Early Diagnosis and Its Benefits in Sepsis Blood Purification Treatment

Mohamed Ghalwash
Computer and Information Sciences
Temple University
Philadelphia, PA 19122
Email: mohamed.ghalwash@temple.edu

Vladan Radosavljevic
Computer and Information Sciences
Temple University
Philadelphia, PA 19122
Email: vladan@temple.edu

Zoran Obradovic
Computer and Information Sciences
Temple University
Philadelphia, PA 19122
Email: zoran.obradovic@temple.edu

***Abstract*—Sepsis is a progressive medical condition characterized as an uncontrolled inflammatory response, which is the leading cause of death in non-coronary intensive care units in the United States. In sepsis treatment, accurate and timely diagnosis is essential for allowing physicians to design appropriate therapeutic strategies at early stages, when therapies are usually the most effective and the least costly. To make an adequate diagnosis, physicians usually rely on manual inspection of a large amount of complex, high-dimensional longitudinal data. We use our recently published data mining method for extracting patterns from such data and show that these patterns can be used to assist physicians in providing early diagnosis. In conducted experiments, we showed that combination of early diagnosis and blood purification therapy can rescue more patients (52%) than standard approach for blood purification therapy (32%). We also propose a hybrid therapy model that combines strengths of early and standard approaches and further improves the percentage of rescued patients. Finally, by correctly classifying 98% of patients who didn't need treatment, MSD method provides opportunity to reduce the total cost of treatments.**

I. Introduction

Sepsis, a medical condition characterized by uncontrolled inflammatory response due to infection, is one of the main causes of deaths in the intensive care units, with over 750,000 cases annually in the United States alone [12]. One of the main reasons for such a high number of death cases lies in limited understanding and knowledge about the complex inflammatory response mechanism, which has led to only a few effective sepsis therapies. The single approved anti-sepsis drug therapy was withdrawn from global markets in fall 2011 following the failure of its worldwide trial to demonstrate improved patient outcome [1]. In the absence of adequate therapy, the patient is treated with standard broad-spectrum antibiotics and/or intravenous fluids with dosages adjusted manually. Inadequate treatment and the fact that sepsis is often diagnosed too late result in a mortality rate of 30-35%, and for every hour that the administration of appropriate treatment is delayed, the mortality rate increases by about 7% [10].

Blood purification has recently been proposed as a potentially beneficial therapy for septic patients [7]. This therapy is based on the dialysis-like principle, where the blood is purified by a device attached to patient (Figure 1). The goal of purification is to remove harmful particles from patient blood, leading the patient to a healthy state. Preliminary studies on animal models indicate the success of blood purification techniques in sepsis treatment [9]. In conducted animal studies clinicians had full control of the exact time of sepsis induction. Based on that information and theoretical assumptions about sepsis progress, clinicians determined start and duration of therapy. For example, in [9] the duration of therapy was fixed to 4 hours while the onset of therapy was set at 18 hours after sepsis induction (the clinicians' assumption was that after the 18^{th} hour sepsis would be fully manifested). The onset of therapy was the same for all animal models involved in the experiment, although more personalized approach to sepsis diagnosis would be far more effective. Also, fixed onset of therapy with respect to sepsis induction does not reflect real clinical practice where clinicians have limited knowledge of sepsis stage at the moment when a patient shows up in an emergency room. Despite its importance, no work has been done to provide correct and timely sepsis diagnosis in conjunction with optimal blood purification therapy.

The time needed for sepsis detection can be significantly reduced by using the information from multivariate longitudinal data collected from the patient over time. However, utilization of those complex high-dimensional longitudinal data to support clinical decisions is still not fully implemented in practice since physicians lack the tools to extract relevant clinical information in a timely manner. Extracting useful temporal patterns early and building accurate predictive models on such data provides a great challenge for the data mining community.

A recent work is proposed to accurately predicting both sepsis risk and septic shock from noisy, intermittently gathered clinical data using clinical and historical variables [6]. The authors have chosen a small set of features such that to minimize the use of laboratory tests and invasive procedures while maintaining comparable performance to other models. Another recent work has proposed a modified version of the binary particle swarm optimization (MBPSO) method for feature selection with the simultaneous optimization of Support Vector Machine (SVM) kernel parameter setting [11]. The MBPSO method is applied to mortality prediction in septic patients. Although the method has been shown to outperform the original particle swarm optimization method, it has some limitations if there is a poor adjustment in the parameters. In addition, both models [6, 11] do not leverage the temporal information.

Recently, a method that utilizes the temporal information for early classification of multivariate time series is proposed [4, 5]. The model is comprised of an integration of Hidden Markov Models (HMM) and SVM that allows for early,

978-0-7695-5089-3/13 $26.00 © 2013 IEEE
DOI 10.1109/ICHI.2013.81

accurate, and patient-specific classification of multivariate time series. Although the method has attained very promising results on several datasets, the method is not interpretable which limits the application of the method on clinical datasets where physicians tend to use interpretable methods rather than black-box methods.

In this work we use our previously published method [3] to make early diagnosis for septic patients. All work is performed on virtual patients generated by a mathematical model that emulates inflammatory response, which is common practice in biomedical research. In conducted experiments we showed that combination of early diagnosis and blood purification therapy can rescue more patients than standard approach for blood purification therapy. We also propose a hybrid therapy model that combines strengths of early and standard approaches and further improves the percentage of rescued patients.

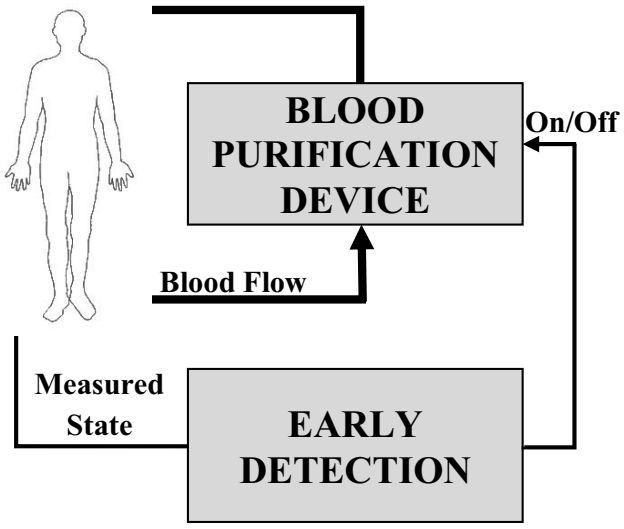

Fig. 1. Schematic diagram of dialysis-like blood purification device accompanied with early detection module.

II. Methodology

We used a machine learning method, called multivariate shapelet detection (MSD), proposed recently to extract interpretable shapelets for early classification of multivariate time series [3]. We used the MSD method for two folds. First, it extracts time series segments from the original time series for early classification. Therefore, the method provides an evidence about the classification of the temporal observation of the patients. Second, the MSD method is proposed for multivariate time series, and it has been shown that it outperforms the state-of-the-art method for univariate time series.

The MSD method has four steps to perform the task of early classification of multivariate time series:

1) Extraction of all multivariate shapelets
2) Ranking the multivariate shapelets
3) Pruning the list of multivariate shapelets
4) Classification using multivariate shapelets

a) Step 1: Extraction all Multivariate Shapelets: Simply, the multivariate shapelet is a time series segment extracted from the original time series. Figure 2 shows an example of a 3-dimensional time series of length 15. It shows an example of an extracted 3-dimensional shapelet of length 4. The shapelet is extracted from the time series from position 6 to position 9.

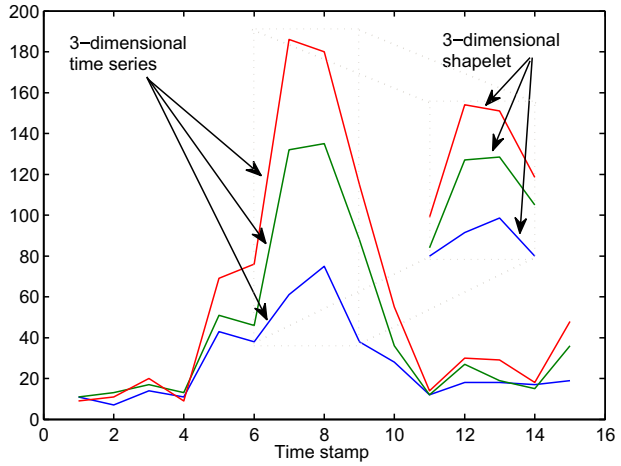

Fig. 2. Illustration of a 3-dimensional shapelet. It shows an example of a 3-dimensional time series (red, green and blue lines) of length 15. An example of an extracted 3-dimensional shapelet of length 4 is illustrated in the right part of the figure. The shapelet is extracted from the time series from position 6 to position 9.

The MSD method iterates over all time series and extracts all possible segments (shapelets) S of all possible lengths l. Figure 2 shows an example of multivariate shapelet of length 4.

Each shapelet is used to discriminate between the classes of the time series. For that purpose, each shapelet has its own distance threshold such that if the distance between the shapelet and the time series is less than or equal to the threshold, the time series is considered to be of the same class as the shapelet.

Details of computing the distance threshold for each shapelet are given in [3] because it exceeds the scope of this paper.

b) Step 2: Ranking the multivariate shapelets: The set of multivariate shapelets extracted from the time series dataset might be exceedingly large. Therefore, it is important to rank the shapelets in order to select a small subset of the shapelets for classification. For this reason, each shapelet has to be assigned a utility score that takes into consideration earliness as well as discrimination among classes.

c) Step 3: Pruning the list of multivariate shapelets: To select a subset of the multivariate shapelets for classification, the shapelets are sorted in descending order using their utility scores. The pruning procedure iterates over the shapelets starting from the highest ranked shapelet. We consider that shapelet and remove all training time series that are covered by that shapelet (the distance between the shapelet and time series is less than or equal to the shapelet's distance threshold). Then, the next highest ranked shapelet is considered. We check

if it covers any of the remaining training time series. If it covers some of them, then we select the shapelet and remove all time series that are covered. Otherwise, we discard it and proceed to the next one. This process of selecting shapelets continues until all training time series are covered.

d) Step 4: Classification using multivariate shapelets: The classification process initially reads l time stamps from the test time series. The highest-ranked shapelet is considered. If the shapelet covers the current stream of the test time series then the time series is classified as the class of the shapelet and the prediction is done. Otherwise, the next shapelet from the ranked list is considered and the process is repeated. If none of the shapelets cover the current stream of the test time series the method reads one more time stamp and continues classifying the time series (Figure 3).

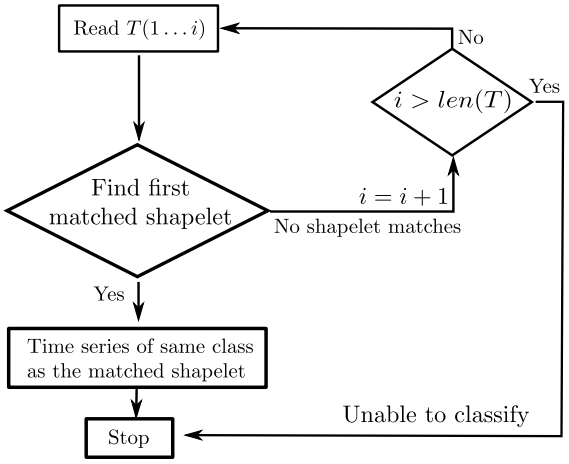

Fig. 3. MSD Classification Process. The MSD method looks at a portion $T(1 \dots i)$ of length i of the unknown time series T. If there is a shapelet matches the current portion of the time series, then the class of the time series is predicted as the class of the matched shapelet. Otherwise, the current portion is extended and the same process applied.

III. Virtual Patients

To significantly reduce the chance of a clinical failure and to save on the costs of clinical trials, biomedical researchers use computer simulations of body processes (often called virtual patients) to perform preliminary tests of hypotheses before they prove them in real patient studies. Virtual patients are generated using a carefully determined mathematical model to simulate the process of interest. A significant advantage of having a virtual patient model for experiments is the possibility of testing different approaches for finding adequate therapies on the same virtual patient and comparing the outcomes. In order to follow a real-life scenario, virtual patient models are accompanied with well-defined constraints in therapy that are in accordance with clinical practice [2, 8, 9].

A. Patient model

The mathematical model for inflammatory response to an infection is derived in [9]. A mathematical model defines the dynamics of concentration of 19 variables (states) among which 8 are observable (Lsel - Lselectin; HMGB1 - high-mobility group protein B-1; CRT - creatinine; ALT - alanine aminotransferase; TNFα - tumor necrosis factor-α; IL-1 - interleukin-1β; IL-6 - interleukin-6; IL-10 - interleukin-10) and 11 are hidden (CLP - cecal ligation and puncture; B - bacteria; Nt - peritoneal neutrophil; Nr - resting blood neutrophil; Np - primed blood neutrophil; Na - activated blood neutrophil; PI - systemic proinflammatory response; AI - systemic anti-inflammatory response; Ns - neutrophil sequestered in lung capillaries; Nl - lung neutrophil).

The model is also capable of modeling interactions between organs. This mathematical model is based on the system of ordinary differential equations (ODE) whose details are presented in [9]. Since it models measurable concentrations of cytokines, the 19-states mathematical model are capable of simulating blood purification system by hemoadsoprtion device (a column packed with beads that adsorb cytokines). It is assumed that a patient's blood is redirected through the hemoadsorption device where pro- and anti-inflammatory particles are removed.

Mathematical model simulates hemoadsorption as a first-order elimination of activated neutrophils (Na) as well as pro- and anti-inflammatory mediators (PI and AI) from the circulating blood. The parameters of the dynamics of adsorption were determined in [9]. We assume that the device has two states - ON and OFF. ON state means that the device is attached to the patient and that it cleans blood with the rate specified in [9]. OFF state means that device is detached from the patient. The ON/OFF states of blood purification device are controllable by clinicians.

Variability in the population of virtual patients is obtained by random initialization of three parameters in ODE and by random initialization of the states' initial conditions. In all of the simulations, t is an hourly step that starts from $t = 0$ when patient state and parameters are initialized. Then, patient state evolves according to ODE through the simulation time of 200 hours. According to [9] there are two possible outcomes at the end of simulation time. A patient is in survival group if (1) the number of bacteria (B) is less than $Bmin$ which was set to $1.0e5$, and (2) the value of systemic inflammation (PI) is less than 0.5. Otherwise, a patient is in non-survival group. Evolution of the patient to the final state can be modulated by applying blood purification device.

B. Real Data

We obtained real data from [9]. Real data contain measurements over time of 8 observable states from mathematical model. As such, real data can be used for calibrating/testing the mathematical model. Experiments to obtain real data were designed to evaluate long-term (one week) survival rate in a model of sepsis that resulted in a mortality rate similar to that observed clinically. The modified cecal ligation and puncture (CLP) protocol, 25% ligated length of cecum and 20-gauge needle, two-puncture, was used by [9] to induce sepsis in 23 rats. Plasma cytokines (tumor necrosis factor (TNF), interleukin(IL)-1b, IL-6 and IL-10), Lselectin (Lsel), high mobility group box1 (HMGB1), creatinine (CRT) and alanine aminotransferase (ALT) were measured from 0.8 ml blood samples at 18, 22, 48, 72, 120, 144, and 168h after CLP. No treatment was applied to any of 23 rats. Seven rats out of these 23 survived up to 7 days, being considered as the

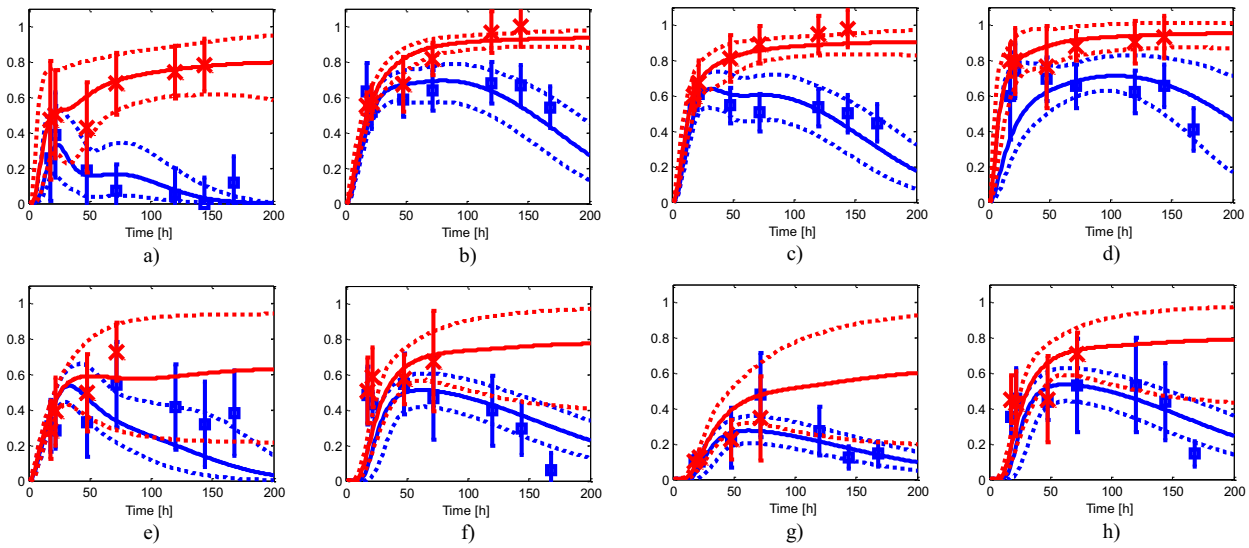

Fig. 4. An agreement between simulated and real data. a) TNFα, b) IL-1, c) IL-6, d) IL-10, e) Lsel, f) HMGB1, g) CRT, h) ALT. Solid lines - mean values of simulation outputs of 5000 virtual patients in survival (blue) and non-survival (red) groups. Dotted lines - region of 95% simulation uncertainty (95% of virtual patients are within the region). Error bars - real observations from animal study experiments.

survivor population; the remaining 16 animals died and were considered as the non-survivor population.

C. Generation of Virtual Patient Population in Agreement with Real Data

We use real data to generate virtual patients that are in accordance with real data as in [9]. We generated each virtual patient according to the following 3-step protocol:

1) We randomly sample parameters of a mathematical model in consistence with valid ranges described in [9].
2) For chosen parameters we simulate the evolution of 19-states over time and determine the outcome (survival or non-survival).
3) We calculated the likelihood [9] that evolution of 8 observable states follows evolution of real data. If the likelihood is high then the virtual patient has been "accepted" as valid. Otherwise, a generated patient has been rejected.

Following this protocol we generated 10000 sham (no treatment) virtual patients. A group of 5000 virtual patients belonged to the survival population, while another group of 5000 virtual patients belonged to the non-survival population. Statistics of simulated data for eight measurable states together with observations from real data are presented in Figure 4.

IV. MSD Training

We sampled randomly 30 patients from the 10000 patients for training the MSD method and used the remaining 9970 patients for testing. We chosen 30 patients to simulate the real-life where small number of patients is provided for training. Then, the MSD method is applied to the 9970 patients. All parameters of the MSD method were optimized using internal 3 cross validation based on the training data.

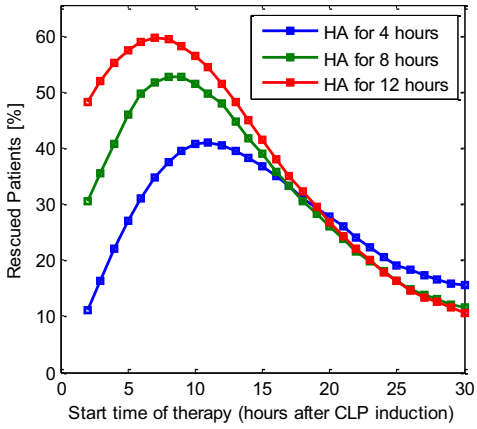

Fig. 5. Blood purification therapy efficacy (percentage of rescued patients) with respect to therapy starting time and the duration of treatment.

We repeated that process, sampling 30 patients for training and applying the trained MSD method on the 9970 patients, three times having three values for each reported statistic. The average and standard deviation of the statistics are reported.

V. Strategies for Blood Purification

A. Analysis of Onset and Duration of Blood Purification Therapy

We would like to analyze the effect of varying onset and duration of application of blood purification device. In [9] was suggested that optimal treatment should start at 18 hours after

sepsis induction (roughly the time point in Figure 4 where non-survival group becomes distinguishable from survival group). It was also suggested that optimal treatment should last for 4 hours. We used a group of 5000 virtual patients that would be in the non-survival population if no treatment was applied.

1) Therapy Onset: We tested how treatment efficacy depends on the starting time of the therapy. We applied the blood purification device continuously for 4 hours, as suggested in [9], with starting time varying from 2 to 30 hours after sepsis induction. We report percentage of rescued patients (patients for whom outcome after treatment was healthy, the higher percentage of rescued the better). The result is presented in Figure 5, blue line. We see that if treatment had been initiated later than 12 hours after sepsis induction, the percentage of rescued patients decreases with every hour. By applying therapy earlier (before 12 hours after sepsis induction), the percentage of rescued patients also decreases with every hour of early therapy. The graph shows that there is a critical time point around 10 hour at which the therapy is the most efficient.

This finding is in strong agreement with theoretical considerations of the sepsis stages and treatment effects. Sepsis treatment requires both a strong pro-inflammatory phase for the clearance of pathogen (Figure 6, area A) and an anti-inflammatory phase for recovery (Figure 6, area C). A stage of an adversary influence of the pro-inflammatory response, which is disproportionate and counterproductive, is presented in Figure 6, area B. An inadequate treatment in either the pro-inflammatory (area A) or the immune-recovering anti-inflammatory phase (area C) might do more harm than good, while delayed treatment when immune response is counter-productive (area B) may significantly reduce the chance of survival.

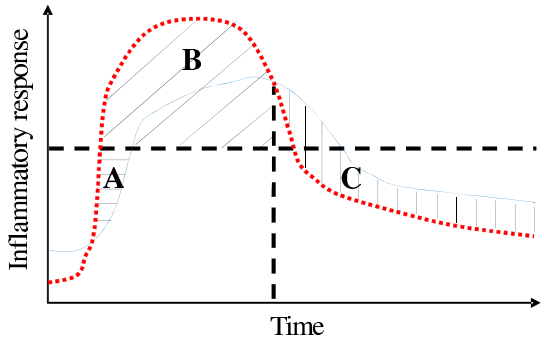

Fig. 6. Theoretical considerations of the sepsis stages and treatment effects. $Time = 0$ - occurrence of an infection; red dotted line - pro-inflammatory response; blue solid line - anti-inflammatory response; black dashed horizontal line - response beyond which the process becomes adversary; black dashed vertical line - a tip-over point beyond which any therapy might be counterproductive; horizontal pattern (area A) - anti-inflammatory therapy likely harmful, pro-inflammatory therapy beneficial; diagonal pattern (area B) - likely maximal benefit from anti-inflammatory therapy; vertical pattern (area C) - anti-inflammatory response restoring patient state, any therapy likely harmful.

2) Therapy Duration: We tested how treatment efficacy depends on the duration of continuous application of blood purification device. In addition to previous experiment, we applied the blood purification device continuously for 8 and 12 hours with varying therapy starting times. The plot in Figure 5 shows that with increased duration of HA application the percentage of rescued patients increases.

B. Treatment with MSD Therapy Onset

We applied our early classification method MSD to the virtual patients generated by the 19-states sepsis model. We used 30 patients (20 survival patients and 10 non-survival patients) for training and patients to test the model. The true positive rate of our method was 100%, meaning that all non-survival patients were predicted correctly and early. We then tested standard 18^{th} hour blood purification therapy versus therapy initiated at MSD suggested time.

1) Known Sepsis Induction Time: This experiment assumes that sepsis induction time is known (common in laboratory conditions but uncommon in clinical practice). We applied 12 hour blood purification therapy on 5000 non-survival patients and successfully rescued 32% of patients that would otherwise die (Figure 7). When 12 hour blood purification therapy was started at MSD suggested time the percentage of rescued patients increased to 52%. This result shows that in laboratory conditions when sepsis induction time is known there is a huge benefit of using MSD to suggest start time of therapy.

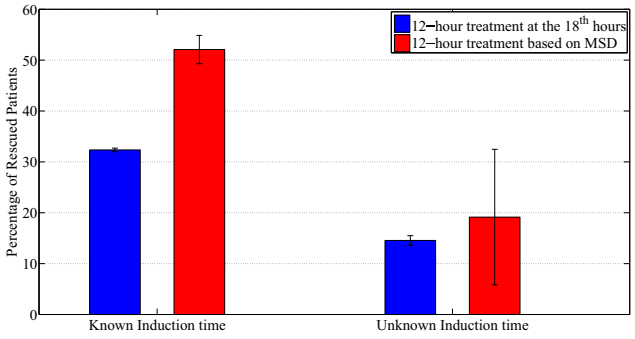

Fig. 7. Percentage of rescued patients. Known sepsis induction time - patients are admitted at the time of the CLP-induced sepsis. Unknown sepsis induction time - patients are admitted with some delay after CLP-induced sepsis (uniformly sampled from 5-12 hours after the induction).

2) Unknown Sepsis Induction Time: In clinical practice the sepsis induction time is unknown. To simulate this the time when the patient visits the ICU was sampled uniformly from 5-12 hours after the sepsis induction.

We applied 12 hour blood purification therapy at 18^{th} hour from time when patient showed up. With this procedure the number of rescued patients is around 14% (Figure 7). When we applied 12 hour therapy starting from MSD suggested time with respect to time when patient showed up we get an improvement to around 19% of rescued patients which again shows the benefit of using MSD to suggest start time of therapy.

3) Cost Reduction by MSD: The application of the standard therapy requires the continuous application of the blood purification device to every patient. Such application of the device is costly. It would be less costly to identify patients who would survive without using that device. Our MSD method was able to classify correctly 98% of the survived patients early before the 18^{th} hour after the sepsis induction (Figure 8). When the

sepsis induction was unknown, the MSD method was able to identify 10% of the survived patient (Figure 8).

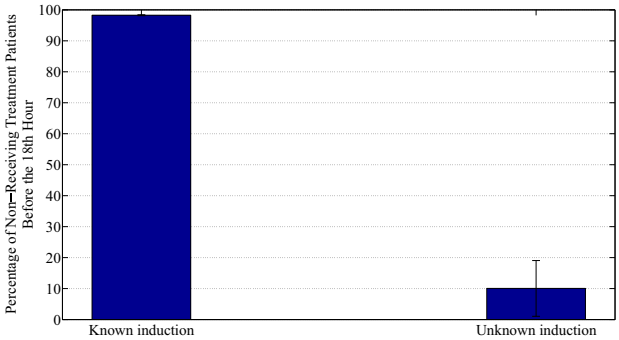

Fig. 8. The percentage of patients who are predicted not to receive treatment when the sepsis induction is known and unknown.

C. Hybrid Treatment Strategy

We realized that patient who were suggested by MSD to receive therapy after 18^{th} hour of sepsis induction (as many as one fourth of total population) may benefit if standard and MSD strategies were combined. Here we propose hybrid treatment strategy:

1) If a patient is recommended for treatment by MSD before the 18^{th} hour, then the treatment is provided at recommended time before the 18^{th} hour.
2) If a patient is recommended for treatment by MSD after the 18^{th} hour, then the treatment is provided at the 18^{th} hour.

In Figure 9, we represent the percentage of MSD-rescued patients with late treatment recommendation when the 12-hour therapy was provided at the recommended time (at 19th hour or after as recommended by our MSD method). Percentage of rescued patients is around 9%. On the other hand, in the same figure we show the percentage of 18th-rescued patients with late treatment recommendation when the 12-hour therapy was provided at the 18^{th} hour. The percentage of 18th-rescued patients is around 27%. Therefore, the hybrid treatment strategy saves more lives when compared to treatment solely based on MSD.

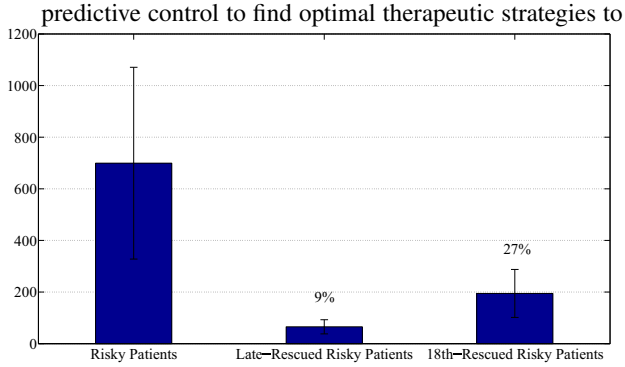

Fig. 9. Hybrid treatment strategy. Risky patients - non survival group, MSD predicts treatment after the 18^{th} hour; Late rescued - rescued when the 12-hour therapy was provided at the recommended time (at 19th hour or after as recommended by our MSD method); 18^{th} rescued - rescued when the 12-hour therapy was provided at the 18^{th} hour.

VI. Conclusion

We applied our method MSD [3] to make early diagnosis for septic patients. We showed that combination of early diagnosis and blood purification therapy can rescue more patients than standard approach for blood purification therapy. We also showed that a hybrid therapy that combines strengths of early and standard approaches and further improves the percentage of rescued patients.

References

[1] D. C. Angus. The search for effective therapy for sepsis: Back to the drawing board? *JAMA*, 306(23):2614–2615, 2011.

[2] G. Clermont, J. Rubin, and J. Day. Using nonlinear model predictive control to find optimal therapeutic strategies to modulate inflammation. *Mathematical Biosciences and Engineering*, 7(4):739–763, 2010.

[3] M. F. Ghalwash and Z. Obradovic. Early classification of multivariate temporal observations by extraction of interpretable shapelets. *BMC Bioinformatics*, 13(195), August 2012.

[4] M. F. Ghalwash, D. Ramljak, and Z. Obradovic. Early classification of multivariate time series using a hybrid hmm/svm model. In *IEEE International Conference on Bioinformatics and Biomedicine*, Philadelphia, PA, Oct 2012.

[5] M. F. Ghalwash, D. Ramljak, and Z. Obradovic. Patient-specific early classification of multivariate observations. *International Journal of Data mining and Bioinformatics*, in press.

[6] J. C. Ho, C. H. Lee, and J. Ghosh. Imputation-enhanced prediction of septic shock in icu patients. In *HI-KDD : ACM SIGKDD Workshop on Health Informatics*, 2012.

[7] T. Rimmele and J. A. Kellum. Clinical review: Blood purification for sepsis. *Crit Care*, 15(1):205, 2011.

[8] K. Ristovski, V. Radosavljevic, and Z. Obradovic. A data mining approach for optimization of acute inflammation therapy. In *IEEE International Conference on Bioinformatics and Biomedicine*, 2012.

[9] S. O. K. Song, J. Hogg, Z.-Y. Peng, R. S. Parker, J. A. Kellum, and G. Clermont. Ensemble models of neutrophil trafficking in severe sepsis. *PLoS Computational Biology*, 8(3), 2012.

[10] S. W. Thiel, J. M. Rosini, W. Shannon, J. A. Doherty, S. T. Micek, and M. H. Kollef. Early prediction of septic shock in hospitalized patients. *Journal of hospital medicine : an official publication of the Society of Hospital Medicine*, 5(1):19–25, 2010.

[11] S. M. Vieiraa, L. F. Mendonçaa, G. J. Farinhaa, and J. ao M.C. Sousaa. Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients. *Applied Soft Computing*, 13(8):3494–3504, 2013.

[12] S. M. Zuev, S. F. Kingsmore, and D. D. Gessler. Sepsis progression and outcome: A dynamical model. *Theoretical Biology and Medical Modelling*, 2006.