35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Multivariate Temporal Symptomatic Characterization of Cardiac Arrest

Yubin Park[1] and Joyce C. Ho[2] and Joydeep Ghosh[3]

***Abstract*— We model the temporal symptomatic characteristics of 171 cardiac arrest patients in Intensive Care Units. The temporal and feature dependencies in the data are illustrated using a mixture of matrix normal distributions. We found that the cardiac arrest temporal signature is best summarized with six hours data prior to cardiac arrest events, and its statistical descriptions are significantly different from the measurements taken in the past two days. This matrix normal model can classify these patterns better than logistic regressions with lagged features.**

I. INTRODUCTION

Cardiac arrest, a sudden failure of the heart, is a life-threatening condition with an in-hospital mortality rate of $\sim 80\%$. Research has shown that $\sim 62\%$ of cardiac arrests could have been prevented based on clinical evidence prior to the event [1], [2]. Studies have also shown that a quick response to cardiac arrest can decrease the mortality rate to 60% [3]. However, prediction systems are unable to accurately identify high-risk patients with sufficient intervention time [2].

Early warning systems use physiologically-based criteria to detect patient deterioration [4], [5], [2]. However, these systems fail to capture temporal patterns in the physiological measurements [6]. Common approaches for incorporating temporal data are lag features [6], encoding temporal patterns within the data [7], [8], and dynamic time series model [9], [10]. We propose a matrix-variate approach to characterize cardiac arrest patients' changes.

II. MATRIX NORMAL DISTRIBUTION

The matrix-variate normal distribution, shortly "matrix normal distribution", is a probability distribution that generalizes the multivariate normal distribution for random matrices. A $n \times p$ random matrix $\mathbf{Y}$ is drawn from a matrix normal distribution, if its probability density function follows:

$$p(\mathbf{Y}|\mathbf{M}, \mathbf{\Omega}, \mathbf{\Sigma}) = \frac{1}{c}\exp(-\frac{1}{2}\text{tr}[\mathbf{\Omega}^{-1}(\mathbf{Y}-\mathbf{M})^T\mathbf{\Sigma}^{-1}(\mathbf{Y}-\mathbf{M})])$$

where $\mathbf{M}_{n\times p}$ is a location matrix, $\mathbf{\Omega}_{p\times p}$ and $\mathbf{\Sigma}_{n\times n}$ are scale matrices. The normalization constant c is $(2\pi)^{np/2}|\mathbf{\Omega}|^{n/2}|\mathbf{\Sigma}|^{p/2}$. The first and second moments are given as $\mathbf{M} = E[\mathbf{Y}]$, $\mathbf{\Sigma} = E[(\mathbf{Y}-\mathbf{M})(\mathbf{Y}-\mathbf{M})^T]$, and $\mathbf{\Omega} \propto E[(\mathbf{Y}-\mathbf{M})^T(\mathbf{Y}-\mathbf{M})]$. As $p(\mathbf{Y}|\mathbf{M}, \mathbf{\Omega}, \mathbf{\Sigma}) = p(\mathbf{Y}|\mathbf{M}, r\mathbf{\Omega}, \mathbf{\Sigma}/r)$ for any $r \neq 0$, $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ are not separately identifiable.

A matrix-normal random matrix $\mathbf{Y}$ drawn from $p(\mathbf{Y}|\mathbf{M}, \mathbf{\Omega}, \mathbf{\Sigma})$ is related to its vectorized form $\text{vec}(\mathbf{Y})_{np\times 1}$ as follows [11]:

$$\text{vec}(\mathbf{Y}) \sim \mathcal{N}(\text{vec}(\mathbf{M}), \mathbf{\Omega} \otimes \mathbf{\Sigma}) \tag{1}$$

where $\otimes$ represents the Kronecker product. The Kronecker product between $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ is defined as:

$$\mathbf{\Omega} \otimes \mathbf{\Sigma} = \begin{pmatrix} \omega_{11}\mathbf{\Sigma} & \cdots & \omega_{1p}\mathbf{\Sigma} \\ \vdots & \ddots & \vdots \\ \omega_{p1}\mathbf{\Sigma} & \cdots & \omega_{pp}\mathbf{\Sigma} \end{pmatrix}_{np\times np} \tag{2}$$

where ω_{ij} is an (i, j) element of $\mathbf{\Omega}$. For a multivariate normal distribution, its corresponding matrix normal distribution is a special case with the covariance structure in Equation (2). This covariance structure models the dependencies between rows and columns. This additional structure reduces the number of unknown covariance parameters in the multivariate normal distribution. For an (np)-dimensional multivariate normal distribution, the number of unknown covariance parameters approximately drops from $(np)^2$ to n^2+p^2 with this constraint. The matrix normal distribution has been applied in various fields such as dynamic models [12], Bayesian analysis [13], matrix regression [14], tensor data [15], and multi-task learning [16].

III. MULTIVARIATE TEMPORAL SIGNATURE

We have a stream of data $\mathcal{D} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_T]$ where $\mathbf{y}_t$ is a p dimensional measurement vector and its subscript t indicates the time of the measurement taken. We define a measurement block $\mathbf{Y}_t$ with n-time sliding window as follows:

$$\mathbf{Y}_t = [\mathbf{y}_{t-n+1}, \mathbf{y}_{t-n+2}, \cdots, \mathbf{y}_t]^T \tag{3}$$

where its dimension is $n \times p$. From its construction, the rows of $\mathbf{Y}_t$ are temporally correlated, and the columns of $\mathbf{Y}_t$ might have correlated measurements. These dependencies can be modeled using the matrix normal distribution.

Our objective is to capture a symptomatic characteristic prior to a cardiac arrest event. We hypothesize that two latent states govern the measurement block process: risky and normal states. We model the measurement block process as a dynamic mixture of matrix normal distributions. Figure 1 shows the idea of this process. K_t represents the latent state at time t where $K_t \in \{\text{risky}, \text{normal}\}$. Based on the state, the corresponding parameters $\{\mathbf{M}_k, \mathbf{\Omega}_k, \mathbf{\Sigma}_k\}$ are chosen, then $\mathbf{Y}_t$ is drawn from the matrix normal distribution $p(\mathbf{Y}_t|\mathbf{M}_k, \mathbf{\Omega}_k, \mathbf{\Sigma}_k)$.

[1]Y. Park is with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712, USA yubin.park at utexas.edu

[2]J.C. Ho is with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712, USA joyceho at utexas.edu

[3]J. Ghosh is the Schlumberger Centennial Chair Professor with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712, USA ghosh at ece.utexas.edu

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

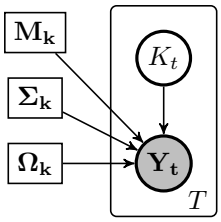

Fig. 1. Graphical model of the mixture of matrix normal distributions

In this paper, we postulate that the risky states appear right before the cardiac arrest events. The normal states are the latent states significantly before the events. For example, a measurement block $\mathbf{Y}_{T-s}$ where $s \gg 0$ is drawn from the normal state distribution. The optimal time window n and the normal state time offset s are obtained through cross-validations (see Section V). Our algorithm is as follows:

1) Estimate $\{\mathbf{M}_k, \mathbf{\Omega}_k, \mathbf{\Sigma}_k\}$ from training data as follows:

$$\hat{\mathbf{M}}_k \approx \text{Avg}_{\text{state}(t)=k} \mathbf{Y}_t \quad (4)$$

$$\hat{\mathbf{\Omega}}_k \approx \text{Avg}_{\text{state}(t)=k} (\mathbf{Y}_t - \hat{\mathbf{M}}_k)^T (\mathbf{Y}_t - \hat{\mathbf{M}}_k) \quad (5)$$

$$\hat{\mathbf{\Sigma}}_k \approx \text{Avg}_{\text{state}(t)=k} (\mathbf{Y}_t - \hat{\mathbf{M}}_k)(\mathbf{Y}_t - \hat{\mathbf{M}}_k)^T \quad (6)$$

2) For a new data block $\mathbf{Y}_{\text{test}}$, calculate the likelihood ratio:

$$\mathcal{L} = \frac{p(\mathbf{Y}_{\text{test}} | \mathbf{M}_{\text{risky}}, \mathbf{\Omega}_{\text{risky}}, \mathbf{\Sigma}_{\text{risky}})}{p(\mathbf{Y}_{\text{test}} | \mathbf{M}_{\text{normal}}, \mathbf{\Omega}_{\text{normal}}, \mathbf{\Sigma}_{\text{normal}})} \quad (7)$$

3) If $\mathcal{L} > \theta_{\text{threshold}}$, label the state of $\mathbf{Y}_{\text{test}}$ as the risky state. Otherwise, label it as the normal state.

The risky state parameters illustrate the multivariate temporal symptomatic characteristics of cardiac arrest events. The mean matrix $\mathbf{M}_{\text{risky}}$ summarizes the overall trajectories of the measurements within n-time window. The temporal covariance matrix $\mathbf{\Sigma}_{\text{risky}}$ describes the individual variations from the global trajectories. The feature dependencies near to the events appear in the feature covariance matrix $\mathbf{\Omega}_{\text{risky}}$. The parameters for the normal state can be similarly interpreted.

IV. EXPERIMENT

A. Data

Our study was conducted on patients at least 18 years of age at time of admission from the MIMIC-II database [17] who had an asystole event. We focused on four common clinical measurements prior to cardiac arrest time: heart rate, respiratory rate, diastolic blood pressure and systolic blood pressure. Data was discretized into 2-hour bins starting when a patient has a least one observation per measurement. Additionally, we required at least 40 discrete time slices ($\sim$ 3 days) for each patient to ensure sufficient number of data points.

We identified 421 cardiac arrest patients with asystole from 27,542 adult hospital admissions. 171 of the 421 cardiac patients met the minimum data requirements. On average, patients had 85 time slices with a standard deviation of 20. We assumed unobserved measurements were missing at random and used the zero-order hold [18], maintaining the last observed value. Figure 2 shows the blood pressure measurements for three patients starting two days prior to the cardiac arrest event.

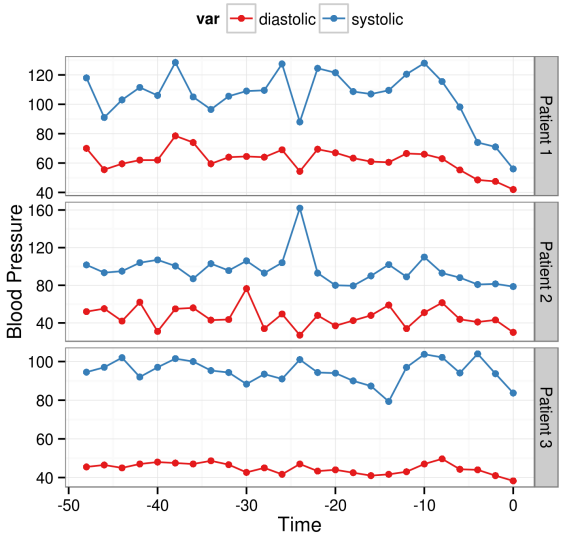

Fig. 2. Examples of blood pressure measurements (mmHg) prior to cardiac arrest.

B. Evaluation Measure

We evaluated the performance of the matrix normal model against logistic regression models with lagged features. Leave-one-out cross validation was used; each model is trained on the remaining 170 patients, and tested on the one hold-out patient. With 171 leave-one-out scores and their true labels, we randomly sampled 171 patients with replacement, and measured the Area Under Receiver Operating Characteristic (AUROC). This procedure is iterated over 100 times; we estimated AUROC using 100 bootstrapped bags.

V. RESULTS

A. Optimal Time Window

We examined four different time windows to characterize the pre-event condition: 2 hr, 4 hr, 6 hr, and 8 hr. Figure 3 shows the differentiability of the risky state from the normal state for each time window. The area under the curve (AUROC) expresses the differentiability between the two states; the higher AUROC, the more differentiable. The two hours window, which is basically one dimensional vector, does not utilize any temporal information, thus it shows the worst performance. As can be seen, the six hours window, which consists of 3 measurement slices, exhibits the best performance (the highest AUROC value) among these four candidates. The eight hours window performs worse than the six hours window. The eight hours may contain more noisy and extra information than the six hours window. The cardiac arrest pre-symptoms are best captured and distinguishable with the six hours data before the event.

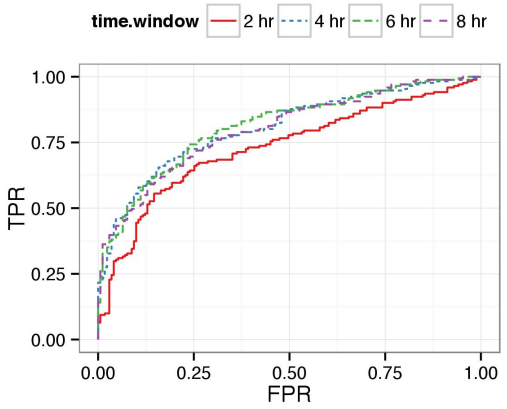

Fig. 3. Classifying the risky state with different time windows.

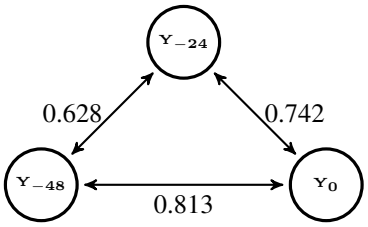

Fig. 4. AUROC scores when classifying two different measurement blocks. $\mathbf{Y}_{-48}$ and $\mathbf{Y}_{-24}$ denote the measurements from two and one days before, respectively.

B. Representative Normal State

In the matrix normal model, the normal states are the latent states significantly before the events, and its offset s needs to be derived from the data. A patient in an Intensive Care Unit (ICU) is mostly with seriously injuries and illness. Any measurement taken in an ICU cannot be the representative normal state.

We estimated the matrix normal parameters from three sets of data: 2-day before, 1-day before, and the day of the event. Figure 4 shows their pair-wise AUROC scores. We picked two datasets, and assigned one of them as the positive class and the other one as the negative class. The pair-wise AUROC scores are measured from the hold-out test data. As can be seen, the data blocks from the past 2-days are the most noticeable and distinguishable from the risky state data blocks. Thus, we set the D-2 data blocks to be the representative normal state data.

C. Multivariate Temporal Signature

Figure 5, 6, and 7 show the estimated matrix normal parameters: $\mathbf{M}_k$, $\mathbf{\Omega}_k$, and $\mathbf{\Sigma}_k$. In Figure 5, we observe that the four measurements in the normal state remain constant over time, while these measurements exhibit descending slopes in the risky state. Noticeably, the heart rate at the risky state starts from 90.5 bpm, which is higher than the normal state. The feature covariance matrices shown in Figure 6 appear approximately the same each other. The dependencies between features are not affected by cardiac arrest events. As can be seen in Figure 7, the normal and

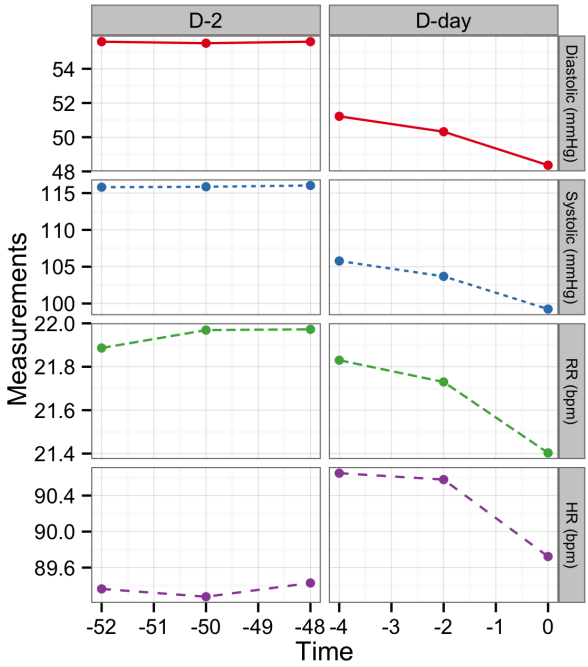

Fig. 5. Mean matrices: $\mathbf{M}_{\text{normal}}$ and $\mathbf{M}_{\text{risky}}$.

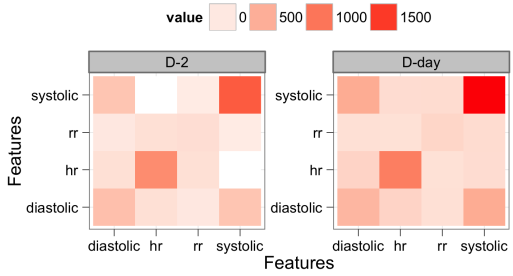

Fig. 6. Feature covariance matrices: $\mathbf{\Omega}_{\text{normal}}$ and $\mathbf{\Omega}_{\text{risky}}$.

risky temporal covariance matrices are significantly different. The measurements in the risky state typically fluctuate with a large amplitude, resulting in this large covariance matrix.

D. Comparison with Logistic models

Logistic regression models can incorporate temporal information using lagged features. We built three types of logistic models as follows:

- Logit.2hr: $\text{logit}(E[\text{state}|\mathbf{y}_t]) = \boldsymbol{\beta}_2^T \mathbf{y}_t$
- Logit.4hr: $\text{logit}(E[\text{state}|\mathbf{y}_{t:t-1}]) = \boldsymbol{\beta}_4^T [\mathbf{y}_t; \mathbf{y}_{t-1}]$
- Logit.6hr: $\text{logit}(E[\text{state}|\mathbf{y}_{t:t-2}]) = \boldsymbol{\beta}_6^T [\mathbf{y}_t; \mathbf{y}_{t-1}; \mathbf{y}_{t-2}]$

Figure 8 shows the leave-one-out AUROC scores for 1) these three logistic models, 2) the matrix normal model, and 3) an ensemble of the matrix normal model and the Logit.4hr model. Logit.6hr is over-fitting to the training data, resulting in the lower AUROC scores than the scores from Logit.4hr. The matrix normal model well captures the dynamic signature of the cardiac arrest pre-conditions. As can be seen, the matrix normal model outperforms these three logistic models. The ROC curves of these three models,

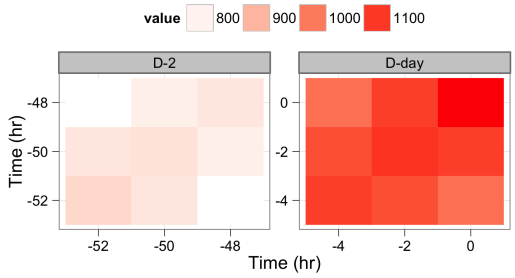

Fig. 7. Temporal covariance matrices: $\mathbf{\Sigma}_{\text{normal}}$ and $\mathbf{\Sigma}_{\text{risky}}$.

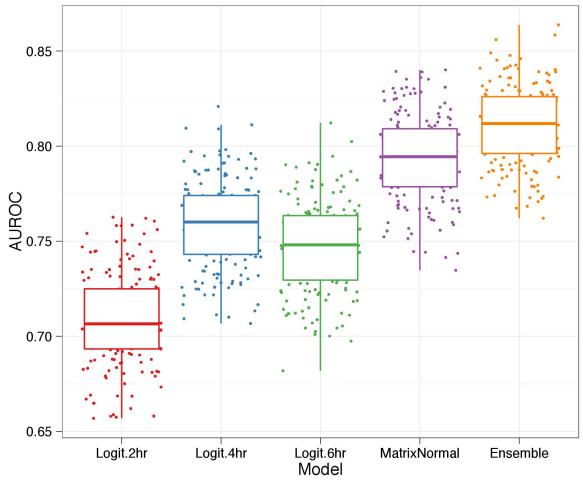

Fig. 8. AUROC scores from five different models. The matrix normal model outperforms the family of logistic regression models with lagged features.

which are not shown due to the page limit, show that the logistic models and the matrix normal model characterize the cardiac arrest from different angles. As a result, the ensemble of these two models improves the AUROC scores from the both models.

VI. DISCUSSION

In this paper, the multivariate temporal signature of cardiac arrest is characterized using the matrix normal model. The presented model explicitly models the dependencies between temporal and feature domains. The estimated parameter matrices intuitively illustrate the measurement process and its underlying states.

This matrix normal model can be enhanced from several aspects. The optimal time window and the representative normal state can be obtained incorporating physicians' domain knowledge. The latent state of the signal blocks can be learned using clustering algorithms. With modeling the possible interventions in ICU's, this model can be further modified to predict the actual event time. Finally, this model can be extended to deal with other types of diseases and conditions, providing the temporal characteristic taxonomy of various diseases.

ACKNOWLEDGMENT

This work is supported by a grant from NSF IIS-1017614 and Janey Slaughter Briscoe Centennial Fellowship from IC^2 Institute.

REFERENCES

[1] C. Sandroni, J. Nolan, F. Cavallaro, and M. Antonelli, "In-hospital cardiac arrest: incidence, prognosis and possible measures to improve survival." *Intensive Care Medicine*, vol. 33, no. 2, pp. 237–245, Feb. 2007.

[2] M. M. Churpek, T. C. Yuen, M. T. Huber, S. Y. Park, J. B. Hall, and D. P. Edelson, "Predicting Cardiac Arrest on the Wards: A Nested Case-Control Study," *Chest*, vol. 141, no. 5, pp. 1170–1176, May 2012.

[3] C. Sandroni, G. Ferro, S. Santangelo, F. Tortora, L. Mistura, F. Cavallaro, A. Caricato, and M. Antonelli, "In-hospital cardiac arrest: survival depends mainly on the effectiveness of the emergency response." *Resuscitation*, vol. 62, no. 3, pp. 291–297, Sep. 2004.

[4] T. J. T. Hodgetts, G. G. Kenward, I. G. I. Vlachonikolis, S. S. Payne, and N. N. Castle, "The identification of risk factors for cardiac arrest and formulation of activation criteria to alert a medical emergency team." *Resuscitation*, vol. 54, no. 2, pp. 125–131, Aug. 2002.

[5] J. McBride, D. Knight, J. Piper, and G. B. Smith, "Long-term effect of introducing an early warning score on respiratory rate charting on general wards," *Resuscitation*, vol. 65, no. 1, pp. 41–44, Apr. 2005.

[6] C. E. Kennedy and J. P. Turley, "Time series analysis as input for clinical predictive modeling: modeling cardiac arrest in a pediatric ICU." *Theoretical biology & medical modelling*, vol. 8, p. 40, 2011.

[7] I. Batal, D. Fradkin, J. Harrison, F. Moerchen, and M. Hauskrecht, "Mining recent temporal patterns for event detection in multivariate time series data," in *KDD '12: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 280–288.

[8] F. Wang, N. Lee, J. Hu, J. Sun, and S. Ebadollahi, "Towards heterogeneous temporal clinical event pattern discovery: a convolutional approach," in *KDD '12: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012.

[9] C. jin Kim, "Dynamic linear models with Markov-switching," *Journal of Econometrics*, vol. 60, pp. 1–22, 1994.

[10] W. J. Freeman, "Simulation of chaotic EEG patterns with a dynamic model of the olfactory system," *Biological Cybernetics*, vol. 56, pp. 139–150, 1987.

[11] A. P. Dawid, "Some matrix-variate distribution theory: Notational considerations and a Bayesian application," *Biometrika*, vol. 68, pp. 265–274, 1981.

[12] C. M. Carvalho and M. West, "Dynamic Matrix-Variate Graphical Models," *Byesian Analysis*, vol. 2, no. 1, pp. 69–98, 2007.

[13] H. Wang and M. West, "Bayesian analysis of matrix normal graphical models," *Biometrika*, vol. 96, no. 4, pp. 821–834, 2009.

[14] O. Stegle, C. Lippert, J. Mooij, and N. Lawrence, "Efficient inference in matrix-variate Gaussian models with iid observation noise," in *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, 2012.

[15] C. Viroli, "Model based clustering for three-way data structures," *Bayesian Analysis*, vol. 6, no. 4, pp. 573–602, 2011.

[16] Y. Zhang and J. Schneider, "Learning Multiple Tasks with a Sparse Matrix-Normal Penalty," in *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, 2010.

[17] M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L.-W. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, and R. G. Mark, "Multiparameter Intelligent Monitoring in Intensive Care II: A public-access intensive care unit database," *Critical Care Medicine*, vol. 39, no. 5, pp. 952–960, May 2011.

[18] A. S. Fialho, F. Cismondi, S. M. Vieira, J. M. C. Sousa, S. R. Reti, and M. D. Howell, "Predicting outcomes of septic shock patients using feature selection based on soft computing techniques," in *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Applications*. Springer Berlin Heidelberg, 2010, pp. 65–74.