machine learning & knowledge extraction

MDPI

Review

AI Advances in ICU with an Emphasis on Sepsis Prediction: An Overview

Charithea Stylianides [1,*], Andria Nicolaou [1], Waqar Aziz Sulaiman [1], Christina-Athanasia Alexandropoulou [2], Ilias Panagiotopoulos [2], Konstantina Karathanasopoulou [2], George Dimitrakopoulos [2], Styliani Kleanthous [1,3], Eleni Politi [2], Dimitris Ntalaperas [4], Xanthi Papageorgiou [4], Fransisco Garcia [5], Zinonas Antoniou [5], Nikos Ioannides [6], Lakis Palazis [6], Anna Vavlitou [6], Marios S. Pattichis [7], Constantinos S. Pattichis [1] and Andreas S. Panayides [1]

1 Centre of Excellence, CYENS, Nicosia 1016, Cyprus; a.nicolaou@cyens.org.cy (A.N.); waqarazizg@gmail.com (W.A.S.); styliani.kleanthous@ouc.ac.cy (S.K.); pattichis.costas@ucy.ac.cy (C.S.P.); a.panayides@cyens.org.cy (A.S.P.)
2 Department of Informatics and Telematics, Harokopio University of Athens, 176 76 Kallithea, Greece; calexand@hua.gr (C.-A.A.); ipanagio@hua.gr (I.P.); kkarathanasopoulou@hua.gr (K.K.); gdimitra@hua.gr (G.D.); politie@hua.gr (E.P.)
3 Faculty of Pure and Applied Sciences, Open University of Cyprus, Latsia 2220, Cyprus
4 UBITECH Limited, Limassol 3071, Cyprus; dimitris.ntalaperas@gmail.com (D.N.); xpapageorgiou@ubitech.eu (X.P.)
5 Research & Development Department, 3aHealth, Strovolos 2020, Cyprus; f.garcia@3ahealth.com (F.G.); z.antoniou@3ahealth.com (Z.A.)
6 State Health Services Organization, Aglantzia 2100, Cyprus; n.ioannides@shso.org.cy (N.I.); l.palazis@shso.org.cy (L.P.); a.vavlitou@shso.org.cy (A.V.)
7 Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87106, USA; pattichi@unm.edu
* Correspondence: c.stylianides@cyens.org.cy

Abstract: Artificial intelligence (AI) is increasingly applied in a wide range of healthcare and Intensive Care Unit (ICU) areas to serve—among others—as a tool for disease detection and prediction, as well as for healthcare resources' management. Since sepsis is a high mortality and rapidly developing organ dysfunction disease afflicting millions in ICUs and costing huge amounts to treat, the area can benefit from the use of AI tools for early and informed diagnosis and antibiotic administration. Additionally, resource allocation plays a crucial role when patient flow is increased, and resources are limited. At the same time, sensitive data use raises the need for ethical guidelines and reflective datasets. Additionally, explainable AI is applied to handle AI opaqueness. This study aims to present existing clinical approaches for infection assessment in terms of scoring systems and diagnostic biomarkers, along with their limitations, and an extensive overview of AI applications in healthcare and ICUs in terms of (a) sepsis detection/prediction and sepsis mortality prediction, (b) length of ICU/hospital stay prediction, and (c) ICU admission/hospitalization prediction after Emergency Department admission, each constituting an important factor towards either prompt interventions and improved patient wellbeing or efficient resource management. Challenges of AI applications in ICU are addressed, along with useful recommendations to mitigate them. Explainable AI applications in ICU are described, and their value in validating, and translating predictions in the clinical setting is highlighted. The most important findings and future directions including multimodal data use and Transformer-based models are discussed. The goal is to make research in AI advances in ICU and particularly sepsis prediction more accessible and provide useful directions on future work.

Academic Editor: Francesco Flammini

Received: 13 September 2024
Revised: 9 December 2024
Accepted: 22 December 2024
Published: 8 January 2025

Citation: Stylianides, C.; Nicolaou, A.; Sulaiman, W.A.; Alexandropoulou, C.-A.; Panagiotopoulos, I.; Karathanasopoulou, K.; Dimitrakopoulos, G.; Kleanthous, S.; Politi, E.; Ntalaperas, D.; et al. AI Advances in ICU with an Emphasis on Sepsis Prediction: An Overview. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 6. https://doi.org/10.3390/make7010006

Keywords: artificial intelligence; emergency department; ethical AI; explainable AI; hospitalization prediction; Intensive Care Unit (ICU); ICU length of stay prediction; sepsis detection; sepsis prediction

1. Introduction

Hospitals and Intensive Care Units (ICUs) continuously manage large volumes of multimodal data. As data are becoming bigger and heterogeneous, dealing with diseases in a timely and informed manner becomes much more complex, especially in the ICU context, where the health status of vulnerable patients can deteriorate rapidly and significantly. The increased complexity associated with disease detection, risk assessment, and personalized treatment according to patient characteristics, can result in suboptimal treatment, yielding delays and complications, which in turn increase length of ICU/hospital stay. The consequences extend beyond the unique treatment plan and outcomes of a single patient, dramatically increasing healthcare expenditures and thus affecting healthcare management and impacting the society economically [1].

Machine Learning (ML) and Artificial intelligence (AI) methods support the development of personalized healthcare systems. Personalized AI algorithm predictions in healthcare refer to the probability of developing a disease, detecting a yet undiagnosed disease, and predicting the prognosis of a current treatment, all considering the patient's unique characteristics towards selecting the optimum treatment plan. This enables the algorithms to predict outcomes for new patients with different clinical characteristics and demographics [2]. Validated algorithms can eventually be used by healthcare professionals, assisting them in decision-making regarding treatment interventions. As a result, increased quality of care and reduced length of stay and rehospitalizations can benefit society socially with increased Quality Adjusted Life Year(s) (QALY), as well as economically, due to decreased healthcare expenditures.

At the same time, healthcare research involves vulnerable populations and is susceptible to potential biases, causing ethical issues and data transparency challenges. These issues are entailed in the AI Clinical Decision Support Systems (CDSS) and highlight the necessity for communicating critical information to the user (the doctor and then, the patient). Explainability or the area of Explainable Artificial Intelligence (XAI) [3,4] aims to achieve confidence, trustworthiness, accessibility, causality, and transferability in predictions, so that health professionals can understand and correlate the results with the clinical practice [5,6].

AI is increasingly used in healthcare in the investigation of sicknesses that are hard to diagnose [7], including cancer, diabetes, cardiovascular diseases, COVID-19, etc., by identifying the most important factors for patients' risk prediction and/or diagnosis. AI predictions can affect several facets of cancer therapy, including drug discovery, development, and clinical validation. Moreover, studies use AI for predicting a diverse set of pathologies, such as the diagnosis of ovarian tumors [8] and the survival of patients with ovarian cancer [9], early breast cancer prediction and diagnosis [10,11], dermatological cancer recognition [12], and lung cancer diagnosis [13]. Additionally, AI methods are cost-effective for reducing ophthalmic complications and preventable blindness associated with diabetes [14]. Studies have further applied AI for the classification of patients into diabetic or non-diabetic [15,16], the risk prediction of developing type 2 diabetes [17,18], diabetes diagnosis [19], and blood glucose level predictions [20]. Regarding cardiovascular diseases, AI has been used to assess the risk of developing a cardiovascular disease [21–27] and identify heart rate severity [28,29]. ML techniques have been used to predict whether

patients are infected by COVID-19 [30], identify its spread [31,32], predict the number of discharged patients and deaths [33], and predict COVID-19 mortality [34,35] to prioritize triage and hospitalization [35]. In any healthcare outcome, the goal is early interventions towards disease prevention or treatment, and, thus, improved patient wellbeing and reduced healthcare costs.

The research community is showing an increasing interest in investigating the effect of AI models in healthcare and ICU and adopted AI systems are already revolutionizing healthcare practice. While similar review studies exist, they focus on a single ICU outcome [36–43], explore studies that use a single type of data [36,38,42], do not elaborate on clinical approaches used in ICU [36–44], and/or do not significantly guide towards future work [36,37,43].

This paper aims to present an overview of clinical approaches for disease assessment (Section 2), and the main applications of ML and DL in healthcare and ICU focusing on predominant clinical outcomes (Section 3). Specifically, the study is focused on AI applications for (a) sepsis prediction/detection and sepsis mortality prediction, as sepsis is a high mortality and high morbidity disease, (b) length of ICU/hospital stay predictions as an indicator of disease severity but also a burden on (human) resource management, and (c) ICU admission/hospitalization probability after emergency department (ED) admission for optimal triage and resource planning. Importantly, critical challenges of AI applications in ICU (Section 4) and the role of XAI (Section 5) are also described. Future work guidance is given in Section 6 and conclusions are made in Section 7.

The main contributions of the study are:

- A holistic overview of existing clinical approaches and AI approaches, both applicable to sepsis prediction and mortality due to sepsis, along with a comparison of their performance,
- An overview of AI approaches in predicting length of ICU/hospital stay and ICU admission/hospitalization probability after ED admission,
- A summary of the most critical challenges of AI applications in ICU with important suggestions on how to address them,
- A summary of Explainable AI methods and an overview of their current applications in healthcare and ICU research,
- Future guidance in healthcare/ICU research based on findings of the study and associate AI advances.

2. Clinical Approaches

This section provides an overview of existing widely used clinical approaches in ICU, in terms of (a) diagnostic biomarkers for infection and sepsis detection, and (b) scoring systems for organ dysfunction detection (that leads to sepsis), mortality prediction including sepsis mortality, and length of stay in the ICU for disease severity assessment.

2.1. Diagnostic Biomarkers

Biomarkers can provide important diagnostic information associated with inflammation and/or infection. To avoid clinical biases applied to the diagnosis of infection as part of 'clinical gestalt' [45–48], biomarkers such as C-Reactive Protein (CRP), Interleukin-6 (IL-6), Lactate Dehydrogenase (LDH), Procalcitonin (PCT) and White Blood Cell Count (WBC) are commonly used in clinical practice. Table 1 summarizes how these biomarkers are used as for infection detection, their usage, associations, and their corresponding diagnosis thresholds according to the literature [49–83]. In what follows, we summarize findings for the most important biomarkers.

C-Reactive Protein (CRP) is used as a biomarker for Systemic Inflammatory Response Syndrome (SIRS), infection and sepsis [49], for the diagnosis of neonatal sepsis [50], and for the differential diagnosis of bacterial versus viral infections [56,57] and their early identification [58,59]. It is a component of the International Patient Summary (IPS) [51,52] and of ICU prognostic blood tests [53]. CRP has been proven to be a predictor of ICU mortality when more than 62.8 mg/L [55] and of severe COVID-19 in patients below 50 years old [49]. CRP can be used as a valuable tool to monitor progress as it responds to therapy against inflammation [54]. It is associated with ICU-acquired infection [51], hospital readmission in patients with heart failure [60], ICU readmission, unexpected mortality after ICU discharge [61–63], as well as ICU mortality [55] and ICU mortality of COVID-19 patients [67]. CRP is also associated with non-infection related aspects, like allergic complications [54], specific drug overdoses [54], obesity, smoking, diabetes mellitus, lack of exercise, hormonal therapy [64], and some hematological therapies [65]. It is also associated with therapeutic interventions (CT-scan, ultrasonography) and flexible endoscopy or (re) laparotomy/thoracotomy in the ICU general surgical population [66]. Higher values of CRP are associated with age below 50 years for predicting COVID-19 [49]. In healthy individuals, the median CRP has been proven to be 0.8 mg/L [69]. A series of studies have identified informative CRP thresholds, as depicted in Figure 1.

Interleukin-6 (IL-6) is used as an inflammation biomarker in septic and non-septic patients [74] and as a predictor of disease severity [75]. Compared to CRP, it is proven that IL-6 can better predict disease severity [75] and it is more associated with organ dysfunction and the need for organ support therapies, like vasopressors/inotropes and/or Renal Replacement Therapy (RRT) [74]. It is also associated with the ability of Simplified Acute Physiologic Score (SAPS) II and Sequential (Sepsis-related) Organ Failure Score (SOFA) to predict 90-day mortality of critically ill patients, with or without sepsis [74]. In healthy individuals, IL-6 ranges from 0 to 7 pg/mL [76], while IL-6 of more than 1 µg/mL is indicative of septic shock [76]. For COVID-19 patients, IL-6 of $\geq$74.98 pg/mL on ICU admission is a predictor of in-hospital mortality [73] (see Figure 1).

Lactate dehydrogenase (LDH) is a marker of COVID-19 virus for all ages [49] and a predictor of severe COVID-19 in patients above 50 years when combined with CRP [49]. However, it is associated with COVID-19 irrespective of age and gender [49], and with severe lung infections [77].

Procalcitonin (PCT) is used as a test for early identification of infections [58,59] and identification of infections in ICU patients [78]. It is also considered a better mortality predictor than CRP [78]. It is associated with both infection and inflammation [48], viral and bacterial infections, and sepsis [79], as well as with the severity of infection and risk of death [78]. When PCT is more than 2 ng/mL, the diagnostic specificity is improved and confirmation of treatment requirement for extracellular bacterial infection is aided [80]. A PCT equal or above 0.56 ng/mL on ICU admission can be a predictor of in-hospital mortality for COVID-19 patients [73] (see Figure 1).

White Blood Cell Count (WBC) can also be used as a Biomarker of infection [48] and is associated with immature granulocytes [81], Neutrophil-Lymphocyte Ratio (NLR) [82,83], non-infectious mimics, drugs, and comorbidities [48].

Table 1. Biomarkers for Infection Detection.

	Usage	Associations	Thresholds
CRP	- SIRS, infection, sepsis biomarker [49] - Neonatal sepsis diagnosis [50] - Diagnosis of bacterial versus viral infections [56,57] - CRP test for early identification of infections [58,59] - IPS component [51] - ICU prognostic blood test component [53] - ICU mortality prediction when combined with APACHE II [55] - Predictor of severe COVID-19 when increased in patients below 50 years [49] - Responds to therapy against inflammation [54]	- SIRS, infection, sepsis [49] - Neonatal sepsis [50] - ICU-acquired infection [51] - Hospital readmission in patients with heart failure [60] - Increased risk of ICU readmission [61–63] - Unexpected mortality after ICU discharge [61–63] - Increased ICU mortality risk [55] - ICU mortality of COVID-19 patients [67] - Age below 50 years of COVID-19 patients [49] - Allergic complications of infections, necrosis, trauma, malignancy conditions [54] - Specific drug overdoses [54] - Obesity, smoking, diabetes mellitus, lack of exercise, hormonal therapy [64] - Some hematological therapies [65] - Therapeutic interventions (CT-scan, ultrasonography) and flexible endoscopy or (re) laparotomy/thoracotomy in the ICU general surgical population [66]	- Healthy: 0.8 mg/L (median) [68] - Inadequate or inappropriate therapy: 22 mg/L [69,70] - Increased risk of ICU readmission, unexpected mortality after ICU discharge: >100 mg/L on the day of discharge [61–63] - Sepsis in patients with trauma three days after trauma: >200 mg/L [71] - Increased probability of ICU mortality: CRP > 62.8 mg/L at ICU admission [55] - Increased risk of ICU readmission and in-hospital mortality in patients with a LOS in ICU of >48 h: CRP $\geq$ 75 mg/L within 24 h before ICU discharge [72] - Diagnosis of bacterial versus viral infections in ICU patients: increase of >41 mg/L from previous days [56,57] - In-hospital mortality predictor of COVID-19 patients: CRP $\geq$ 81 mg/L on ICU admission [73]
IL-6	- Inflammation biomarker in septic and non-septic patients [74] - Predictor of disease severity (better compared to CRP) [75]	- Inflammation in septic and non-septic patients [74] - Ability of SAPS-II or SOFA to predict 90-day mortality in critically ill patients (both septic and non-septic) [74] - Organ dysfunction and need for organ-support therapies, like vasopressors/inotropes and/or RRT (higher association than CRP) [74]	- Healthy: 0 to 7 pg/mL [76] - Septic shock: more than 1 µg/mL [76] - In-hospital mortality predictor of COVID-19 patients: $\geq$74.98 pg/mL on ICU admission [73]
LDH	- Marker of COVID-19 for all ages [49] - Predictor of severe COVID-19 in patients above 50 years when combined with CRP [49]	- Severe lung infections [77] - COVID-19 irrespective of age and gender [49]	
PCT	- PCT test for early identification of infections [58,59] - Used in ICUs to identify infection in patients [78] - Mortality predictor (better than CRP) [78]	- Infection and inflammation [48] - Viral and bacterial infections, sepsis [79] - Severity of infection and risk of death [78]	- Treatment requirement for extracellular bacterial infection: >2 ng/mL [80] - In-hospital mortality predictor of COVID-19 patients: $\geq$0.56 ng/mL on ICU admission [73]
WBC	- Biomarker of infection [48]	- Immature granulocytes [81] - NLR [82,83] - Non-infectious mimics, drugs, and comorbidities [48]	

Refer to Glossary section for acronyms.

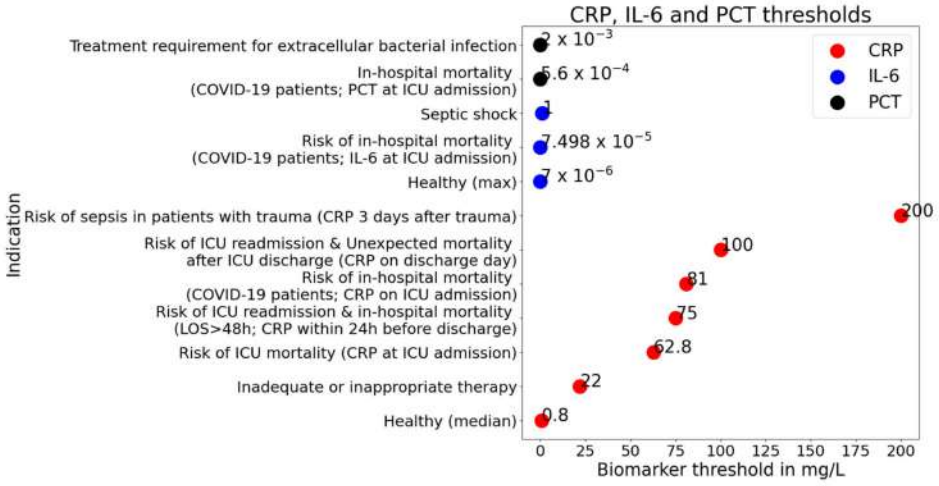

Figure 1. Diagnostic biomarker thresholds according to literature [61–63,68–73,76,80].

2.2. Scoring Systems

Healthcare professionals use predictive scoring systems that describe, assess, and compare the severity of a disease [84], usually as a numerical score, by predicting outcomes like length of stay and mortality rates in patients mainly in ICUs [85]. They are also used to evaluate therapeutic interventions in patients with acute respiratory distress syndrome or sepsis [86,87], and for benchmarking of ICU performance and improvement of quality of care [84]. Commonly used models include the Acute Physiologic and Chronic Health Evaluation (APACHE), the Sequential (Sepsis-related) Organ Failure Score (SOFA), the Simplified Acute Physiologic Score (SAPS) and the Mortality Predictive Model (MPM). An overview of these scoring systems is presented in Table 2.

APACHE is used to assess mortality risk and length of ICU stay. It has the advantage of being widely validated which makes it reproducible and it includes the location of treatment variable not found in other scoring systems. However, it cannot handle comorbidities and includes dynamic parameters that can easily change. Additionally, APACHE III can be applied anytime during ICU stay, while the latest version, APACHE IV, is disease specific considering 115 diseases. The complexity of APACHE IV raises the requirement for additional software and increased costs. APACHE IV is also validated only in the US. Additionally, APACHE II excludes patients below or at the age of 16 years old, and those with burn injuries, coronary artery disease, or a history of cardiac surgery [85,88–95].

SOFA was originally used to understand and describe organ disfunction and the complications caused by it in critically ill patients with sepsis [88,96]. Later it had been validated for use in critically ill patients with non-sepsis related organ disfunction and as a method for predicting mortality rate [88,96]. It is also used in the definition of sepsis [97], while its 'derivative', namely Quick-SOFA, is used as a screening tool to identify sepsis in patients [84]. It does not consider chronic illnesses [84–86,88,96–100].

SAPS assesses mortality risk and can be used to compare resources amongst ICUs. SAPS II was validated only in North America and Europe and excluded patients below or at the age of 16 years old, as well as those with burn injuries, coronary artery disease, or a history of cardiac surgery [84,85,88,94,101–103]. The latest version, SAPS III, is widely validated. SAPS III considers diagnoses, and reflects early severity.

MPM assesses mortality in the form of a probability instead of a score like other scoring systems and it requires less physiological data compared to other scoring systems. However, it excludes cardiac surgery and myocardial infraction patients and MPM II also excludes patients below or at the age of 16 years old, and those with burn injuries and coronary artery disease [57,88,104].

Table 2. Main ICU Scoring Systems Overview.

Scoring System	Outcome	Version	Variables (Obs. Window)	Score	Advantages	Disadvantages
APACHE [85,88–95]	- Mortality - ICU LOS	II	13 + medical history + surgical requirements (first 24 h in ICU)	0–71	- Reproducible/widely validated - More than 1 outcome - Includes location of treatment variable - APACHE III: Applied anytime during ICU stay - APACHE IV: Disease specific (115 diseases)	- Cannot handle comorbidities - Includes dynamic parameters that can be affected - APACHE II: Excludes patients $\leq$ 16 years old, those with burn injuries, coronary artery disease, or a history of cardiac surgery - APACHE IV: Complex, Requires software, Added costs, Validated only in the US
		III	17 (applied anytime in ICU)	0–299		
		IV	142 + 115 disease groups (first 24 h in ICU)	0–286		
SOFA [84–86,88,96–100]	- Organ Disfunction (both septic and non-septic patients) - Sepsis Mortality	-	6 (applied anytime in ICU)	0–4 for each	- Can be used to monitor response to therapy - Used in sepsis definition - Derivative: Quick-SOFA screening tool for sepsis	- Does not consider chronic illnesses
SAPS [84,85,88,94,101–103]	-Mortality	II	17 (first 24 h in ICU)	0–163	- Can be used to compare resources amongst ICUs - SAPS III: Reproducible/widely validated, considers diagnoses, reflects early severity	- SAPS II: Validated only in North America and Europe - SAPS II: Excludes patients $\leq$ 16 years old, those with burn injuries, coronary artery disease, or a history of cardiac surgery
		III	20 (first hour in ICU)	0–217		
MPM [57,88,104]	-Mortality	II	13 (on admission, first 24 h in ICU)	-	- Less physiological data required compared to other scoring systems	- Cardiac surgery and myocardial infraction patients are excluded - MPM II: Excludes patients $\leq$ 16 years old, those with burn injuries, coronary artery disease
		III	13 (preceding 24 h of first 48 h/first 72 h in ICU)	-		

Refer to Glossary section for acronyms.

3. AI Approaches in Healthcare and ICU

This section presents an extensive literature review on how AI is applied to predict (a) sepsis and mortality due to sepsis, (b) length of stay, and (c) hospitalization/ICU admission after ED admission. Studies within each outcome are analyzed in terms of (1) prediction objective, (2) dataset and features, and (3) modelling and evaluation. A comparison of clinical and AI approaches for predicting sepsis and mortality due to sepsis is presented within the first outcome review. An overview of the selection process of the studies is depicted in Appendix A Figure A1.

3.1. Sepsis Prediction & Sepsis Mortality Prediction

AI can be used to predict sepsis. Sepsis is a life-threatening disease affecting up to 30% of ICU patients. Up to 50% of ICU mortality is due to sepsis [104]. Worldwide, an estimated 30 million people are diagnosed with sepsis in ICUs and 6 million people die from sepsis every year. In addition, the hospital's treatment costs increase every year. The study of Nemati et al. (2018) discusses that if the antibiotic treatment is delayed, the mortality is increased every hour [105]. In this context, early recognition of risk factors

and immediate clinical intervention, before any sign of clinical symptoms, are crucial for reducing mortality rates.

Early identification and immediate intervention are keys to sepsis treatment; while scoring systems and diagnostic biomarkers can be insufficient to detect or predict the response to infection and are accompanied by limitations (see Table 2; [48,104,106–109]). While bacterial infections are the most common cause of sepsis, any type of infection, including viral (e.g., influenza, COVID-19), fungal (e.g., candidiasis), and parasitic infections (e.g., malaria), can lead to sepsis. AI models can specifically differentiate between these types of infections by considering critical variations in different vital signs (e.g., temperature, respiratory rate) and lab values (e.g., CRP, PCT). We present a significant body of literature concerning the development of diagnostic and prognostic methods of sepsis through ML and DL methods. These methods intend to enable early identification of patients with any type of sepsis, so that clinicians undertake the most appropriate treatment strategy, confidently, enhancing patient prognosis [105].

An extensive literature review has been performed using GoogleScholar to search for studies published between 2019 and 2024 targeting sepsis and septic shock detection or prediction and sepsis mortality prediction by ML or DL methods. We excluded studies that were published before 2019 (6), that involved patients under 18 years old (2) or were review papers (3). An overview of the 34 final identified studies is presented in Tables 3–6. A technical supplement with open-source code for sepsis prediction is also available in [110] https://github.com/mariehane/ai-gone-astray (accessed on 19 July 2024). Another open-source pipeline that uses a range of databases to predict various clinical outcomes including sepsis is in [111] https://github.com/rvandewater/YAIB (accessed on 19 July 2024).

In what follows, we will provide detailed summaries with links to the literature. We believe our statistical summaries can serve as guides to where most of the research has been focused and for which areas remain under-researched.

3.1.1. Prediction Objective

We define sepsis based on the Sepsis-3 definition given in [98]. Most studies (19; 56%) [112–129] use the Sepsis-3 definition [98]. Most studies (21; 62%) [111,115,117–121,125,128–140] target sepsis prediction. We found that 5 studies (15%) [116,122,124,141,142] target sepsis detection (no prediction window), 4 studies (12%) [113,114,126,127] target sepsis mortality prediction, 3 studies (9%) [123,128,143] target septic shock prediction, while [144,145] target sepsis associated Acute Respiratory Distress Syndrome (ARDS) prediction and septic shock detection, respectively.

Out of 29 prediction tasks, the most frequent (6; 21%) prediction window used is 6 h [118–120,138–140]. Two studies [133,143] conclude that a shorter prediction window increases performance. A study concludes that a longer observation window increases model performance [133].

3.1.2. Dataset & Features

Most of the studies (22; 65%) [110–113,115–122,128–130,133,134,137–140,142,144–146] use data of patients in ICU. The Medical Information Mart for Intensive Care-III (MIMIC-III) [147], is the most common dataset, used by 9 (26%) studies [113–115,121,126,129,133,136,143], followed by the PhysioNet Computing in Cardiology 2019 Challenge dataset [36] used by 6 (18%) studies [118–120,122,134,138]. We note that a freely accessible pipeline for processing EHRs specifically from MIMIC-IV is provided in [148] https://github.com/eyeshoe/cop-e-cat (accessed on 19 July 2024).

In terms of features included, it is shown that kinematics features [122], free-text data [130,141] and a combination of hematological parameters [142] improve performance.

The most important features according to built-in model importance are usually vital signs (e.g., heart rate, respiratory rate) [112,123,135,136,141] and laboratory values (e.g., Platelets, lactate) [112,113,117,119,124,126,141,142,145].

Table 3. Sepsis Detection & Septic Shock Detection.

Task	Ref.	Def.	Ward	Dataset, Samples [a]	Preprocessing	Feats [b]	Obs. Window [c]	Pred. Window [c]	Best Model	Final Remarks
Sepsis Detection	[116]	Sepsis-3	ICU	TED-ICU, 1588	HMV, N, FS, FE	106	12 h	-	XGB AUC: 0.89	XGB outperforms SOFA score.
	[122]	Sepsis-3	ICU	PNCC, 15,515	FS, FE, HMV, ST, N	8	48 h	-	LSTM AUC: 0.835	Kinematics features models show higher performance than vital sign models.
	[141]	HSSC	ED Admission	CHED, 1,059,386	PCA, CB, VI	NM	First 12 h in ED	-	NLP-XGB AUC: 0.97	Free-text data improves performance. IF: vital signs, clinical notes, lab values.
	[124]	Sepsis-3	ED Admission	CMT, 8296	FS, VI	34	-	-	XGB AUC: 0.86	XGB outperforms scoring systems. IF: CRP, Sodium, Lymphocytes (%)
	[142]	ICD-10	ICU	YUSH, 7743 (patients with fever)	FS (SWT, TT, SFS, T-SNE, WL), HMV	17	-	-	LR AUC: 0.86	LR outperforms scoring systems. Combination of hematological parameters increases SEN.
Septic Shock Detection	[145]	CMS, Billing	ICU	GIRB, 45,425 (sepsis patients)	FE, FS, EDA, HMV, HTS, HO, CB (ENN, SMOTE, RU), FE, VI	15	6 h	-	RF AUC: 0.9483	Models based on CMS outperform models based on Billing definition. IF: Lactic acid

[a] Final cohort (training & test set), [b] Final number of features of best model (after feature selection, before feature engineering), [c] Of best model. Refer to Glossary section for acronyms.

Table 4. Sepsis Prediction.

Task	Ref.	Def.	Ward	Dataset, Samples [a]	Preprocessing	Feats [b]	Obs. Window [c]	Pred. Window [c]	Best Model	Final Remarks
Sepsis Prediction	[130]	ICD-10	ICU	SH, 327	CB (SMOTE)	100	-	12 h	NLP-EM AUC: 0.94	Clinical notes improve accuracy.
	[112]	Sepsis-3	ICU	ICUUS, 3596	FEX, HO, HTS, HMV, FE, S, VI	40	NM	4–48 h	NN AUC: 0.953	Online hourly prediction based on alarm. IF: Temperature, WBC, HR
	[115]	Sepsis-3	ICU	MIMIC-III, 7833	HMV, HTS, FS	20	At least 4 h	3 h	CNN AUC: 0.84	CNN outperforms clinical scoring systems.
	[117]	Sepsis-3	ICU	ZUH, 4449	FEX, FS, VI	55	NM	NM	RF AUC: 0.91	IF: neutrophils%, D-dimer, neutrophils.
	[118]	Sepsis-3	ICU	PNCC, 40,336	HMV	40	NM	6 h	TCN AUC: 0.91	Per time-step AUC: 0.98
	[119]	Sepsis-3	ICU	PNCC, 23,711	HMV, CB, FS, FE, VI, SHAP	25	2, 12, 24 h	6 h	LGBM AUC: 0.979	IF: PTT, WBC, platelets.
	[133]	Sepsis-2	ICU	MIMIC-III, 31,575	FEX, HTS, HMV	101	20 h	3 h	RNN AUC: 0.81	Performance increases with increasing observation window and decreasing prediction window.
	[134]	NM	ICU	PNCC, NM	CB, HMV, FS, N	CNN:11 RNN: 40	CNN: Up to 5 h RNN: Up to 11 h	12 h	EM (CNN, RNN) AUC: 0.964	Hourly/real time predictions.
	[121]	Sepsis-3	ICU	MIMIC-III, 6188	FS, CB, HMV	44	Up to 41 h	7 h	DTW-KNN APR: 0.40	Irregularly sampled multivariate time series.
	[120]	Sepsis-3	ICU Admission	PNCC, 40,336	FEX, HMV, N, CB (SMOTE), FS (Z-test, CAn)	6	-	6 h	XGB AUC: 0.98	Only vital signs used.
	[131]	CMS	ED Admission	QAH, 42,979	NM	86	Hourly prediction	4 h	MGP-RNN AUC: 0.882	MGP-RNN outperforms scoring systems.

[a] Final cohort (training & test set), [b] Final number of features of best model (after feature selection, before feature engineering), [c] Of best model. Refer to Glossary section for acronyms.

3.1.3. Modelling & Evaluation

Regarding proposed models, 19 (56%) studies use a DL model, and 15 (44%) studies use a ML model as the best performing one. Most of the ML models are tree-based ones like Extreme Gradient Boosting (XGB) [113,116,120,124,125,141], Random Forest (RF) [117,129,145], Gradient Boosting (GB) [114], Light Gradient Boosting Machine (LGBM) [119], and AdaBoost [144]. DL models proposed are usually temporal ones, like the Long Short Term Memory (LSTM) [122,127,135,143], the Recurrent Neural Network (RNN) [131,133], the Convolutional Neural Network (CNN) [115,140], the Temporal Convolutional Network (TCN) [118], or combinations of them [132,134].

Table 5. Sepsis Prediction & Sepsis Associated ARDS Prediction.

Task	Ref.	Def.	Ward	Dataset, Samples [a]	Preprocessing	Feats [b]	Obs. Window [c]	Pred. Window [c]	Best Model	Final Remarks
Sepsis Prediction	[132]	Sepsis-2	Depart out of ICU	DMM, 3126	HTS, HMV, OHE, S, FE, CB	5030	Up until prediction time	3 h	LSTM-CNN AUC: 0.856	High SEN in departments where sepsis is not common. Representations from raw event sequence used. Patients had not initiated intravenous antibiotics or blood culture at the time of early detection.
	[128]	Sepsis-3	ICU	PNUYH, 21,957	HO, HMV, HTS, FS, N, SHAP	24	24 h	24 h	NN AUC: 0.7888	NN outperforms scoring systems.
	[135]	Sepsis-2	ED	DIIC, 186,575	HMV, HTS, VI	111	NM	4 h	Proposed LSTM-based AUC: 0.892	LSTM outperforms scoring systems. Interpretable. Handles irregular time intervals. IF: RR, pulse, GCS.
	[136]	ICD-9 & SIRS	GW	MIMIC-III, 48,632	HMV, HTS, VI	10	5 h	3 h	NN AUC: 0.86	IF: WBC, RR, DBP.
	[137]	Sepsis-2	ICU	Proprietary EHR, 40,000	OHE, EMB	29	48 h	4 h	PAVE AUC: 0.780	No need to HMV because of EMB. Interpretable.
	[138]	SIRS	ICU	PNCC, 40,336	FEX, FS (RF, AENN, CAn), HMV, CB, FE	15	NM	6 h	LR AUC: 0.614	Anomaly detection semi-supervised framework.
	[129]	Sepsis-3	ICU	MIMIC-III, 685,110	CB (SMOTE), HMV, LE, OHE, HO, VI	31	NM	NM	RF AUC: 0.918	IF: ICU LOS, hospital-to-ICU admission time, O2 saturation.
	[139]	Sepsis-2	ICU	ICUS, 282	HMV, HTS, FE	30	-	6 h	DFSP AUC: 0.92	DFSP outperforms scoring systems.
	[140]	Sepsis-2	ICU	3 hospitals, 40,336	HMV, HTS, HO, S, FE	34	2 h	6 h	ACNN ACC: 0.9318	Classification of features as 'high' or 'low'.
	[125]	ICD-9 & Sepsis-3	Hosp, ED	DAD, 270,438	FEX, HTS, HMV, FE, CB	7	NM	48 h	XGB AUC: 0.827	XGB outperforms scoring systems.
Sepsis associated ARDS Prediction	[144]	ICD-9	ICU	eICU, 19,249 (sepsis patients)	HMV, FEX, FS, FE	14	First 24 h in ICU	NM	AdaBoost AUC: 0.895	3 phenotypes with different therapeutic responses are clustered.

[a] Final cohort (training & test set), [b] Final number of features of best model (after feature selection, before feature engineering), [c] Of best model. Refer to Glossary section for acronyms.

Table 6. Septic Shock Prediction & Sepsis Mortality Prediction.

Task	Ref.	Def.	Ward	Dataset, Samples [a]	Preprocessing	Feats [b]	Obs. Window [c]	Pred. Window [c]	Best Model	Final Remarks
Septic Shock Prediction	[143]	Sepsis-2	Hosp Admission	MIMIC-III, NM	FEX, HO, HMV, S	30	NM	48 h	LSTM AUC: 0.8306	Performance increases with smaller prediction window.
	[123]	Sepsis-3	ED Admission	THC, 604 (AP patients with sepsis)	N, PCA, PCC, FS (KW, ANOVA, RFE, R), CB, VI	11	First 24 h in ED	Up to 28 days	AE AUC: 0.879	AE with PCC and RFE outperforms scoring systems. IF: Disease duration, HR, RR.
	[128]	Sepsis-3	ICU	PNUYH, 23,189 (sepsis & non-sepsis patients)	HO, HMV, HTS, FS, N, SHAP	24	24 h	24 h	NN AUC: 0.8494	NN outperforms scoring systems.
Sepsis Mortality Prediction	[113]	Sepsis-3	ICU	MIMIC-III, 4559 (sepsis patients)	HMV, FS (KST, STT, ANOVA, MWU, KW, χ2, FET, AIC), FE, VI	11	First 24 h in ICU	30 days	XGB AUC: 0.857	XGB outperforms SAPS-II. IF: urine output, lactate-min, BUN-mean
	[114]	Sepsis-3	Hosp	MIMIC-III, 16,688 (sepsis patients)	HMV, FEX, FS, FE	86	First 24 h in ICU	NM	GB AUC: 0.829	SAPS-II has the poorest calibration
	[126]	Sepsis-3	Hosp	MIMIC-III, 9432 (sepsis patients)	FEX, HMV, FS, FE, CB (SMOTE), VI (XGB)	30	NM	NM	NN-GCN ACC: 0.8278	IF: Bicarbonate, age, PH
	[127]	HTDV criteria & Sepsis-3	Hosp	HTDV, 40 (sepsis patients)	FEX, N, HTS, FE, LIME	5	First 24 h after hospitalization	Time to discharge (avg 2 weeks)	LSTM APR: 0.83	Models trained on wearable data outperform models trained on bedside monitor data.

[a] Final cohort (training & test set), [b] Final number of features of best model (after feature selection, before feature engineering), [c] Of best model. Refer to Glossary section for acronyms.

Among the 6 sepsis/septic shock detection papers (see Table 3), 5 (83%) [116,124,141,142,145] of them used a ML model with AUC ranging from 0.86 to 0.97, while just 1 study [122] uses a DL model with an AUC of 0.835. The observation window ranges from 0 to 12 h for the ML models, while the DL model uses LSTM components for detecting sepsis using 48 h of data. Overall, for sepsis detection and septic shock detection, ML models are found to be sufficient.

Among the 29 prediction models, 19 (66%) [112,115,118,123,126–128,130–137,139,140,143] used a DL model with AUC ranging from 0.78 to 0.964, and 10 (34%) [113,114,117,119–121,125,129,138,144] used a ML model with AUC ranging from 0.614 to 0.98. Overall, for sepsis prediction, septic shock prediction and sepsis mortality prediction ML and DL models have been proven equally promising. Specifically, for mortality prediction, 2 studies used a ML model and 2 studies used a DL model with similar performance (AUC 0.8278–0.857). While mortality prediction models did not achieve as high AUC as some sepsis prediction models, their performance interpretation regarding disease severity can be as indicative considering the mortality outcome.

In summary, we differentiate between models used for detecting sepsis presence versus the prediction of sepsis at a future time. Models for detections tend to be much simpler than models used for prediction of sepsis at a future time.

Studies that compare their best performing algorithm with a clinical approach show the superiority of AI in detecting/predicting infection. Performance comparison with at least one clinical scoring system (see Section 2.2) takes place in 11 (32%) of the studies [113,115,116,123–125,128,131,135,139,142], where they are always outperformed by the proposed algorithm. Some studies [132] confirm that some AI methods predicted sepsis before any antibiotics or blood cultures were initiated. This finding suggests that integrating lab values in algorithms, rather than using them based on standalone thresholds (see diagnostic biomarkers in Table 1), leads to early predictions that may help initiate earlier treatment,

potentially preventing sepsis. Clearly, careful clinical studies would have to be performed before initiating a change in clinical practice.

3.2. Length of ICU/Hospital Stay Prediction

Length Of Stay (LOS) prediction of hospitalized patients and ICU patients targeting the duration from ICU/hospital admission to discharge, constitutes an important model outcome for both doctors, management staff and, thus, patients.

It can be used by doctors as a measure of patient acuity indicating illness severity and helping them to avoid overmedication or undertreatment. It also indicates recovery speed, the need for closer monitoring or adjusted treatment plans, minimizing the risk of early discharge and readmission. In addition, predicting LOS of existing admissions helps in resource allocation and management, like ensuring there are enough beds for new admissions. Patient prioritization for discharge and overall scheduling are also handled based on expected LOS for current patients.

However, over-reliance on such models can lead to inadequate monitoring of patients and hence deterioration or readmission, if the model underestimates the length of stay/discharge time. Conversely, an overestimation of the length of stay/discharge time can lead to a waste of allocated resources and inefficient patient prioritization.

A second literature review has been performed using GoogleScholar to identify papers published between 2019 and 2024 predicting patient LOS. We excluded studies that were published before 2019 (3), involved patients under 18 years old (4) or did not use ML/AI modelling (2). An overview of the 39 studies is presented in Tables 7–9. Two downloadable pipelines that predict LOS, among other clinical outcomes, using a range of models and databases are given in [111] https://github.com/rvandewater/YAIB (accessed on 19 July 2024) and [146] https://github.com/yzhao062/PyHealth (accessed on 19 July 2024) which can handle multimodal data.

3.2.1. Prediction Objective

Papers found are split in 3 LOS outcome categories: (i) a continuous outcome in hours/days [149–167] (19, 49%), (ii) a binary outcome of LOS > X days [151,159,168–179] (14, 36%), and (iii) a multiclass outcome of LOS > X days [152,163,180–185] (8, 21%). The first category involves most papers identified for this prediction task. Regarding the second category, there is evidence that in the clinical decision-making the general cut-off point of LOS is 4–5 days [168], while for general ICU patients in the United States the average is 3 days [186]. According to this study, a prespecified value is sometimes used as a threshold [169], while in other cases it is identified based on either the mean [170], median [151,168], 75th percentile [171,172] of the study population outcome, previous studies [173], or clinical importance [174]. Most papers refer to ICU LOS (29, 74%), while the rest refer to hospital LOS (10, 26%).

Table 7. Hospital/ICU Length of Stay Prediction: Continuous.

Ref.	Task	Location	Dataset, Samples [a]	Preprocessing	Feats [b]	Obs. Window [c]	Best Model	Final Remarks
[149]	hours	ICU	MIMIC-III, 6927	FS, HMV, HO, HUM, FE	28	First 24 h in ICU	GA MSE: 43,828	Dispersion tendency statistics (min, max, range), are more suitable for LOS prediction than other FE statistics.
[150]	hours	ICU	HiRID, 21.54 million	HTS, HMV, OHE, S, FE	-	1 week, throughout stay (continuous learning)	LGBM MAE: 56.9	Benchmark result for proposed pipeline. Label distributions contribute to low scores. LGBM-based methods outperform DL methods. FE does not help. https://github.com/ratschlab/HIRID-ICU-Benchmark/ (accessed on 19 July 2024)
[152]	days	ICU	MIMIC-III, 42,276	HTS, HUM	17	Data since ICU admission	LSTM CWK: 0.433	Hourly predictions.
[153]	days	ICU	eICU, 73,389	HMV, HTS, FS, EMB	20	-	BiLSTM R2: 0.643	Positive impact of EMB.
[154]	days	ICU	HHTCM ICU, 17 (COVID-19 survivors)	FS, HO, FE	10	-	LASSO-LNR MAE: 0.723	Prediction achieved before ICU admission.
[151]	days	ICU	MIMIC-III, 44,626	HO, HMV, FE, N	33	First 24 h in ICU	SVM MAE: 2.810	FE improves performance.
[155]	days	ICU	eICU, 89,123	FE, FS, HO, HMV, HTS	38	First 24 h in ICU	LSTM-MPNN MAD: 1.86	Combining LSTM and GNNs improves performance. GNNs provide context for rarer patters of diseases.
[156]	days	ICU	eICU, 89,127	FE, FS, HO, HMV, HTS, CR	38	First 24 h in ICU	GRU (FL-SRC) MAE: 2.21	CR based on local output distribution and local sample size improves FL performance and runtime.
[157]	days	Hospital	UTMB, 805	HMV, FS (LF), FE, OHE, N	59	NM	CoxL RMSE: 5.4834	Censored clinical data are included.
[158]	days	Hospital	YCDTSH, 154	FS (SBFCM), LE, CAn	11	NM	NN MAE: 2.03	Feature selection framework is proposed.
[159]	days	ICU	TRDGU, 23,830	VI	15	-	LNR MAE: 5.0	IF: Ventilation, number of injuries.
[160]	days	Hospital	THI, 5363	FEX, HMV, FS (PFI), FE, N, LE	60	-	SVR MAE: 1.85	Hierarchical Bayesian model outperforms best ML model.
[161]	days	Hospital	MIMIC-IV, 511,741 subjects, 170,934 images	FEX, FS, OHE, HMV, FE, HO, IR, T	52 (tabular data)	NM	DF-Mdl (CNN, LSTM, Att-1DCNN) MAE: 3.8682	Multimodal data (lab results, images, clinical notes, etc.) used.
[162]	days	ICU	MIMIC-IV, 48,367	FEX, IE	30	NM	BT RMSE: 2.863	Uniform incomplete data across all racial features favors performance. No significant impact of imputation method. Small negative impact of missing data quantity for prediction performance.
[163]	days	ICU	ASSIST, 1642 (CHD patients after surgery)	FS, HMV, S, SHAP	93	-	LGBM RMSE: 15.2	Mechanical Ventilation time, patient weight on surgery day: most influential predictors.
[164]	days	Hospital	DHHS, 2.3 million+	HMV, FS (CAn, EDA), VI	34	NM	RF MSE: 5	Patients with diagnoses related to birth complications spent more days in hospital than other diseases. IF: total costs, diagnosis.
[165]	days	Hospital	TMUGH, 168 (FNF patients)	FEX, HMV, IE, OHE, N	38	NM	PCR MAE: 1.525	Postoperative calcium level & lymphocyte%, intraoperative bleeding, glucose & Sodium chloride infusion after surgery, CCI, BMI: most significant predictors
[166]	days	Hospital	ACS-NSQIP, 302,300 (TKA patients)	HMV, N, HO, OHE, IE, S	32	NM	MLP MSE: 0.690	Conventional and deep learning models performed better than mean regressors.
[167]	days	Hospital	SRPH, 4376 (T2DM and HTN patients)	FEX, BCT, FS (IG, ReliefF)	73	NM	RF MAE: 0.935	Patients with primary diseases such as T2DM or HTN may have comorbidities that can prolong inpatient LOS.

[a] Final cohort (training & test set), [b] Final number of features of best model (after feature selection, before feature engineering), [c] Of best model. Refer to Glossary section for acronyms.

3.2.2. Dataset & Features

MIMIC-III [147], is the most common dataset, used by 6 (15%) studies [149,151,152,173,176,184], followed by MIMIC-IV [187] used by 4 (10%) studies [161,162,175,178], and eICU [188] used by 4 (10%) studies [153,155,156,169]. Most of the studies (9, 23%) use the first 24 h in hospital/ICU as an observation window [149,151,155,156,169,174–176,184]. Regarding feature preprocessing, we note that feature engineering [151] and particularly engineering using dispersion statistics [149], as well as embedding [153] can improve performance. The most important features according to built-in model feature importance frequently appear to be vital signs and lab values [165,169,171,174,182].

3.2.3. Modelling & Evaluation

Most studies (26, 67%) are based on a ML model compared to 13 (33%) studies based on DL models. Most ML models are tree-based, and mainly RF [151,164,167,168,173,174,176,178,180–182], XGB [170,172], and LGBM [150,163]. DL models are mostly LSTM-based [152,153,155,161,175].

Within the 14 binary prediction tasks, 2 (14%) [171,175] use a DL model with AUC of 0.76 and 0.915, respectively, while 12 [151,159,168–170,172–174,176–179] use a ML model with AUC ranging from 0.587 to 1.00. It is observed that performance within this binary category is overall increased with increasing classification threshold of length of stay, probably because more days in hospital/ICU can imply disease severity benefiting model discrimination ability. Overall, ML models are more frequently proposed than DL models for LOS prediction, with RF and LSTM-based models being used the most. ML models performed just as well as DL models.

Table 8. Hospital/ICU Length of Stay Prediction: Binary.

Ref.	Task	Location	Dataset, Samples [a]	Preprocessing	Feats [b]	Obs. Window [c]	Best Model	Final Remarks
[151]	LOS > 2.64 days	ICU	MIMIC-III, 44,626	HO, HMV, FE, N	33	First 24 h in ICU	RF AUC: 0.70	FE improves performance.
[175]	LOS > 3 days	ICU	MIMIC-IV v1.0	ICDC, CA, FS, HUM, HO, HMV, HTS	NM	First 24 h in ICU	LSTM-H AUC: 0.76	Benchmark result for proposed pipeline. https://github.com/healthylaife/MIMIC-IV-Data-Pipeline (accessed on 19 July 2024)
[176]	LOS > 3 days	ICU	MIMIC-III, 34,472	FS, FE, HUM, HO, HTS, CA, N, HMV	114	First 24 h in ICU	RF AUC: 0.736	Benchmark result for proposed pipeline. https://github.com/MLforHealth/MIMIC_Extract (accessed on 19 July 2024)
[169]	LOS > 3 days	ICU	eICU, 117,306	FE, HMV, VI	17	First 24 h in ICU	GB AUC: 0.742	IF: Pao2/Fio2 ratio, GCS, SUN.
[178]	LOS > 3 days-2nd admission	ICU	MIMIC-IV, 18,572	FEX, FS (LF), FE, N, VI	220	-	RF AUC: 0.716	IF: LOS of 1st admission, age, Phytonadione and Metoprolol Succinate XL.
[179]	LOS > 3 days	ICU	CCM, 24,876	FS, CB (SMOTE)	NM	NM	EM (GBM, SVM, LR) AUC: 0.587	EM outperforms baseline models.
[168]	LOS > 6 days	ICU	PUMCH ICU, 2224	FS, N, HMV, CB, TT	26	First 6 h in ICU	RF AUC: 0.76	RF outperforms SOFA score.
[176]	LOS > 7 days	ICU	MIMIC-III, 34,472	FS, FE, HUM, HO, HTS, CA, N, HMV	114	First 24 h in ICU	RF AUC: 0.764	Benchmark result for proposed pipeline. https://github.com/MLforHealth/MIMIC_Extract (accessed on 19 July 2024)
[170]	LOS > 7 days	ICU	IHICU, 77 (COVID-19 survivors)	HMV, CB, FS, VI	4	-	XGB AUC: 0.795	IF: Hematocrit and ESR.

Table 8. *Cont.*

[173]	LOS > 7 days	ICU	MIMIC-III, NM (lung cancer patients)	FEX, HMV, D, FS (CS, RFE), CB (ADASYN), SHAP	60	-	RF AUC: 1.00	ADASYN outperforms other CB techniques.
[174]	LOS > 7 days	ICU	CUHICU, 12,133	HMV, FS, VI	91	First 24 h in ICU	RF AUC: 0.881	IF: HR, LDH
[159]	LOS > 7 days	ICU	TRDGU, 108,178	VI	10	-	LR AUC: 0.903	IF: Injury severity, intubation, pre-trauma condition.
[172]	LOS > 9.08 days	ICU	WUHICU, 365 (HT patients)	HMV, FS (ML, CAn, LASSO, PLS-DA), CB, SHAP, VI	6	-	XGB AUC: 0.88	IF: ECMO
[174]	LOS > 14 days	ICU	CUHICU, 12,133	HMV, FS, VI	91	First 24 h in ICU	RF AUC: 0.889	IF: HR, LDH
[177]	LOS > 14 days	ICU	YH, 75 (COVID-19 patients)	HMV, FS (TT, RST, $\chi 2$, FET, AIC)	5	-	LR AUC: 0.848	Elevated PCT significantly associated with hospital LOS > 14 days.
[171]	LOS > 23 days	ICU	THMC, 1417 (TBI patients)	HMV, FS ($\chi 2$, TT), VI	20	-	NN AUC: 0.915 (LOS > 23 days)	IF: Age, initial SBP in ED, ISS

[a] Final cohort (training & test set), [b] Final number of features of best model (after feature selection, before feature engineering), [c] Of best model. Refer to Glossary section for acronyms.

Table 9. Hospital/ICU Length of Stay Prediction: Multiclass.

Ref.	Task	Location	Dataset, Samples [a]	Preprocessing	Feats [b]	Obs. Window [c]	Best Model	Final Remarks
[181]	3 classes in days	ICU	SCUSH ICU, 233	HMV, HO, D, N, FS, FE	31	Data on ICU admission	RF ACC: 0.9199 PR: 0.38	Early resource management and decision making.
[184]	3 classes in days	Hospital	MIMIC-III, 47,796	CB (SMOTE), EMB, VI	NM	First 24 h in Hospital	HAN AUC: 0.82	ARF, CAD, severe sepsis are the highest attention weighted ICD9 diagnosis codes.
[185]	3 classes in days	Hospital	NHCRD, 318,438	FEX, FS, HMV, HR, LE, N, D, SHAP	31	NM	SVM AUC: 0.95	Admission accuracy score and patient history: most significant predictors.
[180]	4 classes in days	ICU	GPCICU, 353 (patients with acute type A aortic dissection)	HMV, FS (KCC)	12	-	RF AUC: 0.991	
[182]	10 classes in days	ICU	KFUH ICU, 895 (COVID-19 patients)	HMV, CB, VI	47	Data on ICU admission	RF ACC: 0.9416	Age, CRP, NOS days: top features related to ICU admission & ICU LOS.
[152]	10 classes in days	ICU	MIMIC-III, 42,276	HTS, HUM	17	Data since ICU admission	LSTM-C-DS CWK: 0.451	Hourly prediction. C and DS improve LSTM performance.
[183]	11 classes in days	Hospital	AVHHA, 455,495	HMV, N, IE	12	NM	NN ACC: 0.408	
[163]	3 classes in days	ICU	ASSIST, 1642 (CHD patients after surgery)	FS, HMV, S, SHAP	93	-	CatBoost AUC: 0.8559	Mechanical Ventilation time, preoperative arterial O2 saturation, VIS: most influential predictors.

[a] Final cohort (training & test set), [b] Final number of features of best model (after feature selection, before feature engineering), [c] Of best model. Refer to Glossary section for acronyms.

3.3. ICU Admission/Hospitalization Prediction After Emergency Department Admission

Comprehensive hospitalization management in Emergency Departments (EDs) is a key indicator of efficient triage and resource prediction/utilization. EDs are continually dealing with large streams in patient traffic and increasing resource demands, making

hospitalization decisions a pivotal factor affecting patient outcomes and the judicious allocation of resources [189]. ML techniques have emerged as effective for ED triage prediction of hospitalization models, achieving high accuracy [190,191]. A popular goal is to identify high-risk patients for hospitalization/ICU admission to help prioritize allocations of medical resources (e.g., beds, staff), both in the unit of transfer and the ED. This ensures smooth and efficient triage flow, avoiding overcrowding in the ED and optimally delivering better quality care both in the hospital/ICU and the ED. Such models, however, should not be over-relied on, but complement triage decision making, as positive predictions for transfers can waste valuable resources. Conversely, negative predictions for transfers can lead to undertreatment, worse patient outcomes and overcrowding in the ED.

A literature review has been performed using Google Scholar to identify papers that predict ICU admission/hospitalization after emergency department admission published between 2018–2024. We excluded 2 studies that included young patients (under 18 years old) and 2 studies predicting ICU admission/hospitalization from departments other than the ED. An overview of the 18 identified papers is presented in Table 10.

3.3.1. Prediction Objective

Most studies (15, 83%) make predictions at triage [191–205], whereas some make predictions at 30 min [206], 1 h [198], 2 h [206,207] and hourly [208] after patient arrival at ED. Most studies (13, 72%) predict hospitalization, while 6 studies (33%) predict ICU admission after ED admission.

Table 10. ICU Admission/Hospitalization Prediction After Emergency Department Admission.

Ref.	Dataset, Samples [a]	Location	Preprocessing	Feats [b]	Prediction Time	Best Model	Final Remarks
[191]	NHAMCS, 135,470	Hospital	HMV	11	At triage	NN, GB AUC:0.82	Algorithms outperform ESI.
[198]	EDCUS, 11,105	Hospital	VI	11	At triage, at 60 min	AutoML AUC: 0.914 (at triage) AUC:0.942 (at 60 min)	IF: previous visit outcomes, triage information.
[192]	NHAMCS, 52,037	Hospital	HMV, VI	13	At triage	GB, NN AUC: 0.8	Higher SPE for all ML models compared to conventional methods. IF: ambulance use, oxygen saturation.
[206]	NEED, 159,499	Hospital	FE, HMV, VI	18	At 30 min, at 2 h	GB AUC: 0.86	Prediction at 30 min after ED admission has similar performances to 2 h one.
[193]	interrail, 2274	Hospital	FS (V), HTS, S, HMV, OHE	723	At triage	GB AUC: 0.8	IF: IVT
[194]	TTH, 282,971	Hospital	HMV, FS	10	At triage	NN AUC: 0.8004	Model performed better in the nontraumatic adult & environmental emergency subgroups.
[208]	DUHS, 418,167	Hospital, ICU	FS, S, HO, N, FE, HMV, OHE, VI	723	Hourly throughout ED stay	LGBM AUC: 0.873 (hospitalization prediction) AUC: 0.951 (ICU admission pred)	Good external validation and online/live performance as well. IF: age, hematocrit, WBC.
[195]	USMH, 42,530	Hospital	VI	8	At triage	XGB AUC: 0.86	XGB is comparatively fast. XGB performance increases with increased data.
[196]	NIH, 107,545	Hospital	FS (χ2, ANOVA), FE, HMV	14	At triage	XGB AUC: 0.859	LR should be considered for interpretability.
[197]	MUSH, 453,664	Hospital	HMV, N, IE, FS, FE, VI	17	At triage	T-ADAB AUC: 0.954	Optimized models outperform ones. IF: O2 Saturation. Accuracy of model does not change with increased data.

Table 10. *Cont.*

Ref.	Dataset, Samples [a]	Location	Preprocessing	Feats [b]	Prediction Time	Best Model	Final Remarks
[207]	THS, 610	ICU	FEX, HTS, FE, FS (BE)	2	At 5 min up to 2 h	GLMM AUC: 0.947	Heart rate variability data used, easily obtained from ECG and PPG sensors.
[199]	MIMIC-IV, 30,206 (AF patients)	ICU	FEX, FS (SHAP, RF), HMV	8	At triage	RF-derived scoring system AUC: 0.737	5 vital signs, ED length of stay, age and arrival transport were used.
[200]	NTUH, 268,716 (retrospective), 1294 (prospective)	Hospital	FEX, HMV, CB (SMOTE, TL)	24	At triage	TabNet, MacBERT ACC: 0.82	Structured and unstructured data included. Interpretability (TabNet, BertViz).
[201]	AMCUS, 19,155 (COVID-19 patients)	ICU	FEX, HMV, N, CB (SMOTE), FS (RFE, SFM, SKB), VI	10	At triage	XGB-SA AUC: 0.892	IF: AKI, age, ARDS
[202]	EDHS, 49,266	Hospital	FEX, IE, FE, FS (V), HMV, N, T, WFC, TF-IDF, SHAP	82	At triage	XGB AUC: 0.922	Including text data improves performance.
[203]	SRHM, 1004 (COVID-19 patients)	ICU	FEX, FS, N, FE, HMV, CB (SMOTE), SHAP	22	At triage	SVM AUC: 0.85	Just 2 demographic features and the CBC test results are required. Low lymphocytes values and high neutrophils values predictive of ICU admission.
[204]	MGB, 3597 (COVID-19 patients)	ICU	FEX, FS, HMV, CB (RU), SHAP, VI	54	At triage	RF AUC: 0.88	IF: CRP, oxygen saturation, and LDH.
[205]	TMUSH, 167,058	Hospital	TP, S, OHE, T, CB	9	At triage	BlueBERT AUC: 0.9014	Translating clinical notes into English and textualizing numerical data into categorical representations improved performance.

[a] Final cohort (training & test set), [b] Final number of features of best model (after feature selection, before feature engineering). Refer to Glossary section for acronyms.

3.3.2. Dataset & Features

Datasets used vary with the National Hospital and Ambulatory Medical Care Survey (NHAMCS) dataset being used twice (11%) [191,192]. Text data are often included [200,202,205], with a study suggesting that the use of text can improve performance [202]. Important features deduced also vary with oxygen saturation [192,197,204] and age [199,201,208] appearing 3 times (17%) each. Other important features are previous visit outcomes [198], ambulance use [192], intravenous therapy (IVT) [193], and different lab values [203,204,208].

3.3.3. Modelling & Evaluation

Most of the studies (15, 83%) propose a ML model achieving AUC in the range of 0.8 to 0.954, while 5 studies (28%) [191,192,194,200,205] propose a DL models achieving AUC performance ranging from 0.80 to 0.9014. Some studies suggest that ML and DL models perform similarly [191,192]. Most of the ML models are tree-based and mainly GB [191–193,206], XGB [195,196,201,202], RF [199,204], LGBM [208], and AdaBoost [197]. A study [191] focused on an improved algorithm for Emergency Severity Index (ESI). For AdaBoost, ref. [197] found that performance did not increase by using a larger dataset. Yet,

for XGB, ref. [195] found that performance did improve by using a larger dataset. Overall, ML and DL models performed similarly.

4. Challenges of AI Applications in ICUs

Despite promising study results in handling ICU outcomes using AI, there are many challenges that are determinant to its successful adoption and impact. We split them into 4 categories: Healthcare Data, Modelling, Clinical Applicability, Ethical Use of Healthcare Data, and we recommend solutions regarding each challenge. If carefully developed, sufficiently validated, and seamlessly integrated, AI can truly improve patient outcomes in clinical settings.

4.1. Healthcare Data

Data flow in ICUs can be diverse due to different patient conditions and machines, leading to signal interruptions and varied granularity. These raise the challenges of irregular time intervals and missing data. They appear in almost all considered papers and are commonly handled with resampling (e.g., averaging values in hourly bins) and imputation (e.g., linear interpolation, last value carried forward approach).

When dealing with disease prediction or mortality prediction, datasets are often imbalanced, with the positive class constituting a minor percentage of the whole data. Oversampling or down sampling data using different class balancing approaches (e.g., SMOTE [120,126,129,130,145], ADASYN [173]) on the training set, are a common preprocessing part, ensuring the model learns to detect both classes.

A high dimensional set of features, that are not easily obtained, as input to an AI model, will challenge its translational potential in the clinical setting. Recurrent Feature Elimination (RFE) [123,173] and statistical tests [113,120,123,142,171,177,180,193,196] are commonly used for feature selection. The most important features for the outcomes of this review based on built-in model importance are usually vital signs and laboratory values. Vital signs and laboratory values, as well as medications and previous medical history, are critical data sources for other medical outcomes, like organ support (such as vasopressors or renal replacement therapy) prediction. Feature selection plays a major role in preprocessing and deploying the models. The lower the number of features, the less complex and the more parsimonious the model, which yields easier implementation in a clinical setting, where a variety of measures exist, and decisions must be timely.

A common challenge of developing AI models for ICU is generalizability. Insufficient data size can lead to overfitting and biased outcomes towards a specific group of patients. In addition, using a single type or modality in AI models can yield insufficient outcomes, lower performance, and inappropriate interventions. Making sure the models are trained on a dataset reflective of the population and of sufficient size (instead of one of specific diseases) and is not overfitted on training data, are all crucial in ensuring validation of predictions in different ICUs. Ideally, ICU data across different institutions must be collected to create a large and diverse data set, following appropriate data sharing guidelines.

Different ways of documenting and sharing data like medications, procedures or vital signs can make them incompatible for an AI model hindering research. In order to allow for more data to be collected efficiently, standardization procedures and interoperability profiles like ICD [209] and FIHR [210] are essential. In addition, integration of multimodal data like text-based clinical notes, time series data, and imaging data means heterogeneous formats that require sophisticated and more complex data preprocessing, as highlighted in the European Health Data Space (EHDS) regulation [211], for the secondary use of health data.

Inconsistency in data storage form poses a risk of loss of information and inadequate data input for the models, potentially leading to inappropriate interventions. Consistency

in electronic and paper storage of data will aid organization and complete integration with AI models.

4.2. Modelling

Apart from limited data, model learning parameters may contribute to overfitting. Evidently, tree-based ML models and temporal DL models are frequently used in sepsis prediction/detection, sepsis mortality prediction, LOS prediction and hospitalization/ICU admission prediction after ED admission. This might be attributed to the ensembling nature, sequential form (GB-based models) and regularization options of tree-based models to avoid overfitting and enhance generalizability. Proposed temporal DL models, like RNN and its variations (e.g., LSTM), are high-performing, probably due to their ability to handle temporal dependencies and the dynamic nature of vital signs and lab values. Importantly, external validation should also take place to verify generalizability of the model.

Detecting diseases early through AI models, even before clinical symptoms appear, is crucial for timely interventions and positive prognosis. Key to the early detection is the size of time series windows. As expected, for sepsis prediction, several studies suggest that a longer observation window [133] and a shorter prediction window [133,143] increase model performance. Unfortunately, a longer observation window could delay treatment intervention or hospitalization. A shorter prediction window may not provide early informative predictions, and/or make interventions less impactful. Optimally, the shorter the observation window and the larger the prediction window, the earlier the detection and intervention. Models should be trained sufficiently to capture the relationship of a relatively short observation window to the clinical outcome.

However, using a model once time series data reaches an observation-window-length in ICU, may not be so informative, delaying predictions and interventions. Some studies [112,134,152] provide real-time predictions, considering data on an hourly basis (hourly labels), within a fixed observation window, providing per-timestep evaluation. Specifically, hourly sepsis predictions enable the model to learn changes in variables like vital signs and lab values by hour. Thus, implementation of such models will hourly capture spikes or falls during the system's response to infection, assessing patient state regularly, and enabling more informed and earlier interventions. Ideally, AI's contribution will lie in its ability to also update in real-time and organize the constantly incoming and changing data to generate an accurate outcome.

Although a large prediction window is crucial for early interventions and survival, before severe symptoms arise, administering antibiotics too early (or too frequently) can be risky and may contribute to antimicrobial resistance (AMR). Premature or unnecessary use of antibiotics, before the causative pathogen of infection is confirmed, can lead to inappropriate broad-spectrum antibiotics administration. Overuse of these antibiotics can also accelerate the development of resistant bacteria and disrupt the body's natural microbiome, killing beneficial bacteria and creating an environment where resistant pathogens can thrive. Balancing timing, correct use and dosage is important for mitigating AMR, undertreatment and overtreatment risks. Ideally, once sepsis is predicted, AI models can be used for its management, by optimizing appropriate antibiotic selection and effective dosage schedules, based on patient data like medical history, vital signs and lab results. They can also predict how a patient will respond to an antibiotic based on factors like age, comorbidities, and severity of infection, ultimately minimizing side effects.

In predicting the onset of a life-threatening disease, like sepsis, or mortality, datasets whose negative class majorly outnumbers the positive one are used, posing a challenge on model evaluation. Wrongly evaluating models on imbalanced test data, can lead to misinterpretation of algorithm performance, causing misinformed or delayed interventions

and adverse effects for the patients. This evaluation leads to a trade-off between sensitivity and specificity of a list of classification thresholds (cut-points). Sensitivity and specificity values that balance this trade-off are the ones corresponding to the optimal threshold. The threshold is 'optimal' when it classifies most of the individuals correctly [212], ensuring maximum sensitivity and specificity. There are different ways for identifying the optimal threshold [213]. A commonly used one is the Youden index (J) method [214]. This method defines the optimal threshold as the point maximizing the Youden function which is the difference between true positive rate (Sensitivity) and false positive rate (1-Specificity) out of all possible thresholds [215]. Evaluation metrics summarizing the performance of the model across all possible thresholds are the AUC and Average Precision. However, depending on the context, emphasis can be placed on recall or specificity, according to how strictly the model is assessed for missing true positives or true negatives, respectively. Higher sensitivity than specificity might be required for a model used for early sepsis diagnosis or mortality prediction, ensuring that most of true positives are identified and interventions begin on time for best outcomes. This is relative to the use case of the algorithm and should possibly be assessed in collaboration with healthcare professionals.

During a public health crisis, applying AI to predict length of stay or hospitalization/ICU admission after ED admission for resource management, or directly predicting resource consumption, poses significant challenges. With overwhelming patient flow in healthcare systems, collected data may be incomplete or inaccurate, while AI models require high-quality, real-time data. Data from different hospitals may also be fragmented making interoperability a challenge. In addition, AI models trained on non-crisis data may not generalize well to novel conditions without retraining or fine-tuning, due to different severity and type of cases in an epidemic. With a disease prevalent in more than 10% of the population and its uncontrollable evolvement, ICU resource demand can be difficult to predict with existing models. New models should be able to adapt to new treatment protocols, variations in disease progression, and shifting resource needs (e.g., ventilators, beds, staff). Furthermore, in critical situations, it is important for healthcare professionals to trust the model's recommendations, so lack of interpretability of AI models would not be helpful. In order for the models to adapt to changing conditions amidst an epidemic, regular retraining with new data from the ongoing epidemic will be essential to capture evolving patterns in patient outcomes and resource usage. Online incremental learning will allow the model to adapt in real-time to shifting patient demographics and treatment protocols. Alternatively, pre-trained models can be fine-tuned on new, epidemic-specific data to adjust to the specific context of the crisis without requiring training from scratch. Finally, ensemble models can help handle uncertainty and improve robustness during the fluctuating epidemic conditions.

4.3. Clinical Applicability

Most of the studies use retrospective data to develop AI models that lack testing in complete real-world scenarios, carrying a high risk of bias. Controlled clinical trials with adequate human comparators are required, to assess the short- and long-term consequences thoroughly, as well as validation of models using prospective data.

Additionally, clinicians lack the skills to use and integrate these algorithms in their everyday job, something that will require time and money. Extensive training on how to run these algorithms and on AI applications in healthcare are required for clinicians and students, respectively. Collaboration between AI experts and clinicians in training and considering the appropriate infrastructure requirements are vital in ensuring smooth clinical applicability of the algorithms.

Moreover, integrating AI models with existing ICU scoring systems (Table 2) presents challenges due to differences in the way traditional scoring systems and AI models function. ICU scoring systems rely on specific, often manually recorded and manually standardized data, which may sometimes be incomplete, inconsistent, inaccurate, require subjective assessment (e.g., Glasgow Coma Scale), or be measured at irregular timesteps (e.g., vital signs). They are also often calculated based on data at a single time point, often at ICU admission, failing to capture the dynamic nature of critical illness. ICU scoring systems are validated in specific settings and populations and for specific diseases limiting their generalizability. They also do not always account for patient-specific characteristics such as genetic factors, comorbidities, or rare conditions, as they are based on a few specific variables. However, AI has the potential to handle time series data, in any given time in the ICU, impute missing values, standardize, resample time series and be trained on various disease data, of different sources and modalities to provide accurate personalized predictions. Therefore, AI will complement existing clinical scores and clinical knowledge.

This difference in the way AI and scoring systems function makes AI-driven sepsis prediction considerably faster and more timely than traditional clinical decision-making processes, like scoring systems. AI-driven sepsis diagnosis is also evidently [113,115,116,123–125,128,131,135,139,142] more accurate compared to scoring systems diagnosis. However, this increased accuracy comes with a risk of false positives. An elevated false positive rate can lead to false interventions and a waste of resources like medication, as well as increased length of stay for the patient. In the case of sepsis prediction, this can also encourage antimicrobial resistance, if antibiotics are administered in falsely detected patients.

The variability in sepsis (and any disease) definitions between hospitals also poses a significant challenge in developing standardized and globally applicable AI models for sepsis detection. Different sepsis definitions imply different protocols and interpretations of what constitutes sepsis in various healthcare settings, and thus different data and patient labels are involved. Global AI sepsis detection models will potentially be able to handle these variations when different institutions collaborate for a consistent use of a standardized sepsis definition (e.g., Sepsis-3). Additionally, training models using data that simulate different definitions of sepsis, from multiple hospitals, can aid model generalizability. Validating models developed on a single sepsis definition on data from different institutions and sepsis definitions can assess applicability to different clinical settings and potentially lead to mitigation of any bias related to local sepsis criteria variability. Furthermore, due to often using time series data, like vital signs and lab values, for the detection of sepsis, models can also adapt to changing sepsis definitions through online updates, using data and labels that comply with different sepsis detection criteria, in real time.

Personalized treatment plans in ICUs are facilitated using a range of data sources, like vital signs, lab results, radiology imaging, genetic variables, patient history of diagnoses, rare conditions etc., to ensure all patient factors are considered. Transformer-based models can ensure maximum and efficient processing of the different data modalities. Pre-trained transformers can adapt to the ICU data without the need to train models from scratch. Multimodal Large Language Models (M-LLMs) leverage different data modalities, ensuring personalized treatment plans output. However, using multimodal data for predictions requires advanced computational resources and complex algorithms due to the large volumes of diverse data. This can be costly, not supported by healthcare systems infrastructure, and computationally expensive for real-time processing and inference. Personalized treatment plans would also require clinical trials for monitoring patient engagement and plan effectiveness. Additionally, if the AI model is explainable and has been trained with

correct inclusive data, then it can be trustworthy and followed by the doctor to create a treatment plan.

AI applications in healthcare and ICUs need to pass rigorous regulatory approvals (e.g., FDA clearance in the U.S.). This process can be time consuming and costly. Additionally, the lack of universal standards for validating AI models in healthcare contributes to uncertainty around their clinical efficacy and safety and delays their deployment and contribution. Global standards and ethics committees, like the EHDS Regulation [211] and the AI Act [216], are needed to establish efficient standardized processes for AI model validation and approval of their implementation in clinical settings.

While AI can assist in decision-making, it should not replace human judgment. Overreliance on AI predictions could lead to waste of resources and insufficient delivery of care, if the models fail to account for all relevant factors, overestimating or underestimating predictions. Human clinicians are essential for interpreting AI predictions within the full clinical context and a human-in-the-loop AI approach is ideal.

4.4. Ethical Use of Healthcare Data

With the rapid development of AI, the discussion on ethics shifted towards the ethical implications of using ML methods for prognosis [217,218]. AI-enabled applications must adhere to the fundamental rights, societal values, and ethical principles of explicability, prevention of harm, fairness, and human autonomy [219]. The development of AI models focuses on helping healthcare professionals better serve those patients at risk, especially for patients in the ICU.

To regulate the development and use of AI models, the European Commission has issued several guidelines [219]. Firstly, if patients believe that their privacy is challenged, they might be hesitant to provide data or trust decisions supported by the AI model. Thus, acquiring consent from either the patient or their relatives is needed for collecting and using their data for both training the algorithms and/or as input to the Clinical Decision Support System (CDSS). Additionally, correctly informing them about privacy regulations, their right to withdraw their data, as well as the benefits of consenting can be effective in collecting the data. Obtained sensitive data should strictly follow privacy regulations. These involve full anonymization of the data and aggregating data into larger datasets. However, in some cases where an individual has an extremely rare condition, it may not be too difficult to deidentify. Hence, legal steps have been made, like the EU General Data Protection Regulation (GDPR) [220], that protects all EU citizens from privacy and data breaches, the EHDS guidelines [211], and the requirement for consent when data are to be reused in other contexts or for other purposes. To ensure that the use of data is also morally acceptable, ethical governance of data is essential. It involves an independent broadly representative group of participants to convene and develop a public statement about how the data, which is being held, is used. It also involves complete audit trails of everyone who has been given access to the data, and the purposes for accessing such data. Limiting data access is achieved through safe havens or formal agreements on the limitations of data use, as well as limited physical access to the databases.

Secondly, non-representative data used in training the algorithms might lead to inequities and biases in the prediction that could exacerbate health disparities and lead to inequitable care. Datasets of clinical or genetic data are determinant in personalizing predictions of ICU outcomes and should reflect the respective population, ensuring that model predictions are not less accurate for underrepresented groups and delivery of care is fair. Thus, to avoid bias, subgroups of certain demographics (e.g., age groups, genders, ethnicities) should be proportionately represented.

Thirdly, the opaqueness of ML approaches makes it difficult for people to trust their outputs and foster accountability of actions. In ICU predictions, healthcare providers should be able to understand how the system ended up making a prediction, and whether this should be trusted (Section 4.3). In ICU conditions, where mortality rates are high, issues of accountability need to be addressed legally and morally. Explainable AI is gaining momentum in bridging the gap between the black-box nature of advanced AI algorithms and the necessity for transparent, understandable, and interpretable decision-making [221]. Explainable AI can justify the model's predictions by indicating which variables, at what values and to what extent have influenced predictions. Clinicians' trust is therefore enhanced and can adopt AI-based clinical decision support systems more confidently. Examples of such applications of XAI can be found in Section 5.

Overall, ethical considerations regarding AI in healthcare and ICU are reflected in the need for patient consent and privacy regulations, inclusive datasets and explainability.

5. Explainable AI

Recently, there is strong growth in the development of explainable AI solutions for medical decision support [221]. Their taxonomy is multifaceted where the common classification criteria include (a) Explanation scope, (b) Explanation stage, and (c) Explanation approach.

First, for explanation scope, methods are either global or local. Global methods (e.g., Shapley Additive Explanations-SHAP [222,223], Feature Importance [224]) are used to describe the overall functioning of the model. Local methods (e.g., Local Interpretable Model-agnostic Explanations-LIME [225], Break Down [226], Ceteris Paribus [226]) explain a single prediction made by the model [221,227].

Second, the explanation stage is concerned with defining the time of the learning process. Pre-hoc methods explain the data used to develop models [227]. Ante-hoc methods (e.g., rule-based [228]) apply explainability during the development and design of the model. Post-hoc methods perform explainability after the development of the model (e.g., LIME, SHAP, Feature Importance, Break Down, Ceteris Paribus) [221,226,227].

Third, for explanation approach, methods are model-specific or model-agnostic. Model-specific methods (e.g., Gradient-weighted Class Activation Mapping-Grad-CAM [229], Sensitivity Analysis [230], Heatmaps [221,227], Instance-wise variable selection (INVASE) [231], and belief rule-based inference methodology [228]) are applied to ML or DL models with a specific structure or architecture. Model-agnostic methods (e.g., SHAP, LIME, Feature Importance, Break Down, Ceteris Paribus) can be applied to any ML algorithm no matter how complicated it is, treating the model as a black box [221,227].

Additional criteria for classifying explainability methods are based on the problem type (e.g., classification, regression), the input data (e.g., numerical, categorical, pictorial, textual, time series, vectors), and the schema or the output format of the explanation (e.g., numerical, rule-based, textual, visual, mixed) [221,227].

According to literature, healthcare model explainability is performed on the tasks this paper focuses on, i.e., sepsis detection and prediction, septic shock detection and prediction, sepsis mortality prediction, length of stay assessment, and hospitalization risk after ED admission, as well as ICU readmission risk, totaling 56 papers.

Several explainable AI methods are applied to sepsis detection and prediction models. For sepsis prediction, most studies use ML built-in feature importance [112,117,119,124,129,135,136,141,232,233] and SHAP [119,128,233–236]. Sensitivity analysis [232,237], LIME [127,238], heatmaps [239] and Grad-CAM [234] are also used. Regarding septic shock detection, a study [145] uses built-in feature importance, and to explain septic shock prediction algorithms studies use built-in feature importance [123] and SHAP [128]. For sepsis mortality prediction models, built-in feature importance

[113,240–242], SHAP [240,241,243–246], LIME [127,245,247], Break Down [241], Ceteris Paribus [241], and INVASE [242] are the explainable AI techniques used in literature.

For length of stay prediction explainability, model built-in feature importance [159,164,169–172,174,178,182,184] and SHAP [163,173,185] are applied on the best performing models. In explaining hospitalization/ICU admission predictions after ED admission, according to studies found in this paper, built-in feature importance [192,195,197,198,201, 204,206,208] and SHAP [202–204] are applied.

Explainability is also performed for ICU readmission prediction. Specifically, studies use SHAP [248,249], LIME [248], and an extended belief rule-based (EBRB) system [250].

Overall, XAI has been applied for ICU readmission, ICU LOS, sepsis onset, mortality, and sepsis mortality predictions. The applied algorithms were mainly post-hoc including model-specific and model-agnostic methods. Most of the studies use model-agnostic methods, like Feature Importance (35, 63%), and SHAP (22, 39%), with some studies using more than one. Appropriate explanations should be considered, as they can lead to confidence and trustworthiness of predictions by healthcare professionals and the ability to translate algorithms in the clinical setting.

6. Future Directions

This overview aims to summarize studies that target reduced sepsis infections and sepsis mortality and improved resource allocation. Hence, guidance from a data science perspective can be deduced to achieve maximum model performance, improved patient care and reduced healthcare cost.

Data size and type play an important role in model performance. Data are getting larger as interoperability increases, more countries and clinical sites share their data [159,188], and dataset(s) are being updated [147,187]. The number of modalities keeps increasing [146,200]. They include text (doctor's notes), images (e.g., X-rays, MRIs, ultrasounds), and video (e.g., echocardiogram, electrocardiogram, electroencephalogram). In this way, a holistic approach to a patient's condition is offered, capturing diverse factors that influence the clinical outcome, potentially leading to more accurate predictions and earlier interventions. Additionally, there is strong interest in using Generative AI [251,252] to generate data for rare disease patients or for cases where text, image, and/or video data are limited. Adopting more and higher-modality data for sepsis patients should improve performance and help identify relevant disease patterns.

Healthcare professionals' input in features included in models should also be prioritized. As comorbidities are common in patients, special approaches, procedures, and/or measurements might be required to manage the disease. Cohorts of patients with multiple diagnoses could be used for predicting multiple conditions, taking into account the doctors' recommendations for optimal study designs/cohort treatment.

Data stratified by initial diagnosis can provide more accurate predictions. As different diseases might require different treatment and length of stay, predicting LOS based on event on admission (e.g., stroke, sepsis) can improve predictions. Sepsis prognosis is associated with different diseases, age groups and time of onset (see Table 1), which make these groups possible cohorts for predictions.

Deep learning techniques also follow an evolution. More commonly used Convolutional Neural Networks, either for time series, image, or video data, are now substituted/complemented by Transformer-based models [202]. They can handle bigger volumes of data and different modalities or combinations of modalities, capturing more complex patterns more efficiently. Transformers excel at capturing long-range dependencies in sequential data, using a self-attention mechanism to consider all parts of the input data at once, allowing for parallelization and significantly faster training times, especially on

GPUs. This makes transformers more scalable and efficient in handling large datasets. Transformers often generalize better compared to traditional models, especially when fine-tuned on domain-specific data. Open access pretrained transformer models can be fine-tuned on smaller datasets, resulting in high performance with fewer training samples compared to traditional methods. This is advantageous when labelled data are limited. This ability to use transfer learning in healthcare allows for quicker deployment and more efficient use of computational resources. Text transformers can be used for NLP, vision transformers (ViTs) can be used for image classification, while time-series transformers are effective for predicting sequential patterns. Their positional encoding mechanism allows them to work with any type of sequence data and retain the relative order of elements, which is not always possible with traditional models. Transformers are used in M-LLMs to handle combinations of different types of data input and analyse, interpret, and generate clinical reports, personalized treatment plans, medical images and videos. In the case of sepsis prediction, transformer-based models can use image data, genomic data, longer and/or higher granularity time series data and clinical notes to explore more dependencies within variables efficiently. Some popular transformer-based models trained specifically on healthcare data are BioBERT [253], MedBERT [254], BEHRT [255], BioGPT [256], Med-Palm [257], Foresight [258], and Gemini [259]. All these models, apart from Med-Palm and Gemini, are open access models.

7. Conclusions

AI capabilities can handle big, heterogeneous, multimodal, and irregularly sampled healthcare and ICU data, providing early predictions for disease prevention and interventions, hugely benefiting patient wellbeing, the society, and the economy at large. A literature overview in predicting sepsis, length of stay and hospitalization/ICU admission after ED arrival is provided to guide new researchers in the area. Critical challenges faced when using healthcare data, developing AI models, and integrating them in clinical settings while considering ethical aspects, are further documented. Explainable AI methods can have a transformative impact on the adoption of AI methods in medicine. To improve model performance, future work is expected to investigate sepsis prediction and other clinical outcomes using multimodal data, Transformer-based models, specific disease cohorts, and be informed and driven by clinical knowledge.

Author Contributions: C.S.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. A.N.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. W.A.S.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. C.-A.A.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. I.P.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. K.K.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. G.D.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. S.K.: Conceptualization, methodology, software, valida-

tion, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. E.P.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. D.N.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. X.P.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. F.G.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. Z.A.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. N.I.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. L.P.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. A.V.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. M.S.P.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. C.S.P.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. A.S.P.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration, funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been performed in the context of the project Hospital Transformation through Artificial Intelligence—HOSPAITAL, CODEVELOP-ICTHEALTH/0322/0071, which is co-financed by the Republic of Cyprus through the Research and Innovation Foundation and the European Regional Development Fund. The APC was funded by CODEVELOP-ICTHEALTH/0322/0071.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: Authors Dimitris Ntalaperas and Xanthi Papageorgiou were employed by the company UBITECH Limited. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The project was fully funded by the Republic of Cyprus through the Research and Innovation Foundation and the European Regional Development Fund. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Glossary

3LSCG	Three-Level Sequential Cascade Generalization
ACC	Accuracy
ACNN	Adaptive Convolutional Neural Network
ACS-NSQIP	American College of Surgeons National Surgical Quality Improvement Program

AdaBoost	Adaptive Boosting
ADASYN	Adaptive Synthetic sampling approach for imbalanced learning
AE	Auto-Encoder
AF	Atrial Fibrillation
AIC	Akaike Information Criterion
AKI	Acute Kidney Infection
AMCUS	Academic Medical Center in the US
ANOVA	Analysis of Variance test
APACHE	Acute Physiologic and Chronic Health Evaluation,
APR	Area under Precision-Recall curve
ARDS	Acute Respiratory Distress Syndrome
ARF	Acute Respiratory Failure
Att-1DCNN	Attention-Embedded 1-Dimention Convolutional Neural Network
AUC	Area Under receiver operating characteristic Curve
AVHHA	Analytics Vidhya Hackathon about Healthcare Analytics
BCT	Box Cox Transformation
BE	Backward Elimination
BertViz	Interactive tool that can visualize attention in transformer language models
BiLSTM	Bidirectional Long Short Term Memory
BlueBERT	Biomedical Language Understanding Evaluation Bidirectional Encoder Representations from Transformers
BMI	Body Mass Index
BT	Bagged Tree
BUN	Blood Urea Nitrogen,
C	Channel-wise
CA	Clinical Aggregates
CAD	Coronary Artery Disease
CAn	Correlation Analysis
CB	Class Balancing,
CBC	Complete Blood Count
CCI	Charlson Comorbidity Index
CCM	Clinical Center in Madrid
CHD	Congenital Heart Defects
CHED	4 clinically heterogeneous academically affiliated emergency departments
CMS	Centers for Medicare & Medicaid Services criteria
CMT	Chi-Mei medical center in southern Taiwan
CNN	Convolutional Neural Network
CR	Client Recruitment
CRP	C-Reactive Protein
CS	Clinical Significance
CT	Computer Tomography
CUHICU	Chiba University Hospital ICU in Japan
CWK	Cohen's Weighted Kappa
D	Discretization
DAD	Dascena Analysis Dataset
DBP	Diastolic Blood Pressure
DF-Mdl:	Data Fusion Model
DFSP	Double Fusion Sepsis Predictor
DHHS	Department of Health & Human Services in the US
DIIC	DII Challenge 2019
DL	Deep Learning
DMM	Danish Municipality Multi-center data outside ICU
DS	Deep Supervision
DTW-KNN	Dynamic Time Warping-K-Nearest Neighbors

DUHS	Duke University Health System
ECG	Electrocardiography
ECMO	Use of extracorporeal membrane oxygenation
ED	Emergency Department,
EDA	Exploratory Data Analysis
EDCUS	5 Emergency Departments in Colorado US
EDHS	Emergency Department Hospital in Seoul in South Korea
EHR	Electronic Health Record
eICU	Collaborative Research Database
ELS	Ensemble Learning Strategy
EM	Ensembling Model
EMB	Embedding
ENN	Edited Nearest Neighbours
ESI	Emergency Severity Index
ESR	Erythrocyte Sedimentation Rate
FE	Feature Engineering
FET	Fisher's Exact Test,
FEX	Feature Extraction
FL	Federated Learning
FL-SRC	Federated Learning-recruited clients make up the federation, 10% of which partake in each training round
FNF	Femoral Neck Fracture
FS	Feature Selection
GA	Genetic Algorithm
GB	Gradient Boosting
GCS	Glasgow Comma Score
GIRB	Geisinger Institutional Review Board,
GLMM	Generalized Linear Mixed Model
GNN	Graph Neural Network
GPCICU	Guangdong Provincial Cardiovascular Institute ICU
GRU	Gated Recurrent Unit
GW	General Ward
HAN	Hierarchical Attention Network
HHTCM	Huangpi Hospital of Traditional Chinese Medicine
HiRID	High time-Resolution ICU Dataset
HMV	Handling Missing Values
HO	Handling Outliers
HR	Heart Rate
HSSC	Health System Sepsis Committee criteria
HTDV	Hospital for Tropical Diseases in Vietnam
HTN	Hypertension
HTS	Handling Time Series
HUM	Handling Units of Measurement
ICD	International Classification of Diseases
ICDC	ICD code Conversion
ICU	Intensive Care Unit
ICUS	Intensive Care Unit department in Shanghai hospital
ICUUS	ICU Department of Hospital in US
IE	Integer Encoding
IF	Important Features
IG	Information Gain
IHICU	Iranian local Hospitals ICU
IL-6	Interleukin-6
IPS	International Patient Summary

IR	Image Resizing
ISS	Injury Severity Score
IVT	Intravenous Therapy ordered or scheduled prior to emergency department visit
KCC	Kendall Correlation Coefficient
KFSH&RC	King Faisal Specialist Hospital & Research Centre hospital in Saudi Arabia
KFUH:	King Fahad University Hospital
KST	Kolmogorov–Smirnov Test
KW	Kruskal–Wallis test
LASSO-LNR	Least Absolute Shrinkage and Selection Operator Linear Regression
LDH	Lactate Dehydrogenase
LE	Label Encoding
LF	Removal of features of Low Frequency
LGBM	Light Gradient Boosting Machine
LIME	Local Interpretable Model-Agnostic Explanation
LNR	Linear Regression
LOS	Length Of Stay
LR	Logistic Regression
LSTM	Long Short-Term Memory
LSTM-C-DS	Long Short Term Memory Channel-wise with Deep Supervision
LSTM-H	Long Short Term Memory-Hybrid
LSTM-MPNN	Long Short Term Memory-Message Passing Neural Networks
MacBERT	Chinese version of bidirectional encoder representations from transformers (BERT),
MAD	MAD: Mean Absolute Difference
MAE	Mean Absolute Error
MGB	Mass General Brigham Healthcare database
MGP-RNN	Multi-output Gaussian Processes and Recurrent Neural Networks
MIMIC	Medical Information Mart for Intensive Care
ML	Machine Learning
MPM	Mortality Predictive Mode
MSE	Mean Squared Error
MUSH	Midwest Hospital in US
MWU	Mann–Whitney U test
N	Normalization
NEED	Netherlands Emergency Department Evaluation Database
NHAMCS	National Hospital and Ambulatory Medical Care Survey
NHCRD	National Hospital Care Research Database
NIH	2 major acute Northern Ireland Hospitals
NLP	Natural Language Processing
NLR	Neutrophil-Lymphocyte Ratio
NM	Not Mentioned
NN	Neural Network
NN-GCN	Neural Network combined with Graph Convolutional Network
NOS	Nasal Oxygen Support
NTUH	National Taiwan University Hospital
OHE	One Hot Encoding
PAVE	Pattern Attention model with Value Embedding
PCA	Principal Component Analysis
PCC	Pearson's Correlation Coefficient
PCR	Principal Component Regression
PCT	Procalcitonin
PFI	Permutation Feature Importance
PNCC	PhysioNet Computing in Cardiology 2019 Challenge
PNUYH	Pusan National University Yangsan Hospital ICU

PPG	Photoplethysmography
PR	Precision
PTT	Partial Thromboplastin Time
PUMCH	Peking Union Medical College Hospital
QAH	Quaternary Academic Hospital
R	Relief
R2	Coefficient of determination
RF	Random Forest
RFE	Recursive Feature Elimination
RNN	Recurrent Neural Network
RR	Respiratory Rate
RRT	Renal Replacement Therapy
RST	Rank Sum Test
RU	Random Undersampling
S	Standardization
SA	Simulated Annealing
SAPS	Simplified Acute Physiologic Score
SBFCM	Statistical-Based Fuzzy Cognitive Maps
SBP	Systolic Blood Pressure
SCUSH	Suez Canal University Specialized Hospital
SEN	Sensitivity
SFM	Selection From Model
SFS	Stepwise-Forward Selection
SH	Singapore government-based Hospital
SHAP	Shapley Additive Explanations
SIRS	Systemic Inflammatory Response Syndrome
SKB	Selection of K-Best
SMOTE	Synthetic Minority Over Sampling Technique
SOFA	Sequential (Sepsis-related) Organ Failure Score
SPE	Specificity
SRHM	San Rafaele Hospital Emergency Department in Milan
SRPH	Dr Soekardjo Regional Public Hospital in Indonesia
ST	Smoothing Time series
STT	Student's t test
SUN	Serum Urea Nitrogen level
SVM	Support Vector Machine
SVR	Support Vector Regression
SWT	Shapiro-Wilk's Test
T	Tokenization
T2DM	Type 2 Diabetes Mellitus
TabNet	Existing encoder
T-ADAB	Adaboost integrated with Tabu Search
TCN	Temporal Convolutional Network
TED-ICU	Taipei Medical University Hospital Electronic Medical Record System
TF-IDF	Term Frequency-Inverse Document Frequency
THC	3 Tertiary care Hospital Eds in China
THI	Tabba Heart Institute in Pakistan
THMC	Trauma center of Hamad Medical Corporation
THS	Tertiary Hospital in Seoul in South Korea
TKA	Total Knee Arthroplasty
TL	Tomek Links algorithm
TMUGH	Tianjin Medical University General Hospital
TMUSH	Taipei Medical University-Shuang Ho Hospital
TP	Text Processing

TRDGU	TraumaRegister of the German Trauma Society
T-SNE	t-distributed Stochastic Neighbor Embedding
TT	t-test
TTH	Teaching Hospital in Tainan Taiwan
USMH	Metropolitan-area Hospital in US
UTMB	University of Texas Medical Branch at Galveston
V	Keeping features according to their Variance
VI	Variable Importance
VIS	Vasoactive Inotropic Score in surgery
WBC	White Blood Cell count
WFC	Word Frequency Counting
WL	Wilks's Lambda
WUHICU	Wuhan Union Hospital ICU
XGB	Extreme Gradient Boosting
YCDTSH	Yedikule Chest Diseases and Thoracic Surgery Training & Research Hospital
YH	Hospital in Yueqing China
YUSH	Yonsei University Severance Hospital in Rep. of Korea
ZUH	First Affiliated Hospital ICU of Zhengzhou University
$\chi 2$	chi-squared test

Appendix A

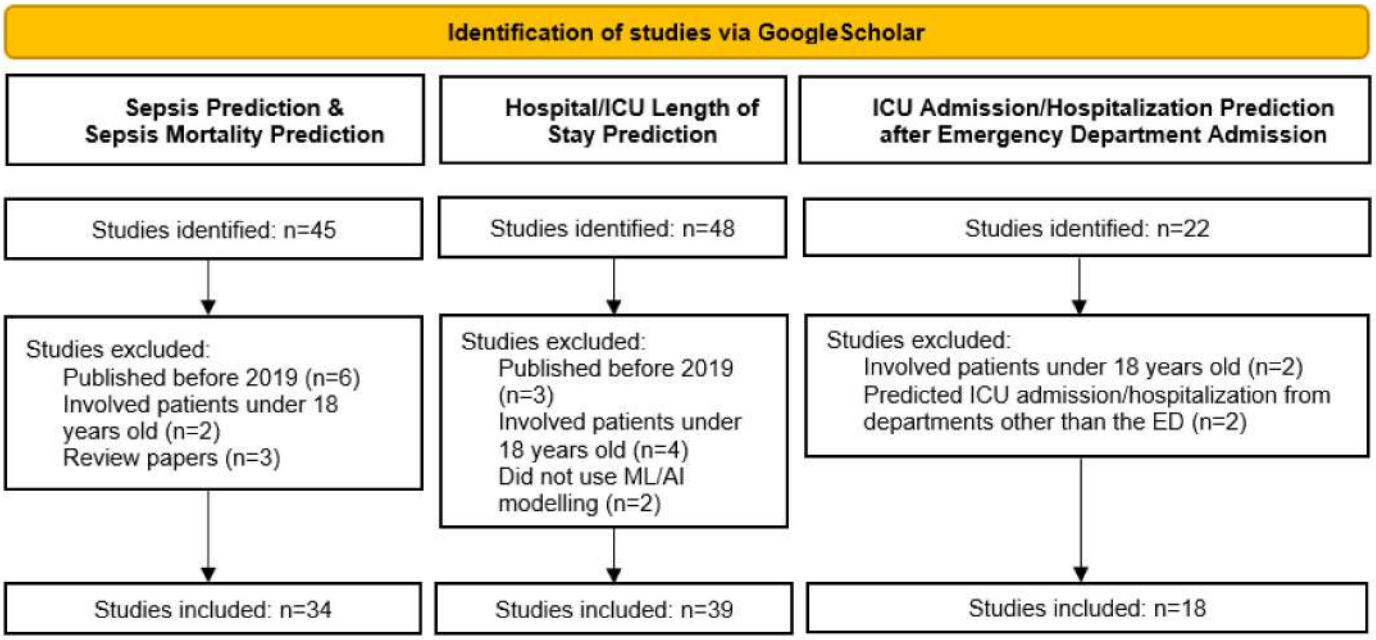

Figure A1. Flow Diagram of inclusion of studies.

References

1. Alexandropoulou, C.-A.; Panagiotopoulos, I.; Kleanthous, S.; Dimitrakopoulos, G.; Constantinou, I.; Politi, E.; Ntalaperas, D.; Papageorgiou, X.; Stylianides, C.; Ioannides, N.; et al. AI-Enabled Solutions, Explainability and Ethical Concerns for Predicting Sepsis in ICUs: A Systematic Review. In Proceedings of the 2023 IEEE 19th International Conference on e-Science (e-Science), Limassol, Cyprus, 9–13 October 2023; pp. 1–9.
2. Panayides, A.S.; Amini, A.; Filipovic, N.D.; Sharma, A.; Tsaftaris, S.A.; Young, A.A.; Foran, D.J.; Do, N.V.; Golemati, S.; Kurc, T.; et al. AI in Medical Imaging Informatics: Current Challenges and Future Directions. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 1837–1857. [CrossRef] [PubMed]
3. Doshi-Velez, F.; Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv* **2017**, arXiv:1702.08608.
4. Haque, A.B.; Islam, A.N.; Mikalef, P. Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. *Technol. Forecast. Soc. Change* **2022**, *186 Pt A*, 122120. [CrossRef]

5. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
6. Holzinger, A.; Biemann, C.; Pattichis, C.S.; Kell, D.B. What do we need to build explainable AI systems for the medical domain? *arXiv* **2017**, arXiv:1712.09923.
7. Memon, M.; Li, J.; Haq, A.; Memon, M. Breast cancer detection in the Iot health environment using modified recursive feature selection. *Wirel. Commun. Mob.* **2019**, *2019*, 5176705. [CrossRef]
8. Akazawa, M.; Hashimoto, K. Artificial Intelligence in Ovarian Cancer Diagnosis. *Anticancer Res.* **2020**, *40*, 4795–4800. [CrossRef]
9. Azar, A.S.; Rikan, S.B.; Naemi, A.; Mohasefi, J.B.; Pirnejad, H.; Mohasefi, M.B.; Wiil, U.K. Application of machine learning techniques for predicting survival in ovarian cancer. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 234. [CrossRef]
10. Khourdifi, Y.; Bahaj, M. Applying Best Machine Learning Algorithms for Breast Cancer Prediction and Classification. In Proceedings of the 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, Morocco, 5–6 December 2018; pp. 1–5.
11. Naji, M.A.; El Filali, S.; Aarika, K.; Benlahmar, E.H.; Abdelouhahid, R.A.; Debauche, O. Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis. *Procedia Comput. Sci.* **2021**, *191*, 487–492. [CrossRef]
12. Vijayalakshmi, M.M. Melanoma Skin Cancer Detection using Image Processing and Machine Learning. *Int.-Natl. J. Trend Sci. Res. Dev. (Ijtsrd)* **2019**, *3*, 780–784.
13. Sun, W.; Zheng, B.; Qian, W. Computer aided lung cancer diagnosis with deep learning algorithms. In Proceedings of the Medical Imaging 2016: Computer-Aided Diagnosis, San Diego, CA, USA, 27 February–3 March 2016; SPIE: Bellingham, WA, USA, 2016; Volume 9785. [CrossRef]
14. Chaki, J.; Ganesh, S.T.; Cidham, S.K.; Theertan, S.A. Machine learning and artificial intelligence based diabetes mellitus detection and self-management: A systematic review. *J. King Saud Univ. Comput. Inf. Sci.* **2020**, *34*, 3204–3225. [CrossRef]
15. Kaur, H.; Kumari, V. Predictive modelling and analytics for diabetes using a machine learning approach. *Appl. Comput. Inform.* **2022**, *18*, 90–100. [CrossRef]
16. Maniruzzaman; Rahman, J.; Ahammed, B.; Abedin, M. Classification and prediction of diabetes disease using machine learning paradigm. *Health Inf. Sci. Syst.* **2020**, *8*, 7. [CrossRef] [PubMed] [PubMed Central]
17. Xiong, X.-L.; Zhang, R.-X.; Bi, Y.; Zhou, W.-H.; Yu, Y.; Zhu, D.-L. Machine Learning Models in Type 2 Diabetes Risk Prediction: Results from a Cross-sectional Retrospective Study in Chinese Adults. *Curr. Med. Sci.* **2019**, *39*, 582–588. [CrossRef] [PubMed]
18. Deberneh, H.M.; Kim, I. Prediction of Type 2 Diabetes Based on Machine Learning Algorithm. *Int. J. Environ. Res. Public Health* **2021**, *18*, 3317. [CrossRef]
19. Xie, Z.; Nikolayeva, O.; Luo, J.; Li, D. Building Risk Prediction Models for Type 2 Diabetes Using Machine Learning Techniques. *Prev. Chronic Dis.* **2019**, *16*, E130. [CrossRef] [PubMed] [PubMed Central]
20. Woldaregay, A.Z.; Årsand, E.; Walderhaug, S.; Albers, D.; Mamykina, L.; Botsis, T.; Hartvigsen, G. Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes. *Artif. Intell. Med.* **2019**, *98*, 109–134. [CrossRef] [PubMed]
21. Narin, A.; Isler, Y.; Ozer, M. Early prediction of Paroxysmal Atrial Fibrillation using frequency domain measures of heart rate variability. In Proceedings of the 2016 Medical Technologies National Congress (TIPTEKNO), Antalya, Turkey, 27–29 October 2016. [CrossRef]
22. Hasan, N.; Bao, Y. Comparing different feature selection algorithms for cardiovascular disease prediction. *Health Technol.* **2020**, *11*, 49–62. [CrossRef]
23. Drożdż, K.; Nabrdalik, K.; Kwiendacz, H.; Hendel, M.; Olejarz, A.; Tomasik, A.; Bartman, W.; Nalepa, J.; Gumprecht, J.; Lip, G.Y.H. Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: A machine learning approach. *Cardiovasc. Diabetol.* **2022**, *21*, 240. [CrossRef]
24. Shah, D.; Patel, S.; Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. *SN Comput. Sci.* **2020**, *1*, 345. [CrossRef]
25. Bhatt, C.M.; Patel, P.; Ghetia, T.; Mazzeo, P.L. Effective Heart Disease Prediction Using Machine Learning Techniques. *Algorithms* **2023**, *16*, 88. [CrossRef]
26. Haq, A.U.; Li, J.P.; Memon, M.H.; Nazir, S.; Sun, R. A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mob. Inf. Syst.* **2018**, *8*, 3860146. [CrossRef]
27. Alotaibi, F.S. Implementation of Machine Learning Model to Predict Heart Failure Disease. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*, 261–268. [CrossRef]
28. Tuli, S.; Basumatary, N.; Gill, S.S.; Kahani, M.; Arya, R.C.; Wander, G.S. HealthFog: An ensemble deep learning based smart healthcare system for automatic diagnosis of heart diseases in integrated IoT and fog computing environments. *Future Gener. Comput. Syst.* **2019**, *104*, 187–200. [CrossRef]
29. Isravel, D.P.; Silas, S.V.P.D. Improved heart disease diagnostic IoT model using machine learning techniques. *Neuroscience* **2020**, *9*, 4442–4446.

30. Painuli, D.; Mishra, D.; Bhardwaj, S.; Aggarwal, M. Forecast and prediction of COVID-19 using machine learning. In *Data Science for COVID-19*; Academic Press: New York, NY, USA, 2021; pp. 381–397. [CrossRef] [PubMed Central]
31. Stylianides, C.; Malialis, K.; Kolios, P.A. Study of Data-Driven Methods for Adaptive Forecasting of COVID-19 Cases. In *International Conference on Artificial Neural Networks*; Springer: Cham, Switzerland, 2023; pp. 62–74.
32. Yahya, B.M.; Yahya, F.S.; Thannoun, R.G. COVID-19 prediction analysis using artificial intelligence procedures and GIS spatial analyst: A case study for Iraq. *Appl. Geomat.* **2021**, *13*, 481–491. [CrossRef]
33. Gupta, M.; Jain, R.; Arora, S.; Gupta, A.; Awan, M.J.; Chaudhary, G.; Nobanee, H. AI-Enabled COVID-19 Outbreak Analysis and Prediction: Indian States Vs. Union Territories. *Comput. Mater. Contin.* **2021**, *67*, 933–950. [CrossRef]
34. Gao, Y.; Cai, G.-Y.; Fang, W.; Li, H.-Y.; Wang, S.-Y.; Chen, L.; Yu, Y.; Liu, D.; Xu, S.; Cui, P.-F.; et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nat. Commun.* **2020**, *11*, 5033. [CrossRef]
35. Pourhomayoun, M.; Shakibi, M. Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health* **2021**, *20*, 100178. [CrossRef]
36. Reyna, M.A.; Josef, C.S.; Jeter, R.; Shashikumar, S.P.; Westover, M.B.; Nemati, S.; Clifford, G.D.; Sharma, A. Early prediction of sepsis from clinical data: The PhysioNet/Computing in Cardiology Challenge 2019. *Crit. Care Med.* **2020**, *48*, 210–217. [CrossRef]
37. Olang, O.; Mohseni, S.; Shahabinezhad, A.; Hamidianshirazi, Y.; Goli, A.; Abolghasemian, M.; Shafiee, M.A.; Aarabi, M.; Alavinia, M.; Shaker, P. Artificial Intelligence-Based Models for Prediction of Mortality in ICU Patients: A Scoping Review. *J. Intensiv. Care Med.* **2024**, 08850666241277134. [CrossRef] [PubMed]
38. Yan, M.Y.; Gustad, L.T.; Nytrø, Ø. Sepsis prediction, early detection, and identification using clinical text for machine learning: A systematic review. *J. Am. Med. Inform. Assoc.* **2022**, *29*, 559–575. [CrossRef] [PubMed]
39. Moor, M.; Rieck, B.; Horn, M.; Jutzeler, C.R.; Borgwardt, K. Early Prediction of Sepsis in the ICU Using Machine Learning: A Systematic Review. *Front. Med.* **2021**, *8*, 607952. [CrossRef] [PubMed]
40. 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.; Thoral, P.; Ercole, A.; et al. Machine learning for the prediction of sepsis: A systematic review and meta-analysis of diagnostic test accuracy. *Intensiv. Care Med.* **2020**, *46*, 383–400. [CrossRef]
41. Yadgarov, M.Y.; Landoni, G.; Berikashvili, L.B.; Polyakov, P.A.; Kadantseva, K.K.; Smirnova, A.V.; Kuznetsov, I.V.; Shemetova, M.M.; Yakovlev, A.A.; Likhvantsev, V.V. Early detection of sepsis using machine learning algorithms: A systematic review and network me-ta-analysis. *Front. Med.* **2024**, *11*, 1491358. [CrossRef]
42. Islam, K.R.; Prithula, J.; Kumar, J.; Tan, T.L.; Reaz, M.B.I.; Sumon, S.I.; Chowdhury, M.E.H. Machine Learning-Based Early Prediction of Sepsis Using Electronic Health Records: A Systematic Review. *J. Clin. Med.* **2023**, *12*, 5658. [CrossRef]
43. Gao, Y.; Wang, C.; Shen, J.; Wang, Z.; Liu, Y.; Chai, Y. Systematic review and network meta-analysis of machine learning algorithms in sepsis prediction. *Expert Syst. Appl.* **2023**, *245*, 122982. [CrossRef]
44. Suresh, V.; Singh, K.K.; Vaish, E.; Gurjar, M.; Am, A.; Khulbe, Y.; Muzaffar, S.; Nambi, A.A. Artificial Intelligence in the Intensive Care Unit: Current Evidence on an Inevitable Future Tool. *Cureus* **2024**, *16*, e59797. [CrossRef]
45. Atkinson, A.J., Jr.; Colburn, W.A.; DeGruttola, V.G.; DeMets, D.L.; Downing, G.J.; Hoth, D.F.; Oates, J.A.; Peck, C.C.; Schooley, R.T.; Biomarkers Definitions Working Group; et al. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* **2001**, *69*, 89–95.
46. Keijzers, G.; Fatovich, D.M.; Egerton-Warburton, D.; Cullen, L.; A Scott, I.; Glasziou, P.; Croskerry, P. Deliberate clinical inertia: Using meta-cognition to improve decision-making. *Emerg. Med. Australas.* **2018**, *30*, 585–590. [CrossRef]
47. A Scott, I.; Soon, J.; Elshaug, A.G.; Lindner, R. Countering cognitive biases in minimising low value care. *Med. J. Aust.* **2017**, *206*, 407–411. [CrossRef] [PubMed]
48. Heffernan, A.J.; Denny, K.J. Host Diagnostic Biomarkers of Infection in the ICU: Where Are We and Where Are We Going? *Curr. Infect. Dis. Rep.* **2021**, *23*, 4. [CrossRef] [PubMed]
49. Astagimath, M.; Aryapu, R.; Patil, V.; Doddamani, S. C-Reactive Protein and Lactate Dehydrogenase in Intensive Care Unit and Nonintensive Care Unit COVID-19 Patients—A Retrospective Study. *APIK J. Intern. Med.* **2023**, *11*, 33–36. [CrossRef]
50. Li, Z.; Wang, H.; Liu, J.; Chen, B.; Li, G. Serum soluble triggering receptor expressed on myeloid cells-1 and pro-calcitonin can reflect sepsis severity and predict prognosis: A prospective cohort study. *Mediat. Inflamm.* **2014**, *2014*, 641039. [CrossRef]
51. Martini, A.; Gottin, L.; Mélot, C.; Vincent, J.L. A prospective evaluation of the Infection Probability Score (IPS) in the intensive care unit. *J. Infect.* **2008**, *56*, 313–318. [CrossRef]
52. Neocleous, A.; Papaioannou, M.; Savva, P.; Miguel, F.; Panayides, A.; Antoniou, Z.; Neofytou, M.; Schiza, E.C.; Neokleous, K.; Constantinou, I.; et al. The International Patient Summary: Proposal for a National Implementation for Cyprus. In Proceedings of the 2022 E-Health and Bioengineering Conference (EHB), Iasi, Romania, 17–19 November 2022; pp. 1–5.
53. Stachon, A.; Becker, A.; Holland-Letz, T.; Friese, J.; Kempf, R.; Krieg, M. Estimation of the Mortality Risk of Surgical Intensive Care Patients Based on Routine Laboratory Parameters. *Eur. Surg. Res.* **2008**, *40*, 263–272. [CrossRef]
54. Ho, K.M.; Lipman, J. An Update on C-reactive Protein for Intensivists. *Anaesth. Intensiv. Care* **2009**, *37*, 234–241. [CrossRef]

55. Qu, R.; Hu, L.; Ling, Y.; Hou, Y.; Fang, H.; Zhang, H.; Liang, S.; He, Z.; Fang, M.; Li, J.; et al. C-reactive protein concentration as a risk predictor of mortality in intensive care unit: A multicenter, prospective, observational study. *BMC Anesthesiol.* **2020**, *20*, 292. [CrossRef]
56. Póvoa, P.; Coelho, L.; Almeida, E.; Fernandes, A.; Mealha, R.; Moreira, P.; Sabino, H. Early identification of intensive care unit-acquired infections with daily monitoring of C-reactive proHpective observational study. *Crit. Care* **2006**, *10*, R63. [CrossRef]
57. Coster, D.; Wasserman, A.; Fisher, E.; Rogowski, O.; Zeltser, D.; Shapira, I.; Bernstein, D.; Meilik, A.; Raykhshtat, E.; Halpern, P.; et al. Using the kinetics of C-reactive protein response to improve the differential diagnosis between acute bacterial and viral infections. *Infection* **2020**, *48*, 241–248. [CrossRef]
58. Yang, A.-P.; Liu, J.; Yue, L.-H.; Wang, H.-Q.; Yang, W.-J.; Yang, G.-H. Neutrophil CD64 combined with PCT, CRP and WBC improves the sensitivity for the early diagnosis of neonatal sepsis. *Clin. Chem. Lab. Med. (CCLM)* **2016**, *54*, 345–351. [CrossRef] [PubMed]
59. Henriquez-Camacho, C.; Losa, J. Biomarkers for Sepsis. *BioMed Res. Int.* **2014**, *2014*, 547818. [CrossRef] [PubMed] [PubMed Central]
60. Alonso-Martínez, J.L.; Llorente-Diez, B.; Echegaray-Agara, M.; Olaz-Preciado, F.; Urbieta-Echezarreta, M.; González-Arencibia, C. C-reactive protein as a predictor of improvement and readmission in heart failure. *Eur. J. Heart Fail.* **2002**, *4*, 331–336. [CrossRef]
61. Lobo, S.M.; Lobo, F.R.; Bota, D.P.; Lopes-Ferreira, F.; Soliman, H.M.; Meélot, C.; Vincent, J.L. C-reactive protein levels correlate with mortality and organ failure in critically ill patients. *Chest* **2003**, *123*, 2043–2049. [CrossRef]
62. Ho, K.M.; Dobb, G.J.; Lee, K.Y.; Towler, S.C.; Webb, S.A. C-reactive protein concentration as a predictor of intensive care unit readmission: A nested case-control study. *J. Crit. Care* **2006**, *21*, 259–265. [CrossRef]
63. Ho, K.M.; Lee, K.Y.; Dobb, G.J.; Webb, S.A.R. C-reactive protein concentration as a predictor of in-hospital mortality after ICU discharge: A prospective cohort study. *Intensiv. Care Med.* **2007**, *34*, 481–487. [CrossRef]
64. Pradhan, A.D.; Manson, J.E.; Rossouw, J.E.; Siscovick, D.S.; Mouton, C.P.; Rifai, N.; Wallace, R.B.; Jackson, R.D.; Pettinger, M.B.; Ridker, P.M. Inflammatory biomarkers, hormone replacement therapy, and incident coronary heart disease: Pro-spective analysis from the Women's Health Initiative observational study. *JAMA* **2002**, *288*, 980–987. [CrossRef]
65. Dornbusch, H.J.; Strenger, V.; Sovinz, P.; Lackner, H.; Schwinger, W.; Kerbl, R.; Urban, C. Non-infectious causes of elevated procalcitonin and C-reactive protein serum levels in pediatric patients with hematologic and oncologic disorders. *Support. Care Cancer* **2008**, *16*, 1035–1040. [CrossRef]
66. Meyer, Z.C.; Schreinemakers, J.M.J.; Mulder, P.G.H.; de Waal, R.A.L.; Ermens, A.A.M.; van der Laan, L. The Role of C-Reactive Protein and the SOFA Score as Parameter for Clinical Decision Making in Surgical Patients during the Intensive Care Unit Course. *PLoS ONE* **2013**, *8*, e55964. [CrossRef]
67. Chandran, R.T.; Vadhul, P.B. Correlation of c-reactive protein (crp) with icu COVID-19 ards mortality in adults. *Chest* **2022**, *162*, A725. [CrossRef]
68. Shine, B.; de Beer, F.; Pepys, M. Solid phase radioimmunoassays for human C-reactive protein. *Clin. Chim. Acta* **1981**, *117*, 13–23. [CrossRef] [PubMed]
69. Schmit, X.; Vincent, J.L. The Time Course of Blood C-reactive Protein Concentrations in Relation to the Response to Initial Antimicrobial Therapy in Patients with Sepsis. *Infection* **2008**, *36*, 213–219. [CrossRef] [PubMed]
70. Reny, J.-L.; Vuagnat, A.; Ract, C.; Benoit, M.-O.; Safar, M.; Fagon, J.-Y. Diagnosis and follow-up of infections in intensive care patients: Value of C-reactive protein compared with other clinical and biological variables*. *Crit. Care Med.* **2002**, *30*, 529–535. [CrossRef] [PubMed]
71. Flores, J.M.; I Jiménez, P.; Rincón, D.; Márquez, J.; Navarro, H.; Muñoz, A.; Murillo, F. C reactive protein as marker of infection among patients with severe closed trauma. *Enfermedades Infecc. Microbiol. Clin.* **2001**, *19*, 61–65.
72. Gülcher, S.S.; Bruins, N.A.; Kingma, W.P.; Boerma, E.C. Elevated C-reactive protein levels at ICU discharge as a predictor of ICU outcome: A retrospective cohort study. *Ann. Intensiv. Care* **2016**, *6*, 5. [CrossRef]
73. Milenkovic, M.; Hadzibegovic, A.; Kovac, M.; Jovanovic, B.; Stanisavljevic, J.; Djikic, M.; Sijan, D.; Ladjevic, N.; Palibrk, I.; Djukanovic, M.; et al. D-dimer, CRP, PCT, and IL-6 Levels at Admission to ICU Can Predict In-Hospital Mortality in Patients with COVID-19 Pneumonia. *Oxidative Med. Cell. Longev.* **2022**, *2022*, 8997709. [CrossRef]
74. Picod, A.; Morisson, L.; de Roquetaillade, C.; Sadoune, M.; Mebazaa, A.; Gayat, E.; Davison, B.A.; Cotter, G.; Chousterman, B.G. Systemic Inflammation Evaluated by Interleukin-6 or C-Reactive Protein in Critically Ill Patients: Results from the FROG-ICU Study. *Front. Immunol.* **2022**, *13*, 868348. [CrossRef]
75. Hunter, C.A.; Jones, S.A. IL-6 as a keystone cytokine in health and disease. *Nat. Immunol.* **2015**, *16*, 448–457, Correction in *Nat. Immunol.* **2017**, *18*, 1271. [CrossRef]
76. Waage, A.; Brandtzaeg, P.; Halstensen, A.; Kierulf, P.; Espevik, T. The complex pattern of cytokines in serum from patients with meningococcal septic shock. Association between interleukin 6, interleukin 1, and fatal outcome. *J. Exp. Med.* **1989**, *169*, 333–338. [CrossRef]
77. Erez, A.; Shental, O.; Tchebiner, J.Z.; Laufer-Perl, M.; Wasserman, A.; Sella, T.; Guzner-Gur, H. Diagnostic and prognostic value of very high serum lactate dehydrogenase in admitted medical patients. *Isr. Med. Assoc. J.* **2014**, *16*, 439–443.
78. Li, Q.; Gong, X. Clinical significance of the detection of procalcitonin and C-reactive protein in the intensive care unit. *Exp. Ther. Med.* **2018**, *15*, 4265–4270. [CrossRef] [PubMed]

79. Zhang, Y.; Zhou, L. Diagnostic value of C-reactive protein and procalcitonin for bacterial infection in acute exacerbations of chronic obstructive pulmonary disease. *Zhong Nan Da Xue Xue Bao. Yi Xue Ban J. Cent. South Univ. Med. Sci.* **2014**, *39*, 939–943. [CrossRef]
80. Goodlet, K.J.; A Cameron, E.; Nailor, M.D. Low Sensitivity of Procalcitonin for Bacteremia at an Academic Medical Center: A Cautionary Tale for Antimicrobial Stewardship. *Open Forum Infect. Dis.* **2020**, *7*, ofaa096. [CrossRef] [PubMed]
81. Cavallazzi, R.; Bennin, C.-L.; Hirani, A.; Gilbert, C.; Marik, P.E. Review of A Large Clinical Series: Is the Band Count Useful in the Diagnosis of Infection? An Accuracy Study in Critically Ill Patients. *J. Intensiv. Care Med.* **2010**, *25*, 353–357. [CrossRef]
82. Ljungström, L.; Pernestig, A.-K.; Jacobsson, G.; Andersson, R.; Usener, B.; Tilevik, D. Diagnostic accuracy of procalcitonin, neutrophil-lymphocyte count ratio, C-reactive protein, and lactate in patients with suspected bacterial sepsis. *PLoS ONE* **2017**, *12*, e0181704. [CrossRef]
83. Westerdijk, K.; Simons, K.S.; Zegers, M.; Wever, P.C.; Pickkers, P.; de Jager, C.P.C. The value of the neutrophil-lymphocyte count ratio in the diagnosis of sepsis in patients admitted to the Intensive Care Unit: A retrospective cohort study. *PLoS ONE* **2019**, *14*, e0212861. [CrossRef]
84. Quintairos, A.; Pilcher, D.; Salluh, J.I. ICU scoring systems. *Intensiv. Care Med.* **2023**, *49*, 223–225. [CrossRef]
85. Jaganath, U. An Overview of Predictive Scoring Systems Used in ICU. 2020. Available online: https://anaesthetics.ukzn.ac.za/wp-content/uploads/2020/07/05-June-2020-An-overview-of-predictive-scoring-systems-used-in-ICU-U-Jaganath.pdf (accessed on 1 July 2024).
86. Kelley, M.A.; Manaker, S.; Finlay, G. Predictive Scoring Systems in the Intensive Care Unit. UpToDate. 2012. Available online: http://www.uptodate.com/online/content/author.do (accessed on 1 July 2024).
87. Kollef, M.H.; Schuster, D.P. Predicting Intensive Care Unit Outcome with Scoring Systems: Underlying Concepts and Principles. *Crit. Care Clin.* **1994**, *10*, 1–18. [CrossRef]
88. Rapsang, A.G.; Shyam, D.C. Scoring systems in the intensive care unit: A compendium. *Indian J. Crit. Care Med. Peer-Rev. Off. Publ. Indian Soc. Crit. Care Med.* **2014**, *18*, 220. [CrossRef]
89. VijayGanapathy, S.; Karthikeyan, V.S.; Sreenivas, J.; Mallya, A.; Keshavamurthy, R. Validation of APACHE II scoring system at 24 hours after admission as a prognostic tool in urosepsis: A prospective observational study. *Investig. Clin. Urol.* **2017**, *58*, 453–459. [CrossRef]
90. Zimmerman, J.E.; Kramer, A.A.; McNair, D.S.; Malila, F.M. Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients*. *Crit. Care Med.* **2006**, *34*, 1297–1310. [CrossRef] [PubMed]
91. Escarce, J.J.; Kelley, M.A. Admission source to the medical intensive care unit predicts hospital death independent of APACHE II score. *JAMA* **1990**, *264*, 2389–2394. [CrossRef] [PubMed]
92. Ayazoglu, T.A. A comparison of APACHE II and APACHE IV scoring systems in predicting outcome in patients admitted with stroke to an intensive care unit. *Anaesth. Pain Intensiv. Care* **2011**, *15*, 7–12.
93. Vasilevskis, E.E.; Kuzniewicz, M.W.; Cason, B.A.; Lane, R.K.; Dean, M.L.; Clay, T.; Rennie, D.J.; Vittinghoff, E.; Dudley, R.A. Mortality probability model III and simplified acute physiology score II: Assessing their value in predicting length of stay and comparison to APACHE IV. *Chest* **2009**, *136*, 89–101. [CrossRef]
94. Jeong, S. Scoring Systems for the Patients of Intensive Care Unit. *Acute Crit. Care* **2018**, *33*, 102–104. [CrossRef]
95. Nagar, V.S.; Sajjan, B.; Chatterjee, R.; Parab, N.M. The comparison of apache II and apache IV score to predict mortality in intensive care unit in a tertiary care hospital. *Int. J. Res. Med. Sci.* **2019**, *7*, 1598–1603. [CrossRef]
96. Haniffa, R.; Isaam, I.; De Silva, A.P.; Dondorp, A.M.; De Keizer, N.F. Performance of critical care prognostic scoring systems in low and middle-income countries: A systematic review. *Crit. Care* **2018**, *22*, 18. [CrossRef]
97. Singer, M.; Deutschman, C.S.; Seymour, C.W.; Shankar-Hari, M.; Annane, D.; Bauer, M.; Bellomo, R.; Bernard, G.R.; Chiche, J.-D.; Coopersmith, C.M.; et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* **2016**, *315*, 801–810. [CrossRef]
98. Ferreira, F.L.; Bota, D.P.; Bross, A.; Mélot, C.; Vincent, J.-L. Serial Evaluation of the SOFA Score to Predict Outcome in Critically Ill Patients. *JAMA* **2001**, *286*, 1754–1758. [CrossRef]
99. Lambden, S.; Laterre, P.F.; Levy, M.M.; Francois, B. The SOFA score—Development, utility and challenges of accurate assessment in clinical trials. *Crit. Care* **2019**, *23*, 374. [CrossRef]
100. Bouch, D.C.; Thompson, J.P. Severity scoring systems in the critically ill. *Contin. Educ. Anaesth. Crit. Care Pain* **2008**, *8*, 181–185. [CrossRef]
101. Le Gall, J.-R.; Lemeshow, S.; Saulnier, F. A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study. *JAMA* **1993**, *270*, 2957–2963. [CrossRef] [PubMed]
102. Moreno, R.P.; Metnitz, P.G.H.; Almeida, E.; Jordan, B.; Bauer, P.; Campos, R.A.; Iapichino, G.; Edbrooke, D.; Capuzzo, M.; Le Gall, J.-R. SAPS 3—From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. *Intensiv. Care Med.* **2005**, *31*, 1345–1355. [CrossRef] [PubMed]
103. Rothen, H.U.; Stricker, K.; Einfalt, J.; Bauer, P.; Metnitz, P.G.H.; Moreno, R.P.; Takala, J. Variability in outcome and resource use in intensive care units. *Intensiv. Care Med.* **2007**, *33*, 1329–1336. [CrossRef] [PubMed]

104. Higgins, T.L.; Teres, D.; Copes, W.S.; Nathanson, B.H.; Stark, M.; Kramer, A.A. Assessing contemporary intensive care unit outcome: An updated Mortality Probability Admission Model (MPM0-III)*. *Crit. Care Med.* **2007**, *35*, 827–835. [CrossRef]
105. Nemati, S.; Holder, A.M.; Razmi, F.; Stanley, M.D.; Clifford, G.D.; Buchman, T.G. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Crit. Care Med.* **2018**, *46*, 547–553. [CrossRef] [PubMed] [PubMed Central]
106. Bailly, S.; Meyfroidt, G.; Timsit, J.-F. What's New in ICU in 2050: Big Data and Machine Learning. *Intensiv. Care Med.* **2018**, *44*, 1524–1527. [CrossRef]
107. Patel, P.A.; Grant, B.J.B. Application of mortality prediction systems to individual intensive care units. *Intensiv. Care Med.* **1999**, *25*, 977–982. [CrossRef]
108. Nassar, A.P., Jr.; Mocelin, A.O.; Nunes, A.L.; Giannini, F.P.; Brauer, L.; Andrade, F.M.; Dias, C.A. Caution when using prognostic models: A prospective comparison of 3 recent prognostic models. *J. Crit. Care* **2012**, *27*, 423.e1–423.e7. [CrossRef]
109. Sinuff, T.; Adhikari, N.K.J.; Cook, D.J.; Schünemann, H.J.; Griffith, L.E.; Rocker, G.; Walter, S.D. Mortality predictions in the intensive care unit: Comparing physicians with scoring systems*. *Crit. Care Med.* **2006**, *34*, 878–885. [CrossRef]
110. Yang, J.; Karstens, L.; Ross, C.; Yala, A. AI gone astray: Technical supplement. *arXiv* **2022**, arXiv:2203.16452.
111. Van De Water, R.; Schmidt, H.; Elbers, P.; Thoral, P.; Arnrich, B.; Rockenschaub, P. Yet another icu benchmark: A flexible multi-center framework for clinical ml. *arXiv* **2023**, arXiv:2306.05109.
112. Shashikumar, S.P.; Wardi, G.; Malhotra, A.; Nemati, S. Artificial intelligence sepsis prediction algorithm learns to say "I don't know". *npj Digit. Med.* **2021**, *4*, 134. [CrossRef]
113. Hou, N.; Li, M.; He, L.; Xie, B.; Wang, L.; Zhang, R.; Yu, Y.; Sun, X.; Pan, Z.; Wang, K. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: A machine learning approach using XGboost. *J. Transl. Med.* **2020**, *18*, 462. [CrossRef] [PubMed]
114. Kong, G.; Lin, K.; Hu, Y. Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 251. [CrossRef] [PubMed] [PubMed Central]
115. Persson, I.; Östling, A.; Arlbrandt, M.; Söderberg, J.; Becedas, D. A Machine Learning Sepsis Prediction Algorithm for Intended Intensive Care Unit Use (NAVOY Sepsis): Proof-of-Concept Study. *JMIR Form. Res.* **2021**, *5*, e28000. [CrossRef] [PubMed] [PubMed Central]
116. Yuan, K.-C.; Tsai, L.-W.; Lee, K.-H.; Cheng, Y.-W.; Hsu, S.-C.; Lo, Y.-S.; Chen, R.-J. The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit. *Int. J. Med. Inform.* **2020**, *141*, 104176. [CrossRef]
117. Wang, D.; Li, J.; Sun, Y.; Ding, X.; Zhang, X.; Liu, S.; Han, B.; Wang, H.; Duan, X.; Sun, T. A Machine Learning Model for Accurate Prediction of Sepsis in ICU Patients. *Front. Public Health* **2021**, *9*, 754348. [CrossRef]
118. Kok, C.; Jahmunah, V.; Oh, S.L.; Zhou, X.; Gururajan, R.; Tao, X.; Cheong, K.H.; Gururajan, R.; Molinari, F.; Acharya, U. Automated prediction of sepsis using temporal convolutional network. *Comput. Biol. Med.* **2020**, *127*, 103957. [CrossRef]
119. Zhao, X.; Shen, W.; Wang, G. Early Prediction of Sepsis Based on Machine Learning Algorithm. *Comput. Intell. Neurosci.* **2021**, *2021*, 6522633. [CrossRef]
120. Ghias, N.; Ul Haq, S.; Arshad, H.; Sultan, H.; Bashir, F.; Ghaznavi, S.A.; Shabbir, M.; Badshah, Y.; Rafiq, M. Using Machine Learning Algorithms to predict sepsis and its stages in ICU patients. *medRxiv* **2022**. [CrossRef]
121. Moor, M.; Horn, M.; Rieck, B.; Roqueiro, D.; Borgwardt, K. Early recognition of sepsis with Gaussian process temporal convolutional networks and dynamic time warping. In Proceedings of the Machine Learning for Healthcare Conference, Ann Arbor, MI, USA, 9–10 August 2019; pp. 2–26.
122. Cruz, M.F.; Ono, N.; Huang, M.; Amin, A.U.; Kanaya, S.; Cavalcante, C.A.M.T. Kinematics approach with neural networks for early detection of sepsis (KANNEDS). *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 163. [CrossRef] [PubMed]
123. Xia, Y.; Long, H.; Lai, Q.; Zhou, Y. Machine Learning Predictive Model for Septic Shock in Acute Pancreatitis with Sepsis. *J. Inflamm. Res.* **2024**, *17*, 1443–1452. [CrossRef] [PubMed]
124. Lin, P.-C.; Chen, K.-T.; Chen, H.-C.; Islam, M.; Lin, M.-C. Machine Learning Model to Identify Sepsis Patients in the Emergency Department: Algorithm Development and Validation. *J. Pers. Med.* **2021**, *11*, 1055. [CrossRef]
125. Burdick, H.; Pino, E.; Gabel-Comeau, D.; Gu, C.; Roberts, J.; Le, S.; Slote, J.; Saber, N.; Pellegrini, E.; Green-Saxena, A.; et al. Validation of a machine learning algorithm for early severe sepsis prediction: A retrospective study pre-dicting severe sepsis up to 48 h in advance using a diverse dataset from 461 US hospitals. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 276. [CrossRef]
126. Yong, L.; Zhenzhou, L. Deep learning-based prediction of in-hospital mortality for sepsis. *Sci. Rep.* **2024**, *14*, 372. [CrossRef]
127. Ghiasi, S.; Zhu, T.; Lu, P.; Hagenah, J.; Khanh, P.N.Q.; Van Hao, N.; Thwaites, L.; Clifton, D.A.; Vital Consortium. Sepsis Mortality Prediction Using Wearable Monitoring in Low–Middle Income Countries. *Sensors* **2022**, *22*, 3866. [CrossRef]
128. Kim, T.; Tae, Y.; Yeo, H.J.; Jang, J.H.; Cho, K.; Yoo, D.; Lee, Y.; Ahn, S.-H.; Kim, Y.; Lee, N.; et al. Development and Validation of Deep-Learning-Based Sepsis and Septic Shock Early Prediction System (DeepSEPS) Using Real-World ICU Data. *J. Clin. Med.* **2023**, *12*, 7156. [CrossRef]
129. Gholamzadeh, M.; Abtahi, H.; Safdari, R. Comparison of different machine learning algorithms to classify patients suspected of having sepsis infection in the intensive care unit. *Inform. Med. Unlocked* **2023**, *38*, 101236. [CrossRef]

130. Goh, K.H.; Wang, L.; Yeow, A.Y.K.; Poh, H.; Li, K.; Yeow, J.J.L.; Tan, G.Y.H. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nat. Commun.* **2021**, *12*, 711. [CrossRef]
131. Bedoya, A.D.; Futoma, J.; E Clement, M.; Corey, K.; Brajer, N.; Lin, A.; Simons, M.G.; Gao, M.; Nichols, M.; Balu, S.; et al. Machine learning for early detection of sepsis: An internal and temporal validation study. *JAMIA Open* **2020**, *3*, 252–260. [CrossRef] [PubMed]
132. Lauritsen, S.M.; Kalør, M.E.; Kongsgaard, E.L.; Lauritsen, K.M.; Jørgensen, M.J.; Lange, J.; Thiesson, B. Early detection of sepsis utilizing deep learning on electronic health record event sequences. *Artif. Intell. Med.* **2020**, *104*, 101820. [CrossRef] [PubMed]
133. Scherpf, M.; Gräßer, F.; Malberg, H.; Zaunseder, S. Predicting sepsis with a recurrent neural network using the MIMIC III database. *Comput. Biol. Med.* **2019**, *113*, 103395. [CrossRef]
134. Li, X.; Ng, G.A.; Schlindwein, F. Convolutional and Recurrent Neural Networks for Early Detection of Sepsis Using Hourly Physiological Data from Patients in Intensive Care Unit. In Proceedings of the 2019 Computing in Cardiology Conference, Singapore, 8–11 September 2019; p. 1.
135. Zhang, D.; Yin, C.; Hunold, K.M.; Jiang, X.; Caterino, J.M.; Zhang, P. An interpretable deep-learning model for early prediction of sepsis in the emergency department. *Patterns* **2021**, *2*, 100196. [CrossRef]
136. Oei, S.P.; van Sloun, R.J.; van der Ven, M.; Korsten, H.H.; Mischi, M. Towards early sepsis detection from measurements at the general ward through deep learning. *Intell. Med.* **2021**, *5*, 100042. [CrossRef]
137. Kamal, S.A.; Yin, C.; Qian, B.; Zhang, P. An interpretable risk prediction model for healthcare with pattern attention. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 307. [CrossRef]
138. Krissaane, I.; Hampton, K.; Alshenaifi, J.; Wilkinson, R. Anomaly Detection Semi-Supervised Framework for Sepsis Treatment. In Proceedings of the 2019 Computing in Cardiology Conference (CinC), Singapore, 8–11 September 2019; p. 1.
139. Duan, Y.; Huo, J.; Chen, M.; Hou, F.; Yan, G.; Li, S.; Wang, H. Early prediction of sepsis using double fusion of deep features and handcrafted features. *Appl. Intell.* **2023**, *53*, 17903–17919. [CrossRef]
140. Al-Mualemi, B.Y.; Lu, L. A Deep Learning-Based Sepsis Estimation Scheme. *IEEE Access* **2020**, *9*, 5442–5452. [CrossRef]
141. Brann, F.; Sterling, N.W.; O Frisch, S.; Schrager, J.D. Sepsis Prediction at Emergency Department Triage Using Natural Language Processing: Retrospective Cohort Study. *JMIR AI* **2024**, *3*, e49784. [CrossRef]
142. Choi, J.S.; Trinh, T.X.; Ha, J.; Yang, M.S.; Lee, Y.; Kim, Y.E.; Choi, J.; Byun, H.G.; Song, J.; Yoon, T.H. Implementation of complementary model using optimal combination of hematological parameters for sepsis screening in patients with fever. *Sci. Rep.* **2020**, *10*, 273. [CrossRef]
143. Fagerström, J.; Bång, M.; Wilhelms, D.; Chew, M.S. LiSep LSTM: A Machine Learning Algorithm for Early Detection of Septic Shock. *Sci. Rep.* **2019**, *9*, 15132. [CrossRef] [PubMed]
144. Bai, Y.; Xia, J.; Huang, X.; Chen, S.; Zhan, Q. Using machine learning for the early prediction of sepsis-associated ARDS in the ICU and identification of clinical phenotypes with differential responses to treatment. *Front. Physiol.* **2022**, *13*, 1050849. [CrossRef] [PubMed]
145. Misra, D.; Avula, V.; Wolk, D.M.; Farag, H.A.; Li, J.; Mehta, Y.B.; Sandhu, R.; Karunakaran, B.; Kethireddy, S.; Zand, R.; et al. Early Detection of Septic Shock Onset Using Interpretable Machine Learners. *J. Clin. Med.* **2021**, *10*, 301. [CrossRef]
146. Zhao, Y.; Qiao, Z.; Xiao, C.; Glass, L.; Sun, J. Pyhealth: A python library for health predictive models. *arXiv* **2021**, arXiv:2101.04209.
147. Johnson, A.E.W.; Pollard, T.J.; Shen, L.; Lehman, L.-W.H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Celi, L.A.; Mark, R.G. MIMIC-III, a freely accessible critical care database. *Sci. Data* **2016**, *3*, 160035. [CrossRef]
148. Mandyam, A.; Yoo, E.C.; Soules, J.; Laudanski, K.; Engelhardt, B.E. COP-E-CAT: Cleaning and organization pipeline for EHR computational and analytic tasks. In Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, Gainesville, FL, USA, 1–4 August 2021; pp. 1–9.
149. Guo, C.; Lu, M.; Chen, J. An evaluation of time series summary statistics as features for clinical prediction tasks. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 48. [CrossRef]
150. Yèche, H.; Kuznetsova, R.; Zimmermann, M.; Hüser, M.; Lyu, X.; Faltys, M.; Rätsch, G. HiRID-ICU-Benchmark–A Comprehensive Machine Learning Benchmark on High-resolution ICU Data. *arXiv* **2021**, arXiv:2111.08536.
151. Alghatani, K.; Ammar, N.; Rezgui, A.; Shaban-Nejad, A. Predicting Intensive Care Unit Length of Stay and Mortality Using Patient Vital Signs: Machine Learning Model Development and Validation. *JMIR Public Health Surveill.* **2021**, *9*, e21347. [CrossRef]
152. Harutyunyan, H.; Khachatrian, H.; Kale, D.C.; Ver Steeg, G.; Galstyan, A. Multitask learning and benchmarking with clinical time series data. *Sci. Data* **2019**, *6*, 96. [CrossRef]
153. Sheikhalishahi, S.; Balaraman, V.; Osmani, V. Benchmarking machine learning models on multi-centre eICU critical care dataset. *PLoS ONE* **2020**, *15*, e0235424. [CrossRef]
154. Dan, T.; Li, Y.; Zhu, Z.; Chen, X.; Quan, W.; Hu, Y.; Tao, G.; Zhu, L.; Zhu, J.; Jin, Y.; et al. Machine learning to predict ICU admission, ICU mortality and survivors' length of stay among COVID-19 patients: Toward optimal allocation of ICU resources. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Bio-Medicine (BIBM), Seoul, Republic of Korea, 16–19 December 2020; pp. 555–561.

155. Rocheteau, E.; Tong, C.; Veličković, P.; Lane, N.; Liò, P. Predicting patient outcomes with graph representation learning. *arXiv* **2021**, arXiv:2101.03940.
156. Scheltjens, V.; Momo, L.N.W.; Verbeke, W.; De Moor, B. Client Recruitment for Federated Learning in ICU Length of Stay Prediction. *arXiv* **2023**, arXiv:2304.14663.
157. Wen, Y.; Rahman, F.; Zhuang, Y.; Pokojovy, M.; Xu, H.; McCaffrey, P.; Vo, A.; Walser, E.; Moen, S.; Tseng, T.-L. Time-to-event modeling for hospital length of stay prediction for COVID-19 patients. *Mach. Learn. Appl.* **2022**, *9*, 100365. [CrossRef] [PubMed]
158. Dogu, E.; Albayrak, Y.E.; Tuncay, E. Length of hospital stay prediction with an integrated approach of statistical-based fuzzy cognitive maps and artificial neural networks. *Med. Biol. Eng. Comput.* **2021**, *59*, 483–496. [CrossRef]
159. Lefering, R.; Waydhas, C.; DGU, T. Prediction of prolonged length of stay on the intensive care unit in severely injured patients—A registry-based multivariable analysis. *Front. Med.* **2024**, *11*, 1358205. [CrossRef]
160. Abdurrab, I.; Mahmood, T.; Sheikh, S.; Aijaz, S.; Kashif, M.; Memon, A.; Ali, I.; Peerwani, G.; Pathan, A.; Alkhodre, A.B.; et al. Predicting the length of stay of cardiac patients based on pre-operative variables—Bayesian models vs. machine learning models. *Healthcare* **2024**, *12*, 249. [CrossRef]
161. Chen, J.; Wen, Y.; Pokojovy, M.; Tseng, T.-L.; McCaffrey, P.; Vo, A.; Walser, E.; Moen, S. Multi-modal learning for inpatient length of stay prediction. *Comput. Biol. Med.* **2024**, *171*, 108121. [CrossRef]
162. Karankot, M.I.; Marceau, M.; Glenn, E.M.; Fowers, R.P.; Hedges, D.M.; Sheehey, B.; Whitaker, B.M. Addressing the Challenge of Missing Medical Data in Healthcare Analytics: A Focus on Machine Learning Predictions for ICU Length of Stay. In Proceedings of the 2024 Intermountain Engineering, Technology and Computing (IETC), Logan, UT, USA, 13–14 May 2024; pp. 147–151.
163. Junior, J.C.; Caneo, L.F.; Turquetto, A.L.R.; Amato, L.P.; Arita, E.C.T.C.; Fernandes, A.M.d.S.; Trindade, E.M.; Jatene, F.B.; Dossou, P.-E.; Jatene, M.B. Predictors of in-ICU length of stay among congenital heart defect patients using artificial intelligence model: A pilot study. *Heliyon* **2024**, *10*, e25406. [CrossRef]
164. Siddiqa, A.; Naqvi, S.A.Z.; Ahsan, M.; Ditta, A.; Alquhayz, H.; Khan, M.A. Robust Length of Stay Prediction Model for Indoor Patients. *Comput. Mater. Contin.* **2022**, *70*, 5519–5536. [CrossRef]
165. Zhong, H.; Wang, B.; Wang, D.; Liu, Z.; Xing, C.; Wu, Y.; Gao, Q.; Zhu, S.; Qu, H.; Jia, Z.; et al. The application of machine learning algorithms in predicting the length of stay following femoral neck fracture. *Int. J. Med. Inform.* **2021**, *155*, 104572. [CrossRef]
166. Abbas, A.; Mosseri, J.; Lex, J.R.; Toor, J.; Ravi, B.; Khalil, E.B.; Whyne, C. Machine learning using preoperative patient factors can predict duration of surgery and length of stay for total knee arthroplasty. *Int. J. Med. Inform.* **2021**, *158*, 104670. [CrossRef] [PubMed]
167. Barsasella, D.; Gupta, S.; Malwade, S.; Aminin; Susanti, Y.; Tirmadi, B.; Mutamakin, A.; Jonnagaddala, J.; Syed-Abdul, S. Predicting length of stay and mortality among hospitalized patients with type 2 diabetes mellitus and hypertension. *Int. J. Med. Inform.* **2021**, *154*, 104569. [CrossRef] [PubMed]
168. Su, L.; Xu, Z.; Chang, F.; Ma, Y.; Liu, S.; Jiang, H.; Wang, H.; Li, D.; Chen, H.; Zhou, X.; et al. Early Prediction of Mortality, Severity, and Length of Stay in the Intensive Care Unit of Sepsis Patients Based on Sepsis 3.0 by Machine Learning Models. *Front. Med.* **2021**, *8*, 664966. [CrossRef]
169. Wu, J.; Lin, Y.; Li, P.; Hu, Y.; Zhang, L.; Kong, G. Predicting Prolonged Length of ICU Stay through Machine Learning. *Diagnostics* **2021**, *11*, 2242. [CrossRef]
170. Saadatmand, S.; Salimifard, K.; Mohammadi, R.; Kuiper, A.; Marzban, M.; Farhadi, A. Using machine learning in prediction of ICU admission, mortality, and length of stay in the early stage of admission of COVID-19 patients. *Ann. Oper. Res.* **2022**, *328*, 1043–1071. [CrossRef]
171. Abujaber, A.; Fadlalla, A.; Nashwan, A.; El-Menyar, A.; Al-Thani, H. Predicting prolonged length of stay in patients with traumatic brain injury: A machine learning approach. *Intell. Med.* **2022**, *6*, 100052. [CrossRef]
172. Wang, K.; Yan, L.Z.; Li, W.Z.; Jiang, C.; Ni Wang, N.; Zheng, Q.; Dong, N.G.; Shi, J.W. Comparison of Four Machine Learning Techniques for Prediction of Intensive Care Unit Length of Stay in Heart Transplantation Patients. *Front. Cardiovasc. Med.* **2022**, *9*, 863642. [CrossRef]
173. Alsinglawi, B.; Alshari, O.; Alorjani, M.; Mubin, O.; Alnajjar, F.; Novoa, M.; Darwish, O. An explainable machine learning framework for lung cancer hospital length of stay prediction. *Sci. Rep.* **2022**, *12*, 607. [CrossRef]
174. Iwase, S.; Nakada, T.-A.; Shimada, T.; Oami, T.; Shimazui, T.; Takahashi, N.; Yamabe, J.; Yamao, Y.; Kawakami, E. Prediction algorithm for ICU mortality and length of stay using machine learning. *Sci. Rep.* **2022**, *12*, 12912. [CrossRef]
175. Gupta, M.; Gallamoza, B.; Cutrona, N.; Dhakal, P.; Poulain, R.; Beheshti, R. An Extensive Data Processing Pipeline for MIMIC-IV. *Mach. Learn. Health* **2022**, *193*, 311–325.
176. Wang, S.; McDermott, M.B.; Chauhan, G.; Ghassemi, M.; Hughes, M.C.; Naumann, T. Mimic-extract: A data extraction, preprocessing, and representation pipeline for mimic-iii. In Proceedings of the ACM Conference on Health, Inference, and Learning, Toronto, ON, Canada, 2–4 April 2020; pp. 222–235.
177. Hong, Y.; Wu, X.; Qu, J.; Gao, Y.; Chen, H.; Zhang, Z. Clinical characteristics of Coronavirus Disease 2019 and development of a prediction model for prolonged hospital length of stay. *Ann. Transl. Med.* **2020**, *8*, 443. [CrossRef] [PubMed]
178. Zhang, M.; Kuo, T.-T. Early prediction of long hospital stay for Intensive Care units readmission patients using medication information. *Comput. Biol. Med.* **2024**, *174*, 108451. [CrossRef] [PubMed]

179. Belmonte, E.M.; Oton-Tortosa, S.; Gutierrez-Martinez, J.-M.; Castillo-Martinez, A. An Intelligent Model and Methodology for Predicting Length of Stay and Survival in a Critical Care Hospital Unit. *Informatics* **2024**, *11*, 34. [CrossRef]
180. Chen, Q.; Zhang, B.; Yang, J.; Mo, X.; Zhang, L.; Li, M.; Chen, Z.; Fang, J.; Wang, F.; Huang, W.; et al. Predicting Intensive Care Unit Length of Stay After Acute Type A Aortic Dissection Surgery Using Machine Learning. *Front. Cardiovasc. Med.* **2021**, *8*, 675431. [CrossRef]
181. Abd-Elrazek, M.A.; Eltahawi, A.A.; Elaziz, M.H.A.; Abd-Elwhab, M.N. Predicting length of stay in hospitals intensive care unit using general admission features. *Ain Shams Eng. J.* **2021**, *12*, 3691–3702. [CrossRef]
182. Alabbad, D.A.; Almuhaideb, A.M.; Alsunaidi, S.J.; Alqudaihi, K.S.; Alamoudi, F.A.; Alhobaishi, M.K.; Alaqeel, N.A.; Alshahrani, M.S. Machine learning model for predicting the length of stay in the intensive care unit for Covid-19 patients in the eastern province of Saudi Arabia. *Inform. Med. Unlocked* **2022**, *30*, 100937. [CrossRef]
183. K, V.; Mushtaq, S. LoSNet: A Tailored Deep Neural Network Framework for Precise Length of Stay Prediction in Disease-Specific Hospitalization. *Procedia Comput. Sci.* **2024**, *235*, 2599–2608. [CrossRef]
184. Harerimana, G.; Kim, J.W.; Jang, B. A deep attention model to forecast the Length Of Stay and the in-hospital mortality right on admission from ICD codes and demographic data. *J. Biomed. Inform.* **2021**, *118*, 103778. [CrossRef]
185. Bhadouria, A.S.; Singh, R.K. Machine learning model for healthcare investments predicting the length of stay in a hospital & mortality rate. *Multimed. Tools Appl.* **2024**, *83*, 27121–27191. [CrossRef]
186. Lilly, C.M.; Zuckerman, I.H.; Badawi, O.; Riker, R.R. Benchmark Data from More Than 240,000 Adults That Reflect the Current Practice of Critical Care in the United States. *Chest* **2011**, *140*, 1232–1242. [CrossRef]
187. Johnson, A.E.; Bulgarelli, L.; Shen, L.; Gayles, A.; Shammout, A.; Horng, S.; Pollard, T.J.; Hao, S.; Moody, B.; Gow, B.; et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci. Data* **2023**, *10*, 1. [CrossRef] [PubMed]
188. Pollard, T.J.; Johnson, A.E.W.; Raffa, J.D.; Celi, L.A.; Mark, R.G.; Badawi, O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci. Data* **2018**, *5*, 180178. [CrossRef] [PubMed]
189. Sulaiman, W.A.; Nicolaou, A.; Prentza, N.; Stylianides, C.; Panayides, A.; Constantinou, I.; Antoniou, Z.; Kakas, A.; Kyriacou, E.; Palazis, L.; et al. Emergency Department Triage Hospitalization Prediction Based on Machine Learning and Rule Ex-traction. In Proceedings of the 2023 IEEE EMBS Special Topic Conference on Data Science and Engineering in Healthcare, Medicine and Biology, St. Julians, Malta, 7–9 December 2023; pp. 147–148.
190. Xie, F.; Zhou, J.; Lee, J.W.; Tan, M.; Li, S.; Rajnthern, L.S.; Chee, M.L.; Chakraborty, B.; Wong, A.-K.I.; Dagan, A.; et al. Benchmarking Emergency Department Prediction Models with Machine Learning and Public Electronic Health Records. Nature News. Available online: https://www.nature.com/articles/s41597-022-01782-9 (accessed on 4 December 2023).
191. Raita, Y.; Goto, T.; Faridi, M.K.; Brown, D.F.M.; Camargo, C.A., Jr.; Hasegawa, K. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit. Care* **2019**, *23*, 64. [CrossRef] [PubMed]
192. Goto, T.; Camargo, C.A.; Faridi, M.K.; Freishtat, R.J.; Hasegawa, K. Machine learning–based prediction of clinical out-comes for children during emergency department triage. *JAMA Netw. Open* **2019**, *2*, e186937. [CrossRef]
193. Mowbray, F.; Zargoush, M.; Jones, A.; de Wit, K.; Costa, A. Predicting hospital admission for older emergency department patients: Insights from machine learning. *Int. J. Med. Inform.* **2020**, *140*, 104163. [CrossRef]
194. Lee, J.-T.; Hsieh, C.-C.; Lin, C.-H.; Lin, Y.-J.; Kao, C.-Y. Prediction of hospitalization using artificial intelligence for urgent patients in the emergency department. *Sci. Rep.* **2021**, *11*, 19472. [CrossRef]
195. Araz, O.M.; Olson, D.; Ramirez-Nafarrate, A. Predictive analytics for hospital admissions from the emergency department using triage information. *Int. J. Prod. Econ.* **2019**, *208*, 199–207. [CrossRef]
196. Graham, B.; Bond, R.; Quinn, M.; Mulvenna, M. Using Data Mining to Predict Hospital Admissions From the Emergency Department. *IEEE Access* **2018**, *6*, 10458–10469. [CrossRef]
197. Ahmed, A.; Ashour, O.; Ali, H.; Firouz, M. An integrated optimization and machine learning approach to predict the admission status of emergency patients. *Expert Syst. Appl.* **2022**, *202*, 117314. [CrossRef]
198. Sills, M.R.; Ozkaynak, M.; Jang, H. Predicting hospitalization of pediatric asthma patients in emergency departments using machine learning. *Int. J. Med. Inform.* **2021**, *151*, 104468. [CrossRef]
199. Hong, T.; Liu, X.; Deng, J.; Li, H.; Sun, M.; Pan, D.; Zhao, Y.; Cai, Z.; Zhao, J.; Yu, L.; et al. The Scoring Model to Predict ICU Stay and Mortality After Emergency Admissions in Atrial Fibrillation: A Retrospective Study of 30,206 Patients. 2024. Available online: https://www.researchsquare.com/article/rs-3903182/v1 (accessed on 4 December 2023).
200. Lin, Y.-T.; Deng, Y.-X.; Tsai, C.-L.; Huang, C.-H.; Fu, L.-C. Interpretable Deep Learning System for Identifying Critical Patients Through the Prediction of Triage Level, Hospitalization, and Length of Stay: Prospective Study. *JMIR Med. Inform.* **2024**, *12*, e48862. [CrossRef] [PubMed]
201. Ahmed, A.; Zengul, F.D.; Khan, S.; Hearld, K.R.; Feldman, S.S.; Hall, A.G.; Orewa, G.N.; Willig, J.; Kennedy, K. Developing a decision model to early predict ICU admission for COVID-19 patients: A machine learning approach. *Intell. Med.* **2024**, *9*, 100136. [CrossRef]

202. Seo, H.; Ahn, I.; Gwon, H.; Kang, H.J.; Kim, Y.; Cho, H.N.; Choi, H.; Kim, M.; Han, J.; Kee, G.; et al. Prediction of hospitalization and waiting time within 24 hours of emergency department patients with un-structured text data. *Health Care Manag. Sci.* **2024**, *27*, 114–129. [CrossRef] [PubMed]
203. Famiglini, L.; Campagner, A.; Carobene, A.; Cabitza, F. A robust and parsimonious machine learning method to predict ICU admission of COVID-19 patients. *Med. Biol. Eng. Comput.* **2022**, 1–13. [CrossRef]
204. Subudhi, S.; Verma, A.; Patel, A.B.; Hardin, C.C.; Khandekar, M.J.; Lee, H.; McEvoy, D.; Stylianopoulos, T.; Munn, L.L.; Dutta, S.; et al. Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19. *npj Digit. Med.* **2021**, *4*, 87. [CrossRef]
205. Chen, T.-Y.; Huang, T.-Y.; Chang, Y.-C. Using a clinical narrative-aware pre-trained language model for predicting emergency department patient disposition and unscheduled return visits. *J. Biomed. Inform.* **2024**, *155*, 104657. [CrossRef]
206. De Hond, A.; Raven, W.; Schinkelshoek, L.; Gaakeer, M.; Ter Avest, E.; Sir, O.; Lameijer, H.; Hessels, R.A.; Reijnen, R.; De Jonge, E.; et al. Machine learning for developing a prediction model of hospital admission of emergency department patients: Hype or hope? *Int. J. Med. Inform.* **2021**, *152*, 104496. [CrossRef]
207. Choi, D.H.; Lee, H.; Joo, H.; Kong, H.-J.; Lee, S.B.; Kim, S.; Shin, S.D.; Kim, K.H. Development of Prediction Model for Intensive Care Unit Admission Based on Heart Rate Variability: A Case–Control Matched Analysis. *Diagnostics* **2024**, *14*, 816. [CrossRef]
208. Fenn, A.; Davis, C.; Buckland, D.M.; Kapadia, N.; Nichols, M.; Gao, M.; Knechtle, W.; Balu, S.; Sendak, M.; Theiling, B. Development and Validation of Machine Learning Models to Predict Admission From Emergency Department to Inpatient and Intensive Care Units. *Ann. Emerg. Med.* **2021**, *78*, 290–302. [CrossRef]
209. World Health Organization. *International Classification of Diseases: [9th] Ninth Revision, Basic Tabulation List with Alphabetic Index*; World Health Organization: Geneva, Switzerland, 1978.
210. Bender, D.; Sartipi, K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. In Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, Porto, Portugal, 20–22 June 2013; pp. 326–331.
211. European Commission. Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. *COM(2022) 197 Final. 2022 May 3.* Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0197 (accessed on 13 December 2024).
212. Perkins, N.J.; Schisterman, E.F. The Inconsistency of "Optimal" Cutpoints Obtained using Two Criteria based on the Receiver Operating Characteristic Curve. *Am. J. Epidemiol.* **2006**, *163*, 670–675. [CrossRef]
213. Unal, I. Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach. *Comput. Math. Methods Med.* **2017**, *2017*, 3762651. [CrossRef] [PubMed]
214. Youden, W.J. Index for rating diagnostic tests. *Cancer* **1950**, *3*, 32–35. [CrossRef] [PubMed]
215. Fluss, R.; Faraggi, D.; Reiser, B. Estimation of the Youden Index and its associated cutoff point. *Biom. J. J. Math. Methods Biosci.* **2005**, *47*, 458–472. [CrossRef]
216. European Parliament and Council of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance). 13 June 2024. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689 (accessed on 13 December 2024).
217. Cohen, I.G.; Amarasingham, R.; Shah, A.; Xie, B.; Lo, B. The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care. *Health Aff.* **2014**, *33*, 1139–1147. [CrossRef]
218. Amini, M.M.; Jesus, M.; Sheikholeslami, D.F.; Alves, P.; Benam, A.H.; Hariri, F. Artificial intel-ligence ethics and challenges in healthcare applications: A comprehensive review in the context of the European GDPR mandate. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1023–1035. [CrossRef]
219. European Commission Ethics Guidelines for Trustworthy AI. 2019. Available online: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed on 4 December 2023).
220. Voigt, P.; Von dem Bussche, A. The eu general data protection regulation (gdpr). In *A Practical Guide*, 1st ed.; Springer International Publishing: Cham, Switzerland, 2017; Volume 10, pp. 10–5555.
221. Prentzas, N.; Kakas, A.; Pattichis, C.S. Explainable AI applications in the Medical Domain: A systematic review. *arXiv* **2023**, arXiv:2308.05411.
222. Lundberg, S.; Lee, S.-I. SHAP: A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **2017**, 1–10.
223. Shapley, L.S. A value for n-person games. In *Contributions to the Theory of Games II*; Kuhn, H., Tucker, A., Eds.; Princeton University Press: Princeton, MJ, USA, 1953; pp. 307–317.
224. Fisher, A.; Rudin, C.; Dominici, F. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.* **2019**, *20*, 1–81.

225. Ribeiro, M.T.; Singh, S.; Guestrin, C. 'Why should I trust you?' Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.
226. Biecek, P.; Burzykowski, T. *Explanatory Model Analysis: Explore, Explain, and Examine Predictive Models*; CRC Press: London, UK, 2021.
227. Saranya, A.; Subhashini, R. A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. *Decis. Anal. J.* **2023**, *7*, 100230. [CrossRef]
228. Yang, J.-B.; Liu, J.; Wang, J.; Sii, H.-S.; Wang, H.-W. Belief rule-base inference methodology using the evidential reasoning Approach-RIMER. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **2006**, *36*, 266–285. [CrossRef]
229. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359. [CrossRef]
230. Saltelli, A.; Tarantola, S.; Campolongo, F. Sensitivity analysis as an ingredient of modeling. *Stat. Sci.* **2000**, *15*, 377–395.
231. Yoon, J.; Jordon, J.; van der Schaar, M. INVASE: Instance-wise variable selection using neural networks. In Proceedings of the 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, 6–9 May 2019; pp. 1–24.
232. Chen, M.; Hernández, A. Towards an Explainable Model for Sepsis Detection Based on Sensitivity Analysis. *IRBM* **2022**, *43*, 75–86. [CrossRef]
233. Nesaragi, N.; Patidar, S. An Explainable Machine Learning Model for Early Prediction of Sepsis Using ICU Data. In *Infections and Sepsis Development*; Neri, V., Huang, L., Li, J., Eds.; IntechOpen: Rijeka, Croatia, 2021.
234. Chakraborty, S.; Kumar, K.; Reddy, B.P.; Meena, T.; Roy, S. An Explainable AI based Clinical Assistance Model for Iden-tifying Patients with the Onset of Sepsis. In Proceedings of the 2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI), Bellevue, WA, USA, 4–6 August 2023; pp. 297–302.
235. Jiang, Z.; Bo, L.; Wang, L.; Xie, Y.; Cao, J.; Yao, Y.; Lu, W.; Deng, X.; Yang, T.; Bian, J. Interpretable machine-learning model for real-time, clustered risk factor analysis of sepsis and septic death in critical care. *Comput. Methods Programs Biomed.* **2023**, *241*, 107772. [CrossRef]
236. Chen, Q.; Li, R.; Lin, C.; Lai, C.; Chen, D.; Qu, H.; Huang, Y.; Lu, W.; Tang, Y.; Li, L. Transferability and interpretability of the sepsis prediction models in the intensive care unit. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 343. [CrossRef]
237. Shashikumar, S.P.; Josef, C.S.; Sharma, A.; Nemati, S. DeepAISE—An interpretable and recurrent neural survival model for early prediction of sepsis. *Artif. Intell. Med.* **2021**, *113*, 102036. [CrossRef]
238. Zhang, T.Y.; Zhong, M.; Cheng, Y.Z.; Zhang, M.W. An interpretable machine learning model for real-time sepsis pre-diction based on basic physiological indicators. *Eur. Rev. Med. Pharmacol. Sci.* **2023**, *27*, 4348–4356.
239. Rosnati, M.; Fortuin, V. MGP-AttTCN: An interpretable machine learning model for the prediction of sepsis. *PLoS ONE* **2021**, *16*, e0251248. [CrossRef]
240. Jiang, Z.; Bo, L.; Xu, Z.; Song, Y.; Wang, J.; Wen, P.; Wan, X.; Yang, T.; Deng, X.; Bian, J. An explainable machine learning algorithm for risk factor analysis of in-hospital mortality in sepsis survivors with ICU readmission. *Comput. Methods Programs Biomed.* **2021**, *204*, 106040. [CrossRef]
241. Lemańska-Perek, A.; Krzyżanowska-Gołąb, D.; Kobylińska, K.; Biecek, P.; Skalec, T.; Tyszko, M.; Gozdzik, W.; Adamik, B. Explainable Artificial Intelligence Helps in Understanding the Effect of Fibronectin on Survival of Sepsis. *Cells* **2022**, *11*, 2433. [CrossRef]
242. Pick, F. Explainable Machine Learning for Predicting Sepsis Outcome. Master's Thesis, Department Computer Science at Swansea University, Swansea, Wales, 2021.
243. Li, S.; Dou, R.; Song, X.; Lui, K.Y.; Xu, J.; Guo, Z.; Hu, X.; Guan, X.; Cai, C. Developing an Interpretable Machine Learning Model to Predict in-Hospital Mortality in Sepsis Patients: A Retrospective Temporal Validation Study. *J. Clin. Med.* **2023**, *12*, 915. [CrossRef] [PubMed]
244. Hu, C.; Li, L.; Huang, W.; Wu, T.; Xu, Q.; Liu, J.; Hu, B. Interpretable Machine Learning for Early Prediction of Prognosis in Sepsis: A Discovery and Validation Study. *Infect. Dis. Ther.* **2022**, *11*, 1117–1132. [CrossRef] [PubMed]
245. Hu, C.; Li, L.; Li, Y.; Wang, F.; Hu, B.; Peng, Z. Explainable Machine-Learning Model for Prediction of In-Hospital Mortality in Septic Patients Requiring Intensive Care Unit Readmission. *Infect. Dis. Ther.* **2022**, *11*, 1695–1713. [CrossRef] [PubMed]
246. Ke, X.; Zhang, F.; Huang, G.; Wang, A. Interpretable Machine Learning to Optimize Early In-Hospital Mortality Prediction for Elderly Patients with Sepsis: A Discovery Study. *Comput. Math. Methods Med.* **2022**, *2022*, 4820464. [CrossRef] [PubMed]
247. Tasnim, N.; Al Mamun, S. Comparative Performance Analysis of Feature Selection for Mortality Prediction in ICU with Explainable Artificial Intelligence. In Proceedings of the 2023 International Conference on Electrical, Computer and Communication Engineering (ECCE), Dubai, United Arab Emirates, 30–31 December 2023; pp. 1–6.
248. Alabdulhafith, M.; Saleh, H.; Elmannai, H.; Ali, Z.H.; El-Sappagh, S.; Hu, J.W.; El-Rashidy, N. A Clinical Decision Support System for Edge/Cloud ICU Read-mission Model Based on Particle Swarm Optimization, Ensemble Machine Learning, and Explainable Artificial Intelligence. *IEEE Access* **2023**, *11*, 100604–100621. [CrossRef]
249. de Sá, A.G.C.; Gould, D.; Fedyukova, A.; Nicholas, M.; Dockrell, L.; Fletcher, C.; Pilcher, D.; Capurro, D.; Ascher, D.B.; El-Khawas, K.; et al. Explainable Machine Learning for ICU Readmission Prediction. *arXiv* **2023**, arXiv:2309.13781.

250. Zheng, J.; Yang, L.-H.; Wang, Y.-M.; Gao, J.-Q.; Zhang, K. An explainable decision model based on extended belief-rule-based systems to predict admission to the intensive care unit during COVID-19 breakout. *Appl. Soft Comput.* **2023**, *149*, 110961. [CrossRef]
251. Qiu, J.; Li, L.; Sun, J.; Peng, J.; Shi, P.; Zhang, R.; Dong, Y.; Lam, K.; Lo, F.P.-W.; Xiao, B.; et al. Large AI Models in Health Informatics: Applications, Challenges, and the Future. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 6074–6087. [CrossRef]
252. Betzalel, E.; Penso, C.; Fetaya, E. Evaluation Metrics for Generative Models: An Empirical Study. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 1531–1544. [CrossRef]
253. Lee, J.; Yoon, W.; Kim, S.; Kim, D.; Kim, S.; So, C.H.; Kang, J. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **2019**, *36*, 1234–1240. [CrossRef]
254. Rasmy, L.; Xiang, Y.; Xie, Z.; Tao, C.; Zhi, D. Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *npj Digit. Med.* **2021**, *4*, 86. [CrossRef] [PubMed]
255. Li, Y.; Rao, S.; Solares, J.R.A.; Hassaine, A.; Ramakrishnan, R.; Canoy, D.; Zhu, Y.; Rahimi, K.; Salimi-Khorshidi, G. BEHRT: Transformer for Electronic Health Records. *Sci. Rep.* **2020**, *10*, 7155. [CrossRef] [PubMed]
256. Luo, R.; Sun, L.; Xia, Y.; Qin, T.; Zhang, S.; Poon, H.; Liu, T.-Y. BioGPT: Generative pre-trained transformer for biomedical text generation and mining. *Brief. Bioinform.* **2022**, *23*, bbac409. [CrossRef] [PubMed]
257. Singhal, K.; Tu, T.; Gottweis, J.; Sayres, R.; Wulczyn, E.; Hou, L.; Clark, K.; Pfohl, S.; Cole-Lewis, H.; Neal, D.; et al. Towards expert-level medical question answering with large language models. *arXiv* **2023**, arXiv:2305.09617. [CrossRef] [PubMed]
258. Kraljevic, Z.; Bean, D.; Shek, A.; Bendayan, R.; Hemingway, H.; Yeung, J.A.; Deng, A.; Balston, A.; Ross, J.; Idowu, E.; et al. Foresight—A generative pretrained transformer for modelling of patient timelines using electronic health records: A retrospective modelling study. *Lancet Digit. Health* **2024**, *6*, e281–e290. [CrossRef]
259. Team, G.; Georgiev, P.; Lei, V.I.; Burnell, R.; Bai, L.; Gulati, A.; Tanzer, G.; Vincent, D.; Pan, Z.; Wang, S.; et al. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv* **2024**, arXiv:2403.05530.