A Correlation Matrix-based Tensor Decomposition Method for Early Prediction of Sepsis from Clinical Data

Naimahmed Nesaragi[a], Shivnarayan Patidar[a,*], Veerakumar Thangaraj[a]

[a]*Dept. of Electronics & Communication Engineering, National Institute of Technology, Goa*

Abstract

Early detection of sepsis can assist in clinical triage and decision-making, resulting in early intervention with improved outcomes. This study aims to develop a machine learning framework to predict the onset of sepsis through EHR data by applying tensor decomposition on correlation matrices of clinical covariates for every record, arranged on an hourly basis for the length of stay(LOS) in intensive care unit.

A third-order tensor $\theta_{x,y,z}$ representing a clinical correlation among selected 24 covariates for a considered time frame of sepsis onset duration of 6 hours, with a stride of 1 hour is formed for each record. Such a fused tensor with dimensions $\theta \in \mathbb{R}^{24\times24\times LOS}$ for every record undergoes Tucker decomposition with an optimal rank of 10. The factor matrices U_1, U_2, U_3 thus obtained after decomposition are excluded and only the core tensor σ with a dimension $\sigma \in \mathbb{R}^{10\times10\times LOS}$ has been retained, and used to provide latent features for prediction of sepsis onset. A five-fold cross-validation scheme is employed wherein the obtained 100 latent features from the reshaped core tensor, are fed to Light Gradient Boosting Machine Learning models (LightGBM) for binary classification, further alleviating the involved class imbalance. The machine-learning framework is designed via Bayesian optimization, yielding an average normalized utility score of 0.4314 on publicly available PhysioNet/Computing in Cardiology Challenge 2019 training data.

The proposed tensor decomposition deciphers the higher-order interrelations among the considered clinical covariates for early prediction of sepsis and the results obtained are on par with existing state-of-the-art performances.

Keywords: Sepsis, Machine Learning, Tensor Factorization, Correlation Matrix, Early Prediction,model-based diagnosis, PhysioNet

*Dr.Shivnarayan Patidar

Email addresses: naimahmed.nesaragi@gmail.com (Naimahmed Nesaragi), shivnarayan.patidar@nitgoa.ac.in (Shivnarayan Patidar), tveerakumar@nitgoa.ac.in (Veerakumar Thangaraj)

1. Introduction

Tensor factorization is the extension of 2-D factorization of matrices in higher-order. It has evolved as a reliable transformation technique to deal with various challenges related to multi-dimensionality of the EHR data with efficient scalability and interpretability [1, 2]. Tensor-factorization assists us to uncover the meaningful underlying hidden low-dimensional structure within tensor, unlike matrix factorization, which generally faces the problem of yielding non-unique solutions. Meaning, the rank-decomposition of matrices is generally highly non-unique. The uniqueness is guaranteed only under the application of very stringent conditions such as orthogonality, as seen by singular value decomposition (SVD). In contrast, there is no requirement of such strong conditions to be imposed on tensor to offer a unique decomposition [3]. The tensor factorization based techniques have been widely deployed in process monitoring,recommendation systems,and social network analysis, etc [4, 5, 6]. In recent times,tensor-based machine-learning models have assisted in providing solutions to various healthcare related problems, which include image-based treatment and diagnosis,phenotype generation [7, 8], retrieval of medical information [9], and precision medicine [10, 11].

The traditional machine learning techniques mostly rely on vectors and matrices based transformation as inputs, whereas in comparison, tensor-based machine learning techniques use tensor-based higher-order transformations for improved detection and identification tasks. Such models based on tensor-factorization have following prominent advantages: First, multi-aspect features in multi-dimensions are utilized;Second, the problem of sparsity is addressed, which is a big concern for various data analysis tasks, in particular with the EHR data; and finally it incorporates knowledge based on medical-domain from physicians and/or knowledge base systems of medicine in structured data to improve the learning of rule-based models for decision-support. These mentioned advantages of tensor factorizations make it a reliable and promising modelling approach for early prediction of diseases using the EHR data, that is mostly sparse-multidimensional data and needs advice from domain experts to ensure validity of the model [1, 12, 13].

Early risk identification facilitates early prevention and disease management, thereby reducing the individual's suffering from acute and chronic diseases [14, 15]. However, this cannot be appreciated until the EHRs are widely adapted for predictive analytics [16, 17, 18]. The harnessing of EHRs for early prediction of diseases has evolved as an emerging area of investigation for precision medicine, critical care,and medical informatics [19, 20, 21]. Predictive analytics models using EHR data assist clinicians in the assessment of the future risk involved for a particular individual towards a particular chronic condition and also identifies the likelihood of such patients acquiring new diseases. Moreover, predictive models have benefits in terms of costs/risks in comparison to conventional manual treatments and measures for prevention. Altogether,allowing personalized individual care [22, 23].

Sepsis is a perplexed clinical condition that results due to adverse reaction of the patient's body to infection and as a consequence develops multiple organ dysfunction. Sepsis can practically affect numerous organ systems however, the organs involved and the degree of dysfunction varies distinctly among patients and may lead to death in some cases [24, 25].

Despite decades of extensive research and clinical trials carried out, to date, there is no trusted therapy brought into clinical practice targeting dysregulated and inflammatory response that completely characterizes sepsis [26]. Several studies claim that Neutrophil activity plays a crucial role in both initiation and propagation of this complex inflammatory environment by migrating into vital organs like lungs and liver leading to the immunopathogenesis of severe sepsis [27, 28, 29]. Most of these studies consistently identified specific alterations of neutrophil function in sepsis patients [30]; while some of them were linked to poor clinical outcomes [31, 32]. Currently, some knowledge gaps limit the clinical application of these sepsis-related neutrophil dysfunctions.

In addition, even the recent update on sepsis definition by the Third International Consensus Definitions for Sepsis and Septic shock (Sepsis-3), states that neutrophil dysfunctions defined on the background of previous clinical definition may not be relevant to the current practice [29].

At present, physicians rely on clinical tools e.g., Sequential Organ Failure Assessment (SOFA) score to distinguish sepsis from infection, even though signs and symptoms of sepsis overlap considerably with other diseases [29].

At the onset of sepsis, its treatment seems to be relatively easy with the abundant availability of broad-spectrum antibiotics [33]. On the contrary when the sepsis is well developed in its later stages,though diagnosis of sepsis becomes much easier but treatment becomes extremely difficult. Therefore, early diagnosis of sepsis is the need of the hour for better clinical management [34].

Current manual assessment of sepsis using screening tools, like SOFA score for ICU-patients, are complex in terms of measured clinical signs and even lack adequate sensitivity [25, 35]. On contrary, artificial intelligence (AI) and machine learning-based automated decision support systems that harness EHR data have reflected a significant improvement in agreement with these treatment protocols in ICUs, guiding physicians through predefined workflows [36, 37, 38]. The current era wherein we have abundant availability of EHRs has brought more feasibility to such automated realizations [39]. The development of such a machine learning method for the prediction of sepsis onset is an active and important area of investigation.

Contemporary research on prediction of sepsis and its subsequent management strategies for prevention has been mainly addressed by development of machine learning models that applies clinical risk descriptors such as biomarkers, demographics, family, and medical history [22, 40, 41]. Machine-learning (ML) techniques are data-driven and provide a comprehensive approach towards diagnosis and prognostication, unlike traditional statistical methods [42, 43, 44, 45]. Previously, ML methods have been used in various clinical scenarios for the detection of sepsis [46]. Even various deep learning techniques have tried to gain insight into clinical data of sepsis by sequence modeling, and have yielded satisfactory predictive performances [47, 48, 49, 50].

However, with hindsight, providing a direct and straightforward comparison among these machine learning studies for prediction of sepsis is a tedious task. In such studies, the context of each problem addressed is different, and variability exists in the definitions of sepsis onset. Even for the validation different metrics are used on different databases [51].

As a part of recent literature, the PhysioNet 2019 Challenge [51] aimed at the development of such various effective signal processing and machine learning schemes for the early prediction of sepsis from EHR data. For instance, as a submission to challenge, a signature transform-based regression model augmented with hand-crafted features is employed to identify the patient's risk of sepsis based on physiological multivariate time series [52]. In another study [53], a Time phAsed Sepsis Prediction (TASP) machine learning model used short term and long-term window features for three different stages of the LOS to predict the sepsis onset. Similar research [54] on developing gradient boosting machine learning models using sliding and non-sliding window-based features is explored to address the said problem. In addition to this, an explainable artificial intelligence predictor using SHapley Additive exPlainations (SHAP) was employed to build XGBoost models that were trained with statistical and differential features [55].

In this regard, our previous work [56] submitted to PhysioNet 2019 challenge presents a novel clinical application of a machine learning method for the prediction of sepsis onset using ratios and power-based descriptors and derivatives of the given clinical covariates. By applying the Genetic Algorithm (GA), clinically significant ratio-power, and derivative features are optimally selected after transforming the given raw patient's covariates. The feature set is fed to the hybrid Random Under-Sampling-Boosting algorithm, called RusBoost for binary classification. Specifically, from the cross-validation data, various sets of clinical features are tried and tested to find the optimal feature set during GA based optimization that eventually maximizes the underlying classification performance. The optimal feature set is then used with an optimal RusBoost classifier architecture for clinical decision making.

Despite the faithful results obtained by these risk assessment models for sepsis onset prediction, the existing literature is void of the noteworthy application of multidimensional signal processing techniques such as tensor factorization for such high dimensional EHR data [57, 51]. This proposed extension work is yet another innovative multidimensional signal processing and machine learning method on tabular EHR data, the first of its kind to apply tensor factorization for early prediction of sepsis. Here, a 3-way tensor representing the clinical correlations among given covariates for the considered time frame of sepsis onset of 6 hours, and maintaining a stride of 1-hour between such time frames, equal to the length of stay is formed for each record. Such a fused tensor for every record undergoes Tucker decomposition with an optimal rank. The core tensor resulting after decomposition is retained to form the feature set while the 3 other unitary factors obtained after decomposition are excluded. The feature set yielding from core tensor is fed to the machine-learning models designed via Bayesian optimization for clinical decision making.

2. Materials and Methods

2.1. Study Population and Data Sources

PhysioNet Challenge 2019 data comprises 3 different EHRs collected from 3 distinct geographically located U.S. hospital systems [51]. Publicly available data from 2 hospital systems A and B are used for validation and model development purposes while, sequestered

data from the A, B, and independent hospital C are used for testing. The training set publically available has two groups of records. Group A has 790,215 records of 20,336 patients. Group B has 761,995 records of 20,000 patients. The combined data has 1,552,210 records for 40,336 patients. Each patient-record includes data collected for every hour in terms of vital signs, laboratory values, and patient descriptions. Particularly, data for every patient record contains 40 clinical covariates i.e. 8 vital signs, 26 laboratory values, and 6 demographic values. The labelling of the patient data was done adhering to Sepsis-3 clinical criteria. Consent of approval from the appropriate institutional review boards was obtained by organizers to collect data [51]. The given EHR data has a high density of missing information. For this study clinical covariates having missing information less than 95% are only considered which corresponds to 24 clinical covariates. Table 1 presents the considered 24 clinical covariates of the EHR data along with their percentage of missing information and statistics.

Table 1: Details of the considered 24 clinical covariates with missing values in percentage and statistics

Sr. No.	**Covariates**	**Units**	**Missing values(%)**	**Mean**	**Median**
1	Heart rate	beats/min	9.8	84.17	83.00
2	O_2Sat	%	13	96.83	98.00
3	Temperature	^oC	66	35.06	36.80
4	SBP	mm Hg	14.5	122.92	120.00
5	MAP	mm Hg	12.45	82.32	80.00
6	DBP	mm Hg	31.34	57.67	58.00
7	Resp	breaths/min	15.35	18.59	18.00
8	Bicarbonate	mmol/L	94.81	10.76	1.94
9	FiO_2	%	91.66	0.24	0.08
10	pH	-	93.06	3.58	0.85
11	$PaCO_2$	mm Hg	94.44	19.01	3.61
12	BUN	mg/dL	93.13	18.74	14.00
13	Calcium	mg/dL	94.11	5.85	8.00
14	Chloride	mmol/L	94.46	50.10	8.80
15	Creatinine	mg/dL	93.90	1.16	0.80
16	Gl	mg/dL	82.89	116.22	116.00
17	Magnesium	mmol/dL	93.68	1.51	1.90
18	Phosphate	mg/dL	93.98	2.08	2.30
19	Potassium	mmol/L	90.68	3.41	3.90
20	Hematocrit	%	91.14	25.98	29.50
21	Hemoglobin	g/dL	92.61	8.53	9.80
22	PTT	s	94.05	17.12	1.98
23	WBC	count/L	93.59	8.90	8.70
24	Platelet count	count/mL	94.05	161.98	161.00
–	SepsisLabel	1=septic and 0 = nonseptic			

Note: O_2Sat: Pulse oximetry, SBP: Systolic Blood Pressure, MAP: Mean arterial pressure, DBP: Diastolic Blood Pressure, Resp: Respiration rate, FiO_2: Fraction of inspired oxygen, $PaCO_2$: Partial pressure of carbon dioxide from arterial blood, BUN: Blood urea nitrogen, Gl: Serum glucose, PTT: Partial thromboplastin time, WBC: Leukocyte count.

2.2. Data Imputation

The proposed algorithm begins with data imputation as a pre-processing step to fill in the missing values using forward fill imputation for the given EHR data. In a real-time setting, the encountered missing values need to be filled with available past measurements. Thus only the past values of EHR data are fetched for imputation of current observation in hand.

In this study, imputation is carried out into two rounds: first locally, for individual records, and then globally for the combined records of data. In the case of local imputation, the trailing missing values in a row for a particular clinical covariate (or feature vector) are forward filled with the nearest past non-missing value in that row locally for the given record. Ipso facto, if the record encounters *'NaN'* values, in the beginning, i.e. for the first alone measurement at *t=0*, they are retained as it is initially and then later replaced with *'global mean'* for that covariate row obtained by combining all records.

As an illustration, let us consider a matrix A with covariates arranged in a row-wise fashion. The missing entries are imputed locally across the columns (time- axis) to obtain matrix F' and then globally to obtain matrix F.

$$A = \begin{bmatrix} NaN & NaN & 5 & 3 & NaN & 5 & 7 & NaN & 9 & NaN \\ 8 & 9 & NaN & 1 & 4 & 5 & NaN & 5 & NaN & 5 \\ NaN & 4 & 9 & 8 & 7 & 2 & 4 & 1 & 1 & NaN \end{bmatrix}$$

$$F' = \begin{bmatrix} NaN & NaN & 5 & 3 & 3 & 5 & 7 & 7 & 9 & 9 \\ 8 & 9 & 9 & 1 & 4 & 5 & 5 & 5 & 5 & 5 \\ NaN & 4 & 9 & 8 & 7 & 2 & 4 & 1 & 1 & 1 \end{bmatrix}$$

$$F = \begin{bmatrix} 5.8 & 5.8 & 5 & 3 & 3 & 5 & 7 & 7 & 9 & 9 \\ 8 & 9 & 9 & 1 & 4 & 5 & 5 & 5 & 5 & 5 \\ 4.1 & 4 & 9 & 8 & 7 & 2 & 4 & 1 & 1 & 1 \end{bmatrix}$$

2.3. Correlation Matrix

The application of correlation operation exploits the fact that the given clinical covariates for EHR data are clinically related to one another; for instance, patients with faster heart rates typically exhibit high body temperature, while HCO_3 exhibits a good indication of BaseExcess. This inherent relation among the given clinical covariates can be captured by the correlation matrix.

A matrix C is a correlation matrix whose entries quantify the correlation between any 2 covariates, x_i and x_j considered at a time. Here $i,j \subseteq \{1,2,\ldots,N\}$ and N=24 selected covariates.Such correlation matrices are then fused to form a 3-way tensor.

$$C = \begin{bmatrix} Corr[x_1x_1] & \cdots & \cdots & Corr[x_1x_N] \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ Corr[x_Nx_1] & \cdots & \cdots & Corr[x_Nx_N] \end{bmatrix}$$

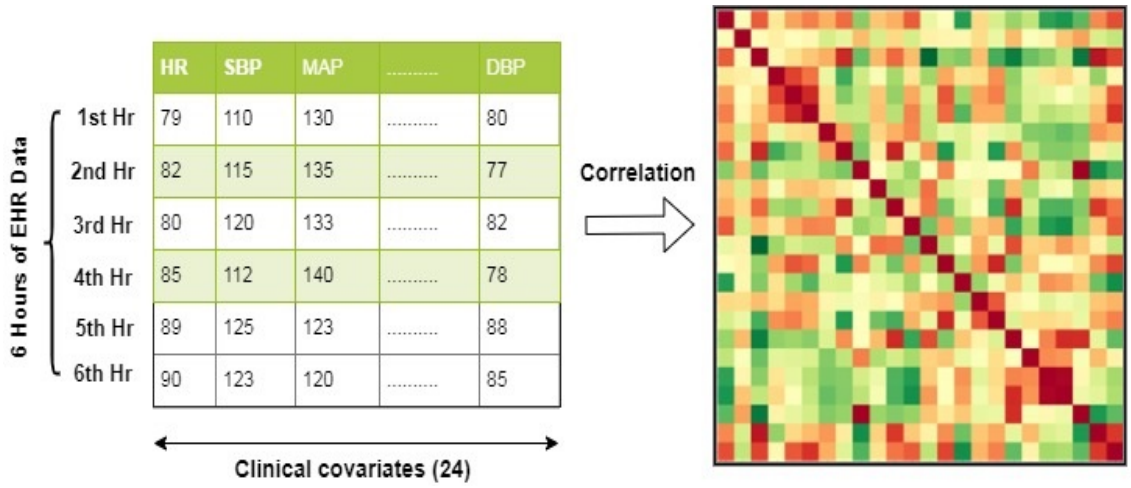

Figure 1: Typical Correlation Matrix of 24 clinical covariates from 6 hours of EHR data

This matrix C can be treated as a vertical slice of the full tensor computed using every 6 hours of EHR data with selected 24 covariates as shown in the Figure 1.

2.4. Correlation Matrix-based Tensor Factorizations for Prediction of Sepsis Onset

The Proposed Tensor-factorization method uses correlation matrices to represent the 2-D relationship between selected clinical covariates for given LOS. Three-way tensors are formulated to envision the high-dimensional interrelations between selected vital signs and laboratory values as clinical covariates of the patients for every hour after admission to ICU.

We collected all the correlation matrices equal to the LOS from an individual record and built a third-order observed tensor $\theta_{x,y,z}$. Entries at (x, y, z) of the tensor θ is the correlation between any 2 covariates, along x and y axis, considered at a time and z axis represents the count of such matrices for every 6 hours of EHR data, computed with a stride of 1 hour, equal to LOS resulting in dimension $\theta \in \mathbb{R}^{24\times24\times LOS}$

In the process of constructing the tensor θ for every record, the first 6 correlation matrices are kept identical which represents the correlation between the selected covariates for the first 6 hours of data. i.e correlation matrices from $T = 1$ to $T = 6$ are identical. Next, the trailing correlation matrices from $T = 7$ (representing the correlation between covariates for next 6 hours of data i.e. from 2^{nd} hour to 7^{th} hour), $T = 8$ (representing the correlation between covariates for next 6 hours of data i.e. from 3^{rd} hour to 8^{th} hour), and so on to $T = LOS$ are unique and are fused to form a three-way tensor as shown in Figure 2.

Such a fused tensor θ for every record undergoes tucker decomposition with an optimal choice of rank equal to 10. Decomposition yields core (σ) and factor matrices (U_1, U_2, U_3) as latent factors with dimensions $\sigma \in \mathbb{R}^{10\times10\times LOS}$, $U_1 \in \mathbb{R}^{24\times10}$, $U_2 \in \mathbb{R}^{24\times10}$, and $U_3 \in \mathbb{R}^{LOS\times LOS}$ respectively as shown in Figure 3. Among the $latentfactors$, only the σ is retained and all other factor matrices are excluded. The extracted σ is unfolded to give 100 latent features as a feature set. In the context of this sepsis onset prediction task, we explored matrix/tensor factorization to extract these latent features from the core of every

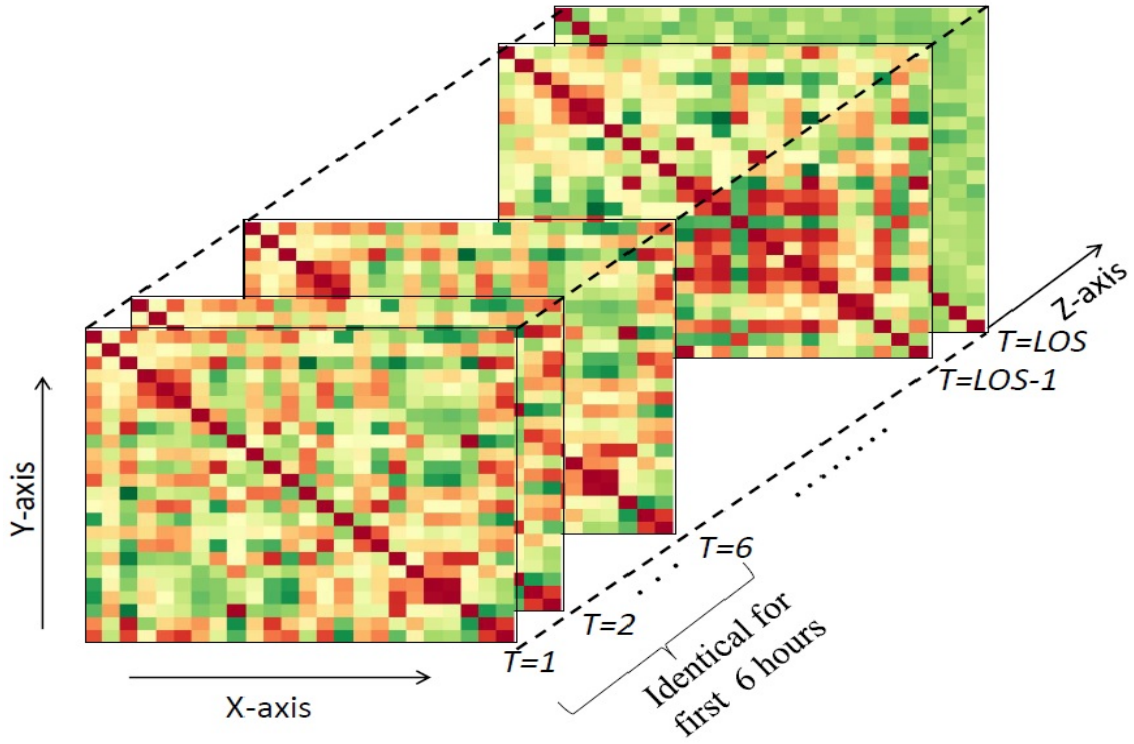

Figure 2: Input data representation for proposed tensor factorization.Each slice along z-axis is a correlation matrix corresponding to every 6 hours of EHR data, arranged with a stride of 1 hour.

patient-tensor representing correlation among covariates. Then, the tensor-factorization based machine learning models are trained with these latent features for the said prediction. Please refer to this recent review [1] for preliminaries and further details on matrix and tensor factorizations.

2.5. Sepsis onset prediction system

Figure 4 presents the work-flow of the proposed approach for the sepsis onset prediction. At the onset, an appropriate set(selected 24 covariates) of EHR data is imputed using forward-fill as described earlier in *subsection* 2.2. The tensors based on the correlation matrices between the selected covariates are constructed and decomposed using tucker decomposition for every record. The extracted cores are unfolded and combined to form the feature set. These features are used to build tensor-based gradient boosting machine learning models in a five-fold cross-validation scheme wherein the model in each fold is trained with, 80% of stratified EHR data, and tested with, hold out 20% as a test set for the prediction.

3. Experiments and Results

The proposed framework performs predictive analytics from the given patient-records to determine the risk of development of sepsis onset before 6 hours. Evaluation is done

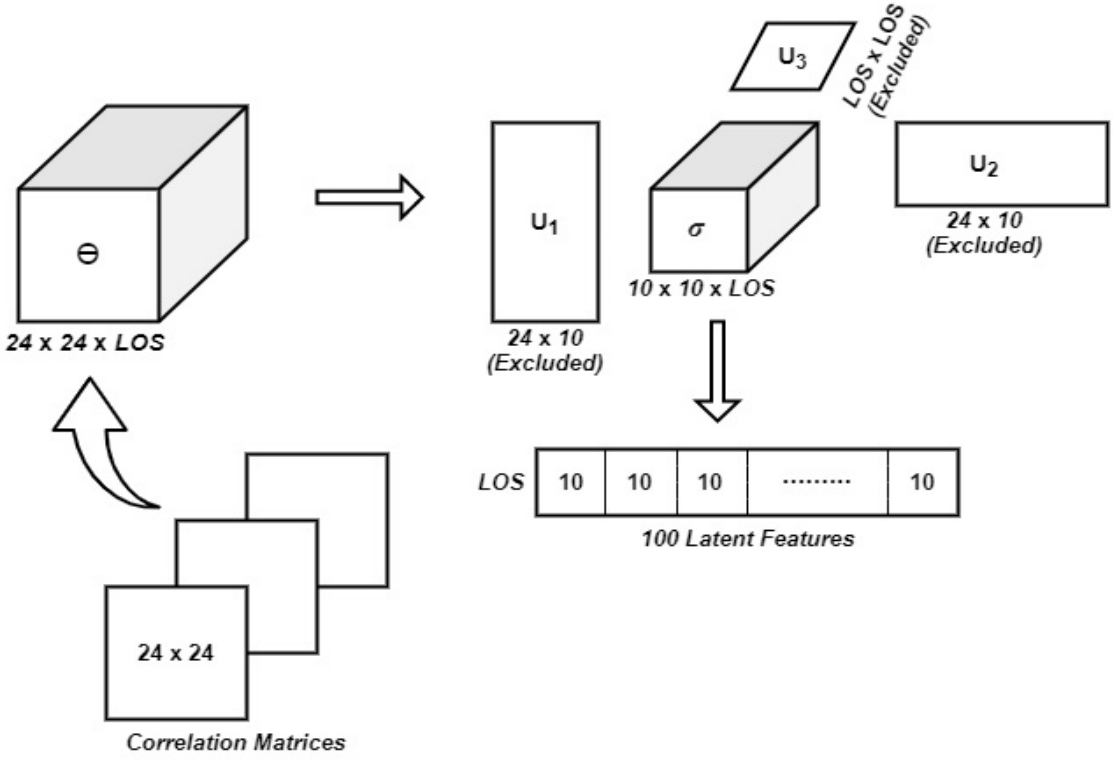

Figure 3: Proposed tensor-factorization for prediction of sepsis onset

using a continuous-valued utility score as defined by PhysioNet 2019 challenge organizers for each prediction. The utility function rewards or penalizes models for their predictions within twelve hours before and three hours after onset of sepsis time and is normalized as described in [51]. Using a five-fold cross-validation scheme five LightGBM models are designed based on patient-wise stratified five-folds each containing unique 20% of the entire training set. The hyper-parameters of the above models that minimize cross-validation loss are obtained by using automatic hyper-parameter optimization utility '*bayesopt*' in Python [58, 59]. The underlying objective function formulated for the optimization is intended to maximize the utility score. The given software utility finds optimal parameters automatically using Bayesian optimization. At the outset, the hyper-parameters values for the best optimized model are listed in the Table 2.

The results obtained by the proposed method on the entire training data under a five-fold cross-validation scheme are summarised in Table 3. Results also include performances on inter-cohorts. To ensure that the models trained are learned only on the underlying pathology and independent of the cohorts, we considered inter-cohort training and testing schemes. i.e models trained with the data of cohort A were scored on cohort B data and vice versa. This certainly avoids the doubt of the over-fitting, thus increasing the robustness of the framework. Moreover, this is supported by Biglarbeigi et.al [60] in their study.Inter-cohort scores of 0.3001 and 0.3602 are obtained for A and B respectively.

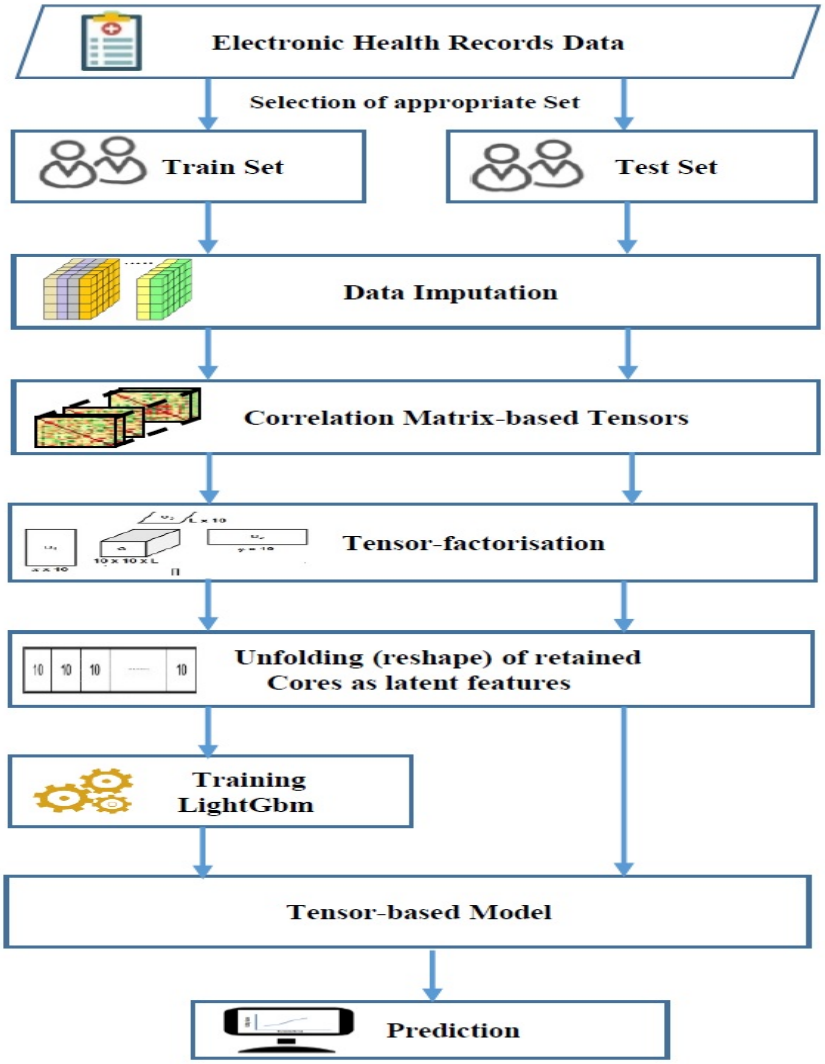

Figure 4: Workflow of sepsis onset prediction system

3.1. Comparison with Baselines

To emphasize and justify the clinical relevance of correlation matrix-based tensor factorization, a comparative analysis of the derived latent features are done with experiments excluding computation of correlation matrix and experiments without tensor-factorization.

As a part of baseline-studies, we performed two well-tuned simpler methods: In the first method, the imputed version of 24 selected covariates data is fed directly to LightGBM models without calculating the correlation matrix and without performing tensor factorization under a five-fold cross-validation scheme. In the second method, the correlation matrices fused tensor with dimensions $\theta \in \mathbb{R}^{24\times24\times LOS}$ is reshaped to $\sigma \in \mathbb{R}^{LOS\times576}$ and fed to LightGBM models without tensor factorization in a five-fold cross-validation scheme. The results of these baseline studies are presented in Tables 4 and 5 accordingly. The performance of

Table 2: Hyper-parameters values for the optimized models.

Hyper-parameters	**Values**
lambda_l1	4.044
lambda_l2	6.72
learning_rate	0.08874
max_depth	17
min_data_in_leaf	52
num_leaves	45
reg_alpha	9.005
reg_lambda	6.675
scale_pos_weight	23.13

Table 3: Summary of results for the proposed method.

Fold	**AUROC**	**F1 Score**	**Utility Score Proposed(Previous)**
1	0.8584	0.1093	0.3766(0.3906)
2	0.8590	0.1557	0.4680(0.3915)
3	0.8664	0.1263	0.4562(0.3982)
4	0.8629	0.1542	0.4580(0.4429)
5	0.8597	0.1075	0.3984(0.3927)
Average	0.8613	0.1307	0.4314 (0.4013)
Training on cohort A and Testing on cohort B			0.3001
Training on cohort B and Testing on cohort A			0.3602

the proposed scheme with tensor factorization based latent features alone was significantly higher as compared to the other baselines.

Performing tensor-factorization, requires an optimal value of rank. Determining the optimal value of the rank is a difficult task [61]. However, in this study ,we empirically experimented with different values of the rank and eventually best experimental results for the utility score were obtained with $rank = 10$. Further, this low value of rank reduces the complexity of the core-tensor-based machine learning model in terms of latent features. Because for each increasing value of rank R, the number of latent features gets increased by order R^2. So we set values of rank in the first and second mode of tucker decomposition to 10. We implemented the proposed method by using *'tensorly toolbox'* in python [62].

4. Discussion

This study explores the use of correlation-matrix from selected clinical covariates as input for the tensor factorization-based method. As a result of factorization, core-tensor-based latent features are generated as output,which are further used for the prediction of sepsis in ICU. The primary contribution of this study is fourfold. First, only 24 clinical covariates out

Table 4: Summary of the baseline results using selected 24 clinical covariates only under a five-fold cross-validation scheme

Fold	**AUROC**	**F1 Score**	**Utility Score**
1	0.7675	0.1121	0.1744
2	0.7053	0.0920	0.1672
3	0.7514	0.1007	0.2333
4	0.7102	0.0957	0.2220
5	0.7410	0.0979	0.1550
Average	0.7351	0.0967	0.1904

Table 5: Summary of the baseline results using correlation matrices of covariates without tensor-factorization under a five-fold cross-validation scheme

Fold	**AUROC**	**F1 Score**	**Utility Score**
1	0.6913	0.0920	0.0902
2	0.6313	0.0665	0.0730
3	0.6844	0.0807	0.1284
4	0.6314	0.0698	0.1135
5	0.6678	0.0754	0.777
Average	0.6612	0.0768	0.0965

of given 40 variables are used to obtain faithful results. Next, higher-order multidimensional interrelations among clinical covariates are captured for classification. Further, it excludes demographics and length of stay clinical covariates that may bias the algorithms, and finally, achieve a significant utility score with fewer latent features as a feature set compared to the contemporary state of arts.

Our previous work [56] was based on capturing static interrelations among the vital signs, rather than considering them independently, which could enhance the clinical potential of the extracted feature set for automated analysis. In literature, such interrelations among the covariates in the form of ratio and power have been found to improve the capabilities of anomaly detection systems [36, 63]. As an extension, this proposed study explored the potential of higher-order interrelations through multidimensional latent information among covariates. This approach is quite efficient compared to the previous work in terms of less number of features with more predictive power. The comparative analysis of the proposed work as evaluated against our previous work developed in [56] is also shown in Table 3. The proposed model outperforms that developed in [56] consistently across various criteria and validation test sets. This shows that the correlation-based tensor-factorization developed in this work helps the early identification of patients who will develop sepsis in near future.

The PhysioNet challenge 2019 results available on training data [51], provided several solutions employing 5 or 10 fold cross-validation schemes, and obtained utility scores in the range of 0.37 to 0.44 as listed in Table 6. The top-ranked team Morrill et al. proposed a signature-based transform scheme and obtained a utility score of 0.434 [52]. Zabihi et al. [54] employed an ensemble of XGBoost models to obtain a utility score of 0.4281 while Yang et al. [55] obtained 0.43 using fusion-based XGBoost learning under 5-fold cross-validation schemes. Lyra et al. [64] obtained a utility score of 0.376 using random forest and Christopher Kok et al. [65] obtained a utility score of 0.43 using temporal convolution network (TCN) under a 10-fold cross-validation scheme. Another study employing TCN was proposed by Chang et al. [66] yielding a utility score of 0.4170. Lee et al. [67] employed graph convolutional networks (GCN) to obtain a utility score of 0.3820. Du et al. [68] introduced a gradient boosting decision tree (GBDT) based classifier to estimate the sepsis onset. Li et al. [53] obtained a utility score of 0.43 using a time-phased model under a 5-fold cross-validation scheme, while He et al. [69] employed ensemble learning of both deep and hand-crafted features under 10-fold cross-validation to obtain a utility score of 0.4010. Recently Rafiei et al. [70] performed a unique predictive analysis study with the LSTM-CNN model for different sepsis onset window durations using this challenge data, reporting best results for 4 hours window with AUROC of 0.92 , specificity and sensitivity of 0.81 and 0.85 respectively. This study did not report the custom-defined challenge-metric utility score.

Apart from challenge submissions, the proposed work is also compared with other studies [42, 48, 47, 49, 50, 46] employing machine and/or deep learning methods for the prediction of sepsis onset using different datasets with different prediction windows. These studies are also summarised in Table 6. Thakur et al. [42] performed an evaluation study to compare the performance of SOFA score with that of binary logistic regression models to predict neonatal sepsis. This study employed 1580 neonates from MIMIC-III database and reported sensitivity and specificity of 0.3125 and 0.9730 respectively. In another study by Svenson

et al. [48] a channelled LSTM model with 10 channels was tasked to predict mortality in sepsis with 48 hours onset against the SOFA score. This study also used MIMIC-III critical care database with 5,784 critical care admissions and reported AUROC and AUPRC of 0.8920 and 0.4780 respectively. Lauritsen et al. [47] employed LSTM-CNN model for early detection of sepsis utilizing EHRs event sequences from multiple Danish hospitals and obtained AUROC 0.7560 (24 hours before onset) and AUROC 0.8560 (3 hours before onset) respectively. Automated prediction of sepsis using Multi-task Gaussian Process (MGP) based TCN and Dynamic Time warping (DTW) was studied by Moor et al. [49, 50] using MIMIC-III database with 570 sepsis, and 5618 normal subjects.

Statistical analysis of certain observational clinical studies [27, 29] claiming the role of neutrophil activity leading to the immunopathogenesis of severe sepsis are also summarised in Table 6. Most of these studies consistently identified specific alterations of neutrophil function in sepsis patients [26, 28, 30]. Further, an interesting review by Castilho et al. [41] using heart rate variability (HRV) analysis showed that SDNN (Standard deviation of the Normal to Normal interval) of HRV can also be an important indicator for the prediction of sepsis onset.

There are a few important limitations to be considered in our proposed study. First, the predictive analytics of the proposed work is limited only to publicly available EHR data from 2 cohorts, a retrospective setting. So the stability and generalization of the proposed model need to be evaluated systematically in prospective settings. Second, we investigated the risk prediction performance using the given task-specific utility function that reward predictions between twelve hours before and three hours after the onset of sepsis (Sepsis-3 guidelines) with a optimum window definition of 6 hours before onset . Moreover, the performance particular to different T hours (where $T = 24$, 18, 12, or 6) needs to be evaluated with the corresponding change in the definition of optimum value and the window for the prediction lying between before and sepsis onset. Lastly, since the dataset used in this study is prepared by sourcing only ICU patients, so even the model which is trained and tested is for ICU patients and does not evaluate for non-ICU patients.

As an alternative to gradient-boosting machine learning schemes, deep learning can be employed to learn such multiplicative interactions derived from proposed transformation schemes to yield improved results . However, this may be achieved at the penalty of high computational cost. Furthermore, literature published related to challenge submissions [54, 71, 64, 72] reported that boosting schemes have emerged as better candidates for handling class-imbalance problems as considered in this study when compared to the neural networks. Therefore, we refrained from a baseline study with deep learning schemes in the current study, and resort to LightGBM based GBDT framework for solving the class-imbalance problem.

In this study, the enhanced results of the proposed method as compared to baselines validates that, factorizing the third moment provably recovers the parameters of the inherent latent temporal associations for the sepsis onset, in a way that only capturing pair-wise interactions, i.e. the second-order moment alone uniquely cannot [73]. Tensor factorization turns out to be a highly effective and applicable tool for learning such embeddings, with increased potential. Leveraging higher-order data assists in encoding new types of latent

Table 6: Summary of the clinical studies for the prediction of sepsis.

Reference	Methodology	Database/participants	Results
Bergmann et al. [27]	Statistical analysis of Animal models	Mice (Jackson Laboratories,Bar Harbor) in order to assess IL-10 production.	Acute(24h) production of IL-10 by neutrophils is associated with sepsis.
Seree-aphinan et al. [29]	Longitudinal observational study. The surface level of C-X-C motif chemokine receptor 2 (CXCR2) and others are measured from peripheral blood neutrophils using flow cytometry.	Tertiary university hospital Thailand. 30 Patients diagnosed with acute infection and presented within 48 h of their symptom onset were recruited.	Reduction in CXCR2 surface level is associated with sepsis.
Fleuren et al. [46]	Systematic Review and Meta-analysis of 8 ICU cohorts	Multiple hospital ICU databases.	AUROC Range: 0.6800–0.9900
Castilho et al. [41]	HRV analysis	Observational studies with multiple databases. Total of 536 patients	SDNN :Vital HRV parameter for sepsis outcome
Thakur et al. [42]	Binary Logistic Regression Models	MIMIC-III database. 1580 neonates	Sensitivity 0.3125 Specificity 0.9730
Svenson et al. [48]	Fully Channelled LSTM (FC-L) with 10 channels	MIMIC-III database. 5,784 critical care admissions	AUROC 0.8920 AUPRC 0.4780 (48 hours before onset)
Lauritsen et al. [47]	LSTM-CNN model	EHRs from multiple Danish hospitals Vital signs: 3129 Full dataset: 52 229 patients	AUROC 0.7560 (24 hours before onset) AUROC 0.8560 (3 hours before onset)
Rafiei et al.[70]	LSTM-CNN model for different sepsis onset window durations	PhysioNet Challenge 2019 dataset	Accuracy 0.8100 AUROC 0.9200 Sensitivity 0.8500 Specificity 0.8100 PPV 0.6200 (4 hours before onset)
Moor et al. [49, 50]	MGP-TCN DTW-KNN	MIMIC-III database Sepsis patients: 570 Normal : 5618 subjects	AUPRC 0.400 AUROC 0.8600 (7 hours before onset)
Morrill et al. [52]	A signature-based transform model	PhysioNet Challenge 2019 dataset	Utility Score 0.4340
Zabihi et al. [54]	An Ensemble of XGBoost models	PhysioNet Challenge 2019 dataset	Utility Score 0.4281 AUROC 0.8333
Yang et al. [55]	Fusion-based XGBoost learning	PhysioNet Challenge 2019 dataset	Utility Score 0.4300 AUROC 0.8400
Chang et al.[66]	Temporal Convolutional Networks	PhysioNet Challenge 2019 dataset	Utility Score 0.4170
Lee at al.[67]	Graph Convolutional Networks	PhysioNet Challenge 2019 dataset	Utility Score 0.3820 AUROC 0.8170
Du at al. [68]	Gradient Boosted Decision Trees	PhysioNet Challenge 2019 dataset	Utility Score 0.4090 AUROC 0.8630
He et al. [69]	Ensemble learning of deep and hand-crafted features	PhysioNet Challenge 2019 dataset	Utility Score 0.4010
Lyra et al. [64]	Using Random forest classification	PhysioNet Challenge 2019 dataset	Utility Score 0.3760 AUROC 0.8100
Li et al. [53]	A Time-phased model	PhysioNet Challenge 2019 dataset	Utility Score 0.4300
Kok et al. [65]	Temporal Convolution Network	PhysioNet Challenge 2019 dataset	Utility Score 0.4300 AUPRC 0.68 AUROC 0.9100 Accuracy 0.955
Nesaragi & Patidar [56]	Ratio and Power-based Rusboost model	PhysioNet Challenge 2019 dataset	Utility Score 0.4013 AUROC 0.8432
Proposed Work	Correlation-based Tensor factorization	PhysioNet Challenge 2019 dataset	Utility Score 0.4314 AUROC 0.8613

information and semantic relationships compared to models that are trained using only pairwise data [74].

5. Conclusion

This work presents an innovative method that applies tensor-based multi-modal signal processing and machine learning technique for the early prediction of sepsis. This work is first of its kind to apply correlation-based tensor analysis on tabular structured EHR data for detection of sepsis onset. This research demonstrates that tensor factorization is an appropriate and reliable modeling technique to encapsulate the inherent quiescent high-dimensional interrelations among given clinical covariates of the patients for the onset prediction. The tensor-based machine learning predictive models thus designed can predict the sepsis onset at the individual level with enhanced utility scores as compared to benchmark machine learning techniques.

Our future research work in the pipeline involves the employment of graph signal processing techniques and kernel-based entropy methods and testing by acquiring data from more cohorts with different optimum onset definitions.

6. Conflict of interest statement

Authors have no competing interests to declare

References

[1] E. E. Papalexakis, C. Faloutsos, N. D. Sidiropoulos, Tensors for data mining and data fusion: Models, applications, and scalable algorithms, ACM Transactions on Intelligent Systems and Technology (TIST) 8 (2016) 1–44.

[2] Y. Luo, Y. Xin, E. Hochberg, R. Joshi, O. Uzuner, P. Szolovits, Subgraph augmented non-negative tensor factorization (santf) for modeling clinical narrative text, Journal of the American Medical Informatics Association 22 (2015) 1009–1019.

[3] S. Rabanser, O. Shchur, S. Günnemann, Introduction to tensor decompositions and their applications in machine learning, arXiv preprint arXiv:1711.10781 (2017).

[4] E. C. Chi, T. G. Kolda, On tensors, sparsity, and nonnegative factorizations, SIAM Journal on Matrix Analysis and Applications 33 (2012) 1272–1299.

[5] T. G. Kolda, B. W. Bader, Tensor decompositions and applications, SIAM review 51 (2009) 455–500.

[6] D. M. Dunlavy, T. G. Kolda, E. Acar, Temporal link prediction using matrix and tensor factorizations, ACM Transactions on Knowledge Discovery from Data (TKDD) 5 (2011) 1–27.

[7] Y. Kim, R. El-Kareh, J. Sun, H. Yu, X. Jiang, Discriminative and distinct phenotyping by constrained tensor factorization, Scientific reports 7 (2017) 1–12.

[8] J. C. Ho, J. Ghosh, J. Sun, Marble: high-throughput phenotyping from electronic health records via sparse nonnegative tensor factorization, in: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 115–124.

[9] H. Wang, Q. Zhang, J. Yuan, Semantically enhanced medical information retrieval system: a tensor factorization based approach, IEEE Access 5 (2017) 7584–7593.

[10] Y. Luo, F. Wang, P. Szolovits, Tensor factorization toward precision medicine, Briefings in bioinformatics 18 (2017) 511–514.

[11] Y. Yang, D. B. Dunson, Bayesian conditional tensor factorizations for high-dimensional classification, Journal of the American Statistical Association 111 (2016) 656–669.

[12] H. Wang, Q. Zhang, H.-Y. So, A. Kwok, Z. S.-Y. Wong, Temporal prediction of in-hospital falls using tensor factorisation, BMJ innovations 4 (2018).
[13] D. Goldfarb, Z. Qin, Robust low-rank tensor recovery: Models and algorithms, SIAM Journal on Matrix Analysis and Applications 35 (2014) 225–253.
[14] D. Yach, C. Hawkes, C. L. Gould, K. J. Hofman, The global burden of chronic diseases: overcoming impediments to prevention and control, Jama 291 (2004) 2616–2622.
[15] C. N. Vasconcelos, B. N. Vasconcelos, Experiments using deep learning for dermoscopy image analysis, Pattern Recognition Letters (2017).
[16] R. Kohli, S. S.-L. Tan, Electronic health records: how can is researchers contribute to transforming healthcare?, Mis Quarterly 40 (2016) 553–573.
[17] G. Shmueli, O. R. Koppius, Predictive analytics in information systems research, MIS quarterly (2011) 553–572.
[18] K. He, S. Huang, X. Qian, Early detection and risk assessment for chronic disease with irregular longitudinal data analysis, Journal of biomedical informatics 96 (2019) 103231.
[19] E. A. Ashley, The precision medicine initiative: a new national effort, Jama 313 (2015) 2119–2120.
[20] F. S. Collins, H. Varmus, A new initiative on precision medicine, New England journal of medicine 372 (2015) 793–795.
[21] R. Mirnezami, J. Nicholson, A. Darzi, Preparing for precision medicine, N Engl J Med 366 (2012) 489–491.
[22] J. Finkelstein, I. cheol Jeong, Machine learning approaches to personalize early prediction of asthma exacerbations, Annals of the New York Academy of Sciences 1387 (2017) 153.
[23] R. K. Leung, Y. Wang, R. C. Ma, A. O. Luk, V. Lam, M. Ng, W. Y. So, S. K. Tsui, J. C. Chan, Using a multi-staged strategy based on machine learning and mathematical modeling to predict genotype-phenotype risk patterns in diabetic kidney disease: a prospective case–control cohort analysis, BMC nephrology 14 (2013) 1–9.
[24] L. Giesen, M. Singer, What Is Sepsis?, in: W. J. Wiersinga, C. W. Seymour (Eds.), Handbook of Sepsis, Springer International Publishing, Cham, 2018, pp. 3–14.
[25] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. van der Poll, J.-L. Vincent, D. C. Angus, The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3), JAMA 315 (2016) 801.
[26] P. M. Potey, A. G. Rossi, C. D. Lucas, D. A. Dorward, Neutrophils in the initiation and resolution of acute pulmonary inflammation: understanding biological function and therapeutic potential, The Journal of pathology 247 (2019) 672–685.
[27] C. B. Bergmann, C. E. Salyer, N. Beckmann, C. C. Caldwell, Intraperitoneal neutrophil il-10 production is promoted by interferon γ in a murine model of sepsis model in the acute phase of sepsis, Biochemical and Biophysical Research Communications 530 (2020) 278–284.
[28] S. Y. Ahn, Y.-S. Maeng, Y. R. Kim, Y. H. Choe, H. S. Hwang, Y.-M. Hyun, In vivo monitoring of dynamic interaction between neutrophil and human umbilical cord blood-derived mesenchymal stem cell in mouse liver during sepsis, Stem cell research & therapy 11 (2020) 44.
[29] C. Seree-Aphinan, P. Vichitkunakorn, R. Navakanitworakul, B. Khwannimit, Distinguishing sepsis from infection by neutrophil dysfunction: A promising role of cxcr2 surface level, Frontiers in Immunology 11 (2020) 3339.
[30] X.-F. Shen, K. Cao, J.-p. Jiang, W.-X. Guan, J.-F. Du, Neutrophil dysregulation during sepsis: an overview and update, Journal of cellular and molecular medicine 21 (2017) 1687–1697.
[31] K. Brown, S. Brain, J. Pearson, J. Edgeworth, S. Lewis, D. Treacher, Neutrophils in development of multiple organ failure in sepsis, The Lancet 368 (2006) 157–169.
[32] K. M. Groeneveld, L. Koenderman, B. L. Warren, S. Jol, L. P. Leenen, F. Hietbrink, Early decreased neutrophil responsiveness is related to late onset sepsis in multitrauma patients: an international cohort study, PLoS One 12 (2017) e0180145.
[33] R. P. Dellinger, M. M. Levy, A. Rhodes, D. Annane, H. Gerlach, S. M. Opal, J. E. Sevransky, C. L.

Sprung, I. S. Douglas, R. Jaeschke, et al., Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock, 2012, Intensive care medicine 39 (2013) 165–228.
[34] C. W. Seymour, J. N. Kennedy, S. Wang, C.-C. H. Chang, C. F. Elliott, Z. Xu, S. Berry, G. Clermont, G. Cooper, H. Gomez, D. T. Huang, J. A. Kellum, Q. Mi, S. M. Opal, V. Talisa, T. van der Poll, S. Visweswaran, Y. Vodovotz, J. C. Weiss, D. M. Yealy, S. Yende, D. C. Angus, Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis, JAMA 321 (2019) 2003.
[35] J.-L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C. Reinhart, P. Suter, L. G. Thijs, The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure, 1996.
[36] K. E. Henry, D. N. Hager, P. J. Pronovost, S. Saria, A targeted real-time early warning score (trewscore) for septic shock, Science translational medicine 7 (2015) 299ra122–299ra122.
[37] S. Horng, D. A. Sontag, Y. Halpern, Y. Jernite, N. I. Shapiro, L. A. Nathanson, Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning, PLOS ONE 12 (2017) e0174708.
[38] T. Desautels, J. Calvert, J. Hoffman, M. Jay, Y. Kerem, L. Shieh, D. Shimabukuro, U. Chettipally, M. D. Feldman, C. Barton, D. J. Wales, R. Das, Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach, JMIR Medical Informatics 4 (2016) e28.
[39] K. Kawamoto, C. A. Houlihan, E. A. Balas, D. F. Lobach, Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success, BMJ 330 (2005) 765.
[40] S. M. Oh, K. M. Stefani, H. C. Kim, Development and application of chronic disease risk prediction models, Yonsei medical journal 55 (2014) 853.
[41] F. M. de Castilho, A. L. P. Ribeiro, V. Nobre, G. Barros, M. R. de Sousa, Heart rate variability as predictor of mortality in sepsis: A systematic review, PloS one 13 (2018) e0203487.
[42] J. Thakur, S. K. Pahuja, R. Pahuja, Performance comparison of systemic inflammatory response syndrome with logistic regression models to predict sepsis in neonates, Children 4 (2017) 111.
[43] G. Wardi, M. Carlile, A. Holder, S. Shashikumar, S. R. Hayden, S. Nemati, Predicting progression to septic shock in the emergency department using an externally generalizable machine-learning algorithm, Annals of emergency medicine 77 (2021) 395–406.
[44] G. Wardi, S. Shashikumar, M. Carlile, M. Krak, S. Hayden, A. Holder, S. Nemati, 301 use of transfer learning to improve external validity of a machine-learning algorithm to predict septic shock in the emergency department, Annals of Emergency Medicine 76 (2020) S116.
[45] S. Nemati, A. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, T. G. Buchman, An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU:, Critical Care Medicine 46 (2018) 547–553.
[46] L. M. Fleuren, T. L. Klausch, C. L. Zwager, L. J. Schoonmade, T. Guo, L. F. Roggeveen, E. L. Swart, A. R. Girbes, P. Thoral, A. Ercole, et al., Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy, Intensive care medicine 46 (2020) 383–400.
[47] S. M. Lauritsen, M. E. Kalør, E. L. Kongsgaard, K. M. Lauritsen, M. J. Jørgensen, J. Lange, B. Thiesson, Early detection of sepsis utilizing deep learning on electronic health record event sequences, Artificial Intelligence in Medicine 104 (2020) 101820.
[48] P. Svenson, G. Haralabopoulos, M. T. Torres, Sepsis deterioration prediction using channelled long short-term memory networks, in: International Conference on Artificial Intelligence in Medicine, Springer, pp. 359–370.
[49] M. Moor, M. Horn, B. Rieck, D. Roqueiro, K. Borgwardt, Early recognition of sepsis with gaussian process temporal convolutional networks and dynamic time warping, in: Machine Learning for Healthcare Conference, PMLR, pp. 2–26.
[50] M. Moor, M. Horn, B. Rieck, D. Roqueiro, K. Borgwardt, Temporal convolutional networks and dynamic time warping can drastically improve the early prediction of sepsis, arXiv preprint arXiv:1902.01659 (2019).

[51] M. A. Reyna, C. S. Josef, R. Jeter, S. P. Shashikumar, M. B. Westover, S. Nemati, G. D. Clifford, A. Sharma, Early Prediction of Sepsis From Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019, Critical care medicine 48 (2020) 210–217. Publisher: Lippincott Williams & Wilkins.
[52] J. H. Morrill, A. Kormilitzin, A. J. Nevado-Holgado, S. Swaminathan, S. D. Howison, T. J. Lyons, Utilization of the signature method to identify the early onset of sepsis from multivariate physiological time series in critical care monitoring, Critical Care Medicine 48 (2020) e976–e981.
[53] X. Li, X. Xu, F. Xie, X. Xu, Y. Sun, X. Liu, X. Jia, Y. Kang, L. Xie, F. Wang, et al., A time-phased machine learning model for real-time prediction of sepsis in critical care, Critical Care Medicine 48 (2020) e884–e888.
[54] M. Zabihi, S. Kiranyaz, M. Gabbouj, Sepsis prediction in intensive care unit using ensemble of xgboost models, in: 2019 Computing in Cardiology (CinC), pp. Page 1–Page 4.
[55] M. Yang, C. Liu, X. Wang, Y. Li, H. Gao, X. Liu, J. Li, An explainable artificial intelligence predictor for early detection of sepsis, Critical Care Medicine 48 (2020) e1091–e1096.
[56] N. Nesaragi, S. Patidar, Early prediction of sepsis from clinical data using ratio and power-based features, Critical Care Medicine 48 (2020) e1343–e1349.
[57] B. A. Goldstein, A. M. Navar, M. J. Pencina, J. Ioannidis, Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review, Journal of the American Medical Informatics Association 24 (2017) 198–208.
[58] E. Brochu, V. M. Cora, N. de Freitas, A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning, arXiv:1012.2599 [cs] (2010). ArXiv: 1012.2599.
[59] J. Snoek, H. Larochelle, R. P. Adams, Practical bayesian optimization of machine learning algorithms, Advances in neural information processing systems 25 (2012) 2951–2959.
[60] P. Biglarbeigi, D. McLaughlin, K. Rjoob, A. Abdullah, N. McCallan, A. Jasinska-Piadlo, R. Bond, D. Finlay, K. Ng, A. Kennedy, J. McLaughlin, Early prediction of sepsis considering early warning scoring systems, in: 2019 Computing in Cardiology (CinC), pp. Page 1–Page 4.
[61] C. J. Hillar, L.-H. Lim, Most tensor problems are np-hard, Journal of the ACM (JACM) 60 (2013) 1–39.
[62] J. Kossaifi, Y. Panagakis, A. Anandkumar, M. Pantic, Tensorly: Tensor learning in python, The Journal of Machine Learning Research 20 (2019) 925–930.
[63] R. J. Delahanty, J. Alvarez, L. M. Flynn, R. L. Sherwin, S. S. Jones, Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis, Annals of emergency medicine 73 (2019) 334–344.
[64] S. Lyra, S. Leonhardt, C. H. Antink, Early prediction of sepsis using random forest classification for imbalanced clinical data, in: 2019 Computing in Cardiology (CinC), pp. 1–4.
[65] C. Kok, V. Jahmunah, S. L. Oh, X. Zhou, R. Gururajan, X. Tao, K. H. Cheong, R. Gururajan, F. Molinari, U. R. Acharya, Automated prediction of sepsis using temporal convolutional network, Computers in Biology and Medicine 127 (2020) 103957.
[66] Y. Chang, J. Rubin, G. Boverman, S. Vij, A. Rahman, A. Natarajan, S. 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), IEEE, pp. Page–1.
[67] B. T. Lee, O.-Y. Kwon, H. Park, K.-J. Cho, J.-M. Kwon, Y. Lee, et al., Graph convolutional networks-based noisy data imputation in electronic health record, Critical Care Medicine 48 (2020) e1106–e1111.
[68] J. A. Du, N. Sadr, P. d. Chazal, Automated prediction of sepsis onset using gradient boosted decision trees, in: 2019 Computing in Cardiology (CinC), pp. Page 1–Page 4.
[69] Z. He, L. Du, P. Zhang, R. Zhao, X. Chen, Z. Fang, Early sepsis prediction using ensemble learning with deep features and artificial features extracted from clinical electronic health records, Critical Care Medicine 48 (2020) e1337–e1342.
[70] A. Rafiei, A. Rezaee, F. Hajati, S. Gheisari, M. Golzan, Ssp: Early prediction of sepsis using fully connected lstm-cnn model, Computers in biology and medicine 128 (2021) 104110.
[71] Y. Wang, B. Xiao, X. Bi, W. Li, J. Zhang, X. Ma, Prediction of sepsis from clinical data using long

short-term memory and extreme gradient boosting, in: 2019 Computing in Cardiology (CinC), IEEE, pp. Page–1.
[72] B. Sweely, A. Park, L. Winter, L. Liu, X. Zhao, Time-padded random forest ensemble to capture changes in physiology leading to sepsis development, in: 2019 Computing in Cardiology (CinC), IEEE, pp. 1–4.
[73] E. Bailey, S. Aeron, Word embeddings via tensor factorization, arXiv preprint arXiv:1704.02686 (2017).
[74] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, M. Telgarsky, Tensor decompositions for learning latent variable models (a survey for alt), in: International Conference on Algorithmic Learning Theory, Springer, pp. 19–38.