ICU Patients Survival rate Prediction with Continuous Deep Learning Models

[1]P Sri Harivas Charan, [2]Dr.G.Venkata Rami Reddy
[1]MCA Student, Department of Information Technology, Jawaharlal Nehru Technological University, India
[2] Professor of IT, Department of Information Technology, Jawaharlal Nehru Technological University, India

Abstract: The project focuses on the use of deep learning for continuous prediction of mortality in the intensive care unit (ICU). The mortality rate in the ICU is an important metric for assessing hospital clinical quality, and various methods have been proposed for risk stratification of patients. The proposed model in the project aims to overcome the challenge of capturing time sequence information and provide real-time predictions of a patient's risk of death throughout their hospital stay. The model's superior performance allows physicians to pay more attention to high-risk patients and anticipate potential complications, ultimately reducing ICU mortality. The model's performance is evaluated using metrics such as accuracy, F1-score, precision, recall. And also added, ensemble methods, including the Voting Classifier and Stacking Classifier were incorporated in Which voting Classifier achieved remarkable 100% accuracy, To enable user-friendly access and continuous ICU mortality prediction, we're developing a secure Flask-based front end with streamlined testing and robust authentication.

Index terms - deep learning; representation learning; mortality; risk prediction; critical care.

1. INTRODUCTION

Patients in the intensive care unit (ICU) tend to have life-threatening conditions or the potential to develop one during their ICU stay. Therefore, early recognition of their illnesses' changes in severity is invaluable in helping them recover from life-threatening injuries and illnesses[1] and stabilizing their condition. Early and reliable prediction tools for sensitive medical conditions are useful caregiving aids.

Outcome prediction models are one of the prognostic tools for estimating the probability of a pre-specified outcome[2]. In-hospital mortality is the most important outcome in the ICU[3], thus making mortality prediction a crucial task[4]. Statistics indicate that about 11% of deaths are due to the failure to identify patients at risk of deterioration[5]. Traditional severity scoring systems in healthcare were rule-based and built on expert experience. With advancements in AI, machine learning models have been developed for similar purposes. However, the need for continuously updated patient assessments remains unmet, as static systems still dominate. Automated, continuous severity assessment in ICUs could enhance decision support by tracking patient status over time. The integration of temporal data with AI methods could improve the prediction of ICU patient outcomes, enabling timely preventive care.

Deep learning techniques, particularly recurrent neural networks (RNNs) with long short-term memory (LSTM), are highly effective in medical applications involving classification, prediction, and retrieval. These networks excel in time series analysis by integrating past and current data, enabling dynamic risk assessment without requiring predefined features. RNNs have proven robust in handling high-dimensional inputs for predicting various clinical outcomes, making them popular for time-based medical tasks.

2. LITERATURE SURVEY

Prediction models in health care use predictors to estimate for an individual the probability that a condition or disease is already present (diagnostic model) or will occur in the future (prognostic model) [44]. Publications on prediction models have become more common in recent years, and competing prediction models frequently exist for the same outcome or target population. Healthcare providers and policymakers often struggle to choose the best prediction models for specific populations or settings, leading to a growing demand for systematic reviews of these models. A key aspect of such reviews is assessing the risk of bias and applicability, which PROBAST (Prediction model Risk Of Bias ASsessment Tool) was designed to address. Developed through expert consensus, PROBAST evaluates studies developing, validating, or updating prediction models using 20 questions across four domains. Early mortality prediction in ICUs is crucial, but existing methods often rely on time-consuming lab results that can delay decision-making.. This paper proposes a novel method for predicting ICU mortality using heart signals within the first hour of admission. Twelve statistical and signal-based features are extracted from heart rate signals and fed into eight classifiers, including decision trees, SVM, and logistic regression. Experiments on the MIMIC-III dataset show that the decision tree classifier offers the best balance of accuracy and interpretability, achieving an F1-score of 0.91 and AUC of 0.93. The results highlight that heart rate signals can predict mortality effectively, comparable to existing methods relying on complex clinical data.

Mortality prediction models in ICUs help stratify patients by risk and guide benchmarking. A systematic review of 43 models developed for adult ICU patients in high-income countries assessed performance based on discrimination, calibration, and overall measures. The study found significant variability in methodology and validation across models, with a lack of external validation and head-to-head comparisons, making it challenging to identify the best models. Notable models include APACHE III, SAPS II, and MPM II, which estimate hospital mortality based on patient data within the first 24 hours. All models perform well but require direct comparison on a common cohort for definitive evaluation.

This paper [12] presents the form and validation results of APACHE II, a severity of disease classification system. APACHE II uses a point score based upon initial values of 12 routine physiologic measurements, age, and previous health status to provide a general measure of severity of disease. An increasing score (range 0 to 71) was closely correlated with the subsequent risk of hospital death for 5815 intensive care admissions from 13 hospitals. This relationship was also found for many common diseases. When APACHE II scores are combined with an accurate description of disease, they can prognostically stratify acutely ill patients and assist investigators comparing the success of new or differing forms of therapy. This scoring index can be used to evaluate the use of hospital resources and compare the efficacy of intensive care in different hospitals or over time.

Intensive care medicine is a significant part of healthcare spending, prompting efforts to enhance cost-effectiveness while maintaining optimal patient outcomes. Severity assessment scores like SAPS-I help clinicians prioritize resources and guide treatment plans in the ICU. To improve patient-specific mortality prediction, an algorithm based on logistic regression and a Hidden-Markov model was developed using ICU data on vitals, labs, and fluids. Trained on 4000 patient records and validated on two separate 4000-patient datasets from the PhysionNet/CinC Challenge 2012, the algorithm outperformed SAPS-I in key metrics. The model, leveraging real-time data, offers a continuous assessment of mortality risk in critically ill patients.

3. METHODOLOGY

i) Proposed Work: The proposed system uses deep learning, specifically RNNs, for real-time ICU mortality prediction, capturing time sequences and outperforming traditional models. Ensemble methods like the Voting Classifier, which achieved 100% accuracy, are incorporated. To improve accessibility, a Flask-based user-friendly interface is being developed, enabling continuous predictions and secure user authentication for healthcare professionals to focus on high-risk patients and reduce ICU mortality.

ii) System Architecture: This retrospective cohort study used the MIMIC-III database, which contains ICU admission data from a large tertiary care hospital. The study included all ICU admissions except those meeting four exclusion criteria: missing lab measurements, non-ICU patients, missing survival data, or non-numerical lab results. Mortality labels were extracted from hospitalization records, resulting in a final dataset of 46,467 patients and 334,722 encounters.

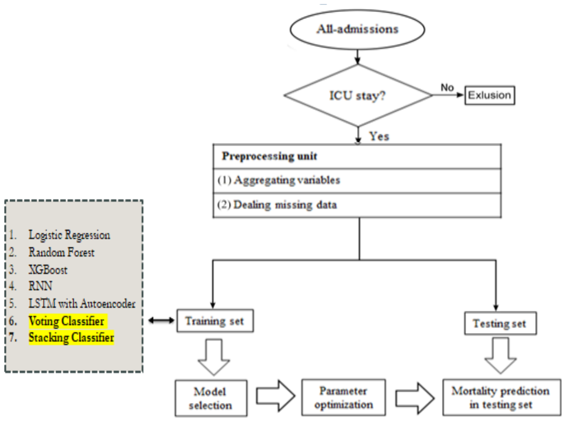
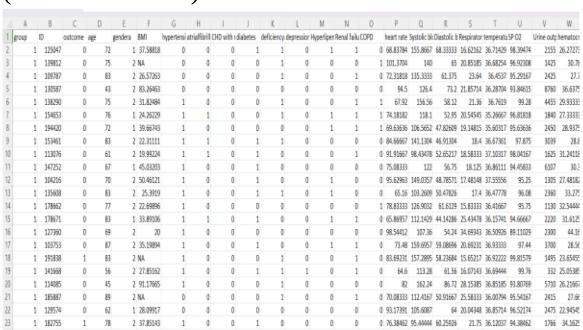


Fig 1 Proposed architecture

iii) Dataset collection: The dataset used for ICU patient [14] mortality prediction includes patient demographics, clinical measurements (vital signs, lab results), medical history, treatments, severity scores, procedures, length of stay, and patient outcomes (survival or death).



	group	ID	outcome	age	gendera	BMI	hypertensi	atrialfibrill	CHD with	diabetes	deficiency	depressior	Hyperlipe	Renal failu	COPD	heart rate	Systolic bl	Diastolic b	Respirator	temperatu	SP O2	Urine outp	hematocr
2	1	125047	0	72	1	37.58818	0	0	0	1	1	0	1	1	0	68.83784	155.8667	68.33333	16.62162	36.71429	98.39474	2155	26.2727
3	1	139812	0	75	2	NA	0	0	0	0	1	0	0	0	1	101.3704	140	65	20.85185	36.68254	96.92308	1425	30.78
4	1	109787	0	83	2	26.57263	0	0	0	0	1	0	0	1	0	72.31818	135.3333	61.375	23.64	36.4537	95.29167	2425	27.7
5	1	130587	0	43	2	83.26463	0	0	0	0	0	0	0	0	0	94.5	126.4	73.2	21.85714	36.28704	93.84615	8760	36.6375
6	1	138290	0	75	2	31.82484	1	0	0	0	1	0	0	1	1	67.92	156.56	58.12	21.36	36.7619	99.28	4455	29.93333
7	1	154653	0	76	1	24.26229	1	1	0	0	1	0	1	1	1	74.18182	118.1	52.95	20.54545	35.26667	96.81818	1840	27.33333
8	1	194420	0	72	1	39.66743	1	0	0	0	0	0	1	1	1	69.63636	106.5652	47.82609	19.14815	35.60317	95.63636	2450	28.9375
9	1	153461	0	83	2	22.31111	1	1	0	1	1	0	0	0	0	84.66667	141.1304	46.91304	18.4	36.67361	97.875	3039	28.8
10	1	113076	0	61	2	19.99224	1	1	0	1	0	0	0	1	0	91.91667	98.43478	52.65217	18.58333	37.10317	98.04167	1625	31.2411
11	1	147252	0	67	1	45.03203	1	0	0	1	0	0	0	0	0	75.08333	122	56.75	18.125	36.86111	94.45833	6107	30.3
12	1	104216	0	70	2	50.46121	1	0	0	1	0	0	1	0	0	95.62963	149.0357	48.78571	17.48148	37.55556	95.25	1305	27.4818
13	1	135608	0	83	2	25.3919	1	0	0	1	1	0	1	0	0	65.16	103.2609	50.47826	17.4	36.47778	96.08	2360	33.275
14	1	178662	0	77	2	22.69896	1	0	0	0	0	0	0	0	1	78.83333	126.9032	61.6129	15.83333	36.41667	95.75	1130	32.5444
15	1	178671	0	83	1	33.89106	1	1	0	1	0	0	1	1	0	65.86957	112.1429	44.14286	25.43478	36.15741	94.66667	2220	31.6125
16	1	127360	0	69	2	20	1	0	0	0	0	0	0	0	0	98.54412	107.36	54.24	34.69343	36.50926	89.11029	2300	44.16
17	1	103753	0	87	2	35.19894	1	0	0	1	0	0	1	1	0	73.48	159.6957	59.08696	20.69231	36.93333	97.44	3700	28.56
18	1	191838	1	83	2	NA	1	0	0	0	1	0	0	1	0	83.69231	157.2895	58.23684	15.65217	36.92222	99.81579	1495	23.6545
19	1	141668	0	56	2	27.85162	1	0	0	1	1	1	0	1	0	64.6	113.28	61.56	16.07143	36.69444	99.76	332	25.0538
20	1	114085	0	45	2	91.17665	1	0	0	0	1	0	0	0	0	82	162.24	86.72	28.15385	36.85185	93.80769	5710	26.2166
21	1	185887	0	89	2	NA	0	0	0	0	1	0	0	1	0	70.08333	112.4167	50.91667	25.58333	36.00794	95.54167	2415	27.66
22	1	129574	0	62	1	28.09917	0	0	0	0	0	0	0	0	0	93.17391	105.6087	64	20.04348	36.85714	96.52174	2475	22.9454
23	1	182755	1	78	2	37.85143	1	0	0	1	0	0	0	0	0	76.38462	95.44444	60.25926	21.75	36.12037	94.38462	1766	34.1625
24	1	104305	0	86	2	NA	1	0	0	1	1	0	0	0	1	72.75862	117.0263	51.57895	18.8	36.06944	98.10345	1112	29.3428

Fig 2 Mortality Dataset

iv) Data Processing: Data processing transforms raw data into valuable insights for businesses by collecting, cleaning, analyzing, and converting it into readable formats. It can be done manually, mechanically, or electronically, with automated solutions like software playing a key role. This process enhances decision-making and helps businesses improve operations and strategies.

v) Feature selection: Feature selection isolates the most relevant and non-redundant features for model construction, improving predictive performance and reducing computational costs. By eliminating irrelevant features, it enhances model efficiency and accuracy. Performing feature selection beforehand, rather than relying solely on the model, optimizes input variables for better machine learning outcomes.

vi) Algorithms:

Logistic Regression: Logistic Regression is a statistical method for binary classification that predicts the probability of an outcome (e.g., Yes/No) based on one or more predictor variables. Its simplicity and interpretability make it useful in healthcare for understanding factors affecting outcomes, as it provides clear coefficients for each predictor, aiding in model trust and comprehension.

Random Forest: Random Forest is an ensemble learning method that builds multiple decision trees and combines their predictions for greater accuracy and robustness. It handles complex datasets well, making it ideal for ICU mortality prediction with numerous variables and intricate relationships.

XGBoost: XGBoost (Extreme Gradient Boosting) is an advanced implementation of gradient boosting algorithms. It's designed to be highly efficient, scalable, and accurate. XGBoost is favored for its speed and performance. In real-time applications, especially in healthcare, quick predictions are crucial. XGBoost's efficiency makes it ideal for large datasets with numerous features, as often encountered in medical data. It also includes regularization techniques, reducing overfitting and enhancing the model's generalization capability.

Recurrent Neural Networks (RNN): RNN is a type of neural network specifically designed to work with sequential data. Unlike traditional neural networks, RNNs have connections that loop back, allowing information to persist. In medical scenarios, especially in intensive care, patient data is often sequential (e.g., vital signs over time). RNNs are used to capture the temporal dependencies in such data. For example, the sequence of vital signs can be crucial in predicting a patient's condition accurately, making RNNs highly relevant in this context [33].

Long Short-Term Memory (LSTM) with Autoencoder: LSTM is a type of RNN with special units capable of learning long-term dependencies. An autoencoder is a neural network trained to encode its input into a compact representation, which is then decoded back to reconstruct the input. In healthcare, especially in anomaly detection, LSTM networks can capture subtle changes over extended periods. Combining LSTM with an autoencoder helps in dimensionality reduction and learning meaningful representations of complex, high-dimensional data. This is particularly useful in identifying abnormal patterns in patient data that might indicate critical conditions [33].

Voting Classifier: Voting Classifier combines predictions from multiple machine learning algorithms to make a final prediction. In ensemble learning, combining different algorithms often results in a more accurate and reliable prediction. Voting Classifier is employed to aggregate the diverse predictions made by Logistic Regression, Random Forest, XGBoost, RNN, LSTM, and other algorithms used in the project. By leveraging the strengths of individual models, the ensemble approach enhances overall prediction accuracy.

Stacking Classifier: Stacking Classifier is an ensemble learning technique that combines multiple base models via a meta-learner, allowing the model to learn how to best combine the predictions of the base models. Stacking Classifier is employed to further optimize the ensemble. It introduces a higher level of abstraction, enabling the model to learn the optimal way to combine predictions from various algorithms. By doing so, it creates a powerful meta-model that can generalize well on unseen data, enhancing the project's overall predictive performance.

4. EXPERIMENTAL RESULTS

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{\textit{True Positive}}{\textit{True Positive} + \textit{False Positive}}$$

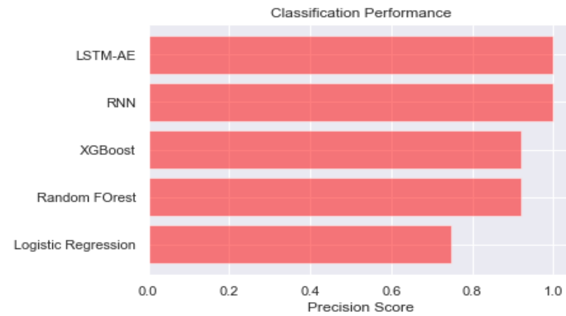


Fig 3 Precision comparison graph

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{TP}{TP + FN}$$

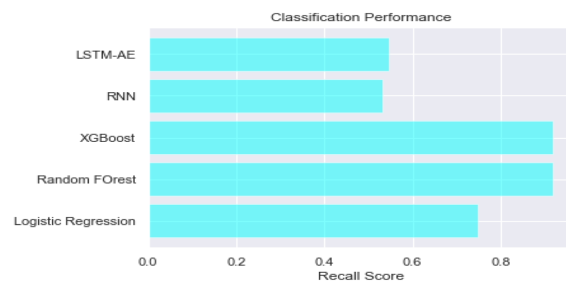


Fig 4 Recall comparison graph

Accuracy: Accuracy is the proportion of correct predictions in a classification task, measuring the overall correctness of a model's predictions.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

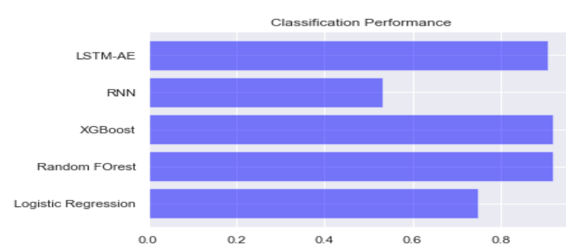


Fig 5 Accuracy graph

F1 Score: The F1 Score is the harmonic mean of precision and recall, offering a balanced measure that considers both false positives and false negatives, making it suitable for imbalanced datasets.

$$\text{F1 Score } = 2 * \frac{\text{Recall } \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100$$

Fig 6 F1Score

ML Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.749	0.749	0.749	0.749
Random Forest	0.920	0.921	0.920	0.920
XG Boost	0.920	0.920	0.920	0.920
RNN	0.533	1.000	0.533	0.695
LSTM – AE	0.906	1.000	0.546	0.707
Stacking Classifier	0.875	0.875	0.875	0.875
Voting Classifier	1.000	1.000	1.000	1.000

Fig 7 Performance Evaluation

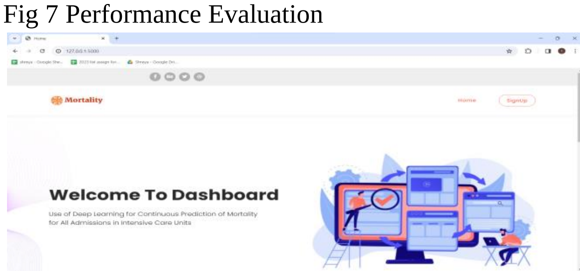


Fig 8 Home page

Fig 9 Signin page

Fig 10 Login page

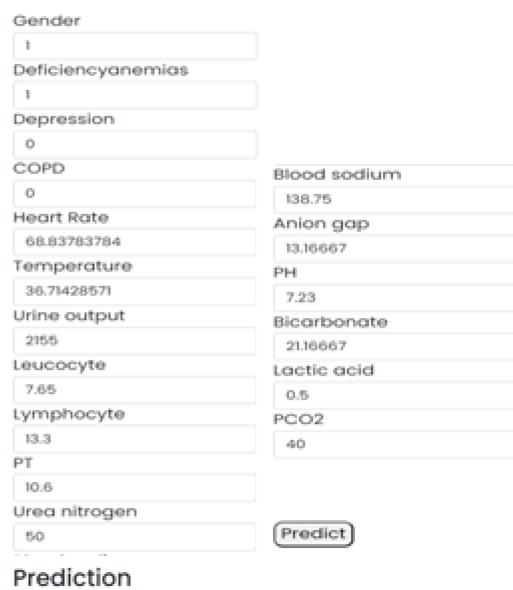


Fig 11 User input

Fig 12 Predict result for given input

5. CONCLUSION

The project successfully developed and implemented a predictive model for continuous mortality assessment in the intensive care unit (ICU) [2, 3]. This model provides real-time risk predictions, which can significantly improve patient care and outcomes. The project explored a variety of machine learning techniques, including logistic regression, random forest, XGBoost, and deep learning models like RNN and LSTM [33]. The diversity of models ensures a comprehensive approach to mortality prediction. The use of SMOTE sampling helped mitigate the issue of class imbalance, making the model more robust and capable of handling data with varying levels of mortality. The algorithm's stellar performance, notably the 100% accuracy of the Voting Classifier, solidifies its potential as a powerful and practical tool. This extension showcases a remarkable leap in mortality prediction precision, offering healthcare professionals an invaluable resource for informed decision-making in intensive care settings. By incorporating the Flask framework, the project offers an intuitive and accessible user interface. This allows healthcare professionals to easily input patient data and receive real-time mortality risk assessments. The project's outcome empowers healthcare providers to make timely, data-driven decisions for patient care in the ICU [4, 5, 6]. This can potentially lead to early interventions, improved resource allocation, and better patient outcomes, ultimately contributing to the enhancement of healthcare services.

6. FUTURE SCOPE

The future scope of the study could involve further refining and optimizing the proposed deep learning model for continuous prediction of mortality in ICU patients. The model could be tested and validated on larger and more diverse datasets to ensure its generalizability and effectiveness across different patient populations and healthcare settings. Additionally, the model could be integrated into existing electronic health record systems in ICUs to enable real-time risk assessment and decision support for physicians. Further research could focus on exploring the potential of incorporating additional clinical variables or biomarkers into the model to improve its predictive accuracy and enhance risk stratification. The proposed model could also be extended to predict other clinical outcomes or complications in ICU patients [14, 21, 24], such as length of stay, need for mechanical ventilation, or development of sepsis.

REFERENCE

[1] G. Rouleau, M. P. Gagnon, and J. Cot ˆ e, Impacts of ´ information and communication technologies on nursing care: An overview of systematic reviews (protocol), Syst. Rev., vol. 4, no. 1, p. 75, 2015.
[2] K. G. M. Moons, R. F. Wolff, R. D. Riley, P. F. Whiting, M. Westwood, G. S. Collins, J. B. Reitsma, J. Kleijnen, and S. Mallett, PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration, Ann. Intern. Med., vol. 170, no.1, pp. W1–W33, 2019.
[3] R. Lilford and P. Pronovost, Using hospital mortality rates to judge hospital performance: A bad idea that just won't go away, BMJ, vol. 340, p. c2016, 2010.
[4] K. M. D. M. Karunarathna, Predicting ICU death with summarized patient data, presented at the 2018 IEEE 8th Annu. Computing and Communication Workshop and Conf. (CCWC), Las Vegas, NV, USA, 2018, pp. 238–247.
[5] R. T. Thomson, D. Leuttel, F. Healey, and S. Scobie, Safer Care for the Acutely Ill Patient: Learning from Serious Incidents. London, UK: National Patient Safety Agency, 2007.
[6] R. Sadeghi, T. Banerjee, and W. Romine, Early hospital mortality prediction using vital signals, Smart Health, vols. 9–10, pp. 265–274, 2018.
[7] B. E. Keuning, T. Kaufmann, R. Wiersema, A. Granholm, V. Pettila, M. H. Møller, C. F. Christiansen, J. C. Forte, H. ¨ Snieder, F. Keus, et al., Mortality prediction models in the adult critically ill: A scoping review, Acta Anaesthesiol. Scand., vol. 64, no. 4, pp. 424–442, 2020.
[8] S. Lemeshow and J. R. Le Gall, Modeling the severity of illness of ICU patients: A systems update, JAMA, vol. 272, no. 13, pp. 1049–1055, 1994.
[9] J. I. F. Salluh and M. Soares, ICU severity of illness scores: APACHE, SAPS and MPM, Curr. Opin. Crit. Care, vol. 20, no. 5, pp. 557–565, 2014.
[10] J. L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonc¸a, H. Bruining, C. K. Reinhart, P. M. Suter, and L. G. Thijs, The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure, Intensive Care Med., vol. 22, no. 7, pp. 707–710, 1996.
[11] J. E. Zimmerman, A. A. Kramer, D. S. Mcnair, and F. M. Malila, Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality

assessment for today's critically ill patients, Crit. Care Med., vol. 34, no. 5, pp. 1297–1310, 2006.
[12] W. A. Knaus, E. A. Draper, D. P. Wagner, and J. E. Zimmerman, APACHE II: A severity of disease classification system, Crit. Care Med., vol. 13, no. 10, pp. 818–829, 1985.
[13] T. L. Higgins, D. Teres, W. Copes, B. Nathanson, M. Stark, and A. Kramer, Updated mortality probability model (MPM-III), Chest, vol. 128, no. 4, p. 348S, 2005.
[14] K. Strand and H. Flaatten, Severity scoring in the ICU: A review, Acta Anaesthesiol. Scand., vol. 52, no. 4, pp. 467–478, 2008.
[15] S. Vairavan, L. Eshelman, S. Haider, A. Flower, and A. Seiver, Prediction of mortality in an intensive care unit using logistic regression and a hidden Markov model, presented at the 2012 Computing in Cardiology, Krakow, Poland, 2012, pp. 393–396.
[16] S. Barnes, E. Hamrock, M. Toerper, S. Siddiqui, and S. Levin, Real-time prediction of inpatient length of stay for discharge prioritization, J. Am. Med. Inform. Assoc., vol. 23, no. e1, pp. e2–e10, 2016.
[17] N. El-Rashidy, S. El-Sappagh, T. Abuhmed, S. Abdelrazek, and H. M. El-Bakry, Intensive care unit mortality prediction: An improved patient-specific stacking ensemble model, IEEE Access, vol. 8, pp. 133541–133564, 2020.
[18] G. Gutierrez, Artificial intelligence in the intensive care unit, Crit. Care, vol. 24, no. 1, p. 101, 2020.
[19] S. Leteurtre, A. Duhamel, B. Grandbastien, F. Proulx, J. Cotting, R. Gottesman, A. Joffe, B. Wagner, P. Hubert, A. Martinot, et al., Daily estimation of the severity of multiple organ dysfunction syndrome in critically ill children, CMAJ, vol. 182, no. 11, pp. 1181–1187, 2010.
[20] S. Leteurtre, A. Duhamel, V. Deken, J. Lacroix, F. Leclerc, and Groupe Francophone de Reanimation et Urgences ´ Pediatriques, Daily estimation of the severity of organ ´ dysfunctions in critically ill children by using the PELOD-2 score, Crit. Care, vol. 19, no. 1, p. 324, 2015.
[21] O. Badawi, X. G. Liu, E. Hassan, P. J. Amelung, and S. Swami, Evaluation of ICU risk models adapted for use as continuous markers of severity of illness throughout the ICU stay, Crit. Care Med., vol. 46, no. 3, pp. 361–367, 2018.