EARLY PREDICTION OF ONSET OF SEPSIS IN CLINICAL SETTING

*Fahim Mohammad¹, Lakshmi Arunachalam¹, Samanway Sadhu¹, Boudewijn Aasman², Shweta Garg³, Adil Ahmed³, Silvie Colman², Meena Arunachalam[†], Sudhir Kulkarni¹, Parsa Mirhaji³

email: fahim.mohammad, lakshmi.arunachalam, samanway.sadhu@intel.com,
baasman@montefiore.org, shweta.garg, adil.ahmed@einsteinmed.edu, scolman@montefiore.org,
meena_aruna@yahoo.com, sudhir.kulkarni@intel.com, parsa.mirhaji@einsteinmed.edu

¹ Intel[®] Corporation, Hillsboro, OR, USA
 ² Montefiore Medical Center, Bronx, NY, USA
 ³ Albert Einstein College of Medicine, Bronx, NY, USA

ABSTRACT

This study proposes the use of Machine Learning models to predict the early onset of sepsis using deidentified clinical data from Montefiore Medical Center in Bronx, NY, USA. A supervised learning approach was adopted, wherein an XGBoost model was trained utilizing 80% of the train dataset, encompassing 107 features (including the original and derived features). Subsequently, the model was evaluated on the remaining 20% of the test data. The model was validated on prospective data that was entirely unseen during the training phase. To assess the model's performance at the individual patient level and timeliness of the prediction, a normalized utility score was employed, a widely recognized scoring methodology for sepsis detection, as outlined in the PhysioNet Sepsis Challenge paper. Metrics such as F1 Score, Sensitivity, Specificity, and Flag Rate were also devised. The model achieved a normalized utility score of 0.494 on test data and 0.378 on prospective data at threshold 0.3. The F1 scores were 80.8% and 67.1% respectively for the test data and the prospective data for the same threshold, highlighting its potential to be integrated into clinical decision-making processes effectively. These results bear testament to the model's robust predictive capabilities and its potential to substantially impact clinical decision-making processes.

Keywords Sepsis Prediction · Septic shock · Sepsis applications · Electronic Health Records · Explainable AI

1 Background

Sepsis is defined as a life-threatening organ dysfunction caused by a dysregulated host response to infection [1]. In normal cases, the body releases chemicals to fight infection. But when sepsis occurs, organs may react differently to these chemicals causing multiple organ failures eventually resulting in death. Early recognition of sepsis is crucial to reduce mortality rates, and administering antibiotics promptly is vital for patient recovery. However, there is currently no single specific test for sepsis detection, so doctors rely on monitoring vital signs, collecting biomarker values and alarming clinical signs like increased heart rate, shortness of breath etc. All the information needs to be gathered and analyzed to define if a patient is septic or non-septic. Various methodologies have been used in the past for predicting early-onset sepsis. In one online challenge, organized by PhysioNet in 2019 [2][3][4], researchers employed variety of techniques, including traditional feature engineering, gradient boosting classifiers, XGBoost [5] ensembles with sliding window features, and other deep learning approaches. Morrill et al [6] achieved promising results by employing Signature features and hand-crafted features, such as shock Index, BUN/CR, Partial SOFA, and SOFA deterioration, to predict sepsis. They emerged as the top performer in the challenge. Du et al [7] proposed a gradient boosting classifier algorithm that utilized comprehensive sets of vital signs, lab features, missing value flags, and changes/variance of

^{*}corresponding author

[†]The author contributed when she was as an employee of Intel[®]. Currently at AMD, Santa Clara.

	Paper 1 (Morrill et al)	Paper 2 (Du et al)	Paper 3 (Zabihi et al)	Paper 4 (Li et al)	Paper 5 (singh et al)	Paper 6 (He et all)	Paper 7 (Lee et al)	Paper 8 (Roussell et al)	Paper 9 (Chang et al)	Paper 10 (Nejedly et al)
Official Ranking	1 st	2 nd	3 rd	4 th	5 th	16 th	22 nd	17 th	8 th	27 th
N. Utility Score (Train/Test)	0.433/0.360	0.455/0.345	0.428/0.339	0.415/0.337	0.404/0.337	0.401/0.313	0.376/0.291	0.393/0.309	0.397/0.328	NA/0.278
Data Imputation	Forward – Fill	Forward - Fill	Linear	Forward Fill	Forward Fill	Forward Fill	decay-to- default imputation and Forward Fill	normalized (0 to 1) & NaN = -1	Recurrent Imputation for Time Series (RITS)	NaN to -1 and rest normalized between 0 to 1 And Forward Fill
Model Techniques	LightGBM	XGBoost	Ensemble of XGBoost	Time phased approach based on ICULOS LightGBM and RNN	XGBoost	Ensemble of XGBoost, GBDT trained on different combination of features	LSTM & Transformer with Auxiliary Loss for Regularization	Embedding module + LSTM (Predit vitals 6hrs ahead)	TCN	LSTM + Genetic Algorithm for HPO
Evaluation Methods	Utility Score	Utility Score	AUPRC, AUROC, F-measure, Utility Score	Utility Score and SHAP <u>values</u>	AUROC , Utility Score	Utility Score	Utility Score, AUROC/AUPRC, Accuracy, F1- score	Utility Score, AUROC/AUPRC, Accuracy, F1- score	Utility score	Utility score
Loss function	Sigmoid & Weighted Log Likelihood	Weighted Cross Entropy	-	Binary Cross Entropy	-	Binary Cross Entropy	Reconstruction loss	Weighted L1 and L2 loss	Utility loss	-

Table 1: Modeling Techniques for Sepsis Prediction. First five papers are the top ranked paper mostly based on tree based boosting algorithms. Paper 6 through 10 employ the Neural Network based models. "-" = not known / NA

vital sign variables. Weighted cross-entropy was chosen as the loss function for their model. Zabihi et al [8] developed an ensemble of XGBoost models, incorporating sliding window and non-sliding window-based features. To handle missing data, they employed linear interpolation, achieving improved imputation accuracy. Li et al [9] introduced a Time-Phased Approach for Sepsis Prediction (TASP), utilizing distinct sub-models for different ICU stages. Their ensemble involved two tree-based methods (LightGBM) and an RNN-based method, each trained on specific data portions with tailored feature sets. Singh et al. [10] used the information missingness along with the XGBoost algorithm and several other variations to achieve good performance.

NN based models: Several researchers explored deep learning methodologies. He et al [11] leveraged a deep feature extractor using pre-trained LSTM [12] and combined it with SOFA and SIRS features, using XGBoost and Gradient Boost Decision Tree models for prediction. Lee et al [13] introduced a decay-to-default imputation technique for handling missing values. They trained an ensemble of Transformers and LSTM models with auxiliary loss for regularization purposes. Roussel et al [14] proposed an embedding module to create compact feature representations. They trained an LSTM model with auxiliary loss to ensure better generalization. Chang et al [15] introduced the TCN model and combined it with the RITS imputation technique to address missing data challenges. Nejedly et al [16] utilized LSTM models and performed hyperparameter optimization to enhance predictive performance. These research works highlight the diverse approaches taken by the scientific community to predict sepsis onset, leveraging a mix of traditional machine learning, ensemble techniques, and deep learning models to achieve accurate and reliable predictions. The Table 1 and Table 2 summarizes the techniques incorporated by above studies.

In Table 1, the five top ranked papers (paper 1 through paper 5) and their corresponding Normalized utility scores on train and test data is shown along with data imputation techniques, modeling techniques, evaluation techniques and loss functions employed by the authors. Paper 6 through paper 10 are mostly based on Neural Network models such as RNN, LSTM or TCN. Table 2 summarized the feature engineering methods used by those papers.

In our current work, we propose the utilization of Machine Learning models to detect sepsis up to 6 hours prior to its onset, leveraging the wealth of clinical data collected at Montefiore Medical Center at Bronx, NY, USA. Based on our review of multiple papers mentioned above, we tried to adopt some of the best features and techniques from those papers. As most of the top-ranked papers used the XGBoost model in some form or other, we used XGBoost to predict the likelihood of a patient developing sepsis within a critical time window. The model was trained on a total of 107 features. We tested the model on 20% of the test data. We also ran an evaluation with prospective data collected by the hospital (completely unseen by the model) at later time point. The following sections outlines the datasets, our method, presents our findings, and offers insights into the implications of early sepsis detection through machine learning.

Feature Engineering	Paper 1 (Morrill et al)	Paper 2 (Du et al)	Paper 3 (Zabihi et al)	Paper 4 (Li et al)	Paper 5 (singh et al)	Paper 6 (He et all)	Paper 7 (Lee et al)	Paper 8 (Roussell et al)	Paper 9 (Chang et al)	Paper 10 (Nejedly et al)
Length Of Stay (LOS)			х	х	х			х		х
Masking		Х			х	Х	х		Х	
MEWS Score					Х					
Age Binning to one hot vector					Х					
Ratio: HR/SBP, SaO2/FiO2	Х			Х						
SOFA	Х			Х		Х		Х		
QSOFA	Х			Х		Х				
Statistical w/ sliding window	х	х	Х	х			х	х		
Delta upto 2 hours	Х			Х	Х					
Worst values of 8 vital signs				Х						
Logarithmic Transformation		х								
Signature Features	х									
Normalization		х						х		
Energy, Shannon Entropy			х							

Table 2: Feature engineering Techniques used for Sepsis Prediction

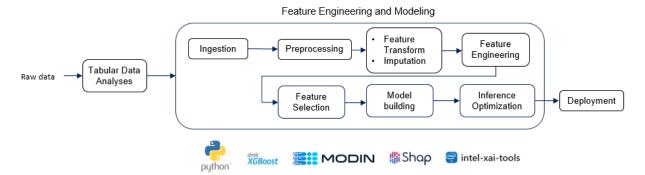


Figure 1: A high-level architecture diagram showing steps.

2 Methods

Our objective is to predict if a patient is septic in timely manner at any time point *t*. To accomplish the objective, we first clean the data, drop unwanted rows and columns, create a mask to capture the missingness in data, impute the data, add time-based statistical features and rule-based generated features to the data. We split the data with 80-20 ratio into train and test subset, train the model until convergence with XGBoost algorithm and test the model on the 20% of retrospective data using metrics like normalized utility score, F1-score etc. We finally test the robustness of model on the prospective data, completely unseen by the model. A high-level architechture diagram is seen in the Figure ??.

2.1 Dataset

Our tabular dataset contains data from 1,039,810 encounters. One encounter (represented as an integer ID) corresponds to a patient's entire stay at the hospital. If a patient is discharged and comes later for hospital admission, then that patient is assigned another encounter ID. Demographics data, vitals and lab results are collected during the entire stay of the patients. These encounters and the associated data are then converted into hourly data, getting a total of 41,355,296 records. The initial data has a total of 85 features shown in Table 3.Most of the lab results and vitals in hourly data has min- and max- components that measure the minimum value and maximum value in that hourly data.

Typically, the Electronic Health Records (EHR) data is noisy, and our first task was to clean the data. This involves renaming some of the columns (Column names have been tidied up to drop spaces between them). The demographic features include encountered, age, sex, and weight. One of the features in earlier studies are shown to be length of stay (LOS) at the hospital or ICU [8] [9] [10]. We calculated LOS by finding the difference between the current timestamp and the hospital admission time. All other time features were dropped. Many vital features and lab test results have a mix of hourly min- and max- values. When multiple readings are taken within an hour, both min- and max- values are

Category	Name of features	Total Features
Demographic features	age, sex, weight	3
Time-base features	Arrival time, Hospital admission time, ICU admission date, Inpatient admission date, Hospital discharge date, ICU discharge date	6
vitals features	Diastolic Blood Pressure, Heart Rate, Mean Arterial Pressure, Respiration Rate, Systolic Blood Pressure, SpO2(o2sat), etCO2 (end tidal), Temperature	8*2 = 16 (min and max)
Lab test results	Alkalinephos, Arterial BE, BUN, Bilirubin direct, Calcium, Chloride, Creatinine-Serum, Creatinine-Urine, FiO2, Fibrinogen, Glucose, HCO3, HCO3 Capillary, Hb, HcT, Lactate, Magnesium, PTT, PaCO2, Phosphate, Platelet Count, Potassium, pH, SGOT (AST), SaO2(ARTERIAL), SaO2(VENOUS), Total Bilirubin, Troponin I, White Blood Cell Count	29*2 = 58 (min and max)
Other calculated features	SOFA base	1
Label	Labels	1
	Total	85

Table 3: List of original features in raw data and their count

recorded. Often, the min- and max- values are the same because these values are not continuously sampled every hour. We measured 8 vitals including Temperature, Heart Rate, Respiration Rate, Systolic Blood Pressure, Mean Arterial Pressure, Diastolic Blood Pressure, SpO2(O2sat) and etCO2 (end tidal). We have min- and max- version of the above vitals making the total number of features to be 16. We are also considering a total of 29 lab test results (58 in total). Additionally, we have the SOFA (Sequential Organ Failure Assessment) score calculated from the raw data. Some post processing features are also included for downstream performance analyses. As we are trying to predict the sepsis 6 hours in advance, all the labels are time-shifted by 6 hours for encouraging early detection of sepsis.

2.2 Data Cleaning

We started with a total of 41,355,296 hourly records. During cleanup we dropped records of patients who stayed less than five hours as we may not have much data for the patient. Most of the patients that are admitted remain in hospital for only few days. However, we do have some patients who stayed in the hospital for several months. To remove some of the outliers, we looked at the percentile values. A 99.99 percentile is close to 695 hours in our data. We round this value to 700 hours which is close to one month of data. We also removed records for patients 3 days after their discharge, some of the test results may arrive after discharge of the patient. We dropped patients older than 105 years old. All the above-mentioned cleaning gave us a total of 36,968,285 records. The Length Of Stay in hours (LOS) is skewed and its distribution can be seen in the Figure 2.

2.3 Removing highly collinear features

The vitals and lab test values contains the min_and max_components, which, most of the time are almost the same. The vitals have a total of 16 (8*2) features while the lab tests have 58 (29*2) features. We systematically remove the highly correlated features from the set of vitals and lab test feature by first finding the correlation matrix which is later converted into distance matrix. The distance matrix is then converted into linkage matrix and then dendrogram is constructed using the hierarchical clustering. Using a Ward distance of 1, we found one representative feature from each cluster. In the end, we got a total of 35 final features from an initial list of a total of 74 vitals and lab test features.

The dataset is highly imbalanced with 5.79% of all patients have sepsis. Finally for each vitals and lab feature, we provide basic statistics in the Table 4.

2.4 Missingness of data and Masking

The vital features have around 20% of missing data. The lab features are sparsely populated with some of them having more than 90% of data missing. Dropping these columns is not a viable option since lab measurements are not taken hourly, and some measurements may not be recorded at all during a patient's entire stay, which is expected. To handle the missing data, a masking technique was employed for both vital signs and lab measurements. This process involved

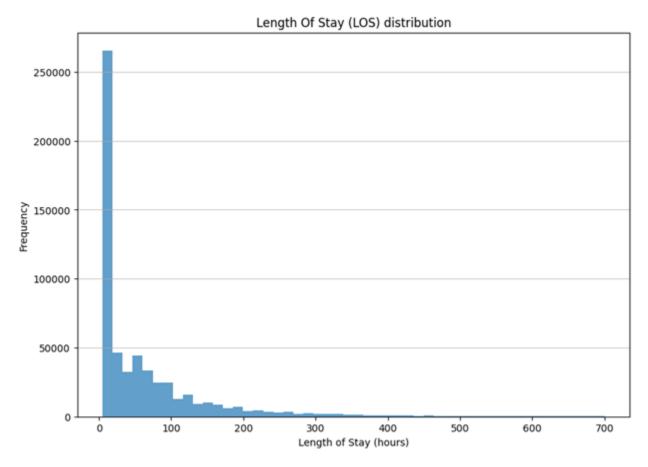


Figure 2: Length of Stay (LOS) of patients in the Hospital

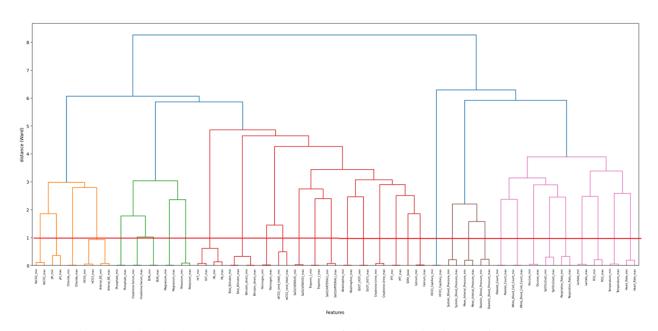


Figure 3: Using dendrogram and ward distance to find representative features from each clusters.

Feature	mean	std	min	25%	50%	75%	max
Alkalinephos	109.13	86.17	35	67	86	117	1000
BUN	21.53	18.05	3	11	16	25	250
Calcium	8.88	0.76	3	8.4	8.9	9.3	19.9
Chloride	102.23	5.33	70	99	103	106	140
Creatinine-Serum	1.45	1.78	0.1	0.72	0.9	1.3	25
Creatinine-Urine	98.38	64.75	5	48.2	81.8	134.7	300
Diastolic Blood Pressure	71.19	13.69	10	62	70	80	236
etCO2 (end tidal)	23.9	12.41	0	15	26	33	50
Fibrinogen	450.59	171.16	70	335	444	562	1000
FiO2	45.39	19.92	21	30	40	50	100
Glucose	157.18	83.25	20	99	129	192	1499
HCO3 Capillary	27.74	5.77	23	24.5	26.8	27.7	40.01
HCO3	25.7	3.97	5	24	26	28	40
НсТ	35.48	6.68	20	30.6	35.8	40.2	65
Heart Rate	79.9	15.51	0	69	79	90	398
Lactate	1.89	1.16	0.5	1.2	1.6	2.2	25
LOS	95.2	104.74	0	23	60	130	700
Magnesium	1.99	0.36	0.5	1.8	2	2.1	14.9
Mean Arterial Pressure	85.38	14.34	5	75	84	94	300
PaCO2	41.6	9.63	10	36	40	45	100
рН	7.27	0.39	5	7.32	7.361	7.4	8
Phosphate	3.53	1.03	1	2.9	3.4	4	15
Platelet Count	245.13	103.62	1	179	232	294	1985
Potassium	4.21	0.59	1.9	3.8	4.2	4.5	8
PTT	37.09	22.33	18	27.6	31	36.8	200.01
Respiration Rate	18.09	2.67	0	18	18	19	120
SaO2(ARTERIAL)	94.34	7.32	9.99	94.7	96.3	97.1	100.01
SaO2(VENOUS)	71.31	15.18	45	58	72	85	99
SGOT (AST)	45.65	132.54	5	19.99	23	36	9911
SOFA base	1.12	1.6	0	0	0	2	17
SpO2(o2sat)	97.84	2.7	10	97	98	100	100
Systolic Blood Pressure	128.77	21.02	10	114	127	142	474
Temperature	98.2	0.74	60.8	97.8	98.1	98.5	109
Total Bilirubin	0.89	2.12	0	0.3	0.5	0.8	68.4
Troponin I	2.12	5.35	0.1	0.15	0.29	1.11	39.84
White Blood Cell Count	8.64	4.37	0	5.9	7.8	10.4	120

Table 4: Basic statistics of the features

creating a binary mask, where present values were marked as 1, and missing values (NaN) were marked as 0. These masked features retains valuable information about the frequency of data collection and may help identify patterns that may be indicative of particular conditions or factors relevant to prognosis (in our case, patient's risk of sepsis). The Masked features can also provide additional information that might not be explicitly captured by the original features. In cases where the presence or absence of specific values in certain features is clinically relevant, the masked features can contribute valuable information to the overall predictive model.

After applying the masking, the missing data was imputed on per patient basis. Two techniques were explored: forward fill and linear interpolation. Forward fill was utilized to impute any missing demographic values. However, based on experimental results, it was found that linear interpolation was a more effective method for estimating missing values in lab features and vital signs. Linear interpolation utilizes a continuous trend between observed data points to make predictions for missing values, preserving the underlying patterns and trends in the data. By combining masking and linear interpolation, we successfully imputed the missing data, making the dataset complete and more suitable for training the predictive model to identify early-onset sepsis effectively.

2.5 Feature Engineering

We performed sepsis prediction using four standard sepsis-related scoring systems: Sequential Organ Failure Assessment (SOFA)[17], quick Sequential Organ Failure Assessment (qSOFA)[18], Modified Early Warning Score (MEWS)[19], and Systemic Inflammatory Response Syndrome (SIRS)[18]. These scoring systems were calculated based on various lab values and vital signs, such as Temperature, Respiratory Rate (Resp), Mean Arterial Pressure (MAP), Bilirubin, and Systolic Blood Pressure (SBP).

To further enhance the predictive capability of the model we also calculated additional indicators, including the Shock Index (Heart Rate divided by Systolic Blood Pressure), the ratio of Blood Urea Nitrogen to Creatinine, and the ratio of Oxygen Saturation (SaO2) to Fraction of Inspired Oxygen (FiO2).

As the ICU Length of Stay was identified as an important feature for sepsis prediction in a related study using PhysioNet data [3]. We calculated this feature for their dataset by deriving the ICU Length of Stay by computing the difference between the ICU admission time and the ICU discharge time for each patient.

To incorporate temporal information and better capture the patient's condition over time, we created sliding windows of the last 6 hours for both the lab values and vital signs. These sliding windows were used as input to their classifier, allowing the model to consider recent trends and changes in the patient's condition.

Additionally, we calculated the differences (deltas) in lab values and vital signs between the current hour and the previous hour. This approach helped assess the stability of a patient's condition by measuring changes between consecutive time points.

Furthermore, we also computed various statistical features, such as the mean, minimum, maximum, median, variance, energy, and slope of the lab values and vital signs within the last 6-hour window. These statistical features provided additional insights into the patient's physiological state and variation over the specified period.

Overall, the combination of standard scoring systems, additional indicators, ICU Length of Stay, sliding windows, delta values, and statistical features aimed to improve the accuracy and effectiveness of sepsis prediction in the study. A list of the definition of generated features are presented in Table 5

2.6 Data Preparation Steps

Following are the steps taken to get the final features from raw data:

- Clean the raw data by renaming columns, removing the features that will not be used and cleaning data by removing outliers or removing patients having insufficient data as describe in preprocessing section.
- Remove highly correlated features explained in the above section. In this step, we started with 74 vitals and lab results features and got 35 features (aka original features) for downstream analyses.
- Generate masked features: Use each of these 35 original features to calculate masked features resulting in 35 additional features. These masked features are binary features and will not have any missing values.
- Perform linear interpolation to fill missing values on the original features. Now the original features will be imputed with the data. Some of the entries will still be missing, such as some lab results whose values are unknown till time point *t*. For patient a lab value is never recorded will have missing entries.
- Generation and selection of statistical features: We calculate the statistical features using including Delta1 (the difference between the current hour and the previous hour), Delta2 (difference between the current hour and

Generated Feature	Definition/Remarks
SOFA	SOFA (Sequential Organ Failure Assessment) is a scoring system used in critical care to evaluate the severity of organ dysfunction. It calculates scores for six organ systems including respiratory (SaO2/FiO2), cardiovascular(MAP), hepatic(bilirubin), coagulation(platelet counts), renal (Creatinine), and neurological (GCS values), providing a comprehensive assessment of a patient's condition. Since our data doesn't have Glasgow Coma Scale (GCS) we just exclude that in SOFA calculation.
qSOFA	qSOFA (quick Sequential Organ Failure Assessment) is a rapid sepsis screening tool used in non-ICU settings. It assesses altered mental status, increased respiratory rate, and low blood pressure, with a score of 2 or more indicating higher risk of sepsis-related complications. Since our data doesn't have Altered ental Status data, we just exclude that in qSOFA calculation.
MEWS	MEWS (Modified Early Warning Score) is a clinical scoring system used to identify and monitor patients at risk of clinical deterioration. It assesses physiological parameters such as heart rate, respiratory rate, blood pressure, temperature, and level of consciousness, assigning scores that indicate the severity of a patient's condition.
SIRS	SIRS (Systemic Inflammatory Response Syndrome) is diagnosed when a patient exhibits at least two of the following criteria: fever or hypothermia, elevated heart rate, increased respiratory rate, or abnormal white blood cell count, serving as indicators of a non-specific inflammatory response to various severe conditions.
Ratio	Ratio of feature "HB/SBP", "SaO2/FiO2", and "BUN/CR"
Delta of window 1	The difference of lab values and vital signs of current hour and previous hour
Delta of window 2	The difference of lab values and vital signs of current hour and two hours before
Variance	Variance is computed by finding the average of squared differences between each data point and the mean. It quantifies the degree of variability or dispersion in the dataset. Variance is calculated over 6 hours of past data
Slope	In time series data, the slope represents the trend's direction and steepness. It is calculated using the formula for linear regression, expressing the rate of change between the independent variable (usually time) and the dependent variable. Here slope is calculated between current hour t and t – 6 data
Energy	Energy is calculated as the sum of squared values, representing the total magnitude or strength of the signal. It quantifies the amplitude of variations in the dataset. Here energy is calculated over past 6 hour data of vital signs and lab values

Table 5: Definition of some of the derived features

two hours before), variance, slope and energy on each of the imputed original features resulting in 175 derived features based on statistical calculations: 35 (delta1) + 35 (delta2) + 35 (variance) + 35 (slope) + 35(Energy) = 175 features. These statistical features are based on time windows to capture the time components in the data. The Variance, Slope and Energy features were calculated on window size = 6. Not all these statistical features significantly contribute to improving the model's performance, so we performed feature selection by fitting histogram classifier until reaching convergence. Following this, we applied permutation importance on the validation set, selecting only the features that exhibited a positive r-value. This exercise reduced the 175 statistical features to 27 features that were considered the most valuable statistical features for our analysis. As randomness is part of the algorithm, the final number of the selected statistical features might vary a little bit on run-to-run basis.

- Next, we calculate some other clinical scoring features which has been of clinical importance and are documented in research papers such as SOFA, MEWS, qSOFA, SIRS, and ratio features such as "HR/SBP", "SaO2/FiO2", and "BUN/CR" resulting in additional 7 features.
- Aggregate all these features to get a total of 107 features (35 original features + 35 masked features + 27 statistical features + 7 other features + 3 demographical features (age, sex, weight)) are the final candidate that goes into training along with the Labels.

3 Model Training

The dataset was split into a training set and a test set using an 80-20% ratio. To ensure the consistency of sepsis prevalence between the two sets, special attention was paid to maintaining the same percentage of sepsis cases in both partitions.



Figure 4: Training Loss curve.

We employed the popular XGBoost algorithm for training the predictive model. Log loss was chosen as the loss function during the training process. Log loss is commonly used in binary classification tasks and is particularly effective when dealing with probabilistic predictions, as it penalizes both false positives and false negatives.

We adopted a dynamic learning rate strategy to optimize the model during training by started with a relatively high learning rate of 0.01 and gradually decreased it as the training proceeded. This approach is known to help the model converge to an optimal solution more effectively.

Training the model was a computationally intensive task, and it took approximately 4 hours on 4th gen Intel Xeon Scalable processors (codenamed Sapphire Rapids) machine. The model was trained for 3000 epochs.

To address the issue of class imbalance in the dataset (which is common in medical datasets where positive cases are often rarer), we experimented with weighted log loss as an alternative loss function. However, observed that using the weighted log loss increased the false positive rate, leading to more false alarms. As false alarms could have serious implications in a medical setting, we reverted to using the regular log loss function.

The training loss significantly decreased in the first 500 epochs and gradually converges around 3000 epochs (see Figure 4). The validation loss is not significantly different from training loss hence we don't see a problem of under or over fitting with validation data.

Cutoff	FPR	PPV	NPV	F1	Accuracy	Time Median	Flag_Rate	Norm. Utility Score
0.1	0.063	0.467	0.997	0.626	0.938	5	0.112	0.641
0.2	0.028	0.653	0.994	0.757	0.968	5	0.076	0.559
0.3	0.015	0.764	0.992	0.808	0.978	4	0.062	0.494
0.4	0.009	0.844	0.989	0.828	0.981	4	0.053	0.435
0.5	0.005	0.896	0.986	0.823	0.982	3	0.047	0.38
0.6	0.003	0.932	0.983	0.802	0.981	3	0.042	0.325
0.7	0.002	0.954	0.979	0.761	0.978	3	0.037	0.267
0.8	0.001	0.972	0.973	0.685	0.973	2	0.03	0.201
0.9	0	0.988	0.964	0.525	0.965	2	0.02	0.118

Table 6: Different metrics on retrospective data

4 Results

4.1 Retrospective Data Analyses

For evaluating the model's performance, we employed the normalized utility score, as defined by the PhysioNet Challenge. The normalized utility score is a comprehensive metric that considers various factors, including sensitivity, specificity, and prediction time. The normalized utility function provides incentives for classifiers that make timely sepsis predictions while imposing penalties for late or missed predictions and for predicting sepsis in non-sepsis cases. The scoring system is described in detail here: [2][4].

- For septic patients (at least one Labels entry of 1), classifiers predicting sepsis within the 12 hours before and 3 hours after the onset are rewarded. where the maximum reward is a parameter (1.0). Conversely, penalties are applied to classifiers failing to predict or predicting sepsis more than 12 hours before (parameter: 0.05 for very early detection, -2.0 for late detection).
- For non-septic patients (all Labels entries of 0), we penalize the False Positives, where the maximum penalty for false alarms is a parameter (0.05; equal to the very early detection penalty).
- We neither reward nor penalize those that do not predict sepsis.

We assess the model's performance in a clinical setting, we examined all predictions for a single encounter. For septic patients, success for the model was deemed by any score greater than the threshold within an six-hour window prior to the time of presentation. A lack of any positive predictions within this window was considered a failure. For non-septic patients, any positive predictions during the encounter were considered a failure, while a lack of any positive predictions between admission to discharge was a success. In addition to precision and recall, to determine optimal use in a clinical setting, we paid particular attention to flag rate (percentage of encounters with at least one positive flag), false positive rate (percentage of encounters with a positive prediction that did not go on to have severe sepsis), and median timeliness (In successful positive cases, median of the hours of first prediction prior to time of presentation). With respect to these metrics, a cutoff of .3 was chosen to be most useful for a live hospital setting. Other commonly used metrics such as FPR (False Positive Rate), PPV (Positive Predictive Value) and NPV (Negative Predictive Values) are also calculated.

Patient level metrics were calculated by varying the confidence threshold to see how the model performed on those confidence levels (Table 6). To determine the optimal performance, we tested different confidence thresholds ranging from 0.1 to 0.9. By analyzing the resulting F1 scores and normalized utility scores, we identified that a threshold of 0.3 provided the most favorable outcome. This threshold yielded a normalized utility score of 0.494.

4.2 Prospective Data Analyses

We wanted to test model's reliability on completely unseen data in hospital settings. We obtained first three months of hospital data from year 2023. The data consisted of total 50,014 encounters (2,030,807 hourly records) of which 6% were septic patients. Out of this, a total of 11,479 encounters have only one hour of data. This can happen in real time scenario, hence to deal with such data we trained. As statistical features based on windowing cannot be obtained

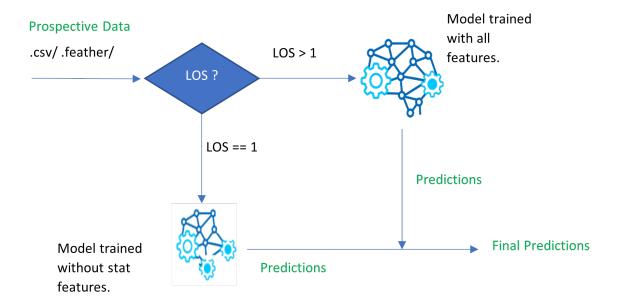


Figure 5: Prospective Data Analyses flow

Cutoff	FPR	PPV	NPV	F1	Accuracy	Time Median	Flag_Rate	Norm. Utility Score
0.1	0.034	0.607	0.985	0.684	0.955	3	0.081	0.554
0.2	0.016	0.731	0.977	0.694	0.964	3	0.057	0.449
0.3	0.009	0.806	0.972	0.671	0.965	2	0.045	0.378
0.4	0.005	0.862	0.968	0.641	0.964	2	0.037	0.323
0.5	0.003	0.906	0.965	0.603	0.963	2	0.032	0.274
0.6	0.002	0.934	0.961	0.556	0.96	2	0.027	0.227
0.7	0.001	0.953	0.957	0.484	0.957	2	0.022	0.177
0.8	0.001	0.965	0.952	0.384	0.952	2	0.016	0.126
0.9	0	0.983	0.946	0.254	0.946	2	0.009	0.067

Table 7: Different metrics on prospective data

for encounters with just one hour of data, we trained an additional XGBoost model only on non-statical features. The remaining data points are sent through the model trained with all the features. In the end we combine all the results together and run our row-level and patient level metrics. See the architecture in 5.

On prospective data, the normalized utility score at 0.378, lower than that of the test data. Similarly the F1 was 67.1%. Despite using this ensembling technique we could see the model performance dropped by 10%, the reason being nearly 40% of the prospective data has only patient record values with less than 5 hours, hence the stats features for these patients couldn't be calculated, as our early prediction window is 6 hours. Since stats features turn out to be important features for sepsis prediction, the model performs poorly in these cases. 7.

5 Explainable AI using SHAP

SHAP (Shapley Additive Explanations) analysis is a powerful technique used in machine learning for interpreting and understanding the predictions made by complex models. Shapley values are calculated relative to a baseline value. This baseline typically represents the expected or average prediction of the model. It provides insights into the contribution of each feature in the prediction of a specific outcome. We used SHAP for better interpretation of the model and plotted

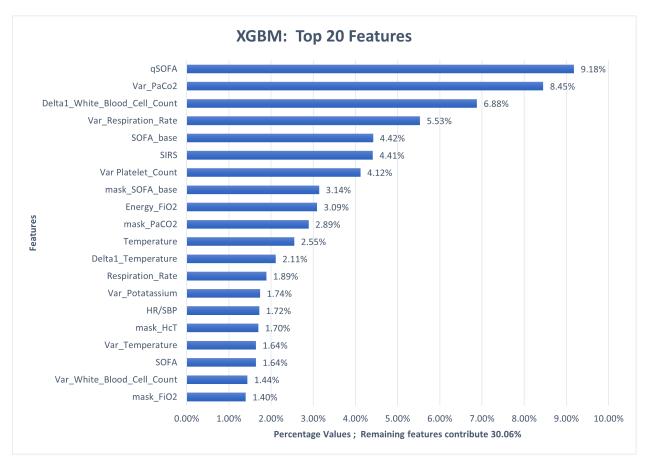


Figure 6: Efficient AI

the feature importance plot (as the percentage of the contribution of individual feature) of the top 20 features that contribute to a model's prediction (See Figure 6)). We observe that the contribution of top 20 features in the prediction of sepsis is almost 69.9%. The remaining 87 features contribute to 30.06%. In our analyses, qSOFA, variance of the PaCO2, SOFA, Delta of WBC, Variation in Respiration, SOFA, SIRS and others to be the most important features.

We looked at some of the scholarly articles to see how the top features are associated with the sepsis.

- qSOFA (quick Sequential Organ Failure Assessment): The qSOFA is used for the rapid assessment of patients suspected of having sepsis [1]. Freund el. al. [20] concluded that the use of qSOFA resulted in greater prognostic accuracy for in-hospital mortality than either SIRS or SOFA, reflected in this plot as well. A similar related study by Fernando et al. [21] shown qSOFA to have poor sensitivity but higher specificity compared to SIRS.
- Var_PaCO2: Monitoring PaCO2 and repsiratory rate is crucial in the care of septic patients to assess respiratory function and guide therapeutic interventions [22]. Abnormal PaCO2 levels can provide important information about the severity of respiratory compromise and the need for respiratory support.
- Delta1_White_Blood_Cell_count: Delta1_White_Blood_Cell_count is the difference in the WBC count of current hour and previous hour. The WBC count is an important component to calculate the SIRS score. The changes in WBC count, body temperature, and heart rate reflect inflammation and the host response to infection [1].
- Var_Respiration_Rate: Respiration rate is an important constituent for calculation of SOFA, qSOFA, SIRS and MEWS. It has been included in the assessment of sepsis according to the Sepsis-3 criteria and other relevant studies [23] [20] [24]. Monitoring respiration rate is crucial in identifying potential respiratory dysfunction associated with sepsis and contributes to the early prediction and management of septic patients. Our model picked up the variance in Respiration rate as the 4th most important feature for sepsis prediction.

- SOFA_base (Sequential Organ Failure Assessment) score: The SOFA score serves as a valuable tool for describing and predicting organ dysfunction in sepsis patients, as validated by studies such as [25] [23] and [26], supporting its widespread adoption in sepsis detection and prognosis. Ferreira et al. [26] concluded that SOFA scores are reliable in evaluating organ dysfunction in the initial days of ICU admission serves as a valuable prognostic indicator. Regardless of the initial score, a rise in SOFA score within the first 48 hours in the ICU is indicative of a mortality rate of at least 50% [26].
- SIRS (Systemic Inflammatory Response Syndrome): SIRS is a generalized inflammatory response, and its progression to sepsis involves evidence of infection, with early recognition crucial for timely intervention [27]. SIRS is a part of historical context and is now replaced by SOFA. Zhang et al. shown that the SIRS criteria are weaker than the SOFA criteria with respect to their predictive efficacy in hospital mortality [28].
- Platelet count: Platelet count is an important hematological parameter in the context of sepsis, as it can reflect the body's response to infection and inflammation. It has been included in the assessment of sepsis to calculate SOFA score, according to the Sepsis-3 criteria [23] [20]. Schupp et al. provided evidence that the platelet count represents a reliable tool for the diagnosis of sepsis and can be effectively used for prognosis of septic shock during the first 10 days of ICU hospitalization [29].
- mask_SOFA_base: masked feature corresponding to the SOFA.
- Temperature and Delta1_Temperature: Monitoring body temperature is essential for recognizing signs of infection and systemic inflammatory response, which are crucial components of sepsis detection. The body temperature is also used to calculate SIRS and MEWS scores. In a study conducted by Khodorkovsky et al. [30], in emergency department patients, hypothermia was correlated with increased time to initial antibiotics, length of stay, rate of ICU admission, and mortality. Another study by Doman et al. [31], shown in a controlled study that the use of temperature does not improve the sepsis prognosis. In our experiment, the predictive contribution of Temperature and Delta1_Temperature, collectively, is 4.66%.
- Respiration_Rate: See Var_Respiration_Rate above.
- Var_Potassium: Potassium imbalance may impact organ functionality, therefore managing Potassium level is crucial in sepsis management. Association between Potassium and prognosis of sepsis has been inadequately investigated. Some machine learning studies found these two to be correlated. Wang et al. [32] found potassium to be the 9th most important feature among a total 55 features in their data and it is closely related to the occurrence of sepsis. Increasing evidence suggests potassium channels play a role in sepsis-related cardiovascular dysfunction, systemic inflammation, and organ damage, potentially influencing sepsis development post-infection [33]. Other studies based on ML algorithms also show the importance of Potassium in sepsis prediction [34] [35].
- HR/SBP (Shock Index): The sepsis shock index is a hemodynamic parameter calculated as the ratio of heart rate (HR) and systolic blood pressure (SBP). A higher sepsis shock index may indicate more severe cardiovascular dysfunction and can be indicative of a worse prognosis in sepsis patients. The HR and/or SBP are used to calculate the early sepsis warning indicators such as MEWS, SIRS and qSOFA. A number of studies found shock index among the most important features [6] [9] [36].
- mask_HcT: As mentioned above, masking is an important feature as they lead to creating additional features that may help retain vauable information about the frequency of data collection. The masked feature of HcT is one of the important features in sepsis prediction. According to Luo et al. [37], The low HcT level is an independent risk factor for the increase of the 30-day mortality in patients with sepsis and can be used as a significant predictor of the clinical outcome of sepsis. Monitoring hematocrit levels is crucial in sepsis prediction as it provides insights into oxygen transport, guiding interventions to optimize tissue perfusion and oxygenation.
- Var Temperature: see Temperature
- SOFA: see SOFA base
- Var_White_Blood_Cell_Count: see Delta1_White_Blood_Cell_count.
- mask_FiO2: FiO2, or Fraction of Inspired Oxygen, is the percentage of oxygen in the air a person breathes. In sepsis detection, monitoring and adjusting FiO2 levels are crucial to ensure optimal oxygenation, addressing potential tissue hypoxia in septic patients. Chicco et al., when analyzing septic shock data, found the FiO2 to be the 5th most important feature when created a model using RandomForest [38].

6 Conclusion and Future Work

In conclusion, this study presents a significant advancement in the early detection of sepsis using Machine Learning models. Sepsis is a life-threatening condition that requires prompt timely intervention for improved patient outcomes, and the absence of a single specific test for sepsis detection makes early recognition challenging. This research demonstrates the potential of leveraging clinical data from Montefiore Medical Center to predict the onset of sepsis up to 6 hours in advance. Early detection of sepsis is paramount in mitigating mortality rates, and our approach offers a valuable tool for timely intervention. By leveraging machine learning techniques, we have shown that it is possible to harness the wealth of clinical data available to enhance sepsis prediction, ultimately leading to improved patient outcomes.

The XGBoost model, trained on a rich dataset obtained at Montefiore Medical Center in Bronx, NY, USA, showcases commendable predictive capabilities. The model achieved a normalized utility score of 0.494 on test data and 0.378 on prospective data with threshold of 0.3, demonstrating its ability to identify sepsis at an early stage accurately. Moreover, the model attained an impressive F1 score of 80.% for test data and 67.1 for the same threshold, highlighting its potential to be integrated into clinical decision-making processes effectively. Overall, this ML/DL technique enable us to use a collection of different scoring techniques, lab values, demographic information and derived features to help prognosis of sepsis.

Moving forward, our work paves the way for further refinement and integration of machine learning models in clinical settings. We are also exploring to extend this solution to multimodal approach by ensembling an XGBoost model and NLP model trained with clinical notes. Continued research and collaboration in this field hold the promise of revolutionizing early sepsis detection and significantly advancing patient care in critical healthcare scenarios.

7 Disclaimer

The information presented in this research paper is intended for educational and informational purposes only and should not be regarded as a substitute for professional medical advice, diagnosis, or treatment. It relies on historical data for training and may not accurately reflect individual cases in real-world scenarios. It is crucial to consult a qualified healthcare professional for precise diagnosis and treatment decisions. The model's predictions serve as supplementary information and should not be considered a definitive diagnosis. Users should be aware of the potential risks associated with misinterpretation or misuse of the model's output. The developers and creators disclaim any liability for damages arising from its use, and users assume full responsibility for the interpretation and application of the model's predictions. Ethical considerations, adherence to patient privacy, and compliance with legal requirements are essential when utilizing this sepsis detection method and model. By using the method and model, users acknowledge and agree to the terms and conditions outlined in this disclaimer.

References

- [1] Mervyn Singer, Clifford S Deutschman, Christopher Warren Seymour, Manu Shankar-Hari, Djillali Annane, Michael Bauer, Rinaldo Bellomo, Gordon R Bernard, Jean-Daniel Chiche, Craig M Coopersmith, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *Jama*, 315(8):801–810, 2016.
- [2] Matthew A Reyna, Christopher S Josef, Russell Jeter, Supreeth P Shashikumar, M Brandon Westover, Shamim Nemati, Gari D Clifford, and Ashish Sharma. Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019. *Critical care medicine*, 48(2):210–217, 2020.
- [3] A Goldberger, L Amaral, L Glass, J Hausdorff, PC Ivanov, R Mark, HE Stanley, and PhysioToolkit PhysioBank. Physionet: Components of a new research resource for complex physiologic signals components of a new research resource for complex physiologic signals. *Circulation*, 101:e215–e220, 2000.
- [4] https://physionet.org/content/challenge-2019/1.0.0/.
- [5] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [6] James Morrill, Andrey Kormilitzin, Alejo Nevado-Holgado, Sumanth Swaminathan, Sam Howison, and Terry Lyons. The signature-based model for early detection of sepsis from electronic health records in the intensive care unit. In *2019 Computing in Cardiology (CinC)*, pages Page–1. IEEE, 2019.
- [7] John Anda Du, Nadi Sadr, and Philip de Chazal. Automated prediction of sepsis onset using gradient boosted decision trees. In 2019 Computing in Cardiology (CinC), pages Page–1. IEEE, 2019.

- [8] Morteza Zabihi, Serkan Kiranyaz, and Moncef Gabbouj. Sepsis prediction in intensive care unit using ensemble of xgboost models. In *2019 Computing in Cardiology (CinC)*, pages Page–1. IEEE, 2019.
- [9] Xiang Li, Yanni Kang, Xiaoyu Jia, Junmei Wang, and Guotong Xie. Tasp: A time-phased model for sepsis prediction. In 2019 Computing in Cardiology (CinC), pages Page–1. IEEE, 2019.
- [10] Janmajay Singh, Kentaro Oshiro, Raghava Krishnan, Masahiro Sato, Tomoko Ohkuma, and Noriji Kato. Utilizing informative missingness for early prediction of sepsis. In 2019 Computing in Cardiology (CinC), pages 1–4. IEEE, 2019.
- [11] Zhengling He, Xianxiang Chen, Zhen Fang, Weidong Yi, Chenshuo Wang, Li Jiang, Zhongkai Tong, Zhongrui Bai, Yueqi Li, and Yichen Pan. Early sepsis prediction using ensemble learning with features extracted from 1stm recurrent neural network. In 2019 Computing in Cardiology (CinC), pages Page–1. IEEE, 2019.
- [12] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural computation, 9(8):1735–1780, 1997.
- [13] ByeongTak Lee, KyungJae Cho, Oyeon Kwon, and Yeha Lee. Improving the performance of a neural network for early prediction of sepsis. In 2019 Computing in Cardiology (CinC), pages Page–1. IEEE, 2019.
- [14] Benjamin Roussel, Joachim Behar, and Julien Oster. A recurrent neural network for the prediction of vital sign evolution and sepsis in icu. In 2019 Computing in Cardiology (CinC), pages Page–1. IEEE, 2019.
- [15] Yale Chang, Jonathan Rubin, Gregory Boverman, Shruti Vij, Asif Rahman, Annamalai Natarajan, and Saman Parvaneh. A multi-task imputation and classification neural architecture for early prediction of sepsis from multivariate clinical time series. In *2019 Computing in Cardiology (CinC)*, pages Page–1. IEEE, 2019.
- [16] Petr Nejedly, Filip Plesinger, Ivo Viscor, Josef Halamek, and Pavel Jurak. Prediction of sepsis using 1stm neural network with hyperparameter optimization with a genetic algorithm. In 2019 Computing in Cardiology (CinC), pages Page–1. IEEE, 2019.
- [17] Faheem Guirgis, Lauren Page Black, and Elizabeth L DeVos. Updates and controversies in the early management of sepsis and septic shock. *Emergency medicine practice*, 20(10):1–28, 2018.
- [18] Robert P Small, Jane L Hendry, Alastair C McKay, Arthur S McPhee, and Gareth E Jones. Is the qsofa more reliable than sirs in detecting post-operative pcnl patients requiring escalation of care? *Journal of Clinical Urology*, 13(1):15–18, 2020.
- [19] Bartosz Mruk, Jerzy Walecki, Piotr Gustaw Wasilewski, Łukasz Paluch, and Katarzyna Sklinda. Interobserver agreement in semi-quantitative scale-based interpretation of chest radiographs in covid-19 patients. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, 27:e931277–1, 2021.
- [20] Yonathan Freund, Najla Lemachatti, Evguenia Krastinova, Marie Van Laer, Yann-Erick Claessens, Aurélie Avondo, Céline Occelli, Anne-Laure Feral-Pierssens, Jennifer Truchot, Mar Ortega, Bruno Carneiro, Julie Pernet, Pierre-Géraud Claret, Fabrice Dami, Ben Bloom, Bruno Riou, Sébastien Beaune, and for the French Society of Emergency Medicine Collaborators Group. Prognostic Accuracy of Sepsis-3 Criteria for In-Hospital Mortality Among Patients With Suspected Infection Presenting to the Emergency Department. *JAMA*, 317(3):301–308, 01 2017. ISSN 0098-7484. doi:10.1001/jama.2016.20329. URL https://doi.org/10.1001/jama.2016.20329.
- [21] Shannon M Fernando, Alexandre Tran, Monica Taljaard, Wei Cheng, Bram Rochwerg, Andrew JE Seely, and Jeffrey J Perry. Prognostic accuracy of the quick sequential organ failure assessment for mortality in patients with suspected infection: a systematic review and meta-analysis. *Annals of internal medicine*, 168(4):266–275, 2018.
- [22] P Nasa, D Juneja, O Singh, R Dang, and V Arora. Severe sepsis and its impact on outcome in elderly and very elderly patients admitted in intensive care unit. *Journal of Intensive Care Medicine*, 27(3):179–183, 2012. doi:10.1177/0885066610397116.
- [23] CW Seymour, VX Liu, TJ Iwashyna, and et al. Assessment of clinical criteria for sepsis: For the third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*, 315(8):762–774, 2016. doi:10.1001/jama.2016.0288.
- [24] Matthew M Churpek, Ashley Snyder, Xuan Han, Sarah Sokol, Natasha Pettit, Michael D Howell, and Dana P Edelson. Quick sepsis-related organ failure assessment, systemic inflammatory response syndrome, and early warning scores for detecting clinical deterioration in infected patients outside the intensive care unit. *American journal of respiratory and critical care medicine*, 195(7):906–911, 2017.
- [25] JL Vincent, R Moreno, J Takala, and et al. The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. *Intensive Care Medicine*, 22(7):707–710, 1996. doi:10.1007/BF01709751.
- [26] Fla'vio L Ferreira, Daniel P Bota, Andreas Bross, Christian Mélot, and Jean-Louis Vincent. Serial evaluation of the sofa score to predict outcome in critically ill patients. *JAMA*, 286(14):1754–1758, 2001. doi:10.1001/jama.286.14.1754.

- [27] Roger C Bone, Robert A Balk, Frank B Cerra, R Phillip Dellinger, Alan M Fein, William A Knaus, Roland M Schein, and William J Sibbald. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Chest*, 101(6):1644–1655, 1992. doi:10.1378/chest.101.6.1644.
- [28] Wei Zhang, Yan Zheng, Xiaoting Feng, Miao Chen, and Yan Kang. Systemic inflammatory response syndrome in sepsis-3: a retrospective study. *BMC infectious diseases*, 19(1):1–10, 2019.
- [29] Tobias Schupp, Kathrin Weidner, Jonas Rusnak, Schanas Jawhar, Jan Forner, Floriana Dulatahu, Lea Marie Brück, Ursula Hoffmann, Maximilian Kittel, Thomas Bertsch, et al. Diagnostic and prognostic role of platelets in patients with sepsis and septic shock. *Platelets*, 34(1):2131753, 2023.
- [30] Boris Khodorkovsky, Elias Youssef, Frosso Adamakos, Tiffany Cina, Amanda Falco, Lauren LaMura, Anthony Marion, Samuel Nathan, and Barry Hahn. Does initial temperature in the emergency department predict outcomes in patients admitted for sepsis? *The Journal of Emergency Medicine*, 55(3):372–377, 2018. ISSN 0736-4679. doi:https://doi.org/10.1016/j.jemermed.2018.06.008. URL https://www.sciencedirect.com/science/article/pii/S0736467918306140.
- [31] Marc Doman, Michael Thy, Julien Dessajan, Mariem Dlela, Hermann Do Rego, Erwann Cariou, Michael Ejzenberg, Lila Bouadma, Etienne de Montmollin, and Jean-François Timsit. Temperature control in sepsis. Frontiers in Medicine, 10, 2023. ISSN 2296-858X. doi:10.3389/fmed.2023.1292468. URL https://www.frontiersin.org/articles/10.3389/fmed.2023.1292468.
- [32] Dong Wang, Jinbo Li, Yali Sun, Xianfei Ding, Xiaojuan Zhang, Shaohua Liu, Bing Han, Haixu Wang, Xiaoguang Duan, and Tongwen Sun. A machine learning model for accurate prediction of sepsis in icu patients. *Frontiers in Public Health*, 9, 2021. ISSN 2296-2565. doi:10.3389/fpubh.2021.754348. URL https://www.frontiersin.org/articles/10.3389/fpubh.2021.754348.
- [33] R Sordi, D Fernandes, BT Heckert, and J Assreuy. Early potassium channel blockade improves sepsis-induced organ damage and cardiovascular dysfunction. *British journal of pharmacology*, 163(6):1289–1301, 2011.
- [34] Suru Yue, Shasha Li, Xueying Huang, Jie Liu, Xuefei Hou, Yumei Zhao, Dongdong Niu, Yufeng Wang, Wenkai Tan, and Jiayuan Wu. Machine learning for the prediction of acute kidney injury in patients with sepsis. *Journal of translational medicine*, 20(1):1–12, 2022.
- [35] Xin Zhao, Wenqian Shen, Guanjun Wang, et al. Early prediction of sepsis based on machine learning algorithm. *Computational Intelligence and Neuroscience*, 2021, 2021.
- [36] Jingyuan Xie, Ting Zhang, Jiandong Gao, Yecheng Liu, Yutong Chen, Mutian Yang, Zetong Liu, Qimin Mei, Zhimao Li, Huadong Zhu, et al. Predicting sepsis within 24 hours at triage stage in emergency department using machine learning. 2023.
- [37] Mengdi Luo, Yang Chen, Yuan Cheng, Na Li, and He Qing. Association between hematocrit and the 30-day mortality of patients with sepsis: a retrospective analysis based on the large-scale clinical database mimic-iv. *PLoS* one, 17(3):e0265758, 2022.
- [38] Davide Chicco and Luca Oneto. Data analytics and clinical feature ranking of medical records of patients with sepsis. *BioData mining*, 14:1–22, 2021.