Table 1: Timeline of clinical scoring systems.

							mmune	Respiratory	Cardiovascular	Renal	Metabolic	Digestive	Neurological	Other
		Acronym	Range		Aim	Features	1	~	0	~	Σ	Q		<u> </u>
1	1974	GCS	3–15	3	Alertness	Eye, verbal, and motor responses							✓	
2	1974	TRISS	0–12	3	Trauma	GCS, RR, SBP		✓	✓				✓	
3	1978	APACHE II	0–71	14	Severity/Mortality	GCS, RR, $\mathrm{FiO}_2, \mathrm{HR}, \mathrm{MAP}, \mathrm{HCT}, \mathrm{CRE}, \mathrm{pH}, \mathrm{Na}, \mathrm{K}, \mathrm{WBC}, \mathrm{BT}, \mathrm{age}, \mathrm{acute}$ renal failure	✓	✓	✓	✓	✓		✓	✓
4	1989	PBS	0–14	7	Severity/Mortality	CNS, RR, SBP, DBP, BT, mechanical ventilation, cardiac arrest		✓	✓				✓	✓
5	1991	APACHE III	0–299	17	Severity/Mortality	HR, MBP, BT, RR, PaO2, HCT, WBC, CRE, BUN, Na, ALB, BIL, Glucose, Age, Urine output,	✓	✓	✓	✓	✓	✓	✓	✓
6	1992	SIRS	0–4	5	Inflammation	HR, BT, RR, PaCO ₂ , WBC	✓	✓	✓					✓
7	1993	SAPS II	0–163	15	Mortality	Age, HR, SBP, BT, PaO2, FiO2, BUN, NA, K, BIL, WBC, Bicarbonate, urine output, chronic disease, type of admission	✓	✓	✓	✓	✓	✓		✓
8	1994	SOFA	0-24	8	Organ Dysfunction	GCS, PaO ₂ , FiO ₂ , MAP, PLT, BIL, CRE, MVENT		✓	✓	✓		✓	✓	
9	1995	MODS	0–21	9	Organ Dysfunction	GCS, PaO_2 , FiO_2 , MAP , PLT , BIL , CRE , CVP , HR		✓	✓	✓		✓	✓	
10	1998	MEWS	0–14	5	Severity	HR, RR, BT, SBP, AVPU		✓	✓				✓	✓
	1999	MEWS II		6	Severity	HR, RR, BT, SBP, CNS, Urine Output		✓	✓	✓			✓	
11	2003	IPS	0–26	12	BSI	BT, HR, RR, WBC, CRP, SOFA	✓	✓	✓	✓		✓	✓	✓
12	2003	MEDS	0–27		Mortality	tachypnea or hypoxia, septic shock, PLT, Bands, Age, CNS, lower respiratory infection, nursing home residence	 							
13	2005	SAPS III	0–217	33	Mortality	GCS, PaO_2 , FiO_2 , BT, HR, SBP, pH, WBC, PLT, BIL, CRE, age, LOS before ICU, comorbidities and others (see supplementary material)	✓	✓	✓	✓	✓	✓		✓
14	2006	APACHE IV	0–286	35	Severity/Mortality	GCS, FiO_2 , PO_2 , PCO_2 , RR , $MVENT$, HR , MAP , HCT , pH , K , BIL , ALB , CRE , $Urea$, WBC , BT , age , $urine$ output, comorbidities and others (see supplementary materia)	✓	✓	✓	✓	✓	✓	✓	✓
15	2008	Shapiro	0–16	11	Bacteremia	BT, SBP, WBC, PLT, CRE, Bands, age, chills, vomiting, suspected endocarditis, indwelling vascular catheter	✓		✓	✓				✓
16	2011	METTS	0–4	7	Severity	RR, HR, BT, SBP, PO ₂ , consciousness, airway	1	✓	✓				✓	✓
17	2015	NEWS	0-20	6	Severity	RR, HR, BT, SBP, O2SAT, oxygen therapy		✓	✓					✓
18	2016	PRESEP	0-11	7	Sepsis	RR, HR, BT, SBP, SaO ₂ , GCS, Glucose	1	✓	✓		✓		✓	✓
19	2016	qSOFA	0–3	3	Organ Dysfunction	GCS, RR, SBP		✓	✓				✓	

 $For \ detailed \ definitions \ of \ the \ clinical \ scoring \ systems, \ please \ refer \ to \ the \ supplementary \ material \ (SM1-Scores).$

GCS: Glasgow Coma Score; TRISS: Trauma-Injury Severity Score; APACHE: Acute Physiology and Chronic Health Evaluation; PBS: Pitt Bacteraemia Score; SIRS: Systemic Inflammatory Response Syndrome; SAPS: Simplified Acute Physiology Score; SOFA: Sequential Organ Failure Assessment; MODS: Multiple Organ Dysfunction Syndrome; MEWS: Modified Early Warning Score; IPS: Infection Probability Score; MEDS: Mortality in Emergency Department for Sepsis; METTS: Rapid Emergency Triage and Treatment System; N: Number of features; NEWS: National Early Warning Score; PRESEP: Prehospital Early Sepsis; CNS: altered mental status; CVP: Central venous pressure; MVENT: Mechanical ventilation

	Year	Acronym	N	Definition	Immune	Respiratory	Cardiovascular	Renal	Metabolic	Digestive	Neurological	Other
		Bacteremia	1	Blood culture with significant pathogen (see WHO $^{\rm ?}$, CDC/NHSN $^{\rm ?}$, or CLSI $^{\rm ?}$)								✓
		BSI	1+	Bacteremia + Signs of infection								✓
19	1991	SEPSIS-1	5	SIRS	✓	✓	✓					✓
19	2001	SEPSIS-2	6	Suspected Infection + SIRS	✓	✓	✓					✓
19	2016	SEPSIS-3	9	Suspected Infection + SOFA		✓	✓	✓		✓	✓	✓

BSI: Blood-stream infection; N: number of clinical variables; WHO: World Health Organization; CDC/NHSN: Centers for Disease Control and prevention / National Healthcare Safety Network; CLSI: Clinical and Laboratory Standards Institute; SIRS: Systemic Inflammatory Response Syndrome; SOFA: Sequential Organ Failure Assessment;

Datasets

Limited public, diverse, and linked datasets

There is a scarcity of publicly available and well-curated healthcare datasets, making it difficult to replicate findings and develop effective models. Additionally, data often lacks diversity and continuity, hindering the use of sequential models.

Handle clinical codes carefully

Clinical classification systems can be useful for modeling and research, but it is important to use them with caution as they can vary between regions and healthcare systems and they undergo regular updates as our understanding of diseases evolves.

Need for continuous data labelling

Existing datasets often lack continuous labelling for time series data, making it challenging to use supervised learning techniques effectively. Information available varies across time points, making it difficult to track disease progression.

Feature selection

Oversight of clinical management pathway

Studies often neglect the clinical management pathway to select the appropriate predictors. Including features aligned with the management pathway and available at each stage could improve model performance and feasibility of implementation in clinical practice.

Be realistic selecting predictors

The selection of features for machine learning models is important, and factors such as relevance, accessibility, data quality, and frequency of updates should be considered. Feature selection should also be mindful of practicality and avoid biases towards unrealistic data availability.

Be careful with data leakage

Numerous studies inadvertently leaked future information that was not available at the time of assessment (time leakage) or used clinical variables that were part of the label definition (target leakage).

Model validation

Embrace structured and thorough reporting

Studies primarily rely on generic metrics like area under the ROC, sensitivity, and specificity which often provide insufficient insights. Aim for a structure and clear reporting, including inclusion criteria, label definitions, experimental setup, and performance metrics in diverse scenarios.

Describe model's behaviour

Feature importance can be a valuable tool, but it is important to scrutinize the results and question them to ensure they are valid. Assess other aspects of model behavior, such as probability calibration and temporal decay.

Capturing dynamics boosts performance

Conventional machine learning models adopting strategies to account for the temporal aspect of data tend to provide better performance indicating that temporal information is relevant.

Existing limitations with sequential models

Research has gravitated to sequential models to take advantage of temporal information, but the drawbacks associated tend to outweigh their advantages. The performance gains are minimal, and in many cases worse when using external datasets. This is often because there is not enough good quality and granular data available to train these data-hungry models properly. Moreover, even if these models worked well, adapting them to existing healthcare systems presents a big obstacle.

Opportunities

Clinical information available varies across time points

To truly capture the dynamic nature of the disease progression, it is necessary to account for the evolving information landscape and use flexible models that can adapt to this changing data ecosystem.

Clinical history matters

Models often overlook patient history information effectively, especially that collected in primary care. This might prove important, particularly in early stages where symptoms are not evident or laboratory results are unavailable.

Need for new data collection methods

New methods are needed to capture real-time data at different stages of the disease progression. Wearable devices and microneedles show promise for collecting granular, real-time data on vital signs and biochemical markers respectively.

Harness the potential of sequential models

We need better data collection and integration into existing healthcare settings if we want to make better use of sequential models in areas like managing blood-related infections.

Need for research on early stages of disease

Existing research mostly focuses on sepsis (the later stages of the disease) and is oriented towards the assessment of organ dysfunction rather than infection diagnosis. More efforts are needed in understanding the early stages to improve diagnosis and decision-making throughout the patient journey. Moreover, there is evidence of the significant clinical benefit of providing support at earlier stages of the disease.

Continual diagnostic beyond treatment

Existing approaches are often useless after treatment is initiated. However, disease is a continuum and antimicrobial therapy is only the beginning of the diagnostic journey. Newly acquired information should be accumulated and used to confirm, refuse, or refine the previous hypothesis.

Table 4: Comparison of static machine learning models.

_	R	Model	LB	LA	FxT	ROC	PR	SENS	SPEC
	Sta	itic							
	?	?, ? Positive culture ^a , 540.871 (4.3%)	daily	profi	les, 6 c	elinical v	variabl	es, All	
		SVM RFC DTC GNB SVM prospective [?]	NA	Oh	6x1	0.83 0.82 0.78 0.81 0.84	0.88 0.88 0.88 0.87	0.75 0.73 0.64 0.73 0.89	0.91 0.91 0.96 0.90 0.63
	?	?, ? Bacteremia, 518.805 (1%) daily p	orofile	es, 6 c	linical	variable	es, All		
		SVM RFC DTC GNB	NA	0h	6x1	0.88 0.85 0.81 0.80	0.85 0.85 0.82 0.79	0.79 0.54 0.49 0.47	0.81 0.92 0.92 0.89

LA: Look-ahead (h); LB: Look-back (h); R: Reference; a Positive microbiology culture in any specimen type.

Table 5: Comparison of non-sequential machine learning

R	Model	LB	LA	FxT	ROC	PR	SENS	SPEC			
Non-Sequential (temporal)											
?	?, ? Bacteremia, 17038 (7.4%) episodes, 18 clinical variables, All										
	GNB (A) RFC SVM MLP128 (B) MLP256 (C) Ensemble (A+B+C)	72h*	Oh	20x1	0.70 0.73 0.70 0.73 0.73 0.73	-	0.68 0.69 0.81 0.81 0.79	0.66 0.55 0.59 0.53 0.57			
?	?, ? BSI, 5030 (38%) episodes, 28	clinical	varial	oles, IC	U						
	XGB SVM MLP LR	72h	24h	32x1	0.82 0.70 0.67 0.66	-	0.71 0.57 0.40 0.61	0.78 0.76 0.81 0.64			
?	?, ? Sepsis-2, 1394 (11.4%) patients, 9 clinical variables, ICU										
	InSight	5h	3h	†	0.92	-	0.90	0.63			
?	?, ? Sepsis-2, 90353 (1.3%) patient	s, 6 clir	nical v	ariable	s, ICU						
	InSight SIRS SOFA	3h	0h	† 5 8	0.92 0.75 0.63	-	0.80 0.82 0.82	0.95 0.32 0.51			
?	?, ? Sepsis-3, 19048 (9.7%) patients, 8 variables, ICU										
	InSight SIRS SOFA qSOFA InSight	2h 2h 2h 2h 2h	Oh Oh Oh Oh 4h	† 5 8 3 †	0.88 0.61 0.77 0.73 0.77	0.60 0.16 0.28 0.28 0.28	0.80 0.72 0.80 0.56 0.80	0.80 0.44 0.48 0.84 0.52			
?	?, ? Sepsis-2, 600 (50%) episodes,	3 varial	oles, I	CU							
	LR SVM ANN	8h	4h	4x1	0.85 0.88 0.85	-	0.92 0.78 0.75	0.82 0.96 0.94			

^{*}Further in time allowed for some specific sporadic clinical variables. † Unclear number of features due to complex feature engineering.

R	Model	LB	LA	FxT	ROC	PR	SENS	SPEC
Se	quential (temporal)							
?	?, ? BSI, 20850 (7.3%) patient	s, 24 clini	ical va	riables, A	A 11			
	LSTM ^c LR	168h NA	Oh Oh	24x7 24x1	0.97 0.74	0.65 0.48	0.93 0.58	0.98 0.95
?	?, ? BSI, 6557 (13.4%), 33 clin	nical varia	ıbles, l	ICU				
	GRU CNN FFN	48h	0h	33x48	0.84 0.83 0.77	0.48 0.52 0.45	-	-
?	?, ? Sepsis-2, 5803 (6.2%) epis	sodes, 9 c	linical	variable	s, ICU			
	DFN LSTM	5h	3h	20x5	0.92 0.93	-	0.89 0.91	0.94 0.94
?	?, ? Sepsis-2, 51697 (21.4%) episodes, 77 clinical variables, ICU							
	MGP + RNN-MI MGP + RNN RNN RFC + FE MEWS MGP + RNN-MI MGP + RNN	All	0h 0h 0h 0h 12h 12h	†	0.90 0.81 0.77 0.86 0.73 0.79		-	-

MGP + KNN-MI	On	1 0.90	0.74	-	-
MGP + RNN	0h	0.81	0.51	-	-
RNN	0h	0.77	0.50	-	-
RFC + FE	0h	0.86	0.67	-	-
MEWS	All Oh	. 0.73	0.48	-	-
MGP + RNN-MI	All 12h	0.79	0.52	-	-
MGP + RNN	12h	0.74	0.42	-	-
RNN	12h	0.72	0.43	-	-
RFC+FE	12h	0.70	0.42	-	-
MEWS	12h	0.65	0.36	-	-
		1			

Sepsis-2, 33084 (4.56%) episodes, 10 clinical variables

Sepsis-3, 27278 (37%) patients, 63 clinical variables

LA: Look-ahead (hours); LB: Look-back (hours); R: Reference; MI: Missing Indicator; FE: Feature Engineering; LR: Logistic Regression; LSTM: Long Short Term Memory; GRU: Gated Recurrent Unit; CNN: Convolutional Neural Network; FFN: Feed Forward Neural Network; MGP: Multi-Output Gaussian Process; RFC: Random Forest Classifier; ATTN: Attention Layer; LGBM: Light Gradient Boosting Machine; **FxT:** Features inputed to the model x Timesteps;

[†] Unclear number of features due to complex feature engineering.

Table 7: Overview of biochemical, physical, and surrogate markers grouped by organ system.

System	Component	Biochemical Marker	Physical Marker	Surrogate Marker		
Immune	Lymphatic	Leukocytes (WBC), Erythrocytes (RBC), Thrombocytes (PLT), Monocytes (MONO), Eosinophils (EOS), Neutrophils (NEUT), Lymphocytes (LYMPH), Basophils (BASO)		Cancer Immunosupressed redness (rubor) swelling (tumour)		
	Inflammatory	C-Reactive Protein (CRP), Procalcitonin (PCT), IL-6		joint pain/stiffness loss of function		
	Lungs	Fraction of inspired oxygen (FiO2) Endtidal carbon dioxide (EtCO2)	Respiratory Rate (RR)	Cough Chest pain		
Respiratory	Blood	Oxygen saturation – arterial (SaO2, O2SAT) Oxygen saturation – venous (SvO2) Partial pressure of oxygen – ambient (PO2) Partial pressure of oxygen – arterial (PaO2) Partial pressure of oxygen – venous (PvO2) Partial pressure of carbon dioxide – ambient PCO2) Partial pressure of carbon dioxide – arterial (PaCO2) Partial pressure of carbon dioxide – venous (PvCO2)	Peripheral capillary oxygen saturation (SPO2)	Crackles Respiratory Distress Shortness of breath Ventilation support Minute-Ventilation Pleural Efussion Hypoxia COPD		
	Heart	Troponin Test	Heart Rate (HR)			
	Blood	Platelets (PLT), Haematocrits (HCT), Haemoglobin (HGB), Fibrinogen (FIB) International Normalized Ratio (INR)		Palpitations Tachycardia Bradycardia Coronary heart disease		
Cardiovascular	Blood vessels	Bands, Prothrombin Time (PT), Partial Thromboplastin Time (PTT), Activated Partial Thromboplastin Time (aPTT), Mean Corpuscular HGB Volume (MCV), Mean Corpuscular HGB Level (MCH), Mean Corpuscular HGB Concentration (MCHC), Red cell Distribution Width (RDW)	Systolic Blood Pressure (SBP) Diastolic Blood Pressure (DBP) Pulse Pressure (PP) Mean Arterial Pressure (MAP)	Ischemic heart disease Use of vasopressors CVD, CHF Hypertension Hypotension Anemia		
Renal	Kidney	Urea Creatinine (CRE) Blood Urea Nitrogen (BUN)	Urine output	Foley Catheter Acute kidney injury Renal failure CKD		
Metabolic	Name	Glucose, Cholesterol, Sodium (Na), Potassium (K), Magnesium (Mg), Calcium (Ca), Lactate (LAC), Lactate Dehydrogenase (LDA), Creatine Kinase (CK), Creatine Kinase MB (CKMB), Total Carbon Dioxide (TCO2), Low-density Lipoprotein (LDL), Blood gas pH (PH)		Obesity Diabetes Hyperglicemia		
Digestive	Liver	Albumin (ALB) ^c , Bilirubin (BIL), Aspartate aminotransferase (AST), Alanine aminotransferase (ALT)		Jaundice, Cirrhosis Fatty liver disease Hepatitis A/B/C		
	Pancreas		Pancreatic injury, I	Pancreatitis		
Digestive	GI		Diarrhoea, Abdominal pain	, Nausea, Vomiting		
Normala	Brain	Cholinesterase (ChEs)	Delirium, Fatigue, Agitation, Depression, PTSD, Ictus			
Neurological	Sensory organs		GCS, AVI	PU		
Other	Thermoregulatory		Body Temperature (BT)	Fever		
Other	Demographics		Age, Gender, Ethnicity, Height, Weight, BMI			

Table 7: Overview of biochemical, physical, and surrogate markers grouped by organ system. (Continued)

Management	Ward, Length of stay (LOS), Blood culture (BC), X-Ray, Electrocardiogram (ECG), Computerized Tomography (CT), Ultrasound, Use of steroids, Use of antimicrobials, Central Venous Cathether (CVC)
Scores	SOFA, SIRS

GI: Gastroinstetinal Tract; **PTSD:** Post-Traumatic Stress Disorder; **CKD:** Chronic Kidney Disease; **CHF:** Congestive Heart Failure; **CVC:** Central Venous Catheter; **COPD:** Chronic Obstructive Pulmonary Disease; **ENDO:** Endotracheal Tube; **CVS:** Cardiovascular Disease; **GCS:** Glasgow Coma Score; **AVPU:** Alert, Voice, Pain, Unresponsive; **BMI:** Body Mass Index; **HD:** Heart Disease; **SOFA:** Sequential Organ Failure Assessment; **SIRS:** Systemic Inflammatory Response Syndrome; **CVD:** Cardiovascular Disease;

References

- [1] G. Teasdale, B. Jennett, Assessment of coma and impaired consciousness: A practical scale, The Lancet 304 (1974) 81–84. Originally published as Volume 2, Issue 7872.
- [2] S. P. Baker, B. o'Neill, W. Haddon Jr, W. B. Long, The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care, Journal of Trauma and Acute Care Surgery 14 (1974) 187–196.
- [3] W. A. Knaus, J. E. Zimmerman, D. P. Wagner, E. A. Draper, D. E. Lawrence, Apache—acute physiology and chronic health evaluation: a physiologically based classification system, Critical care medicine 9 (1981) 591– 597.
- [4] M. N. Al-Hasan, L. M. Baddour, Resilience of the pitt bacteremia score: 3 decades and counting, 2020.
- [5] W. A. Knaus, D. P. Wagner, E. A. Draper, J. E. Zimmerman, M. Bergner, P. G. Bastos, C. A. Sirio, D. J. Murphy, T. Lotring, A. Damiano, F. E. Harrell, The apache iii prognostic system: Risk prediction of hospital mortality for critically iii hospitalized adults, Chest 100 (1991) 1619–1636.
- [6] R. A. Balk, Systemic inflammatory response syndrome (sirs), Virulence 5 (2014) 20–26.
- [7] J.-R. Le Gall, S. Lemeshow, F. Saulnier, A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study, JAMA 270 (1993) 2957–2963.
- [8] A. E. Jones, S. Trzeciak, J. A. Kline, The sequential organ failure assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation, Critical care medicine 37 (2009) 1649.
- [9] J. C. Marshall, D. J. Cook, N. V. Christou, G. R. Bernard, C. L. Sprung, W. J. Sibbald, Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome, Critical care medicine 23 (1995) 1638–1652.
- [10] C. Stenhouse, S. Coates, M. Tivey, P. Allsop, T. Parker, Prospective evaluation of a modified early warning score to aid earlier detection of patients developing critical illness on a general surgical ward, British Journal of Anaesthesia 84 (2000) 663P.
- [11] D. P. Bota, C. Mélot, F. L. Ferreira, J.-L. Vincent, Infection probability score (ips): A method to help assess the probability of infection in critically ill patients, Critical care medicine 31 (2003) 2579–2584.

- [12] N. I. Shapiro, R. E. Wolfe, R. B. Moore, E. Smith, E. Burdick, D. W. Bates, Mortality in emergency department sepsis (meds) score: a prospectively derived and validated clinical prediction rule, Critical care medicine 31 (2003) 670–675.
- [13] P. G. H. Metnitz, R. P. Moreno, E. Almeida, B. Jordan, P. Bauer, R. A. Campos, G. Iapichino, D. Edbrooke, M. Capuzzo, J.-R. Le Gall, o. b. o. t. S. 3 Investigators, Saps 3—from evaluation of the patient to evaluation of the intensive care unit. part 1: Objectives, methods and cohort description, Intensive Care Medicine 31 (2005) 1336–1344.
- [14] J. E. Zimmerman, A. A. Kramer, D. S. McNair, F. M. Malila, Acute physiology and chronic health evaluation (apache) iv: hospital mortality assessment for today's critically ill patients, Critical care medicine 34 (2006) 1297–1310.
- [15] N. I. Shapiro, R. E. Wolfe, S. B. Wright, R. Moore, D. W. Bates, Who needs a blood culture? a prospectively derived and validated prediction rule, The Journal of Emergency Medicine 35 (2008) 255–264.
- [16] B. R. Widgren, M. Jourak, Medical emergency triage and treatment system (metts): a new protocol in primary triage and secondary priority decision in emergency medicine, The Journal of emergency medicine 40 (2011) 623–628.
- [17] G. B. Smith, D. R. Prytherch, P. Meredith, P. E. Schmidt, P. I. Featherstone, The ability of the national early warning score (news) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death, Resuscitation 84 (2013) 465–470.
- [18] O. Bayer, D. Schwarzkopf, C. Stumme, A. Stacke, C. S. Hartog, C. Hohenstein, B. Kabisch, J. Reichel, K. Reinhart, J. Winning, An early warning scoring system to identify septic patients in the prehospital setting: the presep score, Academic emergency medicine 22 (2015) 868–871.
- [19] D. C. Angus, C. W. Seymour, C. M. Coopersmith, C. Deutschman, M. Klompas, M. M. Levy, G. S. Martin, T. M. Osborn, C. Rhee, R. S. Watson, A framework for the development and interpretation of different sepsis definitions and clinical criteria, Critical care medicine 44 (2016) e113.