Journal

IEEE Journal of Biomedical and Health Informatics, 19(3)

ISSN

2168-2194

Authors

Lehman, Li-wei H
Adams, Ryan P
Mayaud, Louis
et al.

Publication Date

2015-05-01

DOI

10.1109/jbhi.2014.2330827

Peer reviewed

A Physiological Time Series Dynamics-Based Approach to Patient Monitoring and Outcome Prediction

Li-Wei H. Lehman, Ryan P. Adams, Louis Mayaud, George B. Moody, Atul Malhotra, Roger G. Mark, and Shamim Nemati

***Abstract*—Cardiovascular variables such as heart rate (HR) and blood pressure (BP) are regulated by an underlying control system, and therefore, the time series of these vital signs exhibit rich dynamical patterns of interaction in response to external perturbations (e.g., drug administration), as well as pathological states (e.g., onset of sepsis and hypotension). A question of interest is whether "similar" dynamical patterns can be identified across a heterogeneous patient cohort, and be used for prognosis of patients' health and progress. In this paper, we used a switching vector autoregressive framework to systematically learn and identify a collection of vital sign time series dynamics, which are possibly recurrent within the same patient and may be shared across the entire cohort. We show that these dynamical behaviors can be used to characterize the physiological "state" of a patient. We validate our technique using simulated time series of the cardiovascular system, and human recordings of HR and BP time series from an orthostatic stress study with known postural states. Using the HR and BP dynamics of an intensive care unit (ICU) cohort of over 450 patients from the MIMIC II database, we demonstrate that the discovered cardiovascular dynamics are significantly associated with hospital mortality (dynamic modes 3 and 9, $p = 0.001$, $p = 0.006$ from logistic regression after adjusting for the APACHE scores). Combining the dynamics of BP time series and SAPS-I or APACHE-III provided a more accurate assessment of patient survival/mortality in the hospital than using SAPS-I and APACHE-III alone ($p = 0.005$ and $p = 0.045$). Our results suggest that the discovered dynamics of vital sign time series may contain additional prognostic value beyond that of the baseline acuity measures, and can potentially be used as an independent predictor of outcomes in the ICU.**

***Index Terms*—Intensive care unit, physiological control systems, switching linear dynamical systems.**

Manuscript received October 31, 2013; revised March 30, 2014; accepted June 1, 2014. This work was supported by the National Institutes of Health Grants R01-EB001659 and R01GM104987 from the National Institute of Biomedical Imaging and Bioengineering, the James S. McDonnell Foundation Postdoctoral Grant, and the DARPA Young Faculty Award N66001-12-1-4219 Grant.

L. H. Lehman, G. B. Moody, and R. G. Mark are with the Massachusetts Institute of Technology, Cambridge, MA 02142 USA (e-mail: lilehman@mit.edu; george@mit.edu; rgmark@mit.edu).

R. P. Adams and S. Nemati are with the Harvard School of Engineering and Applied Sciences, Cambridge, MA 02138 USA (e-mail: rpa@seas.harvard.edu; shamim@seas.harvard.edu).

L. Mayaud is with the Institute of Biomedical Engineering, University of Oxford, Oxford, OX1 3PJ, UK, and also with Hopital Raymond Poincare, Garches 92380, France (e-mail: louis.mayaud@eng.ox.ac.uk).

A. Malhotra is with the Brigham and Women's Hospital & Harvard Medical School, Boston, MA 02115 USA, and also with the UC San Diego Division of Pulmonary & Critical Care Medicine, La Jolla, CA 92037 USA (e-mail: amalhotra1@partners.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2014.2330827

I. Introduction

Modern clinical data acquisition systems are capable of continuously monitoring and storing measurements of patient vital signs, such as heart rate (HR) and blood pressure (BP), over multiple days of hospitalization [1]. Despite this continuous feed of data, commonly used acuity scores, such as APACHE and SAPS [2]–[5], are based on snap-shot values of these vital signs, typically the worst values during a 24-h period. However, physiologic systems generate complex dynamics in their output signals that reflect the state of the underlying control systems [6]–[8]. The objective of this study is to consider an approach to the analysis of critical care bedside monitoring that is based on the dynamical behaviors of vital sign time series.

The time series of vital signs (e.g., HR, BP) are multidimensional, high resolution (from once a second to once a minute), highly coupled due to presence of physiological feedback loops within the body [8], and remarkably nonstationary as a result of internally and externally induced changes in the state of the underlying control systems. For instance, time series of BP can exhibit oscillations on the order of seconds (e.g., due to the variations in sympathovagal balance), to minutes (e.g., as a consequence of fever, blood loss, or behavioral factors), to hours (e.g., due to humoral variations, sleep-wake cycle, or circadian effects) [9], [10]. A growing body of the literature is pointing to the clinical utility of vital signs time series dynamics to inform prognosis [11]–[17], and to provide early predictors of potentially life-threatening conditions in the intensive care unit (ICU) [18].

Techniques for modeling and analysis of cardiovascular and respiratory time series can be broadly classified into linear mechanistic models [19], [20] and nonlinear descriptive indices [6], [7], [21]. The linear techniques commonly used (often based on variants of autoregressive modeling) have the advantage of revealing the individual relationships among the observed variables (e.g., the noninvasive measures of baroreflex gain describes the relationship between HR and BP, excluding the possible influence of respiration). On the other hand, nonlinear indices of complexity are capable of capturing a richer set of dynamical behaviors, with less emphasis on physiological interpretability in terms of specific underlying mechanisms.

In this paper, we assume that although the underlying dynamical system may be nonlinear and nonstationary, and the stochastic noise components can be non-Gaussian, the dynamics can be approximated by a mixture of linear dynamical systems. Each such linear "dynamic" (or mode) is a time-dependent rule that describes how the future state of the system evolves from its

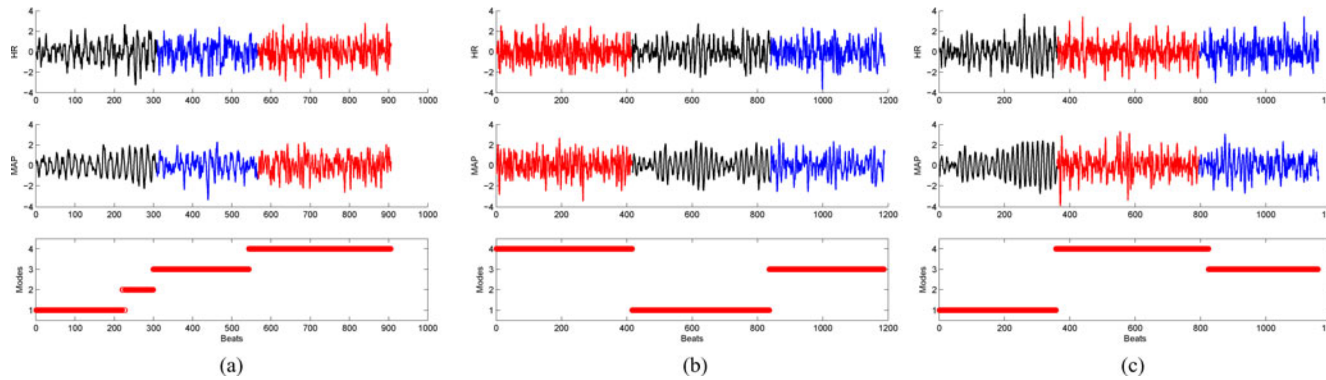


Fig. 1. Simulation study of the cardiovascular system. Three examples (out of the ten simulated time series) of HR and BP (after filtering) are shown in panels a, b, and c. In each case, the actual dynamics are color coded. The horizontal red lines show the inferred segmentation. The algorithm consistently assigned modes 4 and 3 to the dynamics color coded as red and blue, respectively, across all the simulated time series. The black dynamics are represented by modes 1 and 2. (a) Simulated subject 1, (b) Simulated subject 2, and (c) Simulated subject 3.

current state, centered around a given system equilibrium point. Therefore, an ideal algorithm would be able to identify time series segments that follow a "similar" dynamic, and would switch to a different mode upon a change in the state of the underlying system.

To formalize these objectives, we employed a switching vector autoregressive (SVAR) framework [22], [23]. Given a collection of time series from a cohort, the proposed SVAR framework allows for simultaneous learning of the underlying dynamic behaviors or modes, and segmentation of the time series in terms of the most likely dynamic describing the time series evolution at any given point in time. The proposed framework enables characterization of patients in terms of the dynamical modes (e.g., the average time spent within the different modes), and can potentially be used to capture changes in the underlying cardiovascular control systems of human subjects in response to internal (such as onset of infection) and external perturbations (such as postural changes). Furthermore, we hypothesize that when applied to vital sign time series of patients in a critical care setting, the proposed technique can be used to discover dynamical modes with prognostic values for predicting clinical outcomes of interests.

A preliminary version of this study was presented at the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '12) [14]. Here, we extend on our previous work to include a series of validation studies, and a more comprehensive assessment of the utility of the time series dynamics within the ICU.

The rest of this paper is organized as follows. We validated the proposed technique using HR and BP time series from a simulation dataset, and a human laboratory study of subjects undergoing a tilt-table test, where the timing of the occurrence of the different dynamics and the sharing of the dynamics across multiple time series/subjects were known *a priori*. To test the prognostic value of the discovered vital sign dynamics, we applied the proposed approach to the HR and BP dynamics of an ICU cohort from the MIMIC II database [1] during the first 24 h of their ICU stays, and tested whether cardiovascular dynamics during the first 24 h of ICU admission are predictive of survival and mortality after adjusting for the existing acuity scores, such as SAPS-I and APACHE.

II. Materials and Methods

This section describes the utilized datasets, as well as the proposed technique for discovery of shared dynamics among patients, and assessment of risks and outcomes.

A. Datasets

1) Cardiovascular Simulation: We simulated a cardiovascular control system with bivariate time series of HR and BP. The model is based on a delay recruitment model of HR and BP regulation, as described in Fowler and McGuinness [24], and McSharry *et al.* [25]. The model included a coupled system of nonlinear delayed differential equations, controlling HR and BP, with respiration as an exogenous input. We simulated ten different multivariate time series of HR and mean arterial BP, each including three different dynamics that become dominant in a random order and last for a variable length of time. The three dynamics (color-coded as red, blue, and black, respectively, in Fig. 1) approximate aging-related autonomic changes; a progressive reduction in parasympathetic gain (from 0.40 to 0.13 to 0.07 in normalized units; see [24]) and an increase in sympathetic delay (from 3 to 5 s). To be consistent, we used the same preprocessing step as the tilt-table experiment to remove the steady-state baseline and any oscillation in the time series slower than 100 beats/cycle (see below for details).

2) Tilt-Table Experiment: Time series of HR and BP were acquired from ten healthy subjects (five males, five females) undergoing a tilt-table test of orthostatic tolerance [26], [27]. The mean age was 28.7 $\pm$ 1.2 years. The details of the protocol are described by Heldt in [27]. Briefly, subjects were placed in a supine position. Tilting was performed from the horizontal position to the vertical position and back to supine. The study was approved by MIT's Committee on the Use of Humans as Experimental Subjects and the Advisory Board of the MIT-MGH General Clinical Research Center [27]. Volunteers gave written, informed consent prior to participation

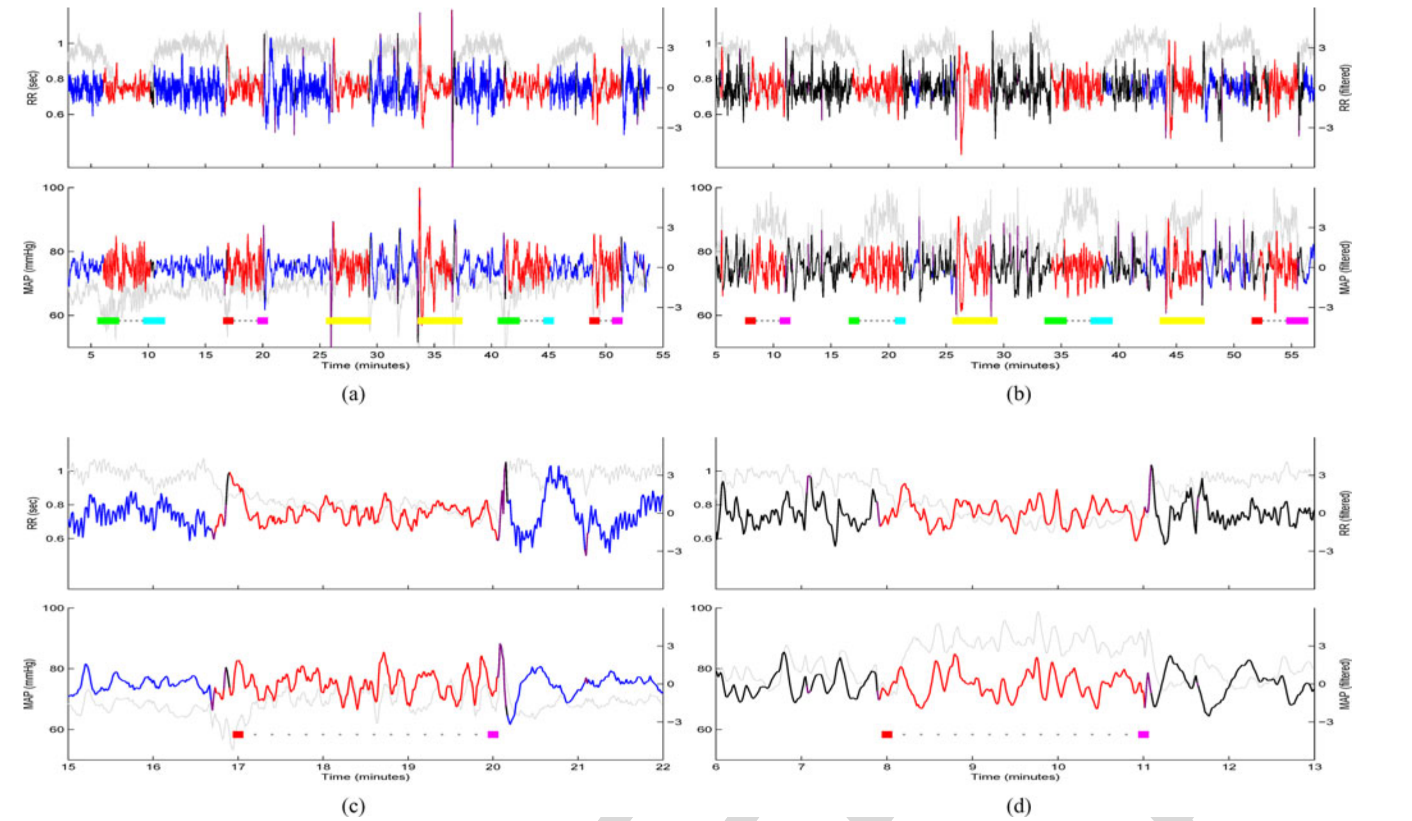


Fig. 2. Tilt-table study modeled using four dynamic modes—1 (Blue), 2 (Red), 3 (Black), and 4 (Purple). Two examples out of the ten recordings of HR and BP from the tilt-table experiment are shown in Panels a and b. Panels c and d show a zoomed in 7-min recording of HR and BP, while the subjects transition to/from supine to nonsupine positions after a fast tilt procedure. Actual values are in gray (*Y*-axis on left) and filtered values (*Y*-axis on right) are color coded based on the inferred dynamical modes. Note that Subjects 1 and 2 shared the same inferred nonsupine dynamics (in red); the algorithm consistently assigns the red mode to the nonsupine position for both subjects. The supine position for Subjects 1 and 2 are captured by the modes in blue and black, respectively. The purple mode seems to capture the high-frequency noise components of the time series. In each case, annotations for the actual tilt procedures performed are plotted as horizontal bars on the bottom of each figure and are color coded (green to cyan: slow tilt up and down to supine; red to pink: rapid tilt up and down to supine; yellow: stand up and back to supine). (a) Tilt Subject 1, (b) Tilt Subject 2, (c) Tilt Subject 1 (zoomed in), and (d) Tilt Subject 2 (zoomed in).

in the study. Since we were interested in the dynamics of interaction between HR and BP in the frequency range pertinent to sympathetic and parasympathetic regulation [28], time series of HR and BP were high-pass filtered to remove the steady-state baseline and any oscillation in the time series slower than 100 beats/cycle. This filtering was done using a seventh-order Butterworth digital filter with a cutoff frequency of 0.01 cycles/beat. Example time series from before and after filtering are shown in Fig. 2.

3) MIMIC II Dataset: The MIMIC II database [1], publicly available via PhysioNet [29], includes clinical (laboratory values, IV medications, etc.) and physiological data (HR, BP, oxygen saturation, etc.) collected from the bedside monitors (Component Monitoring System Intellivue MP-70; Philips Healthcare, Andover, MA, USA) in ICUs of the Beth Israel Deaconess Medical Center (BIDMC) in Boston. The MIMIC II waveform database (version 2) includes approximately 4000 records of high-resolution physiological waveforms of adult ICU patients with associated minute-by-minute (averages of the calculated numerics during the previous minute) vital sign trends. Data collection for the MIMIC II database was approved by the Institutional Review Boards of BIDMC and the Massachusetts Institute of Technology (Cambridge, MA, USA). Individual patient consent was waived because the study did not impact clinical care and protected health information was deidentified.

This study includes adult patients from the MIMIC II waveform database with at least 8 h of continuous minute-by-minute HR and invasive arterial BP trends during the first 24 h in the ICU. Patients with more than 15% of missing or invalid samples (i.e., outside physiologically plausible bounds of 20 to 200 mmHg for mean pressures) were excluded from this study, as were patients with missing SAPS I and APACHE scores. The dataset contains over 9000 h of minute-by-minute HR and invasive mean arterial BP measurements (over 20 h per patient on average) from 453 adult patients collected during the first 24 h in the ICU. HR and BP time series were detrended. Gaussian noise was used to fill in the missing or invalid values. The median age of this cohort was 69 with an interquartile range of (57, 79). 59% of the patients were male. Approximately 15% (67 out of 453) of patients in this cohort died in the hospital; 28-day mortality of this cohort was approximately 19% (85 out of 453). Distributions of the 453 patients in care units are 21% coronary care unit (CCU), 42% cardiac surgery recovery unit (CSRU), 26% medical intensive care unit (MICU), and 12% surgical intensive care unit.

B. SVAR Modeling of Cohort Time Series

Our approach to discovery of shared dynamics among patients is based on the SVAR model [22]. For the nth patient ($n = 1 \dots N$), let $y_t^{(n)}$ be a $M \times 1$ vector of observed values of the vital signs at time t ($t = 1 \dots T^{(n)}$). We assume that there exists a library of K possible dynamics or *modes*; a set of multivariate autoregressive model coefficient matrices $\{A_p^{(k)}\}_{k=1}^{K}$ of size $M \times M$, with maximal time lag $p = 1 \dots P$, and the corresponding noise covariances $\{Q^{(k)}\}_{k=1}^{K}$. Let s_t be a switching variable, indicating the active dynamic mode at time t, and evolving according to a Markovian dynamic with initial distribution $\pi^{(n)}$ and a $K \times K$ transition matrix Z. Following these definitions, an SVAR model for the nth patient is defined as

$$y_t^{(n)} = \sum_{p=1}^{P} A_p^{(s_t^{(n)})} y_{t-p}^{(n)} + w^{(s_t^{(n)})} \tag{1}$$

where the fluctuation term $w^{(s_t^{(n)})}$ is assumed Gaussian distributed with covariance $Q^{(s_t^{(n)})}$. A collection of related time series can be modeled as switching between these dynamic behaviors which describe a locally coherent linear model that persists over a segment of time. However, in practice, we neither know the set of switching variables (i.e., segmentation of the time series) nor the modes. In this study, we perform expectation–maximization (EM) to find the maximum-likelihood set of model parameters, as well as a factored estimate of the posterior distribution over the latent switching variables. A comprehensive treatment of the EM algorithm for SVAR is presented by Murphy (1998) [22]. Briefly, EM is a two-pass iterative algorithm: 1) in the expectation (E) step, we obtain the expected values of the latent switching variables $\{s_t^{(n)}\}_{t=1}^{T}$ using a forward–backward algorithm [22], and 2) in the maximization (M) step, we update all the model parameters $\{A_p^{(k)}\}$, $\{Q^{(k)}\}$, the Markov dynamics Z, and the initial conditions $\pi^{(n)}$ that maximize the expected complete data log likelihood. In our implementation of the EM algorithm, we achieve *shared* dynamics by pooling together all subjects' inferred variables in the M step. Iteration through several steps of the EM algorithm results in learning a set of K shared modes and a global transition matrix Z for all the patients.

For the simulated and the tilt datasets, we modeled the beat-by-beat HR/BP time series as a switching AR(5) process to model most of the parasympathetic responses and at least some of the sympathetic effects, without introducing an unduly complex model. Minute-by-minute BP time series from MIMIC II were modeled as a switching AR(3) process to capture a real oscillation and a possible trend per mode. The number of dynamic modes ($K = 20$) was determined using the Bayesian information criterion (BIC) [30]. Briefly, we computed the BIC scores from switching-VAR models using 5 to 30 modes. Results presented were based on the model with the minimum BIC scores (20 modes).

1) Parallel Computation for Scalable Learning: One of the advantages of the proposed technique is its scalability to hundreds or thousands of patients, due to the parallel implementation of the inference step of the SVAR learning algorithm via EM [22]. This parallelization strategy is effective since the majority of the computational cost of the SVAR training is in running the forward–backward algorithm, which can be done in parallel for each patient time series. We used MATLAB's parallel computation toolbox in association with 120 nodes on our computer cluster to perform a tenfold cross-validated study (12 cores per fold). Ten SVAR models were learned on the training set of each of the folds, followed by mapping the corresponding mode proportions to outcomes (e.g., hospital mortality) using logistic regression. Next, mode assignments of time series in the test set of each fold were inferred based on the modes learned from the corresponding training set (by running only the inference), and the regression weights from the training fold were used to predict outcomes.

C. Evaluation Methods and Statistical Analysis

Let us define a *mode proportion* $\mathrm{MP}_k^{(n)}$ as the proportion of time the nth patient spends within the kth mode. Given the maximum expected log-likelihood estimates of the switching variables s_t from the EM algorithm, we have

$$\mathrm{MP}_k^{(n)} = \frac{1}{T^{(n)}} \sum_{t=1}^{T^{(n)}} \mathrm{Prob}(s_t^{(n)} = k). \tag{2}$$

For classification and prediction purposes, we characterize each time series with its corresponding mode proportion (a $1 \times K$ feature-vector), and use a logistic regression classifier to make predictions about the outcome variables of interest. For illustration of the algorithm's segmentation performance, each time series sample is assigned to the dynamic mode with the maximum posterior probability.

1) Time Series Classification and Outcome Prediction: For the simulated and the tilt-table experiment, we used the mode proportions within each segment (e.g., supine versus nonsupine) as inputs to a logistic regression classifier, and report the classification performance in discriminating between 1) the three different dynamics (corresponding to different aging-related autonomic changes) in the simulated dataset, and 2) two different postural positions (supine versus nonsupine) in the tilt dataset.

To assess the predictive power of the dynamical modes, we performed a tenfold cross-validation study. Ten SVAR models were learned on the training set of each of the folds, followed by mapping the corresponding mode proportions to outcomes (e.g., hospital mortality) using logistic regression. Next, mode assignments of time series in the test set of each fold was inferred based on the modes learned from the corresponding training set (by running only the inference or the E-step), and the regression weights from the training fold was used to predict outcomes. We compared the mortality prediction performance of our approach using the mode proportion from the top ten most common dynamic modes with the existing acuity metrics, SAPS I [2], APACHE III [4], and APACHE IV [5]. Comparison of AUCs was based on the method described in [31].

2) MIMIC Association Analysis: We used univariate and multivariate logistic regressions to examine the associations between dynamic mode proportions and hospital mortality. We

TABLE I
PERFORMANCE OF MORTALITY PREDICTORS

	Hosp. Mortality (AUC)	28-Days Mortality (AUC)
HR_{dyn}	0.59 (0.54, 0.68)	0.61 (0.51, 0.67)
BP_{dyn}/HR_{dyn}	0.64 (0.61, 0.71)	0.65 (0.64, 0.68)
BP_{dyn}	**0.70 (0.67, 0.77)**	**0.66 (0.61, 0.73)**
SAPS I	0.65 (0.59, 0.71)	0.64 (0.56, 0.70)
BP_{dyn} + SAPS I	**0.77 (0.69, 0.82)**	**0.71 (0.69, 0.79)**
APACHE III	0.80 (0.70, 0.84)	0.79 (0.65, 0.84)
BP_{dyn} + APACHE III	**0.84 (0.79, 0.88)**	**0.79 (0.76, 0.86)**
APACHE IV	0.82 (0.77, 0.85)	0.83 (0.74, 0.86)
BP_{dyn} + APACHE IV	**0.85 (0.80, 0.87)**	**0.82 (0.81, 0.88)**

built a separate multivariate logistic regression model for each of the discovered dynamic modes, with the mode proportion as the primary predictive variable, and APACHE IV as a covariate. For each mode, we reported its p value, odds ratio (OR, with 95% confidence interval), and adjusted OR (after including APACHE IV as a covariate). The Hosmer–Lemeshow p values (HL p values) were reported to assess the model fit. The odds ratios were per 10% increase in the mode proportion. Two-sided p values less than 0.05 were considered statistically significant. The analysis was performed to quantify the mortality risk associated with each dynamic mode; modes with significant ($p < 0.05$) associations with mortality were established as either *low-risk* (OR < 1), or *high-risk* (OR > 1) dynamics depending on their odds ratios. Dynamic modes without statistically significant associations with mortality were *neutral* modes. The test of statistical significance was based on p-values after correcting for the false discovery rate (FDR) using the technique described in [32].

III. RESULTS

A. Simulated Study

Fig. 1 shows two examples of simulated time series with the inferred segmentation. In all ten simulated cases, the algorithm was able to divide each time series into distinct segments corresponding to different underlying actual dynamics. The sharing of the dynamics is consistent across the different time series. Using the mode proportion from each segment for multilabel classification, the algorithm achieved classification accuracy of 100%.

B. Tilt-Table Experiment

Fig. 2 shows the segmentation results for two subjects. Note that the two subjects shared the same inferred nonsupine dynamics (in red); the algorithm consistently assigns the red mode to the nonsupine position for both subjects. The application of logistic regression with tenfold cross-validation yielded a median AUC of 1.00 with an interquartile range of (0.98, 1.00).

C. MIMIC II Database

1) Mortality Prediction: Table I evaluates the prognostic power of HR and BP dynamic features (HR_{dyn} and BP_{dyn}). SAPS I, APACHE III, and APACHE IV are used as the baselines. Median AUCs (from tenfold cross validation) and the interquartile range are shown. Note that the BP dynamics outperformed both the HR and HR and BP combined dynamic features. Subsequent analyses focus on the predictive power of the BP dynamics in comparison to the baseline. For each baseline, we show the performance from the baseline alone, and the combined approach (combining BP dynamics and the baseline).

The application of tenfold cross-validation demonstrated that dynamic features from BP alone achieved a median AUC of 0.70, comparable to 0.65 from SAPS I. In comparison, using standard deviation of the mean arterial BP resulted in a median AUC (IQR) of 0.55 (0.43, 0.63).

Combining dynamic BP features with SAPS I resulted in an improved prediction power both in hospital mortality prediction ($p = 0.005$) and 28-day mortality prediction ($p = 0.002$). Combining dynamic features with APACHE III significantly out-performed APACHE III alone ($p = 0.045$) with an improvement in median AUC from 0.80 to 0.84 in hospital mortality prediction. These results indicate that the dynamic features from vital signs contain complementary information to the SAPS I and APACHE III scores.

State-of-the-art risk score APACHE IV achieved better prediction performance than the BP dynamic features alone ($p = 0.008$). Adding BP dynamics to APACHE IV yielded a slight performance improvement from a median AUC of 0.82 to 0.85, however, the performance gain was not statistically significant.

2) Association Analysis: Table II presents logistical regression analyses to test the associations between the proportion of time patients spent in each of the top ten most common BP dynamics and hospital mortality. See Fig. 3 for illustrations of these dynamic modes. Dynamic modes were numbered based on their prevalence across the entire cohort (i.e., mode 1 is the most common dynamic mode). Our results indicate that six of the modes had significant associations (after FDR correction) with hospital mortality. Specifically, two dynamic modes (modes 3 and 5) were significant "high-risk" modes ($p < 0.001$, $p < 0.001$) in which increased proportions of time in these modes were associated with higher hospital mortality with odds ratios 1.81 (1.41, 2.32), 1.36 (1.15, 1.61) respectively.

Dynamic modes 1, 9, 7, and 2 were "low-risk" modes in which increasing proportions of time in these modes were significantly associated with a decreased risk of hospital mortality, with odds ratios less than one. Table II lists the AR coefficients and covariances of the two high-risk and four low-risk dynamic modes, as well as their respective associations with hospital mortality. Note that the high-risk modes appear to correspond to less variability in their dynamics.

For the multivariate analysis (see the right panel in Table II), each row is a separate multivariate model, in which the mode proportion for a given target mode is the primary predictive variable, and APACHE IV is added as a control variable in the multivariate model. Results from multivariate logistic regression indicate that two of the modes (modes 3 and 9) remain significant predictors of patients' outcome even after adjustment for APACHE IV scores ($p = 0.001$, $p = 0.006$), indicating that the proportion of time patients spent in these two dynamic modes

TABLE II
ASSOCIATIONS OF BP DYNAMIC MODES AND HOSPITAL MORTALITY

Mode	AR Coef	Cov.	P-Val	OR(95%CI)	Adjusted P-Val	Adjusted OR(95%CI)	HL PVAL
3	(0.66, 0.22, 0.12)	0.58	**<0.001**	1.81 (1.41, 2.32)	**0.001**	1.60 (1.21 2.11)	0.64
5	(1.00, 0.00, −0.00)	0.22	**<0.001**	1.36 (1.15, 1.61)	0.426	1.08 (0.89 1.32)	0.16
1	(0.66, 0.16, 0.17)	2.69	**0.002**	0.59 (0.42, 0.82)	0.489	0.88 (0.62 1.26)	0.54
9	(1.50, −0.65, 0.06)	7.26	**0.002**	0.25 (0.10, 0.62)	**0.006**	0.26 (0.10 0.68)	0.80
7	(1.00, −0.01, −0.00)	3.46	**0.003**	0.30 (0.13, 0.67)	0.124	0.54 (0.25 1.18)	0.58
2	(0.79, 0.05, 0.12)	8.81	**0.005**	0.65 (0.48, 0.88)	0.265	0.84 (0.62 1.14)	0.69
10	(1.05, −0.01, −0.02)	0.71	0.032	2.95 (1.10, 7.94)	0.791	1.18 (0.36 3.88)	0.07
8	(0.44, 0.30, 0.24)	1.27	0.373	1.18 (0.82, 1.69)	0.318	1.22 (0.82 1.82)	0.22
4	(0.96, −0.01, 0.04)	1.31	0.417	0.81 (0.48, 1.36)	0.887	0.96 (0.53 1.72)	0.02
6	(0.92, −0.10, 0.07)	46.70	0.419	0.83 (0.53, 1.30)	0.658	0.90 (0.57 1.43)	0.08

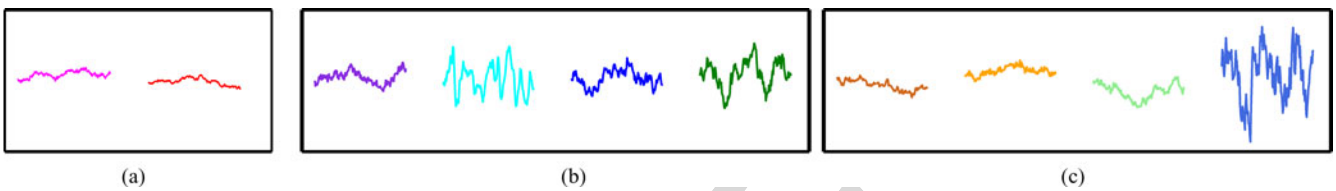


Fig. 3. Discovered dynamic modes of mean arterial BP of 453 patients during the first 24 h in the ICU. Figure shows the top ten most common dynamic modes, simulated using the AR coefficients from each dynamic mode. High-risk dynamic modes (from left to right): 3 (Magenta), 5 (Red). Low-risk dynamic modes: 1 (Violet), 9 (Cyan), 7 (Blue), and 2 (Green). Neutral dynamic modes: 10 (Brown), 8 (Orange), 4 (Light Green), 6 (Royal Blue). All modes were simulated and plotted with the same time duration (150 min) and amplitude scale. (a) High-risk modes, (b) Low-risk modes, and (c) Neutral modes.

during the first 24 h in the ICU are independent risk predictors of hospital mortality.

3) Example Time Series of Patients With Estimated Mortality Risks Over Time: Fig. 3 shows examples of low-risk and high-risk dynamical modes learned using the SVAR technique (see Table II for the odds-ratio associated with each mode). BP time series of four patients are presented in Fig. 4 panels (a) and (b). Hourly risk scores (dark green lines) were computed as the probability of death from the logistic function using a sliding window of 6 h to illustrate that these risk scores could be updated on a continuous basis for real-time monitoring purposes.

Panel (a) shows two of the patients with the highest risk scores (within the test set) at the end of the 24-h period; both patients died in the hospital. Panel (b) shows two patients with a decreasing trend in their risk scores during their first day in the ICU; both patients survived the hospital stay. All four patients were from the same test set, with mode assignment inferred based on dynamic modes learned from the corresponding training set. Note that as time progresses, patients in panel (a) tend to spend more time in the high-risk dynamic modes (mode 3 in magenta, mode 5 in red); their estimated mortality risks rise accordingly over time. In contrast, panel (b) patients show a decreasing trend in mortality risks as they transit to lower-risk dynamic modes over time.

IV. DISCUSSION AND CONCLUSION

We presented a SVAR framework to systematically learn and identify dynamic behaviors from vital sign time series within a patient cohort. We demonstrated that the discovered dynamics may contain prognostic values and can be used for prediction and tracking of a patient's propensity to survive a hospital stay, as well as their 28-days survival. Interestingly, the BP time series dynamics alone had a comparable performance to that of the SAPS I score which uses age and the most extreme values of 13 variables, including systolic BP, HR, temperature, respiratory rate, urinary output, blood nitrogen, hematocrit, white blood cell count, serum glucose, serum potassium, serum sodium, serum bicarbonate, and Glasgow coma score.

Additionally, our results indicate that the BP dynamics may contain complimentary information to existing acuity metrics, which assess the health of multiple organ systems based on a variety of physiological and lab variables. Specifically, combining the dynamics of BP time series and SAPS I or APACHE III provided a more accurate assessment of patient survival/mortality in the hospital ($p = 0.005$ and $p = 0.045$) than using SAPS I and APACHE III alone.

Association analysis of individual dynamic mode and hospital mortality revealed that two of the dynamic modes (modes 3 and 9) remained significant predictors of patients' outcome even after adjusting for APACHE IV scores, indicating that the proportion of time patients spent in these two dynamic modes during the first 24 h in the ICU may contain additional, independent prognostic value beyond that in the APACHE IV acuity score. Future work remains to investigate the prognostic power of these discovered dynamic modes using a larger cohort.

The dynamic features can be calculated in an online manner without delay, and well before the end of the first 24 h of the ICU stay as is required for the standard risk scores. One possible online deployment strategy is to construct a library of dynamic modes on archived patient data, and assign each incoming time series sample (or a sliding window of samples)

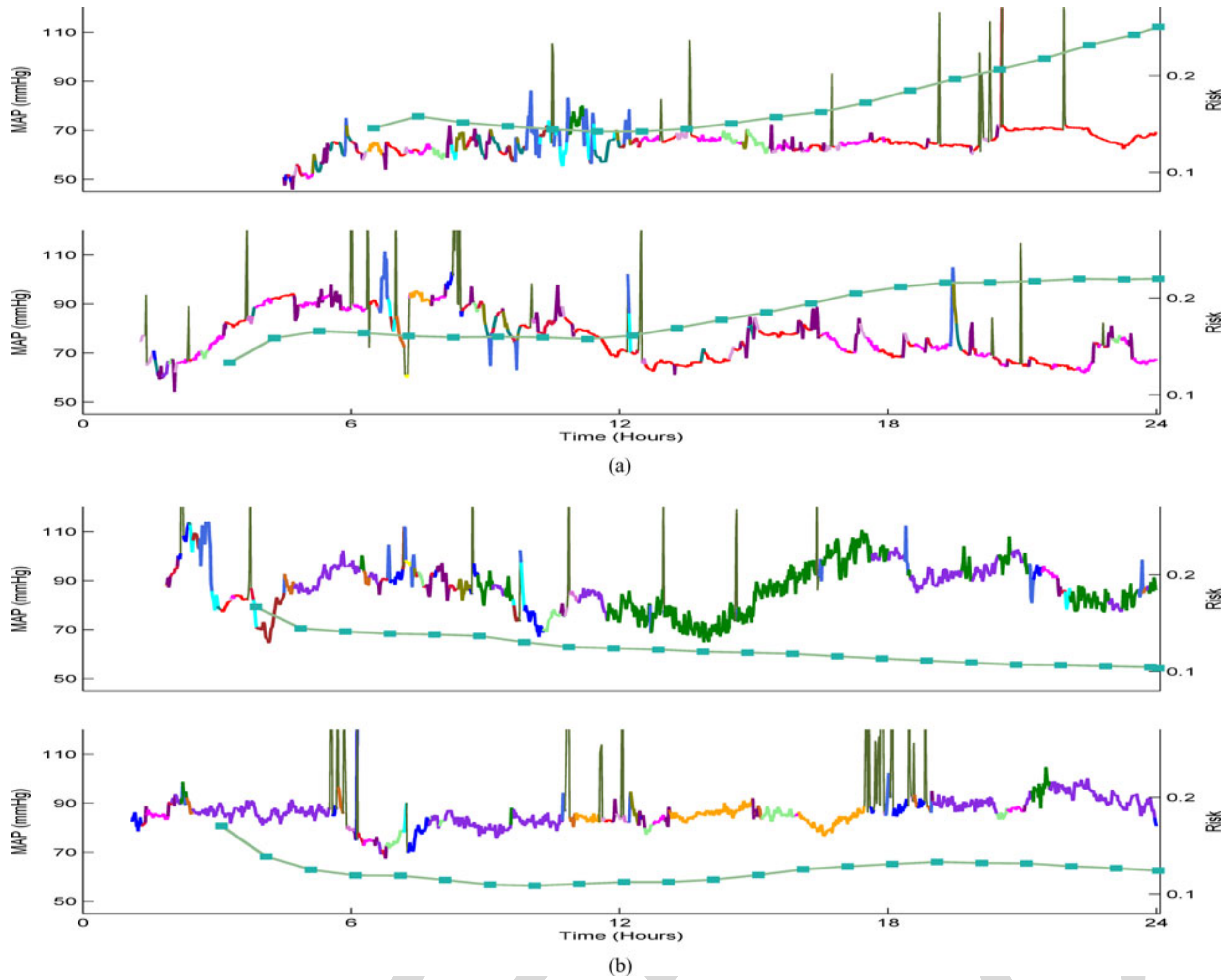


Fig. 4. Mortality risk scores and mean arterial BP of four patients during the first 24 h in the ICU. Samples are color coded by their mode assignment. Mortality risk scores, computed as the probability of death from the logistic regression, were based on mode proportions from a 6-h sliding window by stride of 1 h; estimated risks were plotted as dark green lines with scale indicated by *y*-axes on the right side of each graph. BP measurements plotted in original units (before detrending). All four patients were from the same test set, with dynamic modes and logistic regression parameters learned from the corresponding training set. (a) Patients with the highest ending risk scores at the end of the first day ICU stay. Patients were from MICU (top) and CCU (bottom). Both patients died in the hospital and (b) Patients with decreasing risk scores during their first day ICU stays. Patients were from CSRU (top) and CCU (bottom). Both patients survived the hospital stay.

to the most likely mode in the library (for instance, by using the Viterbi algorithm [16], [22]). Recent studies suggest that therapeutic interventions not only should aim at maintaining the mean BP within an acceptable range, but also should direct the patient's trajectory toward healthy dynamical regimes with enhanced variability [10] . Thus, a real-time implementation of the technique presented here may provide clinicians with a tool for quantification of the effectiveness of such interventions in the ICU.

We showed that changes in the dynamics of HR and BP, either as a result of an altered underlying control system (aging-related changes in the simulated data) or due to external perturbations (positional changes in the tilt-table experiment), can be captured in an automated fashion. Since the proposed framework is built on the dynamical systems framework (which includes the class of vector autoregressive models), the discovered modes can be used to reveal the oscillations that are present within the individual time series, and therefore can be used to extract useful indices of HR and BP variability (assuming beat-to-beat time series). Moreover, given beat-to-beat multivariate time-series of vital-signs, one may use the learned dynamics to derive the directional transfer functions of the system [8] (e.g., baroreflex control of HR and BP).

Association analysis using the minute-by-minute MIMIC-II BP time series revealed that the high-risk modes often correspond to less variable dynamical patterns. It is interesting to note that such low-frequency variability, observed at the minute-to-minute scale, is associated with an enhanced chance of survival, corresponding well to the existing HR/BP variability literature using beat-by-beat vital sign time series [10], [12], [13], [33]. The working hypothesis of our ongoing research is that the observed dynamical patterns are due to patients' underlying physiology, patient-specific response to clinical interventions, and measurement artifacts. Future developments of machine-learning techniques should aim at combining time series dynamics with contextual information pertaining to clinical intervention (administration of fluids, pressors, and titration of medications) to further investigate the clinical and physiological interpretation of the discovered modes.

The SVAR framework allows for defining a notion of "similarity" among multivariate physiological time series based on their underlying shared dynamics. Therefore, one may consider two subjects to be similar if their underlying vital signs time series exhibit similar dynamics in response to external (e.g., tilting of body) or internal perturbations (e.g., onset of blood infection). This approach provides an improvement over time series similarity measures based on trend-detection [34], wavelet-based symbolic representations [35], or Gaussian Mixture modeling [36] due to its compact representation and sharing of the model parameters within and across time series. Prior work using a factorial switching linear dynamical systems for patient monitoring [37] focused on detection of events associated with artifactual measurements and pathological states. Our study, in contrast, jointly models multiple time series across a large patient cohort to identify phenotypic dynamical patterns for patient outcome prediction.

Although we used mortality as our target outcome, there are many other physiological events of significant interest, including timely and successful discontinuation of procedures such as hemodialysis [38] or mechanical ventilation [39], as well as prediction of potentially life-threatening clinical events such as onset of severe sepsis and hypotension [13]. Other short- and long-term outcomes such as probability of readmission to hospital and long-term cognitive impairment beyond ICU [40] also play an important role in closing the gap between the critical care medicine, primary care doctors, and other healthcare providers.

Current and ongoing work involve combining the switching linear dynamical system framework with all available clinical data, including lab tests, medication records, and nursing notes [41] to devise a comprehensive risk score, capable of integrating clinical data of diverse modality over long temporal stretches (order of hours to days). This will allow us to investigate whether continuous patient monitoring based on vital signs dynamics, and other types of sequential data, can alert clinicians to deteriorating patient conditions at an earlier stage than the existing acuity scores, and result in improved patient care and outcome both within ICU and after hospital discharge. Such analysis is likely to provide some insight into the promise of large-scale critical care databases for the future of medicine.

Acknowledgment

The authors would like to thank Dr. T. Heldt (Harvard-MIT Division of Health Sciences and Technology) for kindly providing the tilt-table data analyzed in this study.

References

[1] M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L. H. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, and R. G. Mark, "Multiparameter intelligent monitoring in intensive care (MIMIC II): A public-access intensive care unit database," *Crit. Care Med.*, vol. 39, no. 5, pp. 952–960, May 2011.

[2] J. R. Le Gall, P. Loirat, A. Alperovitch, P. Glaser, C. Granthil, D. Mathieu, P. Mercier, R. Thomas, and D. Villers, "A simplified acute physiology score for ICU patients," *Crit. Care Med.*, vol. 12, no. 11, pp. 975–977, Nov. 1984.

[3] J. R. Le Gall, S. Lemeshow, and F. Saulnier, "A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study," *J. Amer. Med. Assoc.*, vol. 270, no. 24, pp. 2957–2963, 1993.

[4] W. A. Knaus, D. P. Wagner, E. A. Draper, J. E. Zimmerman, M. Bergner, P. G. Bastos, C. A. Sirio, D. J. Murphy, T. Lotring, A. Damiano, and F. Harrell, "The APACHE III prognostic system," *Chest*, vol. 100, pp. 1619–1636, 1991.

[5] J. E. Zimmerman, A. A. Kramer, D. S. McNair, and F. M. Malila, "Acute physiology and chronic health evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients," *Crit. Care Med.*, vol. 34, no. 5, pp. 1297–1310, May 2006.

[6] P. C. Ivanov, L. A. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley, "Multifractality in human heartbeat dynamics," *Nature*, vol. 399, pp. 461–465, 1999.

[7] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, no. 6, p. 068102, 2002. Q2

[8] S. Nemati, B. A. Edwards, S. A. Sands, P. J. Berger, A. Wellman, G. C. Verghese, A. Malhotra, and J. P. Butler, "Model-based characterization of ventilatory stability using spontaneous breathing," *J. Appl. Physiol.*, vol. 111, no. 1, pp. 55–67, 2011.

[9] G. Mancia, "Short-and long-term blood pressure variability present and future," *Hypertension*, vol. 60, no. 2, pp. 512–517, 2012.

[10] G. Parati, J. E. Ochoa, C. Lombardi, and G. Bilo, "Assessment and management of blood-pressure variability," *Nat. Rev. Cardiol.*, vol. 10, pp. 143–155, 2013.

[11] T. G. Buchman, "Nonlinear dynamics, complex systems, and the pathobiology of critical illness," *Current Opinion Crit. Care*, vol. 10, no. 5, pp. 378–382, 2004.

[12] S. Saria, A. K. Rajani, J. Gould, D. Koller, and A. Penn, "Integration of early physiological responses predicts later illness severity in preterm infants," *Sci. Translational Med.*, vol. 2, no. 48, pp. 48–65, 2010.

[13] J. R. Moorman, J. B. Delos, A. A. Flower, H. Cao, B. P. Kovatchev, J. S. Richman, and D. E. Lake, "Cardiovascular oscillations at the bedside: Early diagnosis of neonatal sepsis using heart rate characteristics monitoring," *Physiol. Meas.*, vol. 32, no. 11, pp. 1821–1832, Nov. 2011.

[14] L. H. Lehman, S. Nemati, R. P. Adams, and R. G. Mark, "Discovering shared dynamics in physiological signals: Application to patient monitoring in ICU," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 5939–5942.

[15] J. Wiens, E. Horvitz, and J. V. Guttag, "Patient risk stratification for hospital-associated C. diff as a time-series classification task," in *Adv. Neural Inf. Process. Syst.*, 2012, pp. 476–484.

[16] L. H. Lehman, S. Nemati, R. P. Adams, G. Moody, A. Malhotra, and R. G. Mark, "Tracking progression of patient state of health in critical care using inferred shared dynamics in physiological time series," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2013, pp. 7072–7075.

[17] L. Mayaud, P. S. Lai, G. D. Clifford, L. Tarassenko, L. A. Celi, and D. Annane, "Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension," *Crit. Care Med.*, vol. 41, no. 4, pp. 954–962, 2013.

[18] M. Blount, M. R. Ebling, J. M. Eklund, A. G. James, C. McGregor, N. Percival, K. P. Smith, and D. Sow, "Real-time analysis for intensive care: Development and deployment of the artemis analytic system," *IEEE Eng. Med. Biol. Mag.*, vol. 29, no. 2, pp. 110–118, Mar.–Apr. 2010.

[19] A. Porta, T. Bassani, V. Bari, E. Tobaldini, A. C. M. Takahashi, A. M. Catai, and N. Montano, "Model-based assessment of baroreflex and cardiopulmonary couplings during graded head-up tilt," *Comput. Biol. Med.*, vol. 42, no. 3, pp. 298–305, Mar. 2012.

[20] R. Mukkamala, J. M. Mathias, T. J. Mullen, R. J. Cohen, and R. Freeman, "System identification of closed-loop cardiovascular control mechanisms: Diabetic autonomic neuropathy," *Amer. J. Physiol.-Regulatory, Integrative Comparative Physiol.*, vol. 276, no. 3, pp. R905–R912, 1999.

[21] I. Korhonen, "Multivariate closed-loop model for analysis of cardiovascular dynamics," *Methods Inf. Med.*, vol. 36, pp. 264–267, 1997.

[22] K. P. Murphy, "Switching Kalman filter," Compaq Cambridge Res. Lab., Cambridge, CA, USA, Tech. Rep. 98-10, 1998.

[23] S. Nemati, L. H. Lehman, R. P. Adams, and A. Malhotra, "Discovering shared cardiovascular dynamics within a patient cohort," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 6526–6529.

[24] A. Fowler and M. McGuinness, "A delay recruitment model of the cardiovascular control system," *J. Math. Biol.*, vol. 51, no. 5, pp. 508–526, 2005.

[25] P. McSharry, M. McGuinness, and A. Fowler, "Confronting a cardiovascular system model with heart rate and blood pressure data," in *Proc. Comput. Cardiol.*, 2005, pp. 587–590.

[26] T. Heldt, M. B. Oefinger, M. Hoshiyama, and R. G. Mark, "Circulatory response to passive and active changes in posture," in *Proc. Comput. Cardiol.*, 2003, vol. 30, pp. 263–266.

[27] T. Heldt, "*Computational models of cardiovascular response to orthostatic stress*," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, Sep. 2004.

Q3 [28] A. Camm, M. Malik, J. Bigger, G. Breithardt, S. Cerutti, R. Cohen, P. Coumel, E. Fallen, H. Kennedy, R. Kleiger *et al.*, "Heart rate variability: Standards of measurement, physiological interpretation and clinical use. Task force of the european society of cardiology and the north american society of pacing and electrophysiology," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

Q4 [29] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. (2000 Jun. 13) PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation. [Online]. *101(23)*, pp. e215–e220. Available: http://circ.ahajournals.org/cgi/content/full/101/23/e215 PMID:1085218; doi: 10.1161/01.CIR.101.23.e215

[30] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.

[31] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach," *Biometrics*, vol. 44, pp. 837–845, 1988.

[32] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Statist. Soc.*, vol. 57, no. 1, pp. 289–300, 1995.

[33] W. P. Riordan Jr., P. R. Norris, J. M. Jenkins, and J. A. Morris Jr., "Early loss of heart rate complexity predicts mortality regardless of mechanism, anatomic location, or severity of injury in 2178 trauma patients," *J. Surg. Res.*, vol. 156, no. 2, pp. 283–289, 2009.

[34] R. K. Avent and J. D. Charlton, "A critical review of trend-detection methodologies for biomedical monitoring systems," *Crit. Rev. Biomed. Eng.*, vol. 17, no. 6, pp. 621–659, 1990.

[35] M. Saeed and R. G. Mark, "A novel method for the efficient retrieval of similar multiparameter physiologic time series using wavelet-based symbolic representations," in *Proc. AMIA Annu. Symp.*, 2006, pp. 679–683.

[36] L. H. Lehman, M. Saeed, G. Moody, and R. G. Mark, "Similarity-based searching in multi-parameter time series databases," in *Proc. Comput. Cardiol.*, 2008, pp. 653–656.

[37] J. A. Quinn, C. K. Williams, and N. McIntosh, "Factorial switching linear dynamical systems applied to physiological condition monitoring," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1537–1551, Sep. 2009.

[38] N. Gibney, E. Hoste, E. A. Burdmann, T. Bunchman, V. Kher, R. Viswanathan, R. L. Mehta, and C. Ronco, "Timing of initiation and discontinuation of renal replacement therapy in AKI: Unanswered key questions," *Clin. J. Amer. Soc. Nephrol.*, vol. 3, no. 3, pp. 876–880, 2008.

[39] A. Mikhno and C. M. Ennett, "Prediction of extubation failure for neonates with respiratory distress syndrome using the MIMIC-II clinical database," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 5094–5097.

[40] A. E. Wolters, A. J. Slooter, A. W. van der Kooi, and D. van Dijk, "Cognitive impairment after intensive care unit admission: A systematic review," *Intensive Care Med.*, vol. 39, no. 3, pp. 376–386, 2013.

[41] L. H. Lehman, M. Saeed, W. Long, J. Lee, and R. G. Mark, "Risk stratification of ICU patients using topic models inferred from unstructured progress notes," in *Proc. AMIA Annu. Symp Amer. Med. Informat. Assoc.*, 2012, pp. 505–511.

Authors', photographs and biographies not available at the time of publication.

QUERIES

Q1. Author: Please verify whether the affiliations of all the authors are OK as set.
Q2. Author: Please provide complete page range in Ref. [7].
Q3. Author: Please provide the department name in Ref. [27].
Q4. Author: Please provide names of all the authors in Ref. [28].

A Physiological Time Series Dynamics-Based Approach to Patient Monitoring and Outcome Prediction

Li-Wei H. Lehman, Ryan P. Adams, Louis Mayaud, George B. Moody, Atul Malhotra, Roger G. Mark, and Shamim Nemati

***Abstract*—Cardiovascular variables such as heart rate (HR) and blood pressure (BP) are regulated by an underlying control system, and therefore, the time series of these vital signs exhibit rich dynamical patterns of interaction in response to external perturbations (e.g., drug administration), as well as pathological states (e.g., onset of sepsis and hypotension). A question of interest is whether "similar" dynamical patterns can be identified across a heterogeneous patient cohort, and be used for prognosis of patients' health and progress. In this paper, we used a switching vector autoregressive framework to systematically learn and identify a collection of vital sign time series dynamics, which are possibly recurrent within the same patient and may be shared across the entire cohort. We show that these dynamical behaviors can be used to characterize the physiological "state" of a patient. We validate our technique using simulated time series of the cardiovascular system, and human recordings of HR and BP time series from an orthostatic stress study with known postural states. Using the HR and BP dynamics of an intensive care unit (ICU) cohort of over 450 patients from the MIMIC II database, we demonstrate that the discovered cardiovascular dynamics are significantly associated with hospital mortality (dynamic modes 3 and 9, $p = 0.001$, $p = 0.006$ from logistic regression after adjusting for the APACHE scores). Combining the dynamics of BP time series and SAPS-I or APACHE-III provided a more accurate assessment of patient survival/mortality in the hospital than using SAPS-I and APACHE-III alone ($p = 0.005$ and $p = 0.045$). Our results suggest that the discovered dynamics of vital sign time series may contain additional prognostic value beyond that of the baseline acuity measures, and can potentially be used as an independent predictor of outcomes in the ICU.**

***Index Terms*—Intensive care unit, physiological control systems, switching linear dynamical systems.**

Manuscript received October 31, 2013; revised March 30, 2014; accepted June 1, 2014. This work was supported by the National Institutes of Health Grants R01-EB001659 and R01GM104987 from the National Institute of Biomedical Imaging and Bioengineering, the James S. McDonnell Foundation Postdoctoral Grant, and the DARPA Young Faculty Award N66001-12-1-4219 Grant.

L. H. Lehman, G. B. Moody, and R. G. Mark are with the Massachusetts Institute of Technology, Cambridge, MA 02142 USA (e-mail: lilehman@mit.edu; george@mit.edu; rgmark@mit.edu).

R. P. Adams and S. Nemati are with the Harvard School of Engineering and Applied Sciences, Cambridge, MA 02138 USA (e-mail: rpa@seas.harvard.edu; shamim@seas.harvard.edu).

L. Mayaud is with the Institute of Biomedical Engineering, University of Oxford, Oxford, OX1 3PJ, UK, and also with Hopital Raymond Poincare, Garches 92380, France (e-mail: louis.mayaud@eng.ox.ac.uk).

A. Malhotra is with the Brigham and Women's Hospital & Harvard Medical School, Boston, MA 02115 USA, and also with the UC San Diego Division of Pulmonary & Critical Care Medicine, La Jolla, CA 92037 USA (e-mail: amalhotra1@partners.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2014.2330827

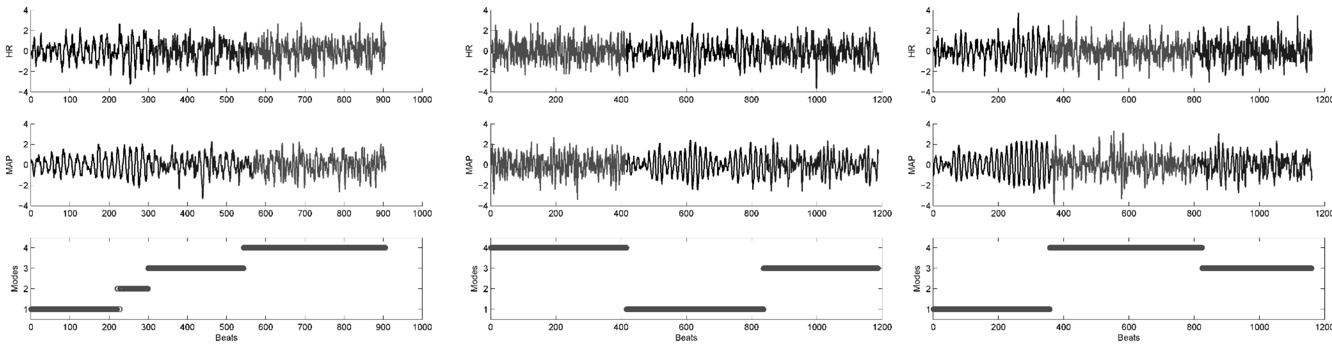
I. Introduction

Modern clinical data acquisition systems are capable of continuously monitoring and storing measurements of patient vital signs, such as heart rate (HR) and blood pressure (BP), over multiple days of hospitalization [1]. Despite this continuous feed of data, commonly used acuity scores, such as APACHE and SAPS [2]–[5], are based on snap-shot values of these vital signs, typically the worst values during a 24-h period. However, physiologic systems generate complex dynamics in their output signals that reflect the state of the underlying control systems [6]–[8]. The objective of this study is to consider an approach to the analysis of critical care bedside monitoring that is based on the dynamical behaviors of vital sign time series.

The time series of vital signs (e.g., HR, BP) are multidimensional, high resolution (from once a second to once a minute), highly coupled due to presence of physiological feedback loops within the body [8], and remarkably nonstationary as a result of internally and externally induced changes in the state of the underlying control systems. For instance, time series of BP can exhibit oscillations on the order of seconds (e.g., due to the variations in sympathovagal balance), to minutes (e.g., as a consequence of fever, blood loss, or behavioral factors), to hours (e.g., due to humoral variations, sleep-wake cycle, or circadian effects) [9], [10]. A growing body of the literature is pointing to the clinical utility of vital signs time series dynamics to inform prognosis [11]–[17], and to provide early predictors of potentially life-threatening conditions in the intensive care unit (ICU) [18].

Techniques for modeling and analysis of cardiovascular and respiratory time series can be broadly classified into linear mechanistic models [19], [20] and nonlinear descriptive indices [6], [7], [21]. The linear techniques commonly used (often based on variants of autoregressive modeling) have the advantage of revealing the individual relationships among the observed variables (e.g., the noninvasive measures of baroreflex gain describes the relationship between HR and BP, excluding the possible influence of respiration). On the other hand, nonlinear indices of complexity are capable of capturing a richer set of dynamical behaviors, with less emphasis on physiological interpretability in terms of specific underlying mechanisms.

In this paper, we assume that although the underlying dynamical system may be nonlinear and nonstationary, and the stochastic noise components can be non-Gaussian, the dynamics can be approximated by a mixture of linear dynamical systems. Each such linear "dynamic" (or mode) is a time-dependent rule that describes how the future state of the system evolves from its



(a) (b) (c)

Fig. 1. Simulation study of the cardiovascular system. Three examples (out of the ten simulated time series) of HR and BP (after filtering) are shown in panels a, b, and c. In each case, the actual dynamics are color coded. The horizontal red lines show the inferred segmentation. The algorithm consistently assigned modes 4 and 3 to the dynamics color coded as red and blue, respectively, across all the simulated time series. The black dynamics are represented by modes 1 and 2. (a) Simulated subject 1, (b) Simulated subject 2, and (c) Simulated subject 3.

current state, centered around a given system equilibrium point. Therefore, an ideal algorithm would be able to identify time series segments that follow a "similar" dynamic, and would switch to a different mode upon a change in the state of the underlying system.

To formalize these objectives, we employed a switching vector autoregressive (SVAR) framework [22], [23]. Given a collection of time series from a cohort, the proposed SVAR framework allows for simultaneous learning of the underlying dynamic behaviors or modes, and segmentation of the time series in terms of the most likely dynamic describing the time series evolution at any given point in time. The proposed framework enables characterization of patients in terms of the dynamical modes (e.g., the average time spent within the different modes), and can potentially be used to capture changes in the underlying cardiovascular control systems of human subjects in response to internal (such as onset of infection) and external perturbations (such as postural changes). Furthermore, we hypothesize that when applied to vital sign time series of patients in a critical care setting, the proposed technique can be used to discover dynamical modes with prognostic values for predicting clinical outcomes of interests.

A preliminary version of this study was presented at the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '12) [14]. Here, we extend on our previous work to include a series of validation studies, and a more comprehensive assessment of the utility of the time series dynamics within the ICU.

The rest of this paper is organized as follows. We validated the proposed technique using HR and BP time series from a simulation dataset, and a human laboratory study of subjects undergoing a tilt-table test, where the timing of the occurrence of the different dynamics and the sharing of the dynamics across multiple time series/subjects were known *a priori*. To test the prognostic value of the discovered vital sign dynamics, we applied the proposed approach to the HR and BP dynamics of an ICU cohort from the MIMIC II database [1] during the first 24 h of their ICU stays, and tested whether cardiovascular dynamics during the first 24 h of ICU admission are predictive of survival and mortality after adjusting for the existing acuity scores, such as SAPS-I and APACHE.

II. Materials and Methods

This section describes the utilized datasets, as well as the proposed technique for discovery of shared dynamics among patients, and assessment of risks and outcomes.

A. Datasets

1) Cardiovascular Simulation: We simulated a cardiovascular control system with bivariate time series of HR and BP. The model is based on a delay recruitment model of HR and BP regulation, as described in Fowler and McGuinness [24], and McSharry *et al.* [25]. The model included a coupled system of nonlinear delayed differential equations, controlling HR and BP, with respiration as an exogenous input. We simulated ten different multivariate time series of HR and mean arterial BP, each including three different dynamics that become dominant in a random order and last for a variable length of time. The three dynamics (color-coded as red, blue, and black, respectively, in Fig. 1) approximate aging-related autonomic changes; a progressive reduction in parasympathetic gain (from 0.40 to 0.13 to 0.07 in normalized units; see [24]) and an increase in sympathetic delay (from 3 to 5 s). To be consistent, we used the same preprocessing step as the tilt-table experiment to remove the steady-state baseline and any oscillation in the time series slower than 100 beats/cycle (see below for details).

2) Tilt-Table Experiment: Time series of HR and BP were acquired from ten healthy subjects (five males, five females) undergoing a tilt-table test of orthostatic tolerance [26], [27]. The mean age was 28.7 ± 1.2 years. The details of the protocol are described by Heldt in [27]. Briefly, subjects were placed in a supine position. Tilting was performed from the horizontal position to the vertical position and back to supine. The study was approved by MIT's Committee on the Use of Humans as Experimental Subjects and the Advisory Board of the MIT-MGH General Clinical Research Center [27]. Volunteers gave written, informed consent prior to participation

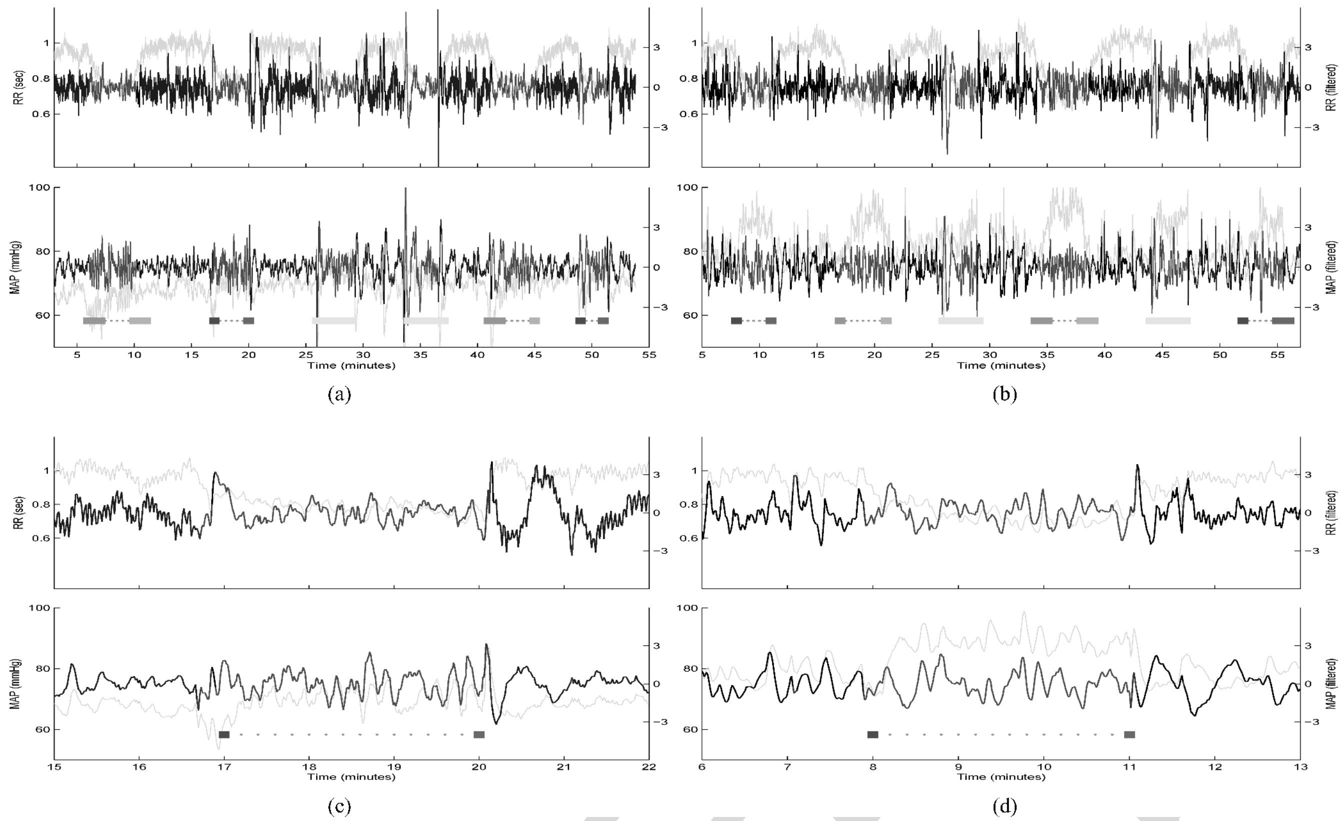


Fig. 2. Tilt-table study modeled using four dynamic modes—1 (Blue), 2 (Red), 3 (Black), and 4 (Purple). Two examples out of the ten recordings of HR and BP from the tilt-table experiment are shown in Panels a and b. Panels c and d show a zoomed in 7-min recording of HR and BP, while the subjects transition to/from supine to nonsupine positions after a fast tilt procedure. Actual values are in gray (*Y*-axis on left) and filtered values (*Y*-axis on right) are color coded based on the inferred dynamical modes. Note that Subjects 1 and 2 shared the same inferred nonsupine dynamics (in red); the algorithm consistently assigns the red mode to the nonsupine position for both subjects. The supine position for Subjects 1 and 2 are captured by the modes in blue and black, respectively. The purple mode seems to capture the high-frequency noise components of the time series. In each case, annotations for the actual tilt procedures performed are plotted as horizontal bars on the bottom of each figure and are color coded (green to cyan: slow tilt up and down to supine; red to pink: rapid tilt up and down to supine; yellow: stand up and back to supine). (a) Tilt Subject 1, (b) Tilt Subject 2, (c) Tilt Subject 1 (zoomed in), and (d) Tilt Subject 2 (zoomed in).

in the study. Since we were interested in the dynamics of interaction between HR and BP in the frequency range pertinent to sympathetic and parasympathetic regulation [28], time series of HR and BP were high-pass filtered to remove the steady-state baseline and any oscillation in the time series slower than 100 beats/cycle. This filtering was done using a seventh-order Butterworth digital filter with a cutoff frequency of 0.01 cycles/beat. Example time series from before and after filtering are shown in Fig. 2.

3) MIMIC II Dataset: The MIMIC II database [1], publicly available via PhysioNet [29], includes clinical (laboratory values, IV medications, etc.) and physiological data (HR, BP, oxygen saturation, etc.) collected from the bedside monitors (Component Monitoring System Intellivue MP-70; Philips Healthcare, Andover, MA, USA) in ICUs of the Beth Israel Deaconess Medical Center (BIDMC) in Boston. The MIMIC II waveform database (version 2) includes approximately 4000 records of high-resolution physiological waveforms of adult ICU patients with associated minute-by-minute (averages of the calculated numerics during the previous minute) vital sign trends. Data collection for the MIMIC II database was approved by the Institutional Review Boards of BIDMC and the Massachusetts Institute of Technology (Cambridge, MA, USA). Individual patient consent was waived because the study did not impact clinical care and protected health information was deidentified.

This study includes adult patients from the MIMIC II waveform database with at least 8 h of continuous minute-by-minute HR and invasive arterial BP trends during the first 24 h in the ICU. Patients with more than 15% of missing or invalid samples (i.e., outside physiologically plausible bounds of 20 to 200 mmHg for mean pressures) were excluded from this study, as were patients with missing SAPS I and APACHE scores. The dataset contains over 9000 h of minute-by-minute HR and invasive mean arterial BP measurements (over 20 h per patient on average) from 453 adult patients collected during the first 24 h in the ICU. HR and BP time series were detrended. Gaussian noise was used to fill in the missing or invalid values. The median age of this cohort was 69 with an interquartile range of (57, 79). 59% of the patients were male. Approximately 15% (67 out of 453) of patients in this cohort died in the hospital; 28-day mortality of this cohort was approximately 19% (85 out of 453). Distributions of the 453 patients in care units are 21% coronary care unit (CCU), 42% cardiac surgery recovery unit (CSRU), 26% medical intensive care unit (MICU), and 12% surgical intensive care unit.

B. SVAR Modeling of Cohort Time Series

Our approach to discovery of shared dynamics among patients is based on the SVAR model [22]. For the nth patient ($n = 1 \ldots N$), let $y_t^{(n)}$ be a $M \times 1$ vector of observed values of the vital signs at time t ($t = 1 \ldots T^{(n)}$). We assume that there exists a library of K possible dynamics or *modes*; a set of multivariate autoregressive model coefficient matrices $\{A_p^{(k)}\}_{k=1}^K$ of size $M \times M$, with maximal time lag $p = 1 \ldots P$, and the corresponding noise covariances $\{Q^{(k)}\}_{k=1}^K$. Let s_t be a switching variable, indicating the active dynamic mode at time t, and evolving according to a Markovian dynamic with initial distribution $\pi^{(n)}$ and a $K \times K$ transition matrix Z. Following these definitions, an SVAR model for the nth patient is defined as

$$y_t^{(n)} = \sum_{p=1}^{P} A_p^{(s_t^{(n)})} y_{t-p}^{(n)} + w^{(s_t^{(n)})} \tag{1}$$

where the fluctuation term $w^{(s_t^{(n)})}$ is assumed Gaussian distributed with covariance $Q^{(s_t^{(n)})}$. A collection of related time series can be modeled as switching between these dynamic behaviors which describe a locally coherent linear model that persists over a segment of time. However, in practice, we neither know the set of switching variables (i.e., segmentation of the time series) nor the modes. In this study, we perform expectation–maximization (EM) to find the maximum-likelihood set of model parameters, as well as a factored estimate of the posterior distribution over the latent switching variables. A comprehensive treatment of the EM algorithm for SVAR is presented by Murphy (1998) [22]. Briefly, EM is a two-pass iterative algorithm: 1) in the expectation (E) step, we obtain the expected values of the latent switching variables $\{s_t^{(n)}\}_{t=1}^T$ using a forward–backward algorithm [22], and 2) in the maximization (M) step, we update all the model parameters $\{A_p^{(k)}\}$, $\{Q^{(k)}\}$, the Markov dynamics Z, and the initial conditions $\pi^{(n)}$ that maximize the expected complete data log likelihood. In our implementation of the EM algorithm, we achieve *shared* dynamics by pooling together all subjects' inferred variables in the M step. Iteration through several steps of the EM algorithm results in learning a set of K shared modes and a global transition matrix Z for all the patients.

For the simulated and the tilt datasets, we modeled the beat-by-beat HR/BP time series as a switching AR(5) process to model most of the parasympathetic responses and at least some of the sympathetic effects, without introducing an unduly complex model. Minute-by-minute BP time series from MIMIC II were modeled as a switching AR(3) process to capture a real oscillation and a possible trend per mode. The number of dynamic modes ($K = 20$) was determined using the Bayesian information criterion (BIC) [30]. Briefly, we computed the BIC scores from switching-VAR models using 5 to 30 modes. Results presented were based on the model with the minimum BIC scores (20 modes).

1) Parallel Computation for Scalable Learning: One of the advantages of the proposed technique is its scalability to hundreds or thousands of patients, due to the parallel implementation of the inference step of the SVAR learning algorithm via EM [22]. This parallelization strategy is effective since the majority of the computational cost of the SVAR training is in running the forward–backward algorithm, which can be done in parallel for each patient time series. We used MATLAB's parallel computation toolbox in association with 120 nodes on our computer cluster to perform a tenfold cross-validated study (12 cores per fold). Ten SVAR models were learned on the training set of each of the folds, followed by mapping the corresponding mode proportions to outcomes (e.g., hospital mortality) using logistic regression. Next, mode assignments of time series in the test set of each fold were inferred based on the modes learned from the corresponding training set (by running only the inference), and the regression weights from the training fold were used to predict outcomes.

C. Evaluation Methods and Statistical Analysis

Let us define a *mode proportion* $\mathrm{MP}_k^{(n)}$ as the proportion of time the nth patient spends within the kth mode. Given the maximum expected log-likelihood estimates of the switching variables s_t from the EM algorithm, we have

$$\mathrm{MP}_k^{(n)} = \frac{1}{T^{(n)}} \sum_{t=1}^{T^{(n)}} \mathrm{Prob}(s_t^{(n)} = k). \tag{2}$$

For classification and prediction purposes, we characterize each time series with its corresponding mode proportion (a $1 \times K$ feature-vector), and use a logistic regression classifier to make predictions about the outcome variables of interest. For illustration of the algorithm's segmentation performance, each time series sample is assigned to the dynamic mode with the maximum posterior probability.

1) Time Series Classification and Outcome Prediction: For the simulated and the tilt-table experiment, we used the mode proportions within each segment (e.g., supine versus nonsupine) as inputs to a logistic regression classifier, and report the classification performance in discriminating between 1) the three different dynamics (corresponding to different aging-related autonomic changes) in the simulated dataset, and 2) two different postural positions (supine versus nonsupine) in the tilt dataset.

To assess the predictive power of the dynamical modes, we performed a tenfold cross-validation study. Ten SVAR models were learned on the training set of each of the folds, followed by mapping the corresponding mode proportions to outcomes (e.g., hospital mortality) using logistic regression. Next, mode assignments of time series in the test set of each fold was inferred based on the modes learned from the corresponding training set (by running only the inference or the E-step), and the regression weights from the training fold was used to predict outcomes. We compared the mortality prediction performance of our approach using the mode proportion from the top ten most common dynamic modes with the existing acuity metrics, SAPS I [2], APACHE III [4], and APACHE IV [5]. Comparison of AUCs was based on the method described in [31].

2) MIMIC Association Analysis: We used univariate and multivariate logistic regressions to examine the associations between dynamic mode proportions and hospital mortality. We

TABLE I
PERFORMANCE OF MORTALITY PREDICTORS

	Hosp. Mortality (AUC)	28-Days Mortality (AUC)
$\mathrm{HR}_{\mathrm{dyn}}$	0.59 (0.54, 0.68)	0.61 (0.51, 0.67)
$\mathrm{BP}_{\mathrm{dyn}}/\mathrm{HR}_{\mathrm{dyn}}$	0.64 (0.61, 0.71)	0.65 (0.64, 0.68)
$\mathbf{BP}_{\mathrm{dyn}}$	**0.70 (0.67, 0.77)**	**0.66 (0.61, 0.73)**
SAPS I	0.65 (0.59, 0.71)	0.64 (0.56, 0.70)
$\mathbf{BP}_{\mathrm{dyn}}$ + SAPS I	**0.77 (0.69, 0.82)**	**0.71 (0.69, 0.79)**
APACHE III	0.80 (0.70, 0.84)	0.79 (0.65, 0.84)
$\mathbf{BP}_{\mathrm{dyn}}$ + APACHE III	**0.84 (0.79, 0.88)**	**0.79 (0.76, 0.86)**
APACHE IV	0.82 (0.77, 0.85)	0.83 (0.74, 0.86)
$\mathbf{BP}_{\mathrm{dyn}}$ + APACHE IV	**0.85 (0.80, 0.87)**	**0.82 (0.81, 0.88)**

built a separate multivariate logistic regression model for each of the discovered dynamic modes, with the mode proportion as the primary predictive variable, and APACHE IV as a covariate. For each mode, we reported its p value, odds ratio (OR, with 95% confidence interval), and adjusted OR (after including APACHE IV as a covariate). The Hosmer–Lemeshow p values (HL p values) were reported to assess the model fit. The odds ratios were per 10% increase in the mode proportion. Two-sided p values less than 0.05 were considered statistically significant. The analysis was performed to quantify the mortality risk associated with each dynamic mode; modes with significant ($p < 0.05$) associations with mortality were established as either *low-risk* (OR < 1), or *high-risk* (OR > 1) dynamics depending on their odds ratios. Dynamic modes without statistically significant associations with mortality were *neutral* modes. The test of statistical significance was based on p-values after correcting for the false discovery rate (FDR) using the technique described in [32].

III. RESULTS

A. Simulated Study

Fig. 1 shows two examples of simulated time series with the inferred segmentation. In all ten simulated cases, the algorithm was able to divide each time series into distinct segments corresponding to different underlying actual dynamics. The sharing of the dynamics is consistent across the different time series. Using the mode proportion from each segment for multilabel classification, the algorithm achieved classification accuracy of 100%.

B. Tilt-Table Experiment

Fig. 2 shows the segmentation results for two subjects. Note that the two subjects shared the same inferred nonsupine dynamics (in red); the algorithm consistently assigns the red mode to the nonsupine position for both subjects. The application of logistic regression with tenfold cross-validation yielded a median AUC of 1.00 with an interquartile range of (0.98, 1.00).

C. MIMIC II Database

1) Mortality Prediction: Table I evaluates the prognostic power of HR and BP dynamic features ($\mathrm{HR}_{\mathrm{dyn}}$ and $\mathrm{BP}_{\mathrm{dyn}}$). SAPS I, APACHE III, and APACHE IV are used as the baselines. Median AUCs (from tenfold cross validation) and the interquartile range are shown. Note that the BP dynamics outperformed both the HR and HR and BP combined dynamic features. Subsequent analyses focus on the predictive power of the BP dynamics in comparison to the baseline. For each baseline, we show the performance from the baseline alone, and the combined approach (combining BP dynamics and the baseline).

The application of tenfold cross-validation demonstrated that dynamic features from BP alone achieved a median AUC of 0.70, comparable to 0.65 from SAPS I. In comparison, using standard deviation of the mean arterial BP resulted in a median AUC (IQR) of 0.55 (0.43, 0.63).

Combining dynamic BP features with SAPS I resulted in an improved prediction power both in hospital mortality prediction ($p = 0.005$) and 28-day mortality prediction ($p = 0.002$). Combining dynamic features with APACHE III significantly out-performed APACHE III alone ($p = 0.045$) with an improvement in median AUC from 0.80 to 0.84 in hospital mortality prediction. These results indicate that the dynamic features from vital signs contain complementary information to the SAPS I and APACHE III scores.

State-of-the-art risk score APACHE IV achieved better prediction performance than the BP dynamic features alone ($p = 0.008$). Adding BP dynamics to APACHE IV yielded a slight performance improvement from a median AUC of 0.82 to 0.85, however, the performance gain was not statistically significant.

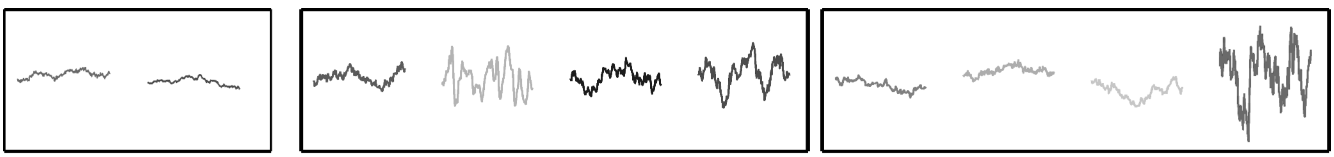
2) Association Analysis: Table II presents logistical regression analyses to test the associations between the proportion of time patients spent in each of the top ten most common BP dynamics and hospital mortality. See Fig. 3 for illustrations of these dynamic modes. Dynamic modes were numbered based on their prevalence across the entire cohort (i.e., mode 1 is the most common dynamic mode). Our results indicate that six of the modes had significant associations (after FDR correction) with hospital mortality. Specifically, two dynamic modes (modes 3 and 5) were significant "high-risk" modes ($p < 0.001$, $p < 0.001$) in which increased proportions of time in these modes were associated with higher hospital mortality with odds ratios 1.81 (1.41, 2.32), 1.36 (1.15, 1.61) respectively.

Dynamic modes 1, 9, 7, and 2 were "low-risk" modes in which increasing proportions of time in these modes were significantly associated with a decreased risk of hospital mortality, with odds ratios less than one. Table II lists the AR coefficients and covariances of the two high-risk and four low-risk dynamic modes, as well as their respective associations with hospital mortality. Note that the high-risk modes appear to correspond to less variability in their dynamics.

For the multivariate analysis (see the right panel in Table II), each row is a separate multivariate model, in which the mode proportion for a given target mode is the primary predictive variable, and APACHE IV is added as a control variable in the multivariate model. Results from multivariate logistic regression indicate that two of the modes (modes 3 and 9) remain significant predictors of patients' outcome even after adjustment for APACHE IV scores ($p = 0.001$, $p = 0.006$), indicating that the proportion of time patients spent in these two dynamic modes

TABLE II
ASSOCIATIONS OF BP DYNAMIC MODES AND HOSPITAL MORTALITY

Mode	AR Coef	Cov.	P-Val	OR(95%CI)	Adjusted P-Val	Adjusted OR(95%CI)	HL PVAL
3	(0.66, 0.22, 0.12)	0.58	$<$**0.001**	1.81 (1.41, 2.32)	**0.001**	1.60 (1.21 2.11)	0.64
5	(1.00, 0.00, −0.00)	0.22	$<$**0.001**	1.36 (1.15, 1.61)	0.426	1.08 (0.89 1.32)	0.16
1	(0.66, 0.16, 0.17)	2.69	**0.002**	0.59 (0.42, 0.82)	0.489	0.88 (0.62 1.26)	0.54
9	(1.50, −0.65, 0.06)	7.26	**0.002**	0.25 (0.10, 0.62)	**0.006**	0.26 (0.10 0.68)	0.80
7	(1.00, −0.01, −0.00)	3.46	**0.003**	0.30 (0.13, 0.67)	0.124	0.54 (0.25 1.18)	0.58
2	(0.79, 0.05, 0.12)	8.81	**0.005**	0.65 (0.48, 0.88)	0.265	0.84 (0.62 1.14)	0.69
10	(1.05, −0.01, −0.02)	0.71	0.032	2.95 (1.10, 7.94)	0.791	1.18 (0.36 3.88)	0.07
8	(0.44, 0.30, 0.24)	1.27	0.373	1.18 (0.82, 1.69)	0.318	1.22 (0.82 1.82)	0.22
4	(0.96, −0.01, 0.04)	1.31	0.417	0.81 (0.48, 1.36)	0.887	0.96 (0.53 1.72)	0.02
6	(0.92, −0.10, 0.07)	46.70	0.419	0.83 (0.53, 1.30)	0.658	0.90 (0.57 1.43)	0.08



(a) (b) (c)

Fig. 3. Discovered dynamic modes of mean arterial BP of 453 patients during the first 24 h in the ICU. Figure shows the top ten most common dynamic modes, simulated using the AR coefficients from each dynamic mode. High-risk dynamic modes (from left to right): 3 (Magenta), 5 (Red). Low-risk dynamic modes: 1 (Violet), 9 (Cyan), 7 (Blue), and 2 (Green). Neutral dynamic modes: 10 (Brown), 8 (Orange), 4 (Light Green), 6 (Royal Blue). All modes were simulated and plotted with the same time duration (150 min) and amplitude scale. (a) High-risk modes, (b) Low-risk modes, and (c) Neutral modes.

during the first 24 h in the ICU are independent risk predictors of hospital mortality.

3) Example Time Series of Patients With Estimated Mortality Risks Over Time: Fig. 3 shows examples of low-risk and high-risk dynamical modes learned using the SVAR technique (see Table II for the odds-ratio associated with each mode). BP time series of four patients are presented in Fig. 4 panels (a) and (b). Hourly risk scores (dark green lines) were computed as the probability of death from the logistic function using a sliding window of 6 h to illustrate that these risk scores could be updated on a continuous basis for real-time monitoring purposes.

Panel (a) shows two of the patients with the highest risk scores (within the test set) at the end of the 24-h period; both patients died in the hospital. Panel (b) shows two patients with a decreasing trend in their risk scores during their first day in the ICU; both patients survived the hospital stay. All four patients were from the same test set, with mode assignment inferred based on dynamic modes learned from the corresponding training set. Note that as time progresses, patients in panel (a) tend to spend more time in the high-risk dynamic modes (mode 3 in magenta, mode 5 in red); their estimated mortality risks rise accordingly over time. In contrast, panel (b) patients show a decreasing trend in mortality risks as they transit to lower-risk dynamic modes over time.

IV. Discussion and Conclusion

We presented a SVAR framework to systematically learn and identify dynamic behaviors from vital sign time series within a patient cohort. We demonstrated that the discovered dynamics may contain prognostic values and can be used for prediction and tracking of a patient's propensity to survive a hospital stay, as well as their 28-days survival. Interestingly, the BP time series dynamics alone had a comparable performance to that of the SAPS I score which uses age and the most extreme values of 13 variables, including systolic BP, HR, temperature, respiratory rate, urinary output, blood nitrogen, hematocrit, white blood cell count, serum glucose, serum potassium, serum sodium, serum bicarbonate, and Glasgow coma score.

Additionally, our results indicate that the BP dynamics may contain complimentary information to existing acuity metrics, which assess the health of multiple organ systems based on a variety of physiological and lab variables. Specifically, combining the dynamics of BP time series and SAPS I or APACHE III provided a more accurate assessment of patient survival/mortality in the hospital ($p = 0.005$ and $p = 0.045$) than using SAPS I and APACHE III alone.

Association analysis of individual dynamic mode and hospital mortality revealed that two of the dynamic modes (modes 3 and 9) remained significant predictors of patients' outcome even after adjusting for APACHE IV scores, indicating that the proportion of time patients spent in these two dynamic modes during the first 24 h in the ICU may contain additional, independent prognostic value beyond that in the APACHE IV acuity score. Future work remains to investigate the prognostic power of these discovered dynamic modes using a larger cohort.

The dynamic features can be calculated in an online manner without delay, and well before the end of the first 24 h of the ICU stay as is required for the standard risk scores. One possible online deployment strategy is to construct a library of dynamic modes on archived patient data, and assign each incoming time series sample (or a sliding window of samples)

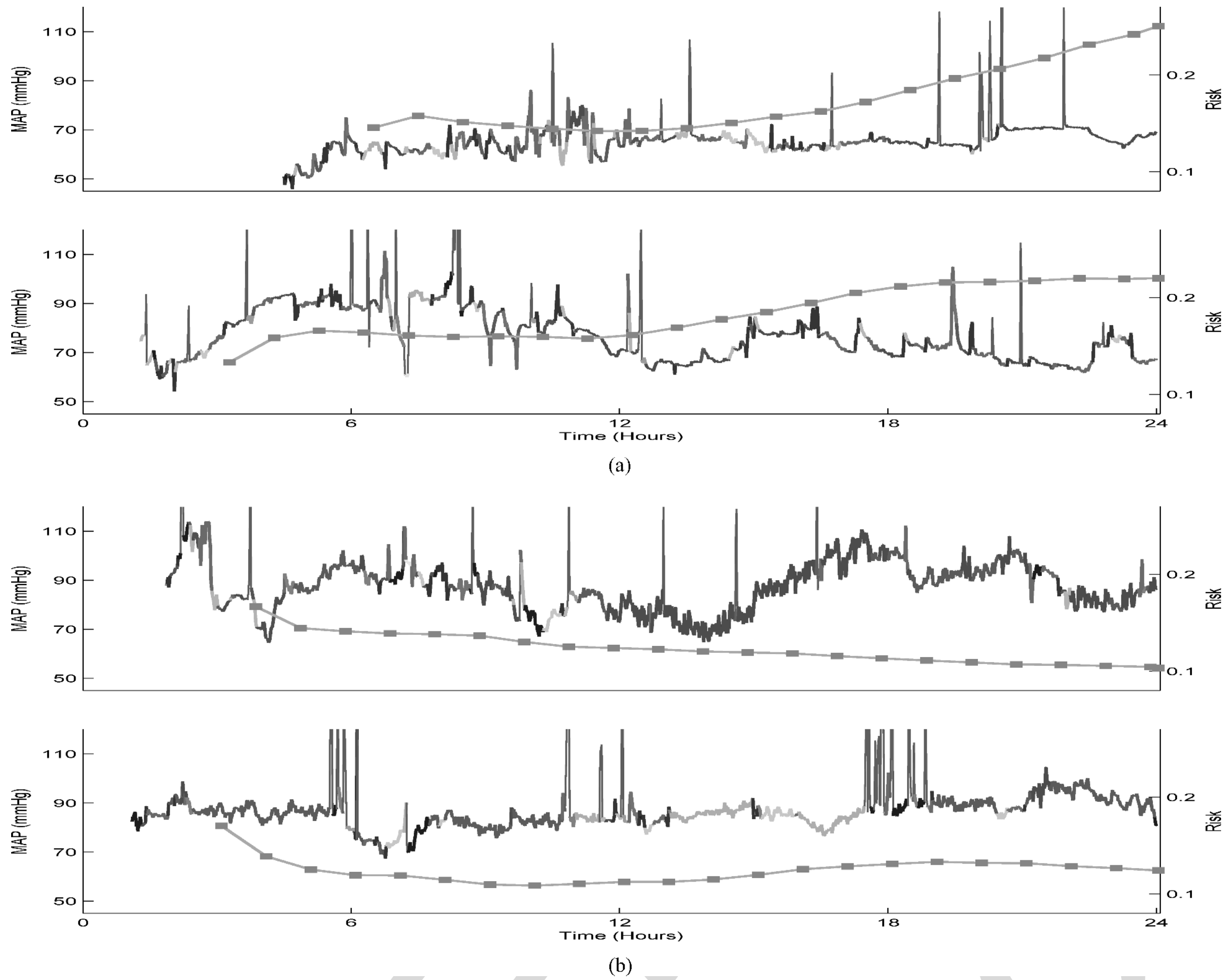


Fig. 4. Mortality risk scores and mean arterial BP of four patients during the first 24 h in the ICU. Samples are color coded by their mode assignment. Mortality risk scores, computed as the probability of death from the logistic regression, were based on mode proportions from a 6-h sliding window by stride of 1 h; estimated risks were plotted as dark green lines with scale indicated by *y*-axes on the right side of each graph. BP measurements plotted in original units (before detrending). All four patients were from the same test set, with dynamic modes and logistic regression parameters learned from the corresponding training set. (a) Patients with the highest ending risk scores at the end of the first day ICU stay. Patients were from MICU (top) and CCU (bottom). Both patients died in the hospital and (b) Patients with decreasing risk scores during their first day ICU stays. Patients were from CSRU (top) and CCU (bottom). Both patients survived the hospital stay.

to the most likely mode in the library (for instance, by using the Viterbi algorithm [16], [22]). Recent studies suggest that therapeutic interventions not only should aim at maintaining the mean BP within an acceptable range, but also should direct the patient's trajectory toward healthy dynamical regimes with enhanced variability [10] . Thus, a real-time implementation of the technique presented here may provide clinicians with a tool for quantification of the effectiveness of such interventions in the ICU.

We showed that changes in the dynamics of HR and BP, either as a result of an altered underlying control system (aging-related changes in the simulated data) or due to external perturbations (positional changes in the tilt-table experiment), can be captured in an automated fashion. Since the proposed framework is built on the dynamical systems framework (which includes the class of vector autoregressive models), the discovered modes can be used to reveal the oscillations that are present within the individual time series, and therefore can be used to extract useful indices of HR and BP variability (assuming beat-to-beat time series). Moreover, given beat-to-beat multivariate time-series of vital-signs, one may use the learned dynamics to derive the directional transfer functions of the system [8] (e.g., baroreflex control of HR and BP).

Association analysis using the minute-by-minute MIMIC-II BP time series revealed that the high-risk modes often correspond to less variable dynamical patterns. It is interesting to note that such low-frequency variability, observed at the minute-to-minute scale, is associated with an enhanced chance of survival, corresponding well to the existing HR/BP variability literature using beat-by-beat vital sign time series [10], [12], [13], [33]. The working hypothesis of our ongoing research is that the observed dynamical patterns are due to patients' underlying physiology, patient-specific response to clinical interventions, and measurement artifacts. Future developments of machine-learning techniques should aim at combining time series dynamics with contextual information pertaining to clinical intervention (administration of fluids, pressors, and titration of medications) to further investigate the clinical and physiological interpretation of the discovered modes.

The SVAR framework allows for defining a notion of "similarity" among multivariate physiological time series based on their underlying shared dynamics. Therefore, one may consider two subjects to be similar if their underlying vital signs time series exhibit similar dynamics in response to external (e.g., tilting of body) or internal perturbations (e.g., onset of blood infection). This approach provides an improvement over time series similarity measures based on trend-detection [34], wavelet-based symbolic representations [35], or Gaussian Mixture modeling [36] due to its compact representation and sharing of the model parameters within and across time series. Prior work using a factorial switching linear dynamical systems for patient monitoring [37] focused on detection of events associated with artifactual measurements and pathological states. Our study, in contrast, jointly models multiple time series across a large patient cohort to identify phenotypic dynamical patterns for patient outcome prediction.

Although we used mortality as our target outcome, there are many other physiological events of significant interest, including timely and successful discontinuation of procedures such as hemodialysis [38] or mechanical ventilation [39], as well as prediction of potentially life-threatening clinical events such as onset of severe sepsis and hypotension [13]. Other short- and long-term outcomes such as probability of readmission to hospital and long-term cognitive impairment beyond ICU [40] also play an important role in closing the gap between the critical care medicine, primary care doctors, and other healthcare providers.

Current and ongoing work involve combining the switching linear dynamical system framework with all available clinical data, including lab tests, medication records, and nursing notes [41] to devise a comprehensive risk score, capable of integrating clinical data of diverse modality over long temporal stretches (order of hours to days). This will allow us to investigate whether continuous patient monitoring based on vital signs dynamics, and other types of sequential data, can alert clinicians to deteriorating patient conditions at an earlier stage than the existing acuity scores, and result in improved patient care and outcome both within ICU and after hospital discharge. Such analysis is likely to provide some insight into the promise of large-scale critical care databases for the future of medicine.

Acknowledgment

The authors would like to thank Dr. T. Heldt (Harvard-MIT Division of Health Sciences and Technology) for kindly providing the tilt-table data analyzed in this study.

References

[1] M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L. H. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, and R. G. Mark, "Multiparameter intelligent monitoring in intensive care (MIMIC II): A public-access intensive care unit database," *Crit. Care Med.*, vol. 39, no. 5, pp. 952–960, May 2011.

[2] J. R. Le Gall, P. Loirat, A. Alperovitch, P. Glaser, C. Granthil, D. Mathieu, P. Mercier, R. Thomas, and D. Villers, "A simplified acute physiology score for ICU patients," *Crit. Care Med.*, vol. 12, no. 11, pp. 975–977, Nov. 1984.

[3] J. R. Le Gall, S. Lemeshow, and F. Saulnier, "A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study," *J. Amer. Med. Assoc.*, vol. 270, no. 24, pp. 2957–2963, 1993.

[4] W. A. Knaus, D. P. Wagner, E. A. Draper, J. E. Zimmerman, M. Bergner, P. G. Bastos, C. A. Sirio, D. J. Murphy, T. Lotring, A. Damiano, and F. Harrell, "The APACHE III prognostic system," *Chest*, vol. 100, pp. 1619–1636, 1991.

[5] J. E. Zimmerman, A. A. Kramer, D. S. McNair, and F. M. Malila, "Acute physiology and chronic health evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients," *Crit. Care Med.*, vol. 34, no. 5, pp. 1297–1310, May 2006.

[6] P. C. Ivanov, L. A. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley, "Multifractality in human heartbeat dynamics," *Nature*, vol. 399, pp. 461–465, 1999.

[7] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, no. 6, p. 068102, 2002. Q2

[8] S. Nemati, B. A. Edwards, S. A. Sands, P. J. Berger, A. Wellman, G. C. Verghese, A. Malhotra, and J. P. Butler, "Model-based characterization of ventilatory stability using spontaneous breathing," *J. Appl. Physiol.*, vol. 111, no. 1, pp. 55–67, 2011.

[9] G. Mancia, "Short-and long-term blood pressure variability present and future," *Hypertension*, vol. 60, no. 2, pp. 512–517, 2012.

[10] G. Parati, J. E. Ochoa, C. Lombardi, and G. Bilo, "Assessment and management of blood-pressure variability," *Nat. Rev. Cardiol.*, vol. 10, pp. 143–155, 2013.

[11] T. G. Buchman, "Nonlinear dynamics, complex systems, and the pathobiology of critical illness," *Current Opinion Crit. Care*, vol. 10, no. 5, pp. 378–382, 2004.

[12] S. Saria, A. K. Rajani, J. Gould, D. Koller, and A. Penn, "Integration of early physiological responses predicts later illness severity in preterm infants," *Sci. Translational Med.*, vol. 2, no. 48, pp. 48–65, 2010.

[13] J. R. Moorman, J. B. Delos, A. A. Flower, H. Cao, B. P. Kovatchev, J. S. Richman, and D. E. Lake, "Cardiovascular oscillations at the bedside: Early diagnosis of neonatal sepsis using heart rate characteristics monitoring," *Physiol. Meas.*, vol. 32, no. 11, pp. 1821–1832, Nov. 2011.

[14] L. H. Lehman, S. Nemati, R. P. Adams, and R. G. Mark, "Discovering shared dynamics in physiological signals: Application to patient monitoring in ICU," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 5939–5942.

[15] J. Wiens, E. Horvitz, and J. V. Guttag, "Patient risk stratification for hospital-associated C. diff as a time-series classification task," in *Adv. Neural Inf. Process. Syst.*, 2012, pp. 476–484.

[16] L. H. Lehman, S. Nemati, R. P. Adams, G. Moody, A. Malhotra, and R. G. Mark, "Tracking progression of patient state of health in critical care using inferred shared dynamics in physiological time series," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2013, pp. 7072–7075.

[17] L. Mayaud, P. S. Lai, G. D. Clifford, L. Tarassenko, L. A. Celi, and D. Annane, "Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension," *Crit. Care Med.*, vol. 41, no. 4, pp. 954–962, 2013.

[18] M. Blount, M. R. Ebling, J. M. Eklund, A. G. James, C. McGregor, N. Percival, K. P. Smith, and D. Sow, "Real-time analysis for intensive care: Development and deployment of the artemis analytic system," *IEEE Eng. Med. Biol. Mag.*, vol. 29, no. 2, pp. 110–118, Mar.–Apr. 2010.

[19] A. Porta, T. Bassani, V. Bari, E. Tobaldini, A. C. M. Takahashi, A. M. Catai, and N. Montano, "Model-based assessment of baroreflex and cardiopulmonary couplings during graded head-up tilt," *Comput. Biol. Med.*, vol. 42, no. 3, pp. 298–305, Mar. 2012.

[20] R. Mukkamala, J. M. Mathias, T. J. Mullen, R. J. Cohen, and R. Freeman, "System identification of closed-loop cardiovascular control mechanisms: Diabetic autonomic neuropathy," *Amer. J. Physiol.-Regulatory, Integrative Comparative Physiol.*, vol. 276, no. 3, pp. R905–R912, 1999.

[21] I. Korhonen, "Multivariate closed-loop model for analysis of cardiovascular dynamics," *Methods Inf. Med.*, vol. 36, pp. 264–267, 1997.

[22] K. P. Murphy, "Switching Kalman filter," Compaq Cambridge Res. Lab., Cambridge, CA, USA, Tech. Rep. 98-10, 1998.

[23] S. Nemati, L. H. Lehman, R. P. Adams, and A. Malhotra, "Discovering shared cardiovascular dynamics within a patient cohort," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 6526–6529.

[24] A. Fowler and M. McGuinness, “A delay recruitment model of the cardiovascular control system,” *J. Math. Biol.*, vol. 51, no. 5, pp. 508–526, 2005.

[25] P. McSharry, M. McGuinness, and A. Fowler, “Confronting a cardiovascular system model with heart rate and blood pressure data,” in *Proc. Comput. Cardiol.*, 2005, pp. 587–590.

[26] T. Heldt, M. B. Oefinger, M. Hoshiyama, and R. G. Mark, “Circulatory response to passive and active changes in posture,” in *Proc. Comput. Cardiol.*, 2003, vol. 30, pp. 263–266.

[27] T. Heldt, “*Computational models of cardiovascular response to orthostatic stress*,” Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, Sep. 2004.

Q3 [28] A. Camm, M. Malik, J. Bigger, G. Breithardt, S. Cerutti, R. Cohen, P. Coumel, E. Fallen, H. Kennedy, R. Kleiger *et al.*, “Heart rate variability: Standards of measurement, physiological interpretation and clinical use. Task force of the european society of cardiology and the north american society of pacing and electrophysiology,” *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

Q4 [29] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. (2000 Jun. 13) PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation. [Online]. *101(23)*, pp. e215–e220. Available: http://circ.ahajournals.org/cgi/content/full/101/23/e215 PMID:1085218; doi: 10.1161/01.CIR.101.23.e215

[30] G. Schwarz, “Estimating the dimension of a model,” *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.

[31] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, “Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach,” *Biometrics*, vol. 44, pp. 837–845, 1988.

[32] Y. Benjamini and Y. Hochberg, “Controlling the false discovery rate: A practical and powerful approach to multiple testing,” *J. Roy. Statist. Soc.*, vol. 57, no. 1, pp. 289–300, 1995.

[33] W. P. Riordan Jr., P. R. Norris, J. M. Jenkins, and J. A. Morris Jr., “Early loss of heart rate complexity predicts mortality regardless of mechanism, anatomic location, or severity of injury in 2178 trauma patients,” *J. Surg. Res.*, vol. 156, no. 2, pp. 283–289, 2009.

[34] R. K. Avent and J. D. Charlton, “A critical review of trend-detection methodologies for biomedical monitoring systems,” *Crit. Rev. Biomed. Eng.*, vol. 17, no. 6, pp. 621–659, 1990.

[35] M. Saeed and R. G. Mark, “A novel method for the efficient retrieval of similar multiparameter physiologic time series using wavelet-based symbolic representations,” in *Proc. AMIA Annu. Symp.*, 2006, pp. 679–683.

[36] L. H. Lehman, M. Saeed, G. Moody, and R. G. Mark, “Similarity-based searching in multi-parameter time series databases,” in *Proc. Comput. Cardiol.*, 2008, pp. 653–656.

[37] J. A. Quinn, C. K. Williams, and N. McIntosh, “Factorial switching linear dynamical systems applied to physiological condition monitoring,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1537–1551, Sep. 2009.

[38] N. Gibney, E. Hoste, E. A. Burdmann, T. Bunchman, V. Kher, R. Viswanathan, R. L. Mehta, and C. Ronco, “Timing of initiation and discontinuation of renal replacement therapy in AKI: Unanswered key questions,” *Clin. J. Amer. Soc. Nephrol.*, vol. 3, no. 3, pp. 876–880, 2008.

[39] A. Mikhno and C. M. Ennett, “Prediction of extubation failure for neonates with respiratory distress syndrome using the MIMIC-II clinical database,” in *Proc. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 5094–5097.

[40] A. E. Wolters, A. J. Slooter, A. W. van der Kooi, and D. van Dijk, “Cognitive impairment after intensive care unit admission: A systematic review,” *Intensive Care Med.*, vol. 39, no. 3, pp. 376–386, 2013.

[41] L. H. Lehman, M. Saeed, W. Long, J. Lee, and R. G. Mark, “Risk stratification of ICU patients using topic models inferred from unstructured progress notes,” in *Proc. AMIA Annu. Symp Amer. Med. Informat. Assoc.*, 2012, pp. 505–511.

Authors’, photographs and biographies not available at the time of publication.

QUERIES

Q1. Author: Please verify whether the affiliations of all the authors are OK as set.
Q2. Author: Please provide complete page range in Ref. [7].
Q3. Author: Please provide the department name in Ref. [27].
Q4. Author: Please provide names of all the authors in Ref. [28].