Network-Based Modeling of Sepsis: Quantification and Evaluation of Simultaneity of Organ Dysfunctions

Ali Jazayeri
Department of Information Science, College of Computing and Informatics, Drexel University
aj629@drexel.edu

Muge Capan
Decision Sciences & MIS Department, LeBow College of Business, Drexel University
mc3922@drexel.edu

Christopher Yang
Department of Information Science, College of Computing and Informatics, Drexel University
chris.yang@drexel.edu

Farzaneh Khoshnevisan
Department of Computer Science, North Carolina State University
fkhoshn@ncsu.edu

Min Chi
Department of Computer Science, North Carolina State University
mchi@ncsu.edu

Ryan Arnold
College of Medicine, Hahnemann University Hospital, Drexel University
rca57@drexel.edu

ABSTRACT

It is shown that appropriate therapeutic management at early stages of sepsis are crucial for preventing further deterioration and irreversible organ damage. Although previous studies considered the cellular and physiological responses as the components of sepsis-related predictive models, temporal connections among the responses have not been widely studied. The objective of this study is to investigate simultaneous changes in cellular and physiological responses represented by 16 clinical variables contributing to seven organ system dysfunctions in patients with sepsis to predict in-hospital mortality. Organ dysfunctions were represented by undirected weighted network models composed of: i) nodes (i.e., 16 clinical variables and three biomarkers including procalcitonin, C-reactive protein, and sedimentation rate), ii) edges (i.e., connection between pair of nodes representing simultaneous dysfunctions), and iii) weights representing the persistence of the co-occurrence of two dysfunctions. Data was collected from 13,367 adult patients (corresponding to 17,953 visits) admitted to the study hospital from July 1, 2013, to December 31, 2015. The study population were categorized based on clinical criteria representing sepsis progression to identify different subpopulations. The findings quantify the optimal window for defining the simultaneity of two dysfunctions, the network properties corresponding to different subpopulations, the discriminatory patterns of simultaneous dysfunctions among subpopulations and in-hospital mortality prediction. The results show that the level of persistence of simultaneous dysfunctions are subpopulation-specific. Insights from this study regarding optimal thresholds of the persistence and combination of simultaneous organ dysfunctions can inform policies to personalize the in-hospital mortality prediction.

ACM-BCB '19, September 7–10, 2019, Niagara Falls, NY, USA

ACM ISBN 978-1-4503-6666-3/19/09…$15.00
https://doi.org/10.1145/3307339.3342160

CCS CONCEPTS

• Applied computing~Health informatics

KEYWORDS

Sepsis prediction; Simultaneous dysfunctions; Network modeling

ACM Reference format:

Ali Jazayeri, Muge Capan, Christopher Yang, Farzaneh Khoshnevisan, Min Chi, Ryan Arnold. 2019. Network-Based Modeling of Sepsis: Quantification and Evaluation of Simultaneity of Organ Dysfunctions. In ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB '19), September 7--10, 2019, Niagara Falls, NY, USA. https://doi.org/10.1145/3307339.3342160.

1. INTRODUCTION

Sepsis is a commonly occurring, yet under-diagnosed disease condition defined as a "life-threatening organ dysfunction caused by a dysregulated host response to infection" [1]. In a retrospective study including more than 2.9 million patients, 21% of adult inpatients with clinical manifestations of sepsis (6% of the study population) died in the hospital or were discharged to hospice [2]. Studies have shown that accurate diagnosis and appropriate therapeutic management at early stages on the sepsis spectrum are crucial for preventing further deterioration and irreversible organ damage [3, 4]. However, early sepsis symptoms are commonly overlooked by providers since the observed clinical manifestations may be vague or may not be specific to sepsis. For example, difficulty breathing or a subtle change in mental status could occur due to a variety of causes. Thus, uncertainty in early stages of sepsis results in increasing need for identifying reliable

clinical indicators to support the recognition of, and response to, sepsis.

Biomarkers (i.e., objectively measured indicators of biological, pathogenic, or pharmacologic processes [5]) present a category of clinical indicators with a potential to inform recognition of sepsis in its early and critical stages when the sepsis-induced deterioration is still reversible. In the last decades, biomarkers have been shown to enhance characterization of the biological phenotype underlying sepsis [6] and therefore promise a new approach to inform timely recognition and management of sepsis. Biomarkers are widely studied for various infection types with the goal to evaluate different stages of sepsis as well as a short- or long-term predictor of patients' outcome. For example, biomarkers are used to identify the severity of sepsis, predict type of infection, inform the anti-infective therapy and provide a measure for predicting organ dysfunction [7]. Studies focusing on biomarkers aim to evaluate the predictive performance or diagnostic and prognostic value of one, or more biomarkers in comparison with other biomarkers or severity of illness scoring systems. Some of the commonly studied blood-based biomarkers associated with sepsis diagnosis and treatment are lactate [8-12], procalcitonin (PCT) [12-14], and C-reactive protein (CRP) [11, 14-16]. However, the proposed biomarkers are not limited to samples from blood. For example, Garcia-Simon et. al. [17] examines the short-term predictive power of 8 metabolites (ethanol, glutamine, methionine, arginine, phenylalanine, glucose, hippurate and an unknown metabolite located at 1.40–1.45 parts per million) form urine samples.

Using different statistical tests, studies focus on the correlation of different biomarkers with different illness severity scores [11, 15, 16, 18]. Furthermore, these studies examine if their proposed biomarkers can be used as an independent or complementary measure with other biomarkers and scores as features for prediction of severity of sepsis or mortality rate using different evaluation metrics such as receiver operating characteristic (ROC) curve and the area under the curve (AUC) [19-21]. However, only a few studies concentrate on the temporal interrelations of biomarkers or their relations with other lab results or vital signs. For example, Jiang, Feng, Gao and Zhang [14] show that procalcitonin at 24 hours, CRP at 48 hours, and copeptin at 72 hours after admission to the intensive care unit (ICU) have been independent risk factors for in-hospital mortality, which implies the differential importance of biomarkers at different time points or different stages of sepsis. The temporal interrelationships among different organ systems and the consequences of their failures on other organ systems can inform the diagnostic and therapeutic decision making. Patterns of inter-organ failures can also inform patients' outcome prediction and consequently implementation of sepsis management procedures.

Various stochastic modeling approaches have been applied to capture the interrelationships among different cellular and physiological responses contributing to sepsis and its progression. For example, Shi, Wu and Ben-Arieh [22] provide a system dynamics mathematical model to represent the dynamics and interactions among variables causing the development of sepsis during the acute inflammatory response (AIR). Due to the practical limitations of traditional analytical models in the analysis of systems composed of large number of variables [23], Shi, Wu and Ben-Arieh [22] utilize agent-based modeling (ABM) framework to model and simulate the AIR progression at the interface of blood vessels and cells. The ABM is a method used for modeling of complex systems in which not only the components of the system (agents) but the relationships and interactions among them should be taken into consideration. These systems cannot be easily modeled and studied using traditional modeling and analytical techniques [24-26]. Using the temporal and dynamic Bayesian networks is another approach which has been used to model the interrelationships among variables contributing to sepsis progression. Bayesian networks are probabilistic graphical representations of a phenomenon of interest in which the nodes of the network are the variables of the system and edges among nodes represents the conditional dependencies among the variables. Orphanou, Stassopoulou and Keravnou [27] provide a survey of the applications of this technique in clinical contexts. Peelen, de Keizer, Jonge, Bosman, Abu-Hanna and Peek [28] use the dynamic Bayesian networks for the investigation of organ failures in patients admitted to the ICU. They develop three models, including temporal changes in severity of organ failure, failure occurrences in specific organs, temporal persistence of an organ failure and temporal development of new organ failures. They also use logistic regression analysis to estimate the transition probabilities between each pair of nodes. Nachimuthu and Haug [29] adopt dynamic Bayesian networks to detect sepsis after arrival to the Emergency Department (ED) using different time-windows of the patients' data. The outcome of their model is the detection of sepsis as a binary variable using the clinical symptoms. Their model shows higher specificity compared to sensitivity, however, they show that as more patient data becomes available, the sensitivity increases. In [30], a decision support system is proposed for prediction of the level of lactate and mortality in patients with sepsis using five variables as the measurements of the condition of the patient and two variables as the outcome of interest (occurrence of sepsis and mortality). To summarize the patients' temporal condition measurement data and decrease the effect of variants of measurements over time, the mean and median values of temperature, respiratory rate (RR), white blood cell count (WBC), mean arterial pressure (MAP), and lactate levels for each patient are used. For each summarized prediction model, a Bayesian network is developed and the best possible combination of edges in the Bayesian network are identified.

Considering that sepsis is a multi-stage disease and patients experience various severities, few of the previously reported studies explicitly considered multiple sepsis stages in network models. Furthermore, the studies discussed above consider the network as an implicit construct to create the model for the prediction of patient outcome. Therefore, properties of the networks and variations in these properties among different subpopulations have not been widely studied. Our study approaches modeling sepsis as a multi-stage stochastic process where the manifestation of the stages (represented by organ failures) are subpopulation-specific. More specifically, we consider four subpopulations; 1) patients with infection, 2) patients with infection who develop sepsis without experiencing septic shock, 3) patients with infection who develop septic shock and survive, and 4) patients with infection who develop septic shock and die. The objectives of this study are to: i) develop network models to analyze the simultaneous dysfunctions

patterns of cellular and physiological responses; ii) compare the patterns and network properties among patient subpopulations; and iii) identify simultaneous failures in different organ systems and categorize these patterns in relation to other clinical observations (e.g., vital signs and lab values) associated with sepsis to predict in-hospital mortality.

This paper is organized as follows. In the next section, the study population is defined, subpopulations definition criteria are explained, then the modeling approach and adopted statistical approach are discussed. In the results section, the findings of the analysis are presented including the properties of the networks and their differences among subpopulations. The significant and discriminatory patterns among subpopulations are shown. In the discussion section, the clinical relevance of the results is discussed in more detail, and finally, the paper concludes with key insights and directions for future work.

2. MATERIAL AND METHODS

2.1 Study population

This study includes 13,367 adult patients admitted from July 1, 2013, to December 31, 2015, corresponding to 17,953 visits. The data is sourced from a single tertiary care health care system made up of 2 hospitals with 1,100 total in-hospital beds. The dataset includes retrospectively collected Electronic Health Records (EHR) data with following inclusion criteria: age ≥18 at arrival, arrival to the hospital between July 2013 and December 2015, and with visit types of inpatient, ED only (outpatient), or observational visits. The study was approved by health systems Institutional Review Board.

2.2 Subpopulations Definition

To investigate the changes in network properties in different subpopulations, we considered four subpopulation-specific networks for patients with infection (i.e., infected population), patients with infection who develop sepsis without experiencing septic shock during hospitalization and survive (i.e., sepsis population), patients with infection who develop septic shock and survive (i.e., septic shock survivors), and patients with infection who develop septic shock and die. Death was defined as any-inhospital death or discharge to hospice care. For patients with multiple visits, if the patient died during the study only the last visit is considered in this subpopulation, and previous visits are categorized in one of the three former subpopulations. The infected population was defined as patients with anti-infective administration for whom the anti-infective administration lasts for at least four days or a positive viral PCR (polymerase chain reaction) test for influenza [2]. The sepsis without septic shock population was defined as a subset of infected population. If the patient is discharged to hospice or died while receiving an anti-infective but before 4 days of anti-infective administration elapsed, this patient would be considered a sepsis non-survivor [2]. To be considered as sepsis, the condition for infected subpopulation should hold in addition to the occurrence of organ dysfunction in the period of 24 hours before first anti-infective administration until the last administration [2]. Organ dysfunction is defined by meeting at least one of the criteria presented in Table 1 or at least one vasopressor administration during hospitalization, where these criteria were developed by synthesizing established Sepsis-3 guidelines [31] and subject matter experts' input. The sepsis with septic shock population is defined as infected patients meeting either at least two cardiovascular organ dysfunction criteria listed in Table 1 with at least 30 minutes between the two abnormal observations (with no normal observation in between) and both observations within maximum 24 hours of each other, or for sepsis patients with any vasopressor administration based on clinical expert input. The number of visits and patients included in each subpopulation is shown in Table 2.

Table 1. Cellular and physiological responses and biomarkers used to create simultaneous network models, and the criteria resulting in the corresponding organ dysfunction.

Organ Dysfunction	Response	Abbreviation	Dysfunction criteria
Cardiovascular	Systolic blood pressure (SBP)	Sb	< 90 mmHg
	SBP_max* - SBP	Sd	> 40 mmHg within an 8-hour period
	Mean arterial pressure (MAP)	Mp	< 65 mmHg
Renal	Creatinine	Cr	> 1.2 mg/dL
	(Creatinine - C_base**)/(C_base)	Cd	> 50% from initial creatinine
	Blood Urea Nitrogen (BUN)	Bu	> 20 mg/dL
Hematopoietic	White Blood Cell Count (WBC)	Wb	< 4,000 cells/mL
	Platelet	Pl	< 100,000 cells/mL
Metabolic	Lactate	La	> 2.0 mmol/L
Gastrointestinal	Bilirubin	Bi	> 2 mg/dL
Respiratory	Fraction of inspired oxygen (FiO_2)	Fi	> 21%
	Pulse oximetry (SpO_2)	Px	< 90%
	SpO_2/FiO_2	Or	< 421
	Oxygen (O_2) Source	Os	Mechanical ventilation required (bilevel positive airway pressure (BiPAP) or continuous positive airway pressure (CPAP) or ventilator)
Central Nervous	Glasgow Comma Score	Gc	< 14
	Glasgow Best Verbal Response	Gv	< 5
Three biomarkers considered	Procalcitonin	pc	> 0.15 ng/mL
	C-Reactive Protein	cr	> 8 mg/L
	Erythrocyte Sedimentation Rate	sr	> 20 mm/hr

*: Maximum systolic blood pressure for each observation within 8-hour windows.
**: Initial creatinine value observed in each visit.

Table 2. Number of visits and patients in each sub-population for creating of network models.

Row No.	Subpopulations	Visits Count	Patients Count
1	Infected	1,574	1,487
2	Sepsis survivor without septic shock	9409	7,536
3	Septic shock survivor	3980	3,432
4	Septic shock non-survivor	1,335	1,335
	Total	**16,298**	**12,172***

*: The number of patients and visits in each row are different because some of the patients have been visited multiple times.

2.3 Network Modeling

To investigate simultaneous changes in cellular and physiological responses, 19 variables available in the dataset were considered. These variables and their failure criteria are shown in Table 1. Before the creation of the network for each subpopulation, several data pre-processing steps were performed, including carrying forward the vital and lab results, management of missing data, and new features creation. For the carry forward, it is assumed that the measured values for vital signs and lab results are valid for 8 and 24 hours, respectively, if there are no observed measurements. The mean arterial pressure (MAP) was calculated using the following formula:

$$MAP = \frac{2 \cdot DBP + SBP}{3} \quad (1)$$

in which, the DBP and SBP represent the diastolic and systolic blood pressure of the patient. Other derived variables are shown in Table 1. The simultaneous failure undirected weighted network models were created with the variables in Table 1 (responses in column three) as nodes, and simultaneous failure of two variables as edges. Failures were considered simultaneous if they co-occur within a 150-minute period. Selection of 150 minutes as the appropriate duration for aggregation will be discussed in the Results section. The failures are defined based on the criteria of Table 1. In these models, nodes refer to the 16 variables and three commonly used biomarkers shown in Table 1, edges refer to the simultaneous failure of each pair of variables, and weights refer to the persistence (as a measure of simultaneous failure of two variables with higher values representing that simultaneous failure of two variables is more common than each one independently) of the co-occurrence of simultaneous failures. For each observation (i.e., measured cellular or physiological response), all the observations occurred in less than 150 minutes before or after the observation were considered. Failures in this time window were considered as simultaneous failures. This was repeated for all the observations in each visit and all the visits in the data. Thus, for each observation there might be none, one or more than one simultaneous failure of variables of Table 1.

In the next step, for different pairs of variables shown in Table 1, the probability of simultaneous failure was calculated as the probability of simultaneous failure of the corresponding pair of variables divided by the probability of each variable's failure, as follows:

$$wt_{ij} = \frac{c_{ij}^{11} \cdot (c_{ij}^{11} + c_{ij}^{10} + c_{ij}^{01})}{(c_{ij}^{11} + c_{ij}^{10}) \cdot (c_{ij}^{11} + c_{ij}^{01})} \quad (2)$$

Here, wt_{ij} represents the weight of simultaneous failure of two variables (i and j represent responses listed in Table 1), and c_{ij}^{kl} $(k,l \in \{0,1\}, 1: failure, 0: no\ failure)$ shows the different combinations of simultaneous states that variables i and j can take (for example c_{SbCr}^{10} shows the count of observations in which the systolic blood pressure is not in the acceptable range, but level of creatinine is). Values of c_{ij}^{00} were not considered in derivation of Equation (2) since the goal was to capture the ratio of simultaneous failures when at least one of the variables fail. In other words, Equation 2 represents the ratio of simultaneous failure of two variables divided by failure of each variable and represent the probability that each pair of responses of Table 1 fails simultaneously. The minimum and maximum values of weights wt_{ij} are zero and one, expressing that two responses never or always fail simultaneously, respectively.

2.4 Statistical analysis

For statistical analysis purposes, we used Levene's test for testing homogeneity of variance, Shapiro-Wilk for normality test, and Kruskal-Wallis and Wilcoxon rank sum tests respectively for testing whether simultaneous failure values are originated from the same distribution for all four subpopulations and for pairwise comparisons.

2.5 Significant Simultaneous Failures

The comparison of networks of different subpopulations identifies the significant simultaneous failures unique to that subpopulation, i.e., simultaneous failures of cellular and physiological responses common in one subpopulation but not in the others. To confirm the significance of the edges detected as significant in network models, we used generalized boosted models from gbm library in R software. This library is an extension and implementation of AdaBoost algorithm [32] and gradient boosting machine [33, 34]. The package technical details can be found in [35]. We implemented the gbm method using the same number of visits in two subpopulations (2,990 visits in each subpopulation, as the number of non-survivors was 2,990 patients), sepsis survivors and sepsis non-survivors. The sepsis non-survivors group includes the fourth row of Table 2 (1,335 visits), and sepsis patients without septic shock who died or discharged to hospice (1,655 visits). For the gbm method, the data were randomly divided into training and testing data sets and 10-fold cross-validation was applied. To implement this method, we adopted two approaches. First, the gbm model was created based on all the possible simultaneous failures among variables of Table 1 as features. Then, we created other gbm models using just the simultaneous failures considered common at different thresholds. The objective of this analysis is to compare the performance of the same method under two conditions; using all the simultaneous failures in comparison with those identified by the network analysis.

3. RESULTS

3.1 Network development

Two failures were considered simultaneous if they co-occur within a 150-minute period. This 150-minute aggregation time was obtained after the following analysis. Ten aggregation windows were tested ranging from 30-minute to 300-minute with 30 minutes intervals. Each observation of the patient was aggregated with other observations of the same patient recorded in the period of t-minutes before or after the observation, where t represents the corresponding aggregation time. The failures of responses listed in Table 1 were collected for aggregated observations. Two networks of sepsis survivors and sepsis non-survivors were created. Next, using the gbm method, the prediction accuracy and area under operating characteristic curve (ROC) were collected. The highest accuracy and AUC combination was observed at 150-minute aggregation window (Figure 1).

Then, four subpopulation-specific networks were created and the weights of simultaneous failure of each pair of variables were calculated. The maximum possible number of edges in each network is 171 (each of 19 variables in Table 1 can connect to 18 other variables). Table 3 shows the mean and standard deviation of weights calculated for the four subpopulations. In situations where there was no observation of simultaneous failure of a pair of variables in the corresponding subpopulation, we assumed that the weight corresponding to the edge connecting the two variables is zero and included that in the calculation of mean and standard deviation, as the lack of observation of the simultaneous failure also might be informative.

Table 3. Mean and Standard deviation (SD) of simultaneous failure weights (wt) of edges calculated for four subpopulations.

Status	Number of edges (wt >0)	Mean	SD
Infected	82	0.121	0.223
Sepsis without septic shock	171	0.252	0.231
Septic shock survivors	171	0.361	0.253
Septic Shock non-survivors	168	0.461	0.286

3.2 Network comparison

Using the weights of simultaneous failures in the networks in different subpopulations, the following statistical analysis was performed. Levene's test rejected that the homogeneity of variances of weights between networks are equal, and the Shapiro-Wilk normality test rejected the hypothesis of normality of weight distribution in each network model (both with $p - value < 0.001$). Next, Kruskal-Wallis and Wilcoxon rank sum tests both rejected that the weights calculated for different subpopulations are originated from the same distribution (Table 4).

Figure 2 shows the variations in simultaneous failures in different subpopulations explicitly for the four subpopulations for several weigth thresholds, i.e. only keeping edges at or above that given threholds value. The variables in the network are color-coded based on the organ systems and biomarkers as it is shown in the legend at the top of the Figure 2. The edges with at least one node including a biomarker are shown in red. The weights of edges range between zero and one. The Figure 2 shows the networks of the four subpopulations for three thresholds in the range of [0,1]. In each row, only the edges with weights greater than or equal to the corresponding threshold are shown. By increasing the thresholds, the number of edges included in all the networks decreases. However, stronger relationships exist between the nodes in networks related to more severe conditions.

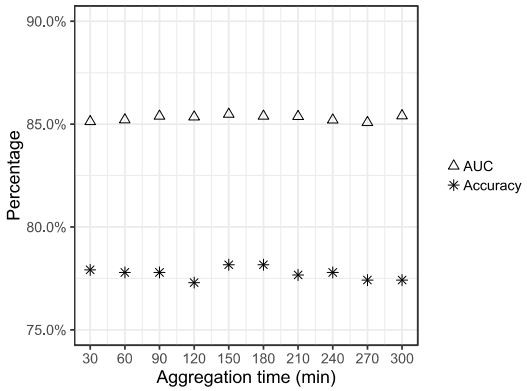

Figure 1: Prediction accuracy for different aggregation times. The aggregation time 150-minute shows the highest prediction accuracy

Table 4. Pairwise comparisons of weights of simultaneous failures in the networks in different subpopulations using Wilcoxon rank sum test.

	Infected	Sepsis without septic shock	Sepsis with septic shock
Sepsis without septic shock	< 0.001	-	-
Septic Shock Survivor	< 0.001	< 0.001	-
Septic Shock Non-survivor	< 0.001	< 0.001	< 0.001

These visualizations also can be translated to network density, defined as the ratio of edges included in the network divided by total possible edges that the network can have. Figure 3 shows the density of the network models for each of the four subpopulations at different weight thresholds. As illustrated, the density of the network decreases as the weight threshold increases for all four subpopulations. However, the reduction rate, the decrease of density by increasing the weights, is different among them. Having the lowest reduction rate in the sepsis non-survivors shows that simultaneous failure are significantly more common in this subpopulation in comparison with other subpopulations. The same explanation can be considered for the comparison of each pair of subpopulations in Figure 3.

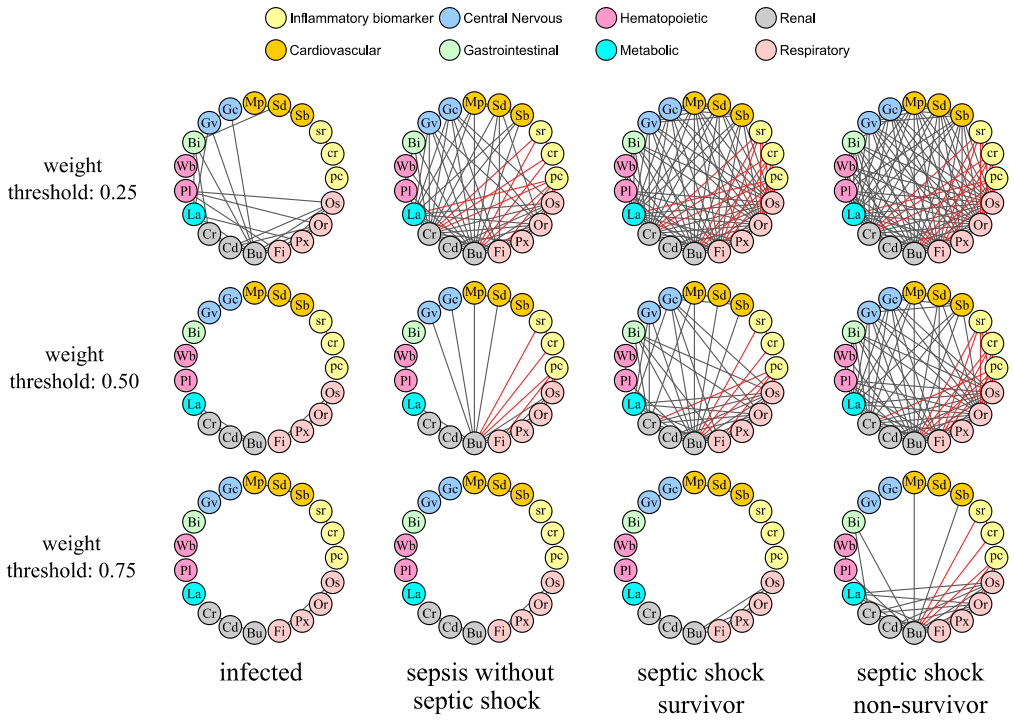

Figure 2: Networks representing simultaneous failure of different cellular and physiological responses in four subpopulations. Each row shows edges in each network that has a weight greater than or equal to the corresponding row threshold, the edges with one end as biomarkers are shown in red (Sb (systolic blood pressure); Sd (Sb (systolic blood pressure); Sd (Sb_max* - SBP); Mp (MAP); Cr (Creatinine); Cd ((Creatinine - C_base)/(C_base)); Wb (WBC); La (Lactate); Pl (Platelet); Bu (BUN); Bi (BiliRubin); Fi (FiO2); Px (SpO2); Or (SpO2/FiO2); Os (Oxygen Source); Gc (Glasgow Coma Score); Gv (Glasgow Best Verbal Response); pc (Procalcitonin); cr (CRP); sr (Sedimentation Rate). *: Maximum systolic blood pressure for each observation within 8-hour windows. **: Initial creatinine value observed in each visit)**

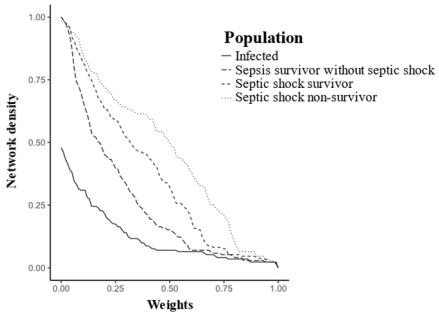

Figure 3: Networks density for the four subpopulations over weights

3.3 Simultaneous failure patterns in subpopulations

Further, the edges among cellular and physiological responses were used to predict the patients' outcome. Thus, to set up the prediction experiment, 70% of visits in each subpopulation are selected as training data. The outcome of interest is survival or death in hospital. Therefore, in the training process, two separate networks corresponding to these groups were created; sepsis survivors and non-survivors. Both networks are composed of the same set of variables and the weights are calculated using equation (2). For prediction of the outcome of interest, we identified the discriminatory performance of simultaneous failures which are unique among non-survivors but not among survivors. Therefore, using the networks created for the two subpopulations, a third network was created which was composed of the edges with weights calculated as the difference between the weights of the corresponding two edges in the networks of survivors and non-survivors (Figure 4). The rightmost network in Figure 4 shows the difference network of the networks of two subpopulations. The difference network highlights the edges with higher weights. These edges represent simultaneous failures which are more common in non-survivors. The weights of edges in the difference network show the difference of weights between the two groups. In other words, the weights in the difference network show the differences in simultaneous failures of two variables in the two subpopulations; higher weights indicate higher probabilities of simultaneous failure of the two corresponding variables in non-survivors in comparison with survivors. The maximum value of weights in the difference network is about 0.4, related to the edge between Oxygen source and C-Reactive Protein nodes when both are out of the acceptable range provided in Table 1.

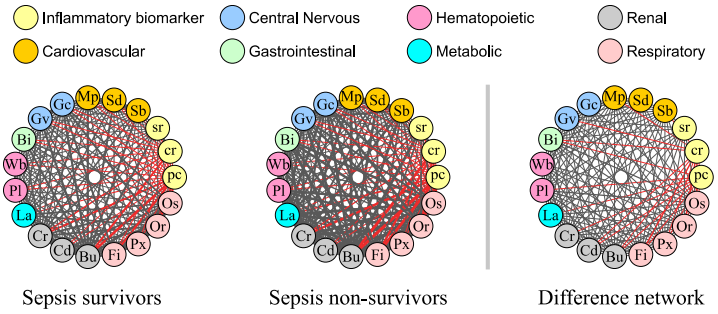

Figure 4: Networks of sepsis survivors and non-survivors at left, and difference network at right. The width of edges represents the relative weight of edges. ((Sb (systolic blood pressure); Sd (Sb (systolic blood pressure); Sd (Sb_max* - SBP); Mp (MAP); Cr (Creatinine); Cd ((Creatinine - C_base)/(C_base)); Wb (WBC); La (Lactate); Pl (Platelet); Bu (BUN); Bi (BiliRubin); Fi (FiO2); Px (SpO2); Or (SpO2/FiO2); Os (Oxygen Source); Gc (Glasgow Coma Score); Gv (Glasgow Best Verbal Response); pc (Procalcitonin); cr (CRP); sr (Sedimentation Rate). *: Maximum systolic blood pressure for each observation within 8-hour windows. **: Initial creatinine value observed in each visit)**

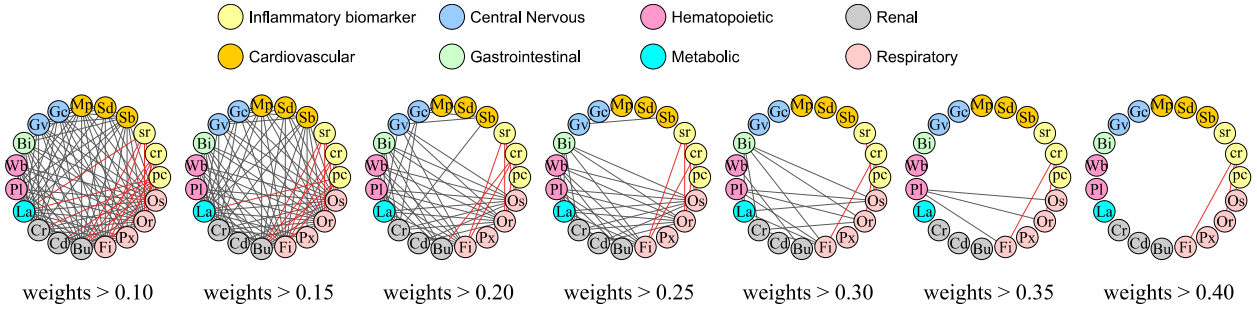

Figure 5: Difference Network created by removing edges with weights less than the values shown under each network, we just keep edges which show simultaneous failures which are more common in patients died/discharged to hospice. ((Sb (systolic blood pressure); Sd (Sb (systolic blood pressure); Sd (Sb_max* - SBP); Mp (MAP); Cr (Creatinine); Cd ((Creatinine - C_base)/(C_base)); Wb (WBC); La (Lactate); Pl (Platelet); Bu (BUN); Bi (BiliRubin); Fi (FiO2); Px (SpO2); Or (SpO2/FiO2); Os (Oxygen Source); Gc (Glasgow Coma Score); Gv (Glasgow Best Verbal Response); pc (Procalcitonin); cr (CRP); sr (Sedimentation Rate). *: Maximum systolic blood pressure for each observation within 8-hour windows. **: Initial creatinine value observed in each visit))**

Figure 5 visualizes the difference network at different weight thresholds. In each network in this figure, the edges with weights more than the corresponding threshold are shown. By creating and using a complete network in which all the 19 variables are connected, a gbm model was developed to predict in-hospital mortality in sepsis population. In addition, the edges in each of the networks in Figure 5 were used to create separate gbm models to predict subpopulation-specific in-hospital mortality. The ROC curves, and AUC and prediction accuracy for the complete network gbm model (composed of all the edges), and subpopulation-specific gbm models are shown in Figures 6 and 7. The accuracy of the gbm models developed for the complete network and the first network in Figure 5 are close, both about 78% with AUC = 0.85 (their corresponding number of edges are 171, and 105).

Using the edges of the network with weights more than 0.2 does not decrease the prediction accuracy (76.5%); however, the number of edges decreases to 51, showing that network models have been able to eliminate 120 insignificant edges without a significant decrease in the prediction accuracy. Using the edges of the network with the weights more than 0.25 decreases the accuracy to 73% using 37 edges. In other words, removing about 130 edges of the network results in 5% drop in accuracy. Using the edges of the network with the weights more than 0.3 corresponds to prediction accuracy above 70% (~71.5% and AUC = 0.77) with only 18 edges. Increasing the weights to 0.35 and 0.4 decreases the prediction accuracy to 60% and 50% respectively with 6 and 3 edges in their corresponding network. The results of the gbm model for the complete network show that edges with the highest relative influence on the prediction accuracy are mostly those representing the simultaneous dysfunctions of respiratory-renal, renal-cardiovascular, and renal-metabolic organ systems (Figure 8).

The edges representing such simultaneous failures are not present in the networks with weight threshold more than 0.35 and 0.4 which explains the lower prediction accuracy in these networks compared to the networks with edges including weights less than 0.3. The network with weights

higher than 0.40 in Figure 5 shows that respiratory responses and C-Reactive Protein have a more than 40% higher probabilities to fail simultaneously in sepsis non-survivors. Although simultaneous failure of respiratory responses and C-Reactive Protein are more common in septic shock non-survivors compared with survivors, they are not co-measured frequently in the study dataset. This might lead to the model not considering their simultaneous failure as a strong predictor of patients' outcome.

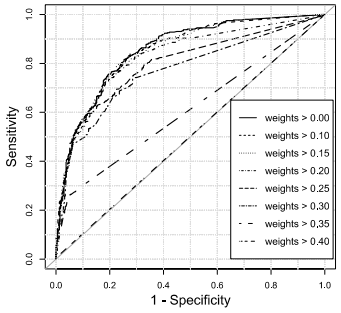

Figure 6: ROC curves for the gbm models developed using all the possible edges (weights > 0), and edges in the networks of Figure 6 with weights > 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, and 0.4

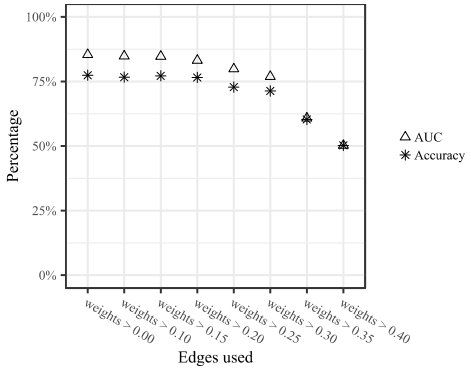

Figure 7: AUC and prediction accuracy for the gbm models developed using all the possible edges (weights > 0), and edges in the networks of Figure 6 with weights > 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, and 0.4

4. DISCUSSION

This study provides an analytical framework for exploring simultaneous abnormalities in cellular and physiological responses representing organ system dysfunctions in patients with sepsis to predict in-hospital mortality. Utilizing novel undirected weighted network models, the key findings quantify: i) optimal window for defining the simultaneity of observed organ system dysfunctions, ii) differences in network properties corresponding to different subpopulations, and iii) discriminatory patterns of simultaneous failures and subpopulations-specific in-hospital mortality prediction. The novelty of the approach is twofold. First, it shows that patterns of simultaneous dysfunctions are subpopulation-specific. Second, it shows that these differences can inform the therapeutic management of sepsis at different stages and in an individualized manner. Also, using different thresholds for the difference network for the purpose of sensitivity analysis, it was shown there are some of the simultaneous dysfunctions which are the main predictors of patients' outcome. In the remainder of this section, we discuss the clinical interpretation and insights derived from the findings, limitations of the study as well as promising future research directions.

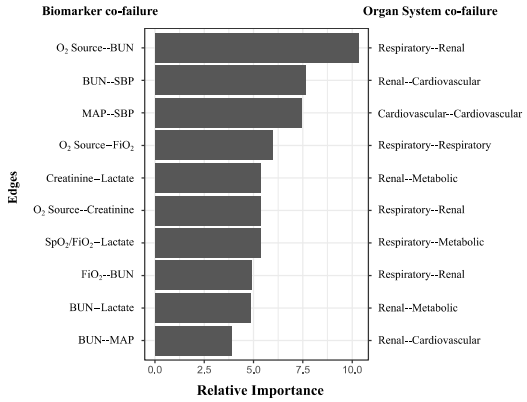

Figure 8: Organ failure-based comparison of the relative importance of the edges contributing to accuracy of predicting in-hospital mortality. Each label in vertical axis represents one edge (top ten most important edges are shown). BUN=Blood Urea Nitrogen; FiO2= Fraction of inspired oxygen; MAP: mean arterial pressure; SBP=systolic blood pressure; SpO2=Pulse oximeter oxygen saturation

4.1 Value of quantifying simultaneous organ dysfunction in sepsis

In previous studies, the cellular and physiological responses have been considered as independent variables for prediction of patients' outcome [11, 12]. Considering that the responses occurring in sepsis progression are not independent events, the temporal relationship between failures of responses can inform the dependency structures between these responses and following care delivery decisions. To the best of our knowledge, organ dysfunction within the sepsis context has not previously been quantified using a temporal relationship, i.e., optimal time window for identifying observed abnormality in responses relative to each other. The time window of 150 minutes to define simultaneity resulted in the highest combined accuracy and AUC (Figure 1). This approach recognizes the importance of considering the temporal patterns of cellular and physiological responses in a data-driven manner.

Key findings of this study show that: i) the simultaneous dysfunctions of renal-respiratory, renal-metabolic, and renal-cardiovascular organ systems exhibit the highest relative influence on the prediction accuracy regarding in-

hospital mortality, and ii) the simultaneous dysfunctions are significantly different between sepsis survivors vs non-survivors. The simultaneous organ systems failures differentiating non-survivors from survivors were not the systems that current standard of care would commonly recognize, which are cardiovascular and respiratory systems. Failure of renal system paired with respiratory, cardiovascular, or metabolic system failure within 150 minutes resulted in the highest associated predictive ability. Furthermore, BUN was the renal system failure biomarker with highest predictive value, and not the commonly described serum creatinine level. BUN biomarker as measure of renal system dysfunction within the sepsis context has been previously described [36, 37], but has not yet been studied in temporal relationship to other responses.

4.2 Value of quantifying simultaneous inflammatory biomarker elevation with organ dysfunction in sepsis

The inflammatory biomarkers are often used in clinical practice to help in diagnosing the presence of infection from other non-infectious causes or to monitor a patient's response to therapy over time. The inflammatory biomarkers (procalcitonin, c-reactive protein (CRP), and erythrocyte sedimentation rate (ESR)) are not routinely assessed in clinical practice, and when they are, are not routinely re-assessed as surrogates for organ system function. This study evaluated of the value of the simultaneous failure of inflammatory and physiologic biomarkers. Key findings of this study showed that simultaneous CRP-respiratory (FiO_2), CRP-respiratory (SpO_2/FiO_2), and PCT-metabolic (lactate) were the strongest associations differentiating non-survivors from survivors compared to any other simultaneous organ system dysfunctions (Figure 5). While assessing the true strength of this association was limited by the relative infrequent measurement the inflammatory biomarkers, these findings can advocate for a more routine assessment of these inflammatory biomarkers in clinical practice.

4.3 Translation of findings into clinical practice

Clinically, multi-organ failure is treated with the goal of addressing shock states with blood pressure support (vasopressor therapy) and supplementing oxygen and breathing ability when deficient. Applying the findings from this study regarding the organ system simultaneous dysfunctions and inflammatory biomarker has the potential to inform early detection of individuals at highest risk of in-hospital death and facilitate the best opportunity for sepsis patients to receive timely care. This work is aligned with a recent study [37] showing that respiratory and renal system dysfunction caries a higher risk of death from sepsis than other organ systems. The use of non-traditional markers such as BUN as a measure for renal system dysfunction has shown great value when it occurs simultaneously with respiratory, cardiovascular or metabolic system dysfunction in differentiating sepsis non-survivors from survivors. Additionally, due to the strength of the signal but relative infrequent assessment of inflammatory biomarkers, a more regular assessment of the tests is warranted in effort to better risk stratify sepsis patients.

Further, severity of illness scoring systems targeting sepsis risk-stratification commonly treat organ system dysfunction as independent events when applying a weighting algorithm to observed dysfunctions. An incorporation of probabilistic weighting as demonstrated in this study could improve the performance of scoring systems, both in improving the sensitivity and specificity, and thus decreasing false alarms. Finally, based on the key findings, our recommendations are to incorporate regular directed assessment for the presence of the following simultaneous (150 minutes) failures in sepsis population: Low FiO_2 and elevated BUN, Low MAP and elevated BUN, elevated lactate and elevated BUN, elevated CRP and low FiO_2, elevated CRP and low SpO_2/FiO_2, and elevated PCT and elevated lactate.

4.4 Limitations and Future Directions

This study is based on a data derived from a single healthcare system which may impact the generalizability of the findings. Future studies can incorporate data from multiple healthcare systems with external validation and improved generalizability purposes. Another limitation is that we used the consensus sepsis definition [1] with a broad definition of organ failure based on assimilation of organ dysfunction thresholds defined by SOFA score [38], PIRO score [36] and clinical expertise. Using a different definition for sepsis may impact the replication of the results. Further, the undirected weighted networks used pairwise failure of responses, i.e. the edges of the networks, as independent variables in predicting in-hospital mortality. However, in some cases, larger subnetworks (composed of more than two responses failing simultaneously) can improve the prediction accuracy. A promising future research area is the application of subgraph mining. Subgraph mining provides an analytical approach to detect relevant and significant subnetworks in a set of networks. Finally, with regards to identifying optimal time windows to define temporal relationships between observed failure of responses, we identified the optimum aggregation time and then aggregated all the observations recorded in a fixed window before and after each specific observation. Another promising future research direction is using other methods of aggregation of temporal relationships between organ system failures, such as dividing the visit into discrete time intervals and studying time interval-specific simultaneous dysfunctions.

ACKNOWLEDGMENTS

We would like to acknowledge that this work has been performed on behalf of the S.E.P.S.I.S (Sepsis Early Prediction Support Implementation System) Collaborative. This work was supported by the National Science Foundation Smart and Connected Health (Award Number: 1833538), and the National Library of Medicine of the National Institutes of Health (Grant Number: 1R01LM012300-01A1, Award

Number: R01LM012300). This work was also supported in part by the National Science Foundation under the Grant NSF-1741306, IIS-1650531, and DIBBs-1443019. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Rhodes, A., Evans, L. E., Alhazzani, W., Levy, M. M., Antonelli, M., Ferrer, R., Kumar, A., Sevransky, J. E., Sprung, C. L. and Nunnally, M. E. 2017. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Intensive care medicine*, 43, 3, 304-377.

[2] Rhee, C., Dantes, R., Epstein, L., Murphy, D. J., Seymour, C. W., Iwashyna, T. J., Kadri, S. S., Angus, D. C., Danner, R. L. and Fiore, A. E. 2017. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014. *Jama*, 318, 13, 1241-1249.

[3] Rivers, E., Nguyen, B., Havstad, S., Ressler, J., Muzzin, A., Knoblich, B., Peterson, E., and Tomlanovich, M. 2001. Early Goal-Directed Therapy Collaborative, G. Early Goal-Directed Therapy in the Treatment of Severe Sepsis and Septic Shock. *The New England Journal of Medicine*, 345, 19, 1368-1377.

[4] Kumar, A., Roberts, D., Wood, K. E., Light, B., Parrillo, J. E., Sharma, S., Suppes, R., Feinstein, D., Zanotti, S., Taiberg, L., Gurka, D., Kumar, A. and Cheang, M. 2006. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical Care Medicine*, 34, 6, 1589-1596.

[5] Group, B. D. W., Atkinson Jr, A. J., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., Oates, J. A., Peck, C. C. and Schooley, R. T. 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, 69, 3, 89-95.

[6] Singer, M. 2013. Biomarkers in sepsis. *Current opinion in pulmonary medicine*, 19, 3, 305.

[7] Pierrakos, C. and Vincent, J.-L. 2010. Sepsis biomarkers: a review. *Critical care*, 14, 1, R15.

[8] Forrest, D. M., Djurdjev, O., Zala, C., Singer, J., Lawson, L., Russell, J. A. and Montaner, J. S. 1998. Validation of the modified multisystem organ failure score as a predictor of mortality in patients with AIDS-related Pneumocystis carinii pneumonia and respiratory failure. *Chest*, 114, 1, 199-206.

[9] Levraut, J., Ichai, C., Petit, I., Ciebiera, J.-P., Perus, O. and Grimaud, D. 2003. Low exogenous lactate clearance as an early predictor of mortality in normolactatemic critically ill septic patients. *Critical care medicine*, 31, 3, 705-710.

[10] Lee, S., Hong, Y., Park, D., Choi, S., Moon, S., Park, J., Kim, J. and Baek, K. 2008. Lactic acidosis not hyperlactatemia as a predictor of inhospital mortality in septic emergency patients. *Emergency Medicine Journal*, 25, 10, 659-665.

[11] Hermans, M., Leffers, P., Jansen, L., Keulemans, Y. and Stassen, P. 2011. The value of the Mortality in Emergency Department Sepsis (MEDS) score, C reactive protein and lactate in predicting 28-day mortality of sepsis in a Dutch emergency department. *Emergency Medicine Journal*, emj. 2010.109090.

[12] Innocenti, F., Bianchi, S., Guerrini, E., Vicidomini, S., Conti, A., Zanobetti, M. and Pini, R. 2014. Prognostic scores for early stratification of septic patients admitted to an emergency department-high dependency unit. *European Journal of Emergency Medicine*, 21, 4, 254-259.

[13] Huang, D. T., Weissfeld, L. A., Kellum, J. A., Yealy, D. M., Kong, L., Martino, M., Angus, D. C. and Investigators, G. 2008. Risk prediction with procalcitonin and clinical rules in community-acquired pneumonia. *Annals of emergency medicine*, 52, 1, 48-58. e42.

[14] Jiang, L., Feng, B., Gao, D. and Zhang, Y. 2015. Plasma concentrations of copeptin, C-reactive protein and procalcitonin are positively correlated with APACHE II scores in patients with sepsis. *Journal of International Medical Research*, 43, 2, 188-195.

[15] Hong, D. Y., Kim, J. W., Paik, J. H., Jung, H. M., Baek, K. J., Park, S. O. and Lee, K. R. 2016. Value of plasma neutrophil gelatinase-associated lipocalin in predicting the mortality of patients with sepsis at the emergency department. *Clinica Chimica Acta*, 452, 177-181.

[16] Lee, C.-C., Chen, S.-Y., Tsai, C.-L., Wu, S.-C., Chiang, W.-C., Wang, J.-L., Sun, H.-Y., Chen, S.-C., Chen, W.-J. and Hsueh, P.-R. 2008. Prognostic value of mortality in emergency department sepsis score, procalcitonin, and C-reactive protein in patients with sepsis at the emergency department. *Shock*, 29, 3, 322-327.

[17] Garcia-Simon, M., Morales, J. M., Modesto-Alapont, V., Gonzalez-Marrachelli, V., Vento-Rehues, R., Jorda-Miñana, A., Blanquer-Olivas, J. and Monleon, D. 2015. Prognosis biomarkers of severe sepsis and septic shock by 1H NMR urine metabolomics in the intensive care unit. *PLoS One*, 10, 11, e0140993.

[18] Liu, B., Chen, Y.-X., Yin, Q., Zhao, Y.-Z. and Li, C.-S. 2013. Diagnostic value and prognostic evaluation of Presepsin for sepsis in an emergency department. *Critical Care*, 17, 5, R244.

[19] Giamarellos-Bourboulis, E. J., Norrby-Teglund, A., Mylona, V., Savva, A., Tsangaris, I., Dimopoulou, I., Mouktaroudi, M., Raftogiannis, M., Georgitsi, M. and Linnér, A. 2012. Risk assessment in sepsis: a new prognostication rule by APACHE II score and serum soluble urokinase plasminogen activator receptor. *Critical care*, 16, 4, R149.

[20] Ju, M., Zhu, D., Tu, G., He, Y., Xue, Z., Luo, Z. and Wu, Z. 2012. Predictive value of N-terminal pro-brain natriuretic peptide in combination with the sequential organ failure assessment score in sepsis. *Chinese medical journal*, 125, 11, 1893-1898.

[21] Kartal, E. D., Karkaç, E., Gülbaş, Z., Alpat, S. N., Erben, N. and Çolak, E. 2012. Several Cytokines and Protein C Levels with the Apache II Scoring System for Evaluation of Patients with Sepsis. *Balkan medical journal*, 29, 2, 174.

[22] Shi, Z., Wu, J. and Ben-Arieh, D. 2014. A Modeling Comparative Study on Sepsis. *IIE Annual Conference. Proceedings. Institute of Industrial and Systems Engineers (IISE)*, 1069.

[23] Sayama, H. 2015. *Introduction to the modeling and analysis of complex systems*. Open SUNY Textbooks.

[24] Macal, C. M. and North, M. J. 2005. *Tutorial on agent-based modeling and simulation*. IEEE, City.

[25] Macal, C. and North, M. 2014. *Introductory tutorial: Agent-based modeling and simulation*. IEEE, City.

[26] Macal, C. M. and North, M. J. 2010. Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4, 3, 151-162.

[27] Orphanou, K., Stassopoulou, A. and Keravnou, E. 2014. Temporal abstraction and temporal Bayesian networks in clinical domains: A survey. *Artificial Intelligence in Medicine*, 60, 3, 133-149.

[28] Peelen, L., de Keizer, N. F., Jonge, E. d., Bosman, R.-J., Abu-Hanna, A. and Peek, N. 2010. Using hierarchical dynamic Bayesian networks to investigate dynamics of organ failure in patients in the Intensive Care Unit. *Journal of Biomedical Informatics*, 43, 2, 273-286.

[29] Nachimuthu, S. K. and Haug, P. J. 2012. *Early detection of sepsis in the emergency department using Dynamic Bayesian Networks*. American Medical Informatics Association, City.

[30] Gultepe, E., Green, J. P., Nguyen, H., Adams, J., Albertson, T. and Tagkopoulos, I. 2013. From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. *Journal of the American Medical Informatics Association*, 21, 2, 315-325.

[31] Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G. R., Chiche, J.-D., Coopersmith, C. M., Hotchkiss, R. S., Levy, M. M., Marshall, J. C., Martin, G. S., Opal, S. M., Rubenfeld, G. D., van der Poll, T., Vincent, J.-L. and Angus, D. C. 2016. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315, 8, 801-810.

[32] Freund, Y. and Schapire, R. E. 1997. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55, 1, 119-139.

[33] Friedman, J. H. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29, 5, 1189-1232.

[34] Friedman, J. H. 2002. Stochastic gradient boosting. *Computational Statistics and Data Analysis*, 38, 4, 367-378.

[35] Ridgeway, G. 2007. Generalized Boosted Models: A guide to the gbm package. *Update*, 1, 1, 2007.

[36] Howell, M. D., Donnino, M. W., Talmor, D., Clardy, P., Ngo, L. and Shapiro, N. I. 2007. Performance of Severity of Illness Scoring Systems in Emergency Department Patients with Infection. *Academic Emergency Medicine*, 14, 8, 709-714.

[37] Capan, M., Hoover, S., Ivy, J. S., Miller, K. E., Arnold, R. and Collaborative, S. E. P. S. I. S. 2018. Not all organ dysfunctions are created equal – Prevalence and mortality in sepsis. *Journal of Critical Care*, 48, 257-262.

[38] Vincent, J. L., Moreno, R., Takala, J., Willatts, S., Mendon a, A. D., Bruining, H., Reinhart, C. K., Suter, P. M. and Thijs, L. G. 1996. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Medicine*, 22, 7, 707-710.