Dhoayan, M. A., Bosaeed, M., al Johani, S., & Arabi, Y. M. (2021). Developing machine-learning prediction algorithm
548 for bacteremia in admitted patients. *Infection and Drug Resistance*, 14, 757–765. <https://doi.org/10.2147/IDR.S293496>
- 549 [27] Garnica, O., Gómez, D., Ramos, V., Hidalgo, J. I., & Ruiz-Giardín, J. M. (2021). Diagnosing hospital bacteraemia in the framework of
550 predictive, preventive and personalised medicine using electronic health records and machine learning classifiers. *EPMA Journal*,
551 12(3), 365–381. <https://doi.org/10.1007/s13167-021-00252-3>
- 552 [28] Murri, R., De Angelis, G., Antenucci, L., Fiori, B., Rinaldi, R., Fantoni, M., Damiani, A., Patarnello, S., Sanguinetti, M., Valentini, V.,
553 Posteraro, B., & Masciocchi, C. (2024). A Machine Learning Predictive Model of Bloodstream Infection in Hospitalized Patients.
554 *Diagnostics (Basel, Switzerland)*, 14(4), 445. <https://doi.org/10.3390/diagnostics14040445>
- 555 [29] Ratzinger, F., Haslacher, H., Perkmann, T., Pinzan, M., Anner, P., Makristathis, A., Burgmann, H., Heinze, G., & Dorffner, G. (2018).
556 Machine learning for fast identification of bacteraemia in SIRS patients treated on standard care wards: a cohort study. *Scientific*
557 *Reports*, 8(1). <https://doi.org/10.1038/s41598-018-30236-9>
- 558 [30] Rahmani, K., Garikipati, A., Barnes, G., Hoffman, J., Calvert, J., Mao, Q., & Das, R. (2022). Early prediction of central line associated
559 bloodstream infection using machine learning. *American journal of infection control*, 50(4), 440–445.
560 <https://doi.org/10.1016/j.ajic.2021.08.017>
- 561 [31] Beeler, C., Dbeibo, L., Kelley, K., Thatcher, L., Webb, D., Bah, A., Monahan, P., Fowler, N. R., Nicol, S., Judy-Malcolm, A., Judy-
562 Malcolm, A., & Azar, J. (2018). Assessing patient risk of central line-associated bacteremia via machine learning. *American Journal*
563 *of Infection Control*, 46(9), 986–991. <https://doi.org/10.1016/j.ajic.2018.02.021>
- 564 [32] Zhou, T., Ren, Z., Ma, Y., He, L., Liu, J., Tang, J., & Zhang, H. (2023). Early identification of bloodstream infection in hemodialysis
565 patients by machine learning. *Heliyon*, 9(7). <https://doi.org/10.1016/j.heliyon.2023.e18263>
- 566 [33] Yoo, J., Kim, S. H., Hur, S., Ha, J., Huh, K., & Cha, W. C. (2021). Candidemia Risk Prediction (CanDETEC) Model for Patients With
567 Malignancy: Model Development and Validation in a Single-Center Retrospective Study. *JMIR medical informatics*, 9(7), e24651.
568 <https://doi.org/10.2196/24651>
- 569 [34] Mooney, C., Eogan, M., Ní Áinle, F., Cleary, B., Gallagher, J. J., O’Loughlin, J., & Drew, R. J. (2021). Predicting bacteraemia in
570 maternity patients using full blood count parameters: A supervised machine learning algorithm approach. *International Journal of*
571 *Laboratory Hematology*, 43(4), 609–615. <https://doi.org/10.1111/ijlh.13434>
- 572 [35] Su, M., Chen, H., Qiu, J., & Huang, J. (2021). Four Biomarkers-Based Artificial Neural Network Model for Accurate Early Prediction
573 of Bacteremia with Low-level Procalcitonin. *Annals of Clinical and Laboratory Science*, 51(3), 408–414.
- 574 [36] Wu, L., Xia, D., & Xu, K. (2023). Multi-Clinical Factors Combined with an Artificial Intelligence Algorithm Diagnosis Model for
575 HIV-Infected People with Bloodstream Infection. *Infection and Drug Resistance*, 16, 6085–6097.
576 <https://doi.org/10.2147/IDR.S423709>
- 577 [37] Choi, D. H., Hong, K. J., Park, J. H., Shin, S. D., Ro, Y. S., Song, K. J., Kim, K. H., & Kim, S. (2022). Prediction of bacteremia at the
578 emergency department during triage and disposition stages using machine learning models. *American Journal of Emergency Medicine*,
579 53, 86–93. <https://doi.org/10.1016/j.ajem.2021.12.065>

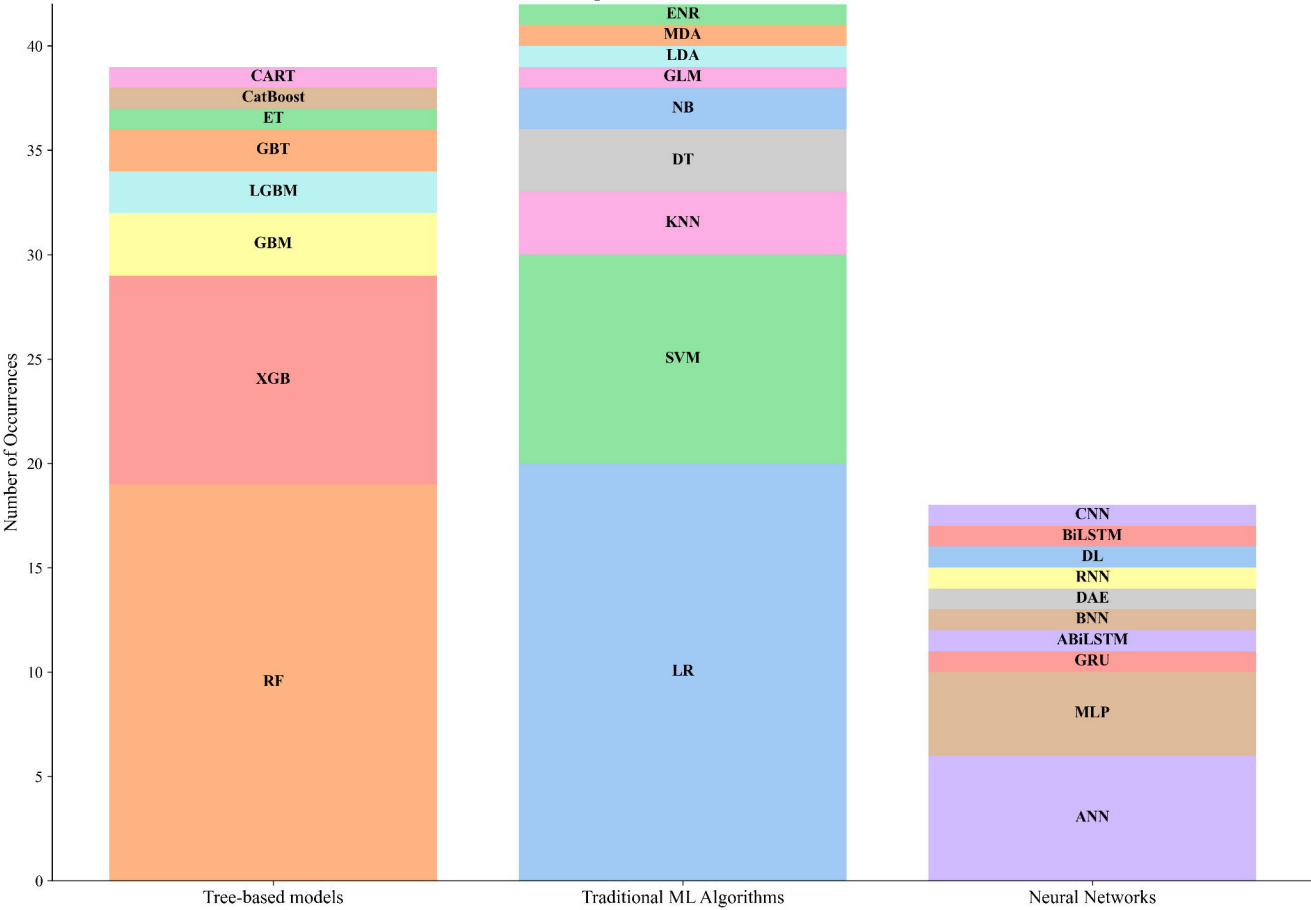
- 580 [38] Choi, D. H., Lim, M. H., Kim, K. H., Shin, S. D., Hong, K. J., & Kim, S. (2023). Development of an artificial intelligence bacteremia
581 prediction model and evaluation of its impact on physician predictions focusing on uncertainty. *Scientific Reports*, 13(1).
582 <https://doi.org/10.1038/s41598-023-40708-2>
- 583 [39] Boerman, A. W., Schinkel, M., Meijerink, L., van den Ende, E. S., Pladet, L. C., Scholtemeijer, M. G., Zeeuw, J., van der Zaag, A. Y.,
584 Minderhoud, T. C., Elbers, P. W. G., Wiersinga, W. J., de Jonge, R., Kramer, M. H., & Nanayakkara, P. W. B. (2022). Using machine
585 learning to predict blood culture outcomes in the emergency department: a single-centre, retrospective, observational study. *BMJ open*,
586 12(1), e053332. <https://doi.org/10.1136/bmjopen-2021-053332>
- 587 [40] Chang, Y.-H., Hsiao, C.-T., Chang, Y.-C., Lai, H.-Y., Lin, H.-H., Chen, C.-C., Hsu, L.-C., Wu, S.-Y., Shih, H.-M., Hsueh, P.-R.,
588 Hsueh, P.-R., & Cho, D.-Y. (2023). Machine learning of cell population data, complete blood count, and differential count parameters
589 for early prediction of bacteremia among adult patients with suspected bacterial infections and blood culture sampling in emergency
590 departments. *Journal of Microbiology, Immunology and Infection*, 56(4), 782–792. <https://doi.org/10.1016/j.jmii.2023.05.001>
- 591 [41] Schinkel, M., Boerman, A. W., Bennis, F. C., Minderhoud, T. C., Lie, M., Peters-Sengers, H., Holleman, F., Schade, R. P., de Jonge,
592 R., Wiersinga, W. J., Wiersinga, W. J., & Nanayakkara, P. W. B. (2022). Diagnostic stewardship for blood cultures in the emergency
593 department: A multicenter validation and prospective evaluation of a machine learning prediction tool. *EBioMedicine*, 82.
594 <https://doi.org/10.1016/j.ebiom.2022.104176>
- 595 [42] Goh, V., Chou, Y.-J., Lee, C.-C., Ma, M.-C., Wang, W. Y. C., Lin, C.-H., & Hsieh, C.-C. (2022). Predicting Bacteremia among Septic
596 Patients Based on ED Information by Machine Learning Methods: A Comparative Study. *Diagnostics*, 12(10).
597 <https://doi.org/10.3390/diagnostics12102498>
- 598 [43] Tsai, W.-C., Liu, C.-F., Ma, Y.-S., Chen, C.-J., Lin, H.-J., Hsu, C.-C., Chow, J. C., Chien, Y.-W., & Huang, C.-C. (2023). Real-time
599 artificial intelligence system for bacteremia prediction in adult febrile emergency department patients. *International Journal of*
600 *Medical Informatics*, 178. <https://doi.org/10.1016/j.ijmedinf.2023.105176>
- 601 [44] Roimi, M., Neuberger, A., Shrot, A., Paul, M., Geffen, Y., & Bar-Lavie, Y. (2020). Early diagnosis of bloodstream infections in the
602 intensive care unit using machine-learning algorithms. *Intensive Care Medicine*, 46(3), 454–462. [https://doi.org/10.1007/s00134-019-](https://doi.org/10.1007/s00134-019-05876-8)
603 [05876-8](https://doi.org/10.1007/s00134-019-05876-8)
- 604 [45] Van Steenkiste, T., Ruysinck, J., De Baets, L., Decruyenaere, J., De Turck, F., Ongenaet, F., & Dhaene, T. (2019). Accurate
605 prediction of blood culture outcome in the intensive care unit using long short-term memory neural networks. *Artificial intelligence in*
606 *medicine*, 97, 38–43. <https://doi.org/10.1016/j.artmed.2018.10.008>
- 607 [46] Boner, Z., & Moore, C. C. (2022). Deep Learning Risk Prediction of Bloodstream Infection in the Intensive Care Unit. In *KDD*
608 *Undergraduate Consortium*, August 14–18, 2022, Washington, D.C... ACM, New York, NY, USA, 7 pages
- 609 [47] Pai, K.-C., Wang, M.-S., Chen, Y.-F., Tseng, C.-H., Liu, P.-Y., Chen, L.-C., Sheu, R.-K., & Wu, C.-L. (2021). An artificial
610 intelligence approach to bloodstream infections prediction. *Journal of Clinical Medicine*, 10(13). <https://doi.org/10.3390/jcm10132901>
- 611 [48] Parreco, J. P., Hidalgo, A. E., Badilla, A. D., Ilyas, O., & Rattan, R. (2018). Predicting central line-associated bloodstream infections
612 and mortality using supervised machine learning. *Journal of Critical Care*, 45, 156–162. <https://doi.org/10.1016/j.jcrc.2018.02.010>
- 613 [49] Yuan, S., Sun, Y., Xiao, X., Long, Y., & He, H. (2021). Using Machine Learning Algorithms to Predict Candidaemia in ICU Patients
614 With New-Onset Systemic Inflammatory Response Syndrome. *Frontiers in Medicine*, 8. <https://doi.org/10.3389/fmed.2021.720926>

- 615 [50] Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for
616 individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ (Clinical research ed.)*, 350, g7594.
617 <https://doi.org/10.1136/bmj.g7594>
- 618 [51] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique.
619 *Journal of artificial intelligence research*, 16, 321-357.
- 620 [52] Little, R., & Rubin, D. (2019). *Statistical analysis with missing data*, third edition. Wiley.
- 621 [53] Younis HA, Eisa TAE, Nasser M, et al. A Systematic Review and Meta-Analysis of Artificial Intelligence Tools in Medicine and
622 Healthcare: Applications, Considerations, Limitations, Motivation and Challenges. *Diagnostics (Basel)*. 2024;14(1):109. Published
623 2024 Jan 4. doi:10.3390/diagnostics14010109
- 624 [54] Dargère, S., Cormier, H., & Verdon, R. (2018). Contaminants in blood cultures: importance, implications, interpretation and
625 prevention. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and
626 Infectious Diseases*, 24(9), 964–969. <https://doi.org/10.1016/j.cmi.2018.03.030>
- 627 [55] Habebh, H., & Gohel, S. (2021). Machine Learning in Healthcare. *Current genomics*, 22(4), 291–300.
628 <https://doi.org/10.2174/1389202922666210705124359>
- 629 [56] Zoabi, Y., Kehat, O., Lahav, D. et al. Predicting bloodstream infection outcome using machine learning. *Sci Rep* 11, 20101 (2021).
630 <https://doi.org/10.1038/s41598-021-99105-2>
- 631 [57] Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. *Artificial Intelligence in Healthcare*,
632 25–60. <https://doi.org/10.1016/B978-0-12-818438-7.00002-2>
- 633 [58] Javaid, M., Haleem, A., Singh, R. P., Suman, R., & Rab, S. (2022). Significance of machine learning in healthcare: Features, pillars
634 and applications. *International Journal of Intelligent Networks*, 3, 58-73. <https://doi.org/10.1016/j.ijin.2022.05.002>
- 635 [59] Bajwa, J., Munir, U., Nori, A., & Williams, B. (2021). Artificial intelligence in healthcare: transforming the practice of medicine.
636 *Future healthcare journal*, 8(2), e188–e194. <https://doi.org/10.7861/fhj.2021-0095>
- 637 [60] Batko, K., & Ślęzak, A. (2022). The use of Big Data Analytics in healthcare. *Journal of big data*, 9(1), 3.
638 <https://doi.org/10.1186/s40537-021-00553-4>
- 639 [61] Alemkere, G., Tenna, A., & Engidawork, E. (2019). Antibiotic use practice and predictors of hospital outcome among patients with
640 systemic bacterial infection: Identifying targets for antibiotic and health care resource stewardship. *PloS one*, 14(2), e0212661.
641 <https://doi.org/10.1371/journal.pone.0212661>
- 642 [62] Giordano, C., Brennan, M., Mohamed, B., Rashidi, P., Modave, F., & Tighe, P. (2021). Accessing Artificial Intelligence for Clinical
643 Decision-Making. *Frontiers in digital health*, 3, 645232. <https://doi.org/10.3389/fgdth.2021.645232>
- 644 [63] Antoniadis AM, Du Y, Guendouz Y, Wei L, Mazo C, Becker BA, Mooney C. Current Challenges and Future Opportunities for XAI in
645 Machine Learning-Based Clinical Decision Support Systems: A Systematic Review. *Applied Sciences*. 2021; 11(11):5088.
646 <https://doi.org/10.3390/app11115088>
- 647 [64] Susanto, A. P., Lyell, D., Widyantoro, B., Berkovsky, S., & Magrabi, F. (2024). How Well Do AI-Enabled Decision Support Systems
648 Perform in Clinical Settings?. *Studies in health technology and informatics*, 310, 279–283. <https://doi.org/10.3233/SHTI230971>
- 649 [65] Iqbal, J., Cortés Jaimes, D. C., Makineni, P., Subramani, S., Hemaida, S., Thugu, T. R., Butt, A. N., Sikto, J. T., Kaur, P., Lak, M. A.,
650 Augustine, M., Shahzad, R., & Arain, M. (2023). Reimagining Healthcare: Unleashing the Power of Artificial Intelligence in Medicine.
651 *Cureus*, 15(9), e44658. <https://doi.org/10.7759/cureus.44658>

- 652 [66] Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Liu, X., Wu, Y., Dong, F., Qiu, C.-W., Qiu, J., Hua, K., Su, W., Wu, J., Xu, H.,
653 Han, Y., Fu, C., Yin, Z., Liu, M., Roepman, R., ... Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research.
654 *The Innovation*, 2(4), Article 100179. <https://doi.org/10.1016/j.xinn.2021.100179>
- 655 [67] Myllyaho, L., Raatikainen, M., Männistö, T., Mikkonen, T., & Nurminen, J. K. (2021). Systematic literature review of validation
656 methods for AI systems. *Journal of Systems and Software*, 181, Article 111050. <https://doi.org/10.1016/j.jss.2021.111050>
- 657 [68] de Hond, A.A.H., Leeuwenberg, A.M., Hooft, L. et al. Guidelines and quality criteria for artificial intelligence-based prediction
658 models in healthcare: a scoping review. *npj Digit. Med.* 5, 2 (2022). <https://doi.org/10.1038/s41746-021-00549-7>
- 659 [69] Stevens, L. M., Mortazavi, B. J., Deo, R. C., Curtis, L., & Kao, D. P. (2020). Recommendations for Reporting Machine Learning
660 Analyses in Clinical Research. *Circulation. Cardiovascular quality and outcomes*, 13(10), e006556.
661 <https://doi.org/10.1161/CIRCOUTCOMES.120.006556>
- 662 [70] Rieke, N., Hancox, J., Li, W. et al. The future of digital health with federated learning. *npj Digit. Med.* 3, 119 (2020).
663 <https://doi.org/10.1038/s41746-020-00323-1>
- 664 [71] Moor, M., Banerjee, O., Abad, Z.S.H. et al. Foundation models for generalist medical artificial intelligence. *Nature* 616, 259–265
665 (2023). <https://doi.org/10.1038/s41586-023-05881-4>
- 666 [72] Antoniou, T., & Mamdani, M. (2021). Evaluation of machine learning solutions in medicine. *CMAJ : Canadian Medical Association*
667 *journal = journal de l'Association medicale canadienne*, 193(36), E1425–E1429. <https://doi.org/10.1503/cmaj.210036>
- 668 [73] Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. *Artificial*
669 *Intelligence in Healthcare*, 295–336. <https://doi.org/10.1016/B978-0-12-818438-7.00012-5>
- 670 [74] Naik, N., Zeeshan, H. B. M., Shetty, D. K., Swain, D., Shah, M., Paul, R., Aggarwal, K., Ibrahim, S., Patil, V., Komal, S., Shetty, S.,
671 Rai, B. P., Chlosta, P., & Somani, B. K. (2022). Legal and ethical consideration in artificial intelligence in healthcare: Who takes
672 responsibility? *Frontiers in Surgery*, 9. <https://doi.org/10.3389/fsurg.2022.862322>
- 673 [75] Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare:
674 Techniques and applications. *Computers in Biology and Medicine*, 158, Article 106848.
675 <https://doi.org/10.1016/j.combiomed.2023.106848>
- 676 [76] Murdoch, B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethics* 22, 122
677 (2021). <https://doi.org/10.1186/s12910-021-00687-3>
- 678
679
680
681
682
683
684
685
686



ML Algorithms for BSI Prediction



Prevalence of Performance Metrics in ML Studies for BSI Prediction

