Journal of Pioneering Medical Sciences
Received June 14, 2023 Accept October 23, 2023 Publish December 23, 2023.
Volume 12, Issue 3, Pages 45-54

DOI https://doi.org/10.61091/jpms202312310

Artificial Intelligence in Laboratory Technologies for Early Detection and Prognostication of Sepsis: A Systematic Review

Mohsen Bakouri[1,*], Nasser M. Alqahtani[2], Othman M. Alhussain[3], Nawaf Alrashidi[3], Sulaiman N. Almutairi[4], Ahmed O. Alabdulwahab[5], Badr S. Alaskar[5], and Megren A. Alqahtani[6]

[1]Department of Medical Equipment Technology, College of Applied Medical Science, Majmaah University, Majmaah City 11952, Saudi Arabia.
[2]Department of Medical Education, College of Medicine, Majmaah University, Majmaah City 11952, Saudi Arabia.
[3]College of Medicine, Majmaah University, Majmaah City 11952, Saudi Arabia.
[4]Clinical Skills Centre, College of Medicine, Majmaah University, Majmaah City 11952, Saudi Arabia.
[5]University Medical Center, Majmaah University, Majmaah City 11952, Saudi Arabia.
[6]King Khalid Hospital, Majmaah City 11952, Saudi Arabia.

Corresponding author: Mohsen Bakouri (e-mail: m.bakouri@mu.edu.sa).

Abstract **Background:** Sepsis a complex clinical syndrome represents life-threatening organ dysfunction instigated by an infection's dysregulated host response. Early detection and accurate prognostication of sepsis are crucial; they pave the way for timely intervention, ultimately enhancing patient outcomes. The rise in interest towards Artificial Intelligence (AI) applications within laboratory technologies is directly related to its potential for improving early detection and prognosis forecasting in sepsis cases; this interest comes as AI continues its advancement. **Methods:** We conducted a systematic review of studies utilizing AI algorithms in laboratory settings for early sepsis detection and prognostication: our methods entailed searching relevant databases for research published until October 2023. Our inclusion criteria spanned original articles; these applied machine learning (ML) and deep learning (DL) techniques to laboratory data with the aim being sepsis prediction. We assessed the quality of the studies, extracted and synthesized data on AI model performance metrics - including: area under receiver operating characteristic curve (AUROC), sensitivity, specificity, and accuracy. **Results:** The review encompassed eight studies meeting the inclusion criteria; AI models showcased exceptional predictive capabilities evidenced by a range of AUROC values from 0.799 to 0.9213, signifying noticeably acceptable performance. However, there was wide variation in sensitivity and specificity among these analyses; an indicator of heterogeneity in model performance. Superior prognostic accuracy and potential for real-time monitoring of patients' early sepsis signs emerged in several models; notably, within the first 12 hours of patient admission their highest predictive period. The models frequently outperformed traditional scoring systems. **Conclusion:** Laboratory technology's AI applications significantly promise sepsis' early detection and prognostication. Reviewed studies suggest AI models may surpass traditional methods, offering potential integration into clinical workflows for rapid sepsis identification aid. Nevertheless, we also acknowledged both the variability in model performance and necessity of additional validation across diverse clinical settings. **Future research:** it must concentrate on two key aspects–the refinement of AI algorithms to enhance sensitivity and precision; furthermore, it should delve into evaluating the clinical impact of tools for sepsis prediction that are assisted by AI.

Key Words Artificial Intelligence, Sepsis, Early Detection, Prognostication, Laboratory Technologies, Machine Learning, Deep Learning

1. Introduction

A dysregulated host response to infection causes sepsis, a life-threatening organ dysfunction that presents as global healthcare challenge with high morbidity and mortality rates. Detecting sepsis early, accurately predicting its outcomes are critical for improving patient results; however, these tasks persist in complexity filled with clinical uncertainty [1]. The promise of a significant revolution in current sepsis management landscape emerges from integrating artificial intelligence (AI) into laboratory technologies [2].

AI processes vast datasets with unparalleled speed and precision, harnessing this capability offers a groundbreaking edge in early sepsis identification [3]. More specifically, machine learning algorithms-a subset of AI-analyzes intricate

Abbreviation	Full form
AI	Artificial Intelligence
AUROC	Area Under the Receiver Operating Characteristic
CNN	Convolutional Neural Network
DOR	Diagnostic Odds Ratio
ICU	Intensive Care Unit
LASSO	Least Absolute Shrinkage and Selection Operator
LSTM	Long Short-Term Memory
ML	Machine Learning
MLD	Machine Learning Derived
NPV	Negative Predictive Value
PBIs	Presumed Bacterial Infections
PPV	Positive Predictive Value
SOFA	Sequential Organ Failure Assessment
TSS	Traumatic Sepsis Score
XGBoost	Extreme Gradient Boosting

Table 1: Abbreviations used in the review

laboratory results; it identifies patterns that may not be immediately evident to human clinicians. Consequently, potential sepsis cases receive swift flags compared to conventional methods. Markedly increasing the risk of mortality is every hour's delay in treating sepsis; hence this swift identification becomes crucial [4].

AI applications in laboratory medicine, moreover, extend beyond mere detection: they provide prognostic insights; these can direct clinical decision-making [5]. By employing predictive analytics- a tool of AI- we stratify patients based on risk and forecast their trajectories; this enables us to tailor treatment strategies for each individual–potentially enhancing patient outcomes while optimizing resource utilization [6].

AI holds the potential to revolutionize laboratory technologies in sepsis through its remarkable computational capability and perpetual learning aptitude. Over time, we can train AI systems on new data which enhances their diagnostic and prognostic capabilities [7]. Coupled with advancements in laboratory techniques like genomics and proteomics, this dynamic nature of AI could drive earlier interventions with greater precision; potentially transforming patient care pathways as well as outcomes [8].

The contemporary medical lexicon characterizes sepsis as a critical condition: it originates from an infection-induced maladaptive response of the host, and this can lead to life-threatening organ dysfunction; furthermore, its definition established in 2016 supersedes the earlier one–predominantly based on systemic inflammation markers delineated in 1992. This redefined perspective not only refines our understanding of sepsis' underlying pathophysiological mechanisms but also boosts precision regarding diagnostic benchmarks. In favor of a more nuanced recognition of sepsis-related organ dysfunction, we have rendered the term 'severe sepsis' obsolete. We designate sepsis that progresses to encompass circulatory collapse as septic shock; this represents the most critical manifestation of the syndrome.

The evolution of the sepsis definition has concomitantly prompted updates in diagnosis protocols, which underscore routine microbiological cultures' necessity–specifically blood ones. These are crucially obtained before initiating antimicrobial therapy for patients with presumptive sepsis or septic shock; this emphasis stands contingent upon a vital condition: such diagnostic efforts should not markedly impede the onset of antimicrobial intervention [8].

Improving sepsis outcomes [9], integrally hinges on the principle of early detection and timely therapeutic intervention. Indeed, we have a well-established correlation: prompt management in prehospital and emergency department settings is directly related to positive patient outcomes [10]. The criticality of the initial hours following symptom onset receives underscored recommendations; however–challenges persist: accelerating patient transfer from emergency departments to intensive care units remains an ongoing issue [11]. Overlapping clinical presentations often present clinicians with the challenge of distinguishing sepsis from other acute illnesses; this necessitates a high degree of vigilance and clinical acumen [12].

This systematic review aims to collate and evaluate the current evidence on the application of AI in laboratory technologies for the early detection and prognostication of sepsis. By doing so, it seeks to understand the extent to which AI has been integrated into clinical practice, identify the benefits and challenges associated with its use, and provide a clear picture of its efficacy and reliability.

2. Materials and Methods

Reporting Standards

This review was reported in accordance with the PRISMA guidelines [13], and the review protocol was registered with PROSPERO prior to the initiation of this review. The results of the study selection process as per PRISMA has been shown through Figure 1.

PECO Strategy

Population- The population of interest for this review comprised patients of any age group who were admitted to any healthcare setting (e.g., emergency departments, intensive care units, general wards) with suspected or confirmed sepsis.

Exposure- The exposure of interest was the application of AI within laboratory technologies. This included the use of machine learning algorithms, deep learning models, and other AI-related approaches to analyze laboratory test results for the early detection and prognostication of sepsis.

Comparator- A comparator was not deemed necessary but was not excluded; when present, it would have been the standard of care without the use of AI-enhanced laboratory technologies. This might have included traditional methods of sepsis detection, such as manual review of laboratory results and clinical assessments.

Outcome- The primary outcomes were the accuracy of AI-enhanced laboratory technologies in the early detection of sepsis (sensitivity, specificity, positive predictive value, negative predictive value) and the prognostic performance in

predicting sepsis-related outcomes (e.g., mortality, length of hospital stay, readmission rates).

Search Strategy

Table 1 shows the abbrevations and full forms of terms which are used is reviews.

A comprehensive literature search was conducted across several electronic databases, including PubMed, EMBASE, Scopus, Web of Science, Google Scholar and IEEE Xplore, to identify relevant studies published up to the current date. The search strategy was designed to encompass terms related to AI (e.g., "machine learning," "deep learning," "neural networks"), laboratory technologies (e.g., "laboratory tests," "biomarkers"), and sepsis (e.g., "sepsis," "septicemia," "systemic inflammatory response"). Boolean operators (AND, OR) were used to combine search terms, as shown in Table 2.

Inclusion and Exclusion Criteria

Table 3 shows the inclusion and exclusion criteria that were devised for this review

Data Extraction

We employed a standardized data extraction form to guarantee consistency and accuracy; two independent reviewers undertook this critical process. The information they extracted from the included studies encompassed author details, publication year, study design–and detailed descriptions of patient populations. In addition to scrutinizing the bibliographic and methodological details, reviewers honed in on the specific AI technology each study employed. They documented a variety of laboratory tests analyzed via AI technologies along with their diverse measured outcomes. The primary interest underscored was how well these AI technologies performed diagnostically and prognostically in early sepsis detection; this encompassed metrics like sensitivity, specificity–positive predictive value versus negative predictive value–as well as any results tied to prognosis related to sepsis: mortality rates, hospital stay duration or even re-admission frequency rates.

Quality Assessment

The methodological quality of the included studies was assessed using appropriate tools based on study design. The studies were evaluated using the Newcastle-Ottawa Scale (NOS) which assessed bias across multiple domains [14], as elucidated through Figure 2. Quality assessment was independently conducted by two reviewers, with disagreements resolved by consensus or by a third reviewer.

3. Results

Schematics of Article Selection

We initially identified potential studies by combing through databases and registers, which yielded a total of 392 records from the former; however, we found no additional findings in the latter. After removing duplicate records–a task that reduced our count by 68–we began screening with an eligible pool totaling to 293 after excluding those marked as ineligible due to automation tools: specifically, 31 of them. The automation tools determined that all remaining records, excluding those removed due to duplication or ineligibility, required screening. Consequently, we pursued the retrieval of these 293 records. Despite our efforts; however-55 reports eluded retrieval-a circumstance which diminished the total number of assessed eligible reports to 238. Upon eligibility assessment, we excluded a total of 230 reports based on several criteria: specifically, 86 didn't respond to the predefined PECO criteria; an additional amount–consisting of 59 –comprised animal-based studies and yet another group–literature reviews contributed significantly with their count reaching up to 63. Moreover, the unavailability of full texts led to the exclusion of 22 reports. Following this comprehensive screening and eligibility assessment, we deemed only 8 studies [15]–[22] suitable for inclusion in our review.

Assessed Bias Across Domains

The study by Calvert et al. [15] demonstrated a notably robust methodology across all domains, with a 'Low' bias rating in each category, suggesting a high level of confidence in the results. Similarly, studies by Lauritsen et al. [17], Lind et al. [18], Lu et al. [19], Mao et al. [20], and Nemati et al. [21], all showed a 'Low' level of bias in each domain, indicating a strong methodological framework and minimal bias. However, Khojandi et al. [16], had some areas with greater potential for bias; specifically, there was a 'Moderate' bias noted in the domains assessing deviations from intended interventions (D4) and measurement of outcomes (D6). While the study maintained a 'Low' bias in the other domains, these moderate concerns affected the overall assessment, resulting in a 'Moderate' overall bias rating. Tang et al. [22], also presented a 'Moderate' bias in two domains. The first was bias due to confounding (D1), which raises questions about other variables possibly influencing the outcomes. The second was bias due to missing data (D5), indicating that some relevant data might not have been accounted for or reported. Despite these moderate concerns, the remaining domains were rated 'Low' for bias, leading to an overall 'Moderate' bias rating. Overall, though, the majority of the studies included in the review exhibited a 'Low' bias across all domains.

Examined Sample Size and Criteria

Table 4 presents the selected studies [15]–[22] that applied various forms of AI to the detection and prognostication of sepsis and other associated assessments.

Calvert et al. [15] conducted a study that scrutinized an extensive dataset of 122,672 hospital stays; they selectively concentrated on patients aged 45 and above who had experienced a minimum hospitalization period of 96 hours. The researchers presumably chose this specific patient group because prolonged stays in hospitals and advancing age are correlated with an escalated risk of sepsis. Khojandi et al. [16] conducted research where they analyzed an even more

Database	Search string
PubMed	("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural networks") AND ("laboratory tests" OR "biomarkers") AND "sepsis"
EMBASE	("machine intelligence" OR "computational learning" OR "AI") AND ("laboratory diagnostics" OR "clinical markers") AND "septicemia"
Scopus	("predictive analytics" OR "AI" OR "neural computing") AND ("diagnostic tests" OR "laboratory data") AND "systemic inflammatory response"
Web of Science	("data mining" OR "algorithmic learning" OR "artificial neural networks") AND ("laboratory findings" OR "molecular markers") AND "sepsis syndrome"
Google Scholar	("algorithm-based" OR "machine prediction" OR "deep computation") AND ("test results" OR "lab values") AND "sepsis"
IEEE Xplore	("intelligent systems" OR "learning systems" OR "deep machine learning") AND ("biomarker identification" OR "lab diagnostics") AND "septic shock"

Table 2: Search strings utilised across the assessed databases

Criteria type	Inclusion criteria	Exclusion criteria
Study design	- Cohort studies - Case-control studies - Cross-sectional studies - In-vitro studies - Experimental studies	- Conference abstracts - Case reports - Commentaries - Editorials -Literature reviews
Study protocol	- Studies describing AI application in lab technologies for early detection or prognostication of sepsis	- Studies not focusing on AI - Studies not related to laboratory technologies - Studies not addressing sepsis detection or prognostication
Population	- Patients of any age group admitted to healthcare settings with suspected or confirmed sepsis	-
Outcome	- Accuracy of AI in lab technologies for early sepsis detection - Prognostic performance for sepsis-related outcomes	-

Table 3: Selection criteria devised for this review

extensive dataset of 332,006 entries. The data focused on two key periods: the hours immediately after admission and the interval preceding sepsis' clinical manifestation - aiming for early detection critical to enhance patient outcomes via prompt treatment. Lauritsen et al. [17] took an approach of retrospective analysis, exploring 3,126 contacts from Danish hospitals over a seven-year period; this longitudinal dataset offered insights into sepsis evolution and the extended timespan potentiality of AI for early detection.

Lind et al. [18] conducted a specialized study, examining an 8,131 patient-cohort of adult recipients of allo-HCT at the Fred Hutchinson center. This population's medical treatment nature heightens their susceptibility to sepsis; this is well-known. Lu et al. [19] selected a focused cohort of 684 trauma patients as their study subject due to sepsis's high risk and rapid progression following traumatic injuries; Mao et al., conducted another relevant study. Its vast dataset-encompassing 684,443 hospital encounters-distinguished [20], through a comprehensive evaluation of AI's role in a heterogenous patient population within the hospital.

In contrast to the aforementioned studies, Nemati et al. [21], utilized a development dataset of 27,527 patient encounters and a validation dataset exceeding 52,000. The research deliberately omitted patients who met Sepsis-3 criteria within four hours of their ICU admission; this focused on evaluating AI's predictive capability for sepsis beyond an immediate timeframe post-ICU entry. Tang et al. [22], directed their study towards a dataset of 2,453 COVID-19 patients; this timely and critical patient group–given the pandemic situation–faces elevated risks for developing sepsis as a secondary complication.

AI Protocols Assessed

Calvert et al. [15], employed gradient-boosted decision trees, specifically XGBoost, an ensemble learning method renowned for its predictive accuracy. This method was likely chosen for its ability to handle large datasets and complex feature interactions, which were characteristic of the 122,672 hospital stays analyzed in their study. Khojandi et al. [16] utilized random forest classifiers, a robust ensemble learning technique that combines the predictions of multiple decision trees to improve predictive accuracy and control over-fitting. The study's expansive dataset of 332,006 entries indicated

that random forest classifiers were well-suited to manage the substantial variability within the data. Lauritsen et al. [17], adopted a CNN-LSTM combination model, an AI architecture that leverages both the feature extraction capabilities of CNNs and the sequential data processing strengths of LSTMs. This choice was particularly appropriate for analyzing the 3,126 contacts from Danish hospitals, capturing both spatial and temporal patterns within the data.

Lind et al. [18] implemented a SuperLearner algorithm, designed to optimize the area under the AUC. The SuperLearner is an ensemble method that uses cross-validation to find the optimal combination of prediction algorithms, which was applied to the study of 8,131 PBIs in allo-HCT recipients. Lu et al. [19], chose the LASSO, a regression analysis method that performs both variable selection and regularization to enhance the prediction accuracy and interpretability of the statistical model. LASSO was applied to a cohort of 684 trauma patients, facilitating the identification of the most relevant predictors of sepsis in a high-dimensional dataset. Mao et al. [20], utilized gradient tree boosting, another ensemble method that builds models in a stage-wise fashion and is known for its predictive power and flexibility. The vast dataset of 684,443 hospital encounters analyzed in their study could benefit from the method's capacity to model complex interactions between variables.

Nemati et al. [21], developed the AISE, an AI model whose details, while not specified in the provided context, were indicative of a tailored solution designed to address the specific challenges of sepsis detection in over 79,527 patient encounters. Tang et al. [22], also applied XGBoost to a dataset of 2,453 patients with COVID-19. XGBoost's application to this dataset was likely due to its strong performance in classification tasks, which was essential for identifying sepsis in COVID-19 patients, where the underlying patterns could be highly complex and non-linear.

Sensitivity and Specificity Values Observed

Calvert et al. [15], reported an AUROC of 0.917, which indicated a high degree of discriminative ability for their AI model, XGBoost. The model's sensitivity and specificity were also relatively high, at 0.799 and 0.860 respectively, demonstrating a balanced ability to identify true positives and true negatives. The accuracy of the model stood at 0.848, suggesting that the model correctly classified a high percentage of cases. Khojandi et al. [16] did not provide an exact AUROC value but indicated a range for sensitivity and specificity, up to 67% and 63% respectively. This suggests some variability in the model's performance, with a tendency toward lower detection rates. The accuracy was reported with a wide range, from 65% to 98.63%, indicating that in some scenarios, the model performed exceptionally well, while in others, its performance was closer to chance.

Lauritsen et al. [17], reported a wide range for sensitivity, from 0.09 to 1.00, and for specificity, from 0.10 to 0.93. The wide range may reflect variations in the model's performance across different settings or thresholds used to define sepsis. However, the absence of an AUROC or accuracy value in the provided data made it difficult to assess the overall performance of the CNN-LSTM combination model used in the study. Lind et al. [18] provided an AUROC value of 0.85, indicating good discriminative power for the SuperLearner optimized AUC model. However, the study did not report sensitivity, specificity, or accuracy, which limits the ability to appraise the model's performance comprehensively. Lu et al. [19], reported an AUROC of 0.799, sensitivity of 64.0%, and specificity of 82.0%. These metrics suggest that the model, LASSO, had reasonable discriminative ability but tended to miss a higher proportion of true positive cases (lower sensitivity) while correctly identifying a majority of true negatives (higher specificity).

Mao et al. [20], achieved an AUROC of 0.92, which suggests excellent performance by the gradient tree boosting model. However, the study did not provide sensitivity, specificity, or accuracy, precluding a full assessment of the model's performance. Nemati et al. [21], reported an AUROC that varied between 0.83 and 0.85, with a fixed sensitivity of 85% and a specificity range of 64% to 72%. These findings indicate that the Artificial Intelligence Sepsis Expert (AISE) was quite effective at detecting true positive cases but had a moderate rate of false positive outcomes. Tang et al. [22], presented an AUROC of 0.9213 for their XGBoost model, with a high sensitivity of 97.17% and a specificity of 82.05%. These values denote an excellent predictive performance with a very high true positive rate and a good true negative rate, suggesting that the model was particularly effective in identifying sepsis among COVID-19 patients.

4. Discussion

Our findings revealed varied performance across different studies, with each model exhibiting unique strengths and limitations. Calvert et al. [15] showcased a model with a high degree of accuracy and an excellent balance between sensitivity and specificity, implying a robust overall performance in the identification of sepsis. In contrast, Khojandi et al. [16], presented a model with less consistency, as evidenced by the wide range of reported accuracy and the absence of an exact AUROC value, which may indicate variability in performance across different clinical scenarios or patient populations.

The model analyzed by Lauritsen et al. [17], utilizing a CNN-LSTM approach, demonstrated significant variability in its sensitivity and specificity ranges, suggesting that its performance might be highly dependent on the specific operational thresholds and clinical settings, yet the lack of comprehensive performance metrics such as AUROC and accuracy impeded a definitive conclusion regarding its efficacy. Lind et al. [18], reported an AI model with good discriminative power as suggested by its AUROC value, but the study's omission of sensitivity, specificity, and accuracy data limited a full evaluation of its clinical utility.

Lu et al. [19], presented a model with moderate discriminative ability, as reflected by its AUROC, but with a tendency

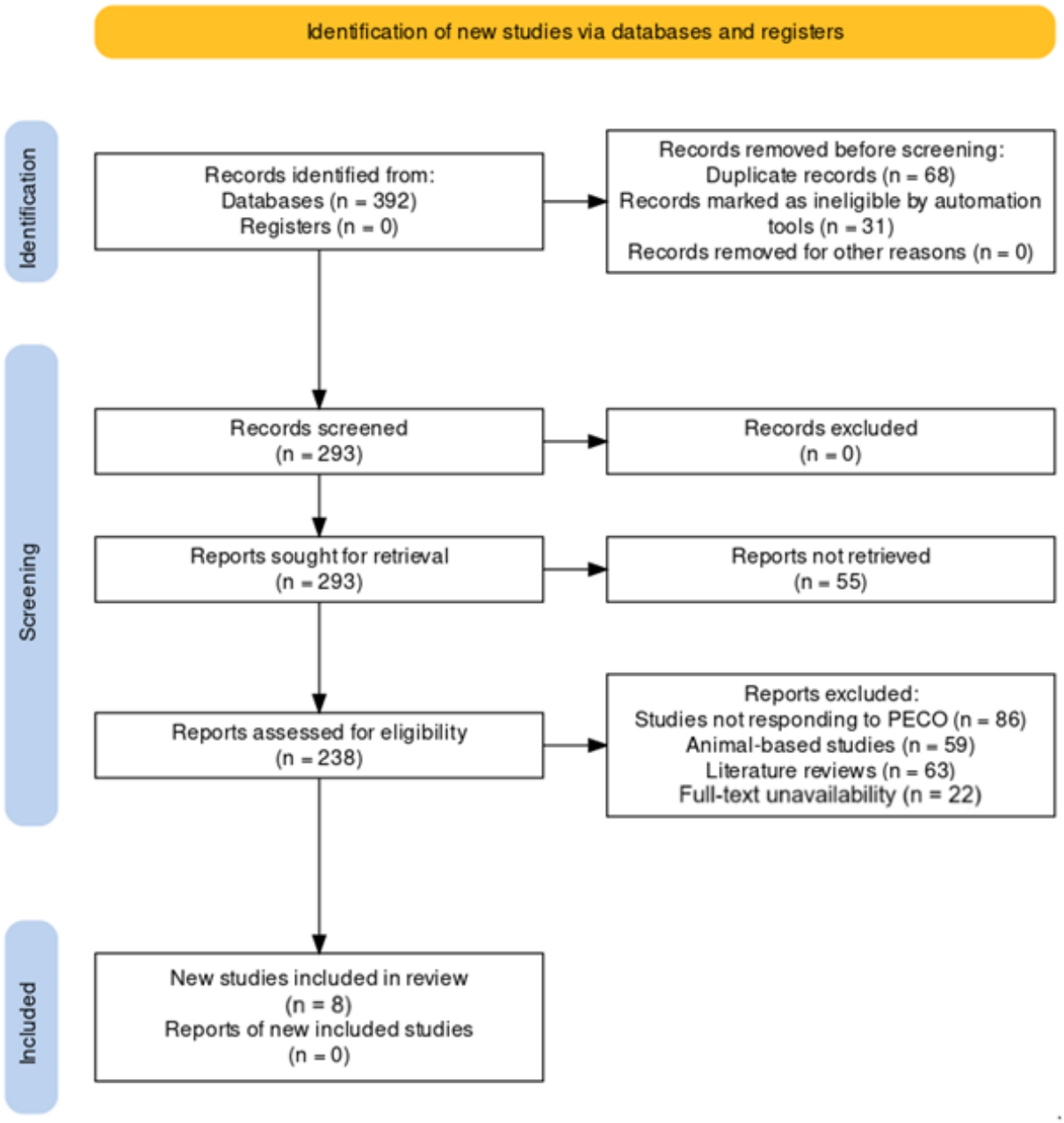

Figure 1: PRISMA protocol representation of the study selection protocol for the review

to miss a higher number of true positives, which could limit its application in clinical settings where high sensitivity is paramount. Mao et al. [20], reported on a model with excellent discriminative ability, yet the lack of reported sensitivity, specificity, and accuracy left gaps in understanding the model's practical performance.

Nemati et al. [21], offered insights into a model that was adept at identifying true positives with a high sensitivity, although the model's specificity indicated a moderate rate of false positives, which could potentially lead to over-treatment or alarm fatigue in clinical practice. Tang et al. [22] presented a model with superior performance in both sensitivity and specificity, indicating an exceptional capability in discerning sepsis, particularly in the context of COVID-19 patients, which could have important implications for managing this specific patient cohort.

Our review intersects with the observations of Yang et al. [23], in the recognition of AI's capabilities to augment the early detection and precise treatment of sepsis, as well as the prognostic assessment of the condition. Both studies concur on the pivotal role of high-quality data, typically abundant in ICU settings, as a critical enabler for AI's efficacy. Where our paths diverge, however, is in the scope of AI's integration in sepsis management. Yang et al. [23], explore AI's extensive

Study	Dataset size	Patient criteria	AI model type	AUROC	Sensitivity	Specificity	Accuracy	Overall inferences pertaining to performance of AI
Calvert et al [15]	122,672 stays	Patients aged 45+ with hospital stays of at least 96 hours	Gradient-boosted decision trees (XGBoost)	0.917	0.799	0.860	0.848	- MLD superior in specificity, PPV, NPV, DOR, accuracy - Strong diagnostic capability for early sepsis detection
Khojandi et al [16]	332,006	Early hours post-admission and period leading up to sepsis	Random forest classifiers	N/A	Up to 67%	Up to 63%	65-98.63%	- High potential of ML for continuous monitoring - Highest predictive accuracy within first 12 hours post-admission
Lauritsen et al [17]	3126 contacts	Retrospective data from Danish hospitals over 7 years	CNN-LSTM combination	N/A	0.09-1.00	0.10-0.93	Unspecified	- Models show significant performance variations - Some require improvements in sensitivity and precision
Lind et al [18]	8131 PBIs	Adult allo-HCT recipients at Fred Hutchinson	SuperLearner optimized AUC	0.85	N/A	N/A	N/A	- Superior prognostic accuracy for outcomes - Potential for timely sepsis detection among allo-HCT recipients
Lu et al [19]	684 patients	Trauma patients	LASSO	0.799	64.0%	82.0%	N/A	- TSS shows good predictive capability (AUC 0.799) -Better than individual predictors and SOFA score - Stratifies patients into risk categories correlated with sepsis incidence - Demonstrates good calibration and reliability for clinical use
Mao et al [20]	684,443 encounters	Hospital patients	Gradient Tree Boosting	0.92	N/A	N/A	N/A	- InSight algorithm exceptional in detecting and predicting sepsis - High AUROC scores, especially for septic shock prediction four hours before onset - Robust to missing data and performs consistently across multiple institutions
Nemati et al [21]	Development: 27,527; Validation: 52,000+	Excluded patients who met Sepsis-3 prior to or within 4 hours of ICU admission.	Artificial Intelligence Sepsis Expert	0.83-0.85	Fixed at 85%	64%-72%	-	- Good discriminative ability with slight decline over time - Higher prediction accuracy closer to event onset
Tang et al [22]	2,453	Patients with COVID-19	XGBoost	0.9213	97.17%	82.05%	-	- Coagulation function indicators highly predictive of viral sepsis caused by SARS-CoV-2 - Early warning of sepsis in COVID-19 patients was possible due to the ML model

Table 4: Studies included in the review and their observed assessments

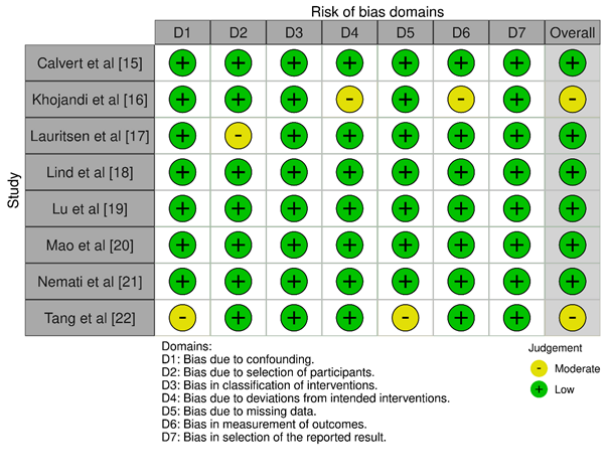

Figure 2: Bias assessment of the selected studies across different domains

transformative potential within the field and its implications for future healthcare practices, encompassing subtyping analysis and precision medicine, areas that our study may not have probed as deeply. Additionally, while Yang et al. [23], advocate for AI's fluid integration into healthcare, our study may have steered toward a more granular investigation of AI's specific performance metrics and applications.

Correspondingly, we find commonality with Yan et al. [24], in the assertion that the integration of unstructured clinical text with structured data can significantly enhance the performance of machine learning models in detecting sepsis early on. The necessity for comprehensive datasets that amalgamate various clinical data modalities is a recurring theme in both studies. On the other hand, Yan et al. [24], focus intently on the utilization of unstructured clinical text and how it bolsters the predictive capabilities of AI models. They also discuss the lack of inclusion of longitudinal patient data extending beyond the current care episode, which may contrast with our study if we have included such temporal data. Moreover, Yan et al. [24], discuss the issues surrounding the application of models developed in ICU settings to general wards, an area that may differ from our study's findings if our focus was on a narrower AI application or a different care setting.

Comparing our findings with the review by Hassan et al. [25], there is a synergy in our mutual exploration of the predictors used in AI algorithms for predicting sepsis, emphasizing the import of specific clinical features. This synergy allows us to juxtapose the average sensitivity and specificity metrics from their review against the performance of our study's AI models. The divergence could be evident if our study also evaluated the impact of different predictors on the models' predictive timeframes and power. Should our study have employed a novel set of predictors or adopted alternative AI techniques, this would mark a salient departure from the findings of Hassan et al. [25], who provide a meta-analysis of sensitivity and specificity across multiple studies and discuss the influence of predictor types on predictive outcomes.

Contemporary protocols for the management of sepsis underscore the imperative of prompt therapeutic interventions, while an overarching principle remains the axiom that prophylactic measures are superior to remedial strategies. In this context, an array of AI frameworks have been architected to prognosticate the advent of sepsis [26]. Longitudinal analyses have substantiated that AI-enabled surveillance of patient clinical streams can presage sepsis with a temporal lead, achieving predictive accuracies that approach the 90th percentile mark, thereby markedly surpassing the prognostic capacities of conventional clinical acuity indices [17], [26]–[29].

To surmount this limitation, investigative efforts have co-

alesced around the exploitation of clinical parameters that are routinely captured across diverse medical environments, including ambulatory care settings and emergency departments, yielding consistently affirmative scholarly outcomes [30]–[32]. Syntheses of extant literature have illuminated that AI-facilitated early warning systems manifest a heightened impact within the precincts of emergency care and general inpatient wards as opposed to intensive care contexts [33]. In an innovative departure from traditional clinical datasets, specific investigations have extended the predictive architecture by integrating biomarker data gleaned from genomic profiling, thereby fostering the construction of algorithmic models proficient in the identification of patients at elevated risk for post-surgical infections or sepsis in the initial triad of postoperative days [34].

The heterogeneity inherent in infection loci and the idiosyncratic physiological responses of individual patients pose formidable obstacles to the accurate nosological characterization of sepsis [35]. Scholarly discourse suggests that the deployment of analytical tools predicated on expansive datasets and sophisticated machine learning techniques can amplify both the sensitivity and precision of sepsis diagnostics [36]–[38]. In a departure from traditional structured clinical datasets, which typically encapsulate patient vitals and laboratory test results [39], diagnostic frameworks that integrate unstructured narrative data have shown potential. Such paradigms have been reported to augment the early diagnosis of sepsis by a third and to curtail the incidence of false-positive determinations by a sixth [37]. For instance, algorithmic systems trained on radiographic imagery such as chest X-rays have achieved diagnostic concordance for acute respiratory distress syndrome in approximately nine out of ten patients [40].

A seminal, multicentric, prospective cohort investigation has recently divulged a substantive association between the proactive deployment of AI-driven sepsis alert mechanisms and the attenuation of in-patient mortality rates, the frequency of organ dysfunction, and the duration of hospitalization [41]. These investigative findings collectively reinforce the notion that AI holds considerable promise for the stringent early detection of sepsis and the consequent enhancement of patient prognoses. Notwithstanding, extant models encapsulate but a subset of potential clinical data variables, leaving a vast expanse of pertinent medical data untapped. This recognition signals an expansive potential for refining the diagnostic acumen of AI in medical practice. For AI-derived predictive analytics to gain traction in clinical utility, the output of such models necessitates a degree of interpretability. Model predictions must be rendered in a format that is intelligible and actionable by healthcare providers, engendering their confidence in the technology and enabling the discernment of potentially anomalous predictions [42]. To this end, strides have been made in the development of interpretable AI frameworks, wherein the logic underpinning predictions is rendered transparent, accessible, and amenable to visualization [38]–[42], thus potentiating the clinical valor of such technologies.

Study-Specific Limitations

Calvert et al. [15], reported a high area under the AUROC of 0.917 for their XGBoost model, reflecting a superior diagnostic ability. The model's sensitivity and specificity, at 0.799 and 0.860 respectively, along with an accuracy of 0.848, pointed to an effective balance in detecting true positive and true negative cases. However, the study did not offer insights into the model's performance across different patient subgroups or settings, which could be relevant for its generalizability. Khojandi et al. [16], while not providing an AUROC, disclosed sensitivity and specificity rates up to 67% and 63%, respectively. These rates suggested room for improvement in the model's sensitivity and precision. The absence of an AUROC value limited a holistic understanding of the model's discriminative power.

Lauritsen et al. [17], offered a broad range for sensitivity and specificity, but the omission of AUROC and accuracy values precluded a comprehensive evaluation of the model's overall effectiveness. Without these metrics, the relative performance of the model compared to other benchmarks remained unclear. Lind et al. [18] presented an AUROC of 0.85, indicative of reliable model performance. Nonetheless, the lack of sensitivity, specificity, and accuracy metrics restricted a thorough assessment of the model's diagnostic utility, particularly in terms of its ability to correctly classify individual cases.

The study by Lu et al. [19], reported an AUROC of 0.799, suggesting a moderate to high capability in distinguishing between sepsis cases and non-cases. The sensitivity and specificity values indicated a propensity for the model to miss some true positive cases while correctly identifying a majority of true negatives. A detailed analysis of the circumstances leading to missed cases and false positives was not provided, which could be essential for clinical application. Mao et al. [20] reported a high AUROC of 0.92, signalling excellent model performance. However, the lack of additional performance metrics such as sensitivity, specificity, and accuracy did not allow for a full evaluation of the model's diagnostic capabilities, particularly in different clinical contexts or patient populations.

Nemati et al. [21], revealed an AUROC range between 0.83 and 0.85, consistent sensitivity of 85%, and a specificity range from 64% to 72%. While these findings suggested a strong capacity for identifying true positive cases, the variability in specificity underscored the need for refinement in reducing false positives. Tang et al. [22], reported an outstanding AUROC of 0.9213, with high sensitivity and good specificity, underscoring the model's potential in timely and accurate sepsis detection. However, the study focused on a specific patient population-those with COVID-19-which might limit the applicability of the findings to the broader population of patients at risk for sepsis.

Recommendations

Based on our findings, the following recommendations can be made to enhance the application of artificial intelligence models in the detection of sepsis;

1) Models demonstrating high AUROC values, such as those reported in several studies, should be further validated in diverse clinical settings to confirm their robustness and generalizability. This is especially critical for models that have shown superior diagnostic capabilities with balanced sensitivity and specificity.
2) For models with lower sensitivity and specificity rates, efforts should be directed toward improving their diagnostic accuracy. This may involve refining algorithms, incorporating additional relevant features, or employing more advanced machine learning techniques.
3) The importance of reporting a full set of performance metrics, including AUROC, sensitivity, specificity, and accuracy, cannot be overstated. Future research should ensure these metrics are included to provide a comprehensive evaluation of the model's performance.
4) Models that show variability in performance metrics suggest the potential need for customization or adjustment according to specific clinical scenarios or patient populations.
5) Continuous improvement should be pursued, especially for models that show a high rate of false positives or false negatives. This could involve iterative training with larger datasets, cross-validation across different patient cohorts, and refinement of feature selection processes.
6) Given the promising results of AI models in identifying sepsis among specific populations, such as COVID-19 patients, further research should explore the applicability of these models to other patient groups at risk for sepsis.
7) The integration of AI models into clinical workflows should be done cautiously, with ongoing monitoring and evaluation to ensure that they support, rather than hinder, clinical decision-making processes. The ultimate goal is to leverage AI to improve patient outcomes while maintaining patient safety and care quality.

5. Conclusion

The findings from the array of studies included in the review consistently demonstrated that AI models can achieve high discriminative performance, as evidenced by generally AUROC values across the board. These models varied in terms of sensitivity and specificity, with some achieving a commendable balance, thereby potentially offering significant clinical value in identifying true sepsis cases while minimizing false positives. It was observed that the models' performance was not uniform, with certain AI approaches displaying excellent predictive abilities in specific settings, such as within the critical initial hours of patient admission. However, the variability in sensitivity and specificity among the different models highlighted the intricate nature of the sepsis detection challenge. AI models tended to differ in their ability to generalize across diverse clinical scenarios, which could be attributed to differences in patient populations, data heterogeneity, and the specificities of the algorithms employed. The studies suggested that the integration of AI into clinical workflows could potentially enhance the timely identification and treatment of sepsis, which is crucial for improving patient outcomes. Nevertheless, the noted variability and the absence of comprehensive performance metrics in some studies underscore the need for standardization in reporting and further validation of these AI tools. Future research should focus on addressing these gaps by refining AI algorithms for improved accuracy, sensitivity, and specificity, and by conducting rigorous evaluations of their real-world clinical impact.

Acknowledgment

The authors extend their appreciation to the Deanship of Scientific Research-Majmaah University for supporting this research work under project number R-2023-887.

Conflict of Interest

The authors declare no conflict of interests. All authors read and approved final version of the paper.

Authors Contribution

All authors contributed equally in this paper.

References

[1] Peiffer-Smadja, N., Delliere, S., Rodriguez, C., Birgand, G., Lescure, F. X., Fourati, S., & Ruppe, E. (2020). Machine learning in the clinical microbiology laboratory: Has the time come for routine practice? Clinical Microbiology and Infection , 26(10), 1300-1309.

[2] Fleischmann, C., Scherag, A., Adhikari, N. K., Hartog, C. S., Tsaganos, T., Schlattmann, P., ... & Reinhart, K. (2016). International forum of acute care trialists. Assessment of Global Incidence and Mortality of Hospital-treated Sepsis, 193(3), 259-272.

[3] Rivert, E., Nguyen, B., & Havstad, S. (2001). Early goal-directed therapy in the treatment of severe sepsis and septic shock. New England Journal of Medicine, 345(19), 1368-1377. 1368-1377.

[4] Kumar, A., Roberts, D., Wood, K. E., Light, B., Parrillo, J. E., Sharma, S., ... & Cheang, M. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Critical Care Medicine, 34(6), 1589-1596.

[5] Israelachvili, J. (1997). The different faces of poly (ethylene glycol). Proceedings of the National Academy of Sciences, 94(16), 8378-8379.

[6] Iskander, K. N., Osuchowski, M. F., Stearns-Kurosawa, D. J., Kurosawa, S., Stepien, D., Valentine, C., & Remick, D. G. (2013). Sepsis: multiple abnormalities, heterogeneous responses, and evolving understanding. Physiological Reviews, 93(3), 1247-1288.

[7] Jawad, I., Luksic, I., Rafnsson, S. B. (2012). Assessing available information on the burden of sepsis: Global estimates of incidence, prevalence and mortality. Journal of Global Health , 2(1), 010404.

[8] Islam, M. M., Nasrin, T., Walther, B. A., Wu, C. C., Yang, H. C., & Li, Y. C. (2019). Prediction of sepsis patients using machine learning approach: a meta-analysis. Computer Methods and Programs in Biomedicine, 170, 1-9.

[9] Schinkel, M., Paranjape, K., Panday, R. N., Skyttberg, N., & Nanayakkara, P. W. (2019). Clinical applications of artificial intelligence in sepsis: a narrative review. Computers in Biology and Medicine, 115, 103488.

[10] Wulff, A., Montag, S., Marschollek, M., & Jack, T. (2019). Clinical decision-support systems for detection of systemic inflammatory response syndrome, sepsis, and septic shock in critically ill patients: a systematic review. Methods of Information in Medicine, 58(S 02), e43-e57.

[11] Teng, A. K., Wilcox, A. B. (2020). A review of predictive analytics solutions for sepsis patients. Applied Clinical Informatics , 11(3), 387-398.

[12] Fleuren, L. M., Klausch, T. L., Zwager, C. L., Schoonmade, L. J., Guo, T., Roggeveen, L. F., ... & Elbers, P. W. (2020). Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Medicine, 46, 383-400.

[13] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. International Journal of Surgery, 88, 105906.

[14] Sterne, J. A., Hernan, M. A., Reeves, B. C., Savovic, J., Berkman, N. D., Viswanathan, M., ... & Higgins, J. P. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ, 355, i4919.

[15] Calvert, J., Saber, N., Hoffman, J., Das, R. (2019). Machine-Learning-Based Laboratory Developed Test for the Diagnosis of Sepsis in High-Risk Patients. Diagnostics , 9, 20.

[16] Khojandi, A., Tansakul, V., Li, X., Koszalinski, R. S., Paiva, W. (2018). Prediction of Sepsis and In-Hospital Mortality Using Electronic Health Records. Methods of Information in Medicine , 57(4), 185-193.

[17] Lauritsen, S. M., Kalor, M. E., Kongsgaard, E. L., Lauritsen, K. M., Jorgensen, M. J., Lange, J., & Thiesson, B. (2020). Early detection of sepsis utilizing deep learning on electronic health record event sequences. Artificial Intelligence in Medicine, 104, 101820.

[18] Lind, M. L., Mooney, S. J., Carone, M., Althouse, B. M., Liu, C., Evans, L. E., ... & Phipps, A. I. (2021). Development and validation of a machine learning model to estimate bacterial sepsis among immunocompromised recipients of stem cell transplant. JAMA Network Open, 4(4), e214514-e214514.

[19] Lu, H. X., Du, J., Wen, D. L., Sun, J. H., Chen, M. J., Zhang, A. Q., & Jiang, J. X. (2019). Development and validation of a novel predictive score for sepsis risk among trauma patients. World Journal of Emergency Surgery, 14(1), 1-8.

[20] Mao, Q., Jay, M., Hoffman, J. L., Calvert, J., Barton, C., Shimabukuro, D., ... & Kerem, Y. (2015). Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open, 44(5), 390-396.

[21] Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., Buchman, T. G. (2018). An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. Critical Care Medicine , 46(4), 547-553.

[22] Tang, G., Luo, Y., Lu, F., Li, W., Liu, X., Nan, Y., ... & Sun, Z. (2021). Prediction of sepsis in COVID-19 using laboratory indicators. Frontiers in Cellular and Infection Microbiology, 10, 586054.

[23] Yang, Jie & Hao, Sicheng & Huang, Jiajie & Chen, Tianqi & Liu, Ruoqi & Zhang, Ping & Feng, Mengling & He, Yang & Xiao, Wei & Hong, Yucai & Zhang, Zhongheng. (2023). The application of artificial intelligence in the management of sepsis. Medical Review. 3. 10.1515/mr-2023-0039.

[24] Yan, M. Y., Gustad, L. T., & Nytro, O. (2022). Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review. Journal of the American Medical Informatics Association, 29(3), 559-575.

[25] Hassan, N., Slight, R., Weiand, D., Vellinga, A., Morgan, G., Aboushareb, F., & Slight, S. P. (2021). Preventing sepsis; how can artificial intelligence inform the clinical decision-making process? A systematic review. International Journal of Medical Informatics, 150, 104457.

[26] Islam, M. M., Nasrin, T., Walther, B. A., Wu, C. C., Yang, H. C., Li, Y. C. (2019). Prediction of sepsis patients using machine learning approach: A meta-analysis. Computational Methods and Programs in Biomedicine , 170, 1-9.

[27] Schinkel, M., Paranjape, K., Panday, R. S. N., Skyttberg, N., Nanayakkara, P. W. B. (2019). Clinical applications of artificial intelligence in sepsis: A narrative review. Computers in Biology and Medicine , 115, 103488.

[28] Mollura, M., Lehman, L.-W. H., Mark, R. G., Barbieri, R. (2021). A novel artificial intelligence-based intensive care unit monitoring system: Using physiological waveforms to identify sepsis. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences , 379, 20200252.

[29] Chang, Y. H., Hsiao, C. T., Chang, Y. C., Lai, H. Y., Lin, H. H., Chen, C. C., ... & Cho, D. Y. (2023). Machine learning of cell population data, complete blood count, and differential count parameters for early prediction of bacteremia among adult patients with suspected bacterial infections and blood culture sampling in emergency departments. Journal of Microbiology, Immunology and Infection, 56, 782-792.

[30] Goh, K. H., Wang, L., Yeow, A. Y. K., Poh, H., Li, K., Yeow, J. J. L., & Tan, G. Y. H. (2021). Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. Nature communications, 12(1), 711.

[31] Wang, D., Li, J., Sun, Y., Ding, X., Zhang, X., Liu, S., ... & Sun, T. (2021). A machine learning model for accurate prediction of sepsis in ICU patients. Frontiers in Public Health, 9, 754348.

[32] Fleuren, L. M., Klausch, T. L., Zwager, C. L., Schoonmade, L. J., Guo, T., Roggeveen, L. F., ... & Elbers, P. W. (2020). Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Medicine, 46(3), 383-400.

[33] Shashikumar, S. P., Wardi, G., Malhotra, A., Nemati, S. (2021). Artificial intelligence sepsis prediction algorithm learns to say "I don't know". NPJ Digital Medicine , 4, 134.

[34] Zhang, Z., Chen, L., Xu, P., Wang, Q., Zhang, J., Chen, K., ... & Hong, Y. (2022). Effectiveness of automated alerting system compared to usual care for the management of sepsis. NPJ Digital Medicine, 5(1), 101.

[35] Delahanty, R. J., Alvarez, J., Flynn, L. M., Sherwin, R. L., Jones, S. S. (2019). Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. Annals of Emergency Medicine , 73(4), 334-344.

[36] Hassan, N., Slight, R., Weiand, D., Vellinga, A., Morgan, G., Aboushareb, F., & Slight, S. P. (2021). Preventing sepsis; how can artificial intelligence inform the clinical decision-making process? A systematic review. International Journal of Medical Informatics, 150, 104457.

[37] Sjoding, M. W., Taylor, D., Motyka, J., Lee, E., Claar, D., McSparron, J. I., ... & Gillies, C. E. (2021). Deep learning to detect acute respiratory distress syndrome on chest radiographs: a retrospective study with external validation. The Lancet Digital Health, 3(6), e340-e348.

[38] Adams, R., Henry, K. E., Sridharan, A., Soleimani, H., Zhan, A., Rawat, N., ... & Saria, S. (2022). Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. Nature Medicine, 28(7), 1455-1460.

[39] Wang, F., Kaushal, R., & Khullar, D. (2020). Should health care demand interpretable artificial intelligence or accept "black box" medicine?. Annals of Internal Medicine, 172(1), 59-60.

[40] Lauritsen, S.M., Kristensen, M., Olsen, M.V. et al. Explainable artificial intelligence model to predict acute critical illness from electronic health records. Nat Commun 11, 3852 (2020).

[41] Li, X., Xu, X., Xie, F., Xu, X., Sun, Y., Liu, X., ... & Xie, G. (2020). A time-phased machine learning model for real-time prediction of sepsis in critical care. Critical Care Medicine, 48(10), e884-e888.

[42] Yang, M., Liu, C., Wang, X., Li, Y., Gao, H., Liu, X., & Li, J. (2020). An explainable artificial intelligence predictor for early detection of sepsis. Critical Care Medicine, 48(11), e1091-e1096.