Discovering Shared Dynamics in Physiological Signals: Application to Patient Monitoring in ICU

Li-wei H. Lehman[1], Shamim Nemati[1], Ryan P. Adams[2] and Roger G. Mark[1]

Abstract— **Modern clinical databases include time series of vital signs, which are often recorded continuously during a hospital stay. Over several days, these recordings may yield many thousands of samples. In this work, we explore the feasibility of characterizing the "state of health" of a patient using the physiological dynamics inferred from these time series. The ultimate objective is to assist clinicians in allocating resources to high-risk patients. We hypothesize that "similar" patients exhibit similar dynamics and the properties and duration of these states are indicative of health and disease. We use Bayesian nonparametric machine learning methods to discover shared dynamics in patients' blood pressure (BP) time series. Each such "dynamic" captures a distinct pattern of evolution of BP and is possibly recurrent within the same time series and shared across multiple patients. Next, we examine the utility of this low-dimensional representation of BP time series for predicting mortality in patients. Our results are based on an intensive care unit (ICU) cohort of 480 patients (with 16% mortality) and indicate that the dynamics of time series of vital signs can be an independent useful predictor of outcome in ICU.**

I. INTRODUCTION

Intensive care units (ICUs) are among the most important components of the health care system. With the ubiquity of inexpensive high-capacity storage and recording devices, it is becoming possible to continuously gather patient vital signs, such as heart rate and blood pressure [1]. Despite this continuous feed of data, the commonly used acuity scores, such as APACHE and SAPS, are based on snap-shot observations of the patient [2], [3], [4]. However, physiologic systems generate complex dynamics in their output signals that reflect the state of the underlying control systems [6], [7], [8], [9]. Discovering and understanding these dynamical behaviors are of both fundamental and clinical importance [5]. The objective of the current investigation is to consider an approach to the analysis of ICU bed-side monitoring data that is driven by dynamical considerations. In particular, we seek to determine whether the dynamic evolution of blood pressure (BP) contains usefully-predictive information for survival and mortality, beyond that contained in the SAPS-I [2] score alone.

[1]L. H. Lehman, S. Nemati, and R. G. Mark are with the Massachusetts Institute of Technology, 45 Carleton Street, Cambridge, MA 02142, USA. lilehman@mit.edu

[2]R. P. Adams is with Harvard School of Engineering and Applied Sciences, 33 Oxford Street, Cambridge, MA 02138, USA

Time series of BP can exhibit oscillations on the order of seconds (e.g., due to the baroreflex control mechanism, variations in sympathovagal balance, etc.) to hours (e.g., as a consequence of fever, blood loss, myocardial infarction, etc.). While clinicians often use snap-shot measurements of BP to diagnose abnormally high or low values (typically in comparison to population-level averages), tracking the patient-specific fluctuations in BP may provide extra prognostic value. Moreover, one may reasonably assume that the dynamic patterns in BP are driven by responses of the underlying control systems to external (e.g., drugs) and internal (e.g., disease) perturbations. It would therefore follow that patients who are subjected to similar perturbations will exhibit similar dynamics.

In this work, we utilize a Bayesian nonparametric technique [10] to discover shared dynamical patterns within a patient cohort. An inherent assumption of the technique is that individual patients can take on a variable number of dynamical patterns, some of which are specific to the patient and some of which are shared across the population.

II. METHODOLOGY

We assume that at the population level patients share a number of dynamical behaviors (or modes). We learn these shared temporal structures in the BP time series of 480 ICU patients using a Bayesian nonparametric learning approach. The proportion of time an individual spends within each dynamic mode depends on his/her underlying "state of health", and thus is predictive of the eventual mortality and survival outcome.

A. Patient Selection and Data Preparation

The MIMIC II waveform database (version 2) [1], available via PhysioNet [11], [12], includes approximately 4,000 sets of high resolution physiological waveforms with associated minute-by-minute vital sign trends. This study includes adult patients from the MIMIC II waveform database (version 2) with clinical information, and with at least 8 hours of continuous minute-by-minute invasive BP trends during the first 24 hours of their ICU stays. Patients with more than 15% of missing or invalid (i.e., outside physiologically plausible bounds of 40 to 250 mmHg for systolic pressures) BP samples were excluded. For the rest of the subjects, linear interpolation was used to fill in the missing values.

Invasive BP measurements were extracted for 480 adult patients (meeting the above criteria) from the first 24 hours of their ICU stays. The data contain approximately 9,700 hours of minute-by-minute systolic BP measurements (20.2

hours per patient on average). 16% of patients in this cohort died before hospital discharge.

B. Bayesian nonparametric multiple time series modeling

Our approach to discovery of shared dynamics among patients is based on the beta process autoregressive HMM (BP-AR-HMM) technique of Fox et al. [10]. Briefly, the technique assumes that there exists a library of possible dynamic behaviors (i.e., a set of AR coefficients and the associated noise covariance) among a set of related time series, and each time series can take on a subset of these behaviors. The beta process provides a systematic Bayesian framework for encouraging dynamics to be shared across the patient cohort.

We modeled each BP time series as a switching AR(6) process. We use the same notations as in [10] to describe our parameter settings. Following [10], we used a Gamma(1,1) hyperprior on γ, which specifies the concentration parameter for the symmetric Dirichlet prior on each time series' Markov switching dynamics; the κ hyperparameter determines the preference for self-transition and it was given a Gamma(100,1) prior. The hyperparameter α, which specifies the distribution over the total number of modes, was given a Gamma(0.1,1) prior. For Markov chain Monte Carlo (MCMC) inference with Metropolis–Hastings (following [10]), the gamma proposals used $\sigma_\gamma^2=1$ and $\sigma_\kappa^2=50$. We ran five different initializations of MCMC for 1000 iterations, and evaluated each model using the output from the 1000th iteration.

C. Evaluation Methods and Statistical Analysis

For each patient, we constructed a feature vector based on the top ten most common (most number of samples) dynamic modes discovered by the BP-AR-HMM. Specifically, the proportion of time a patient spent in each of the top ten dynamic modes ("mode proportions" from now on) was used to construct a feature vector for predicting a patient's underlying "state of health".

Univariate and multivariate logistical regressions were performed to find correlations between mode proportions and hospital mortality. For multivariate analysis, SAPS-I and nine co-morbidity variables (presence of congestive heart failure, cardiac arrhythmia, valvular disease, liver disease, lymphoma, metastatic cancer, obesity, weight loss, and fluid electrolyte) were added as potential confounding factors. For the final multivariate logistic regression model, a forward search algorithm was performed to find the optimal model in which all contained variables were statistically significant predictors of mortality (i.e., with a p value less the 0.05). One patient whose co-morbidity variables could not be determined was excluded from the multivariate analysis.

For all logistic regression analysis, odds ratios for the mode proportion variables were per 10% increase in mode proportions. The overall area under the receiver operating characteristic curve (AUC) of the univariate/multivariate model, and the Hosmer-Lemeshow p value (HL p value) were reported to assess the model fit.

Hospital mortality prediction performance was evaluated as the average AUC from 10 different random initializations of 10-fold cross validations using mode proportions as well as SAPS-I as inputs to multivariate logistic regression. We report the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) corresponding to the point on the receiver operating characteristic (ROC) curve that had the minimum distance to the maximum sensitivity and specificity.

III. RESULTS

We ran the BP-AR-HMM 5 times with different initializations for 1000 iterations as described in the Methodology section. The average hospital mortality prediction performance (10-fold cross-validated AUC averaged across 10 different random test/training set configurations) from these five different runs of BP-AR-HMM ranged from 0.63 ($\pm$ 0.01) to 0.68 ($\pm$ 0.02) using mode proportions alone. In the following, we report the performance from the model that generated the median performance (AUC 0.65 $\pm$ 0.01) using mode proportions alone. The total number of modes generated by this model was 22.

A. Univariate Analysis

Table I summarizes the results of univariate analysis using each of the mode proportion as a predictor variable for hospital mortality; univariate analysis using SAPS-I was presented as a comparison. Within our patient cohort, the SAPS-I score yielded an AUC of 0.66. Notably, one of the modes (mode 2) single handedly performed equally as well as the SAPS-I score. Moreover, five of the modes had a significant predictive power ($p < 0.05$), two of which had a greater than one odds ratio and therefore were associated with an increased chance of mortality ("unhealthy modes"). The remaining three had a smaller than one odds ratio and thus were associated with survival ("healthy modes").

Mode	P-Val	Odds Ratio (95% CI)	AUC
SAPS-1	0.0000	1.15 (1.09 1.22)	0.66
2	0.0000	1.45 (1.25 1.69)	0.66
9	0.0001	1.41 (1.19 1.68)	0.64
4	0.0031	0.63 (0.47 0.86)	0.61
10	0.0066	0.43 (0.23 0.79)	0.60
6	0.0411	0.66 (0.44 0.98)	0.59
3	0.0781	0.78 (0.59 1.03)	0.56
7	0.1019	0.74 (0.52 1.06)	0.59
1	0.1792	0.88 (0.73 1.06)	0.54
8	0.4650	1.14 (0.80 1.63)	0.51
5	0.9372	1.01 (0.74 1.38)	0.50

TABLE I

UNIVARIATE ANALYSIS.

B. Multivariate Analysis

We first performed multivariate analysis on the SAPS-I score and the co-morbidities described in section II C, which yielded an AUC of 0.75 (HL Pvalue = 0.18). Next, we built a separate multivariate logistic regression model for each of

the top ten dynamic modes adjusted for SAPS-I and co-morbidities; for each mode, we report the adjusted p value, odds ratios (OR, with 95% confidence interval). The results presented in Table II indicate that even after adjustment for SAPS-I score and co-morbidities, three of the modes (modes 2, 10, and 9) were significant predictors of patients' outcome. In particular, increased proportions of modes 2 and 9 were associated with higher hospital mortality. Increased proportion of mode 10 was associated with a decreased risk of hospital mortality.

Mode	P-Val	Odds Ratio (95% CI)	AUC	HL P-val
2	0.0002	1.38 (1.16 1.65)	0.78	0.88
10	0.0021	0.34 (0.17 0.68)	0.77	0.82
9	0.0163	1.28 (1.05 1.56)	0.76	0.40
4	0.0625	0.74 (0.53 1.02)	0.76	0.27
3	0.1551	0.80 (0.59 1.09)	0.75	0.37
6	0.2332	0.77 (0.50 1.18)	0.75	0.23
7	0.3001	0.81 (0.54 1.21)	0.75	0.03
8	0.4425	1.17 (0.79 1.74)	0.74	0.07
5	0.7199	0.94 (0.66 1.33)	0.74	0.16
1	0.7407	0.96 (0.78 1.20)	0.75	0.08

TABLE II

MULTIVARIATE ANALYSIS. EACH ROW IS A SEPARATE MULTIVARIATE MODEL.

	P-Val	Odds Ratio (95% CI)
SAPS-I	0.0000	1.17 (1.10 1.24)
Mode 2	0.0004	1.35 (1.14 1.60)
Lymphoma	0.0006	12.86 (3.00 55.11)
Mode 10	0.0040	0.37 (0.19 0.73)

TABLE III

FINAL MULTIVARIATE MODEL. AUC=0.769, HL PVAL=0.76.

Table III summarizes the variables selected by the forward search technique for inclusion in our final multivariate logistic regression model. Two mode proportion variables were included in the final model, as were SAPS-I and the presence of lymphoma. The AUC of the final multivariate model was 0.769 (with Hosmer-Lemeshow p value of 0.76, indicating a good model fit).

Application of ten-fold cross-validation demonstrated that the performance of the mode proportions is similar to that of SAPS-I, and furthermore combining features from the two predictors results in an improved prediction power (see Table IV). Notably, the combined predictor have higher specificity and PPV (i.e., have a lower false alarm rate) than SAPS-I.

	SAPS-I	Mode Prop	Mode Prop + SAPS-I
Sensitivity	0.64 (± 0.04)	0.59 (± 0.05)	0.63 (± 0.02)
Specificity	0.57 (± 0.04)	0.70 (± 0.06)	0.76 (± 0.03)
PPV	0.22 (± 0.01)	0.28 (± 0.02)	0.34 (± 0.02)
NPV	0.89 (± 0.01)	0.90 (± 0.01)	0.91 (± 0.00)
AUC	0.65 (± 0.00)	0.65 (± 0.01)	0.73 (± 0.01)

TABLE IV

PREDICTION PERFORMANCE FOR HOSPITAL MORTALITY.

C. Identifying High-Risk Patient Subgroups Using Mode Proportions

We applied hierarchical clustering to categorize patients according to the proportion of time they spent within each of the top ten modes. As shown in Fig. 1, six distinct subgroups were discovered. Table V describes mortality risk for each subgroup with respect to the overall population.

We sorted the patients based on the proportions of time they spent in the top "healthiest" (mode 4) and "unhealthiest" (mode 2) dynamics (based on univariate analysis). Figures 2 and 3 show example BP time series and mode assignments (color-coded) for three patients with the highest proportions of "unhealthy" and "healthy" modes respectively. One patient (patient 871) with high proportions of the unhealthy mode in Fig. 2 expired at the end of the hospital stay. Interestingly, it appears that the "healthy" mode (e.g., in green) exhibits higher variability than the "unhealthy" mode (e.g., in blue).
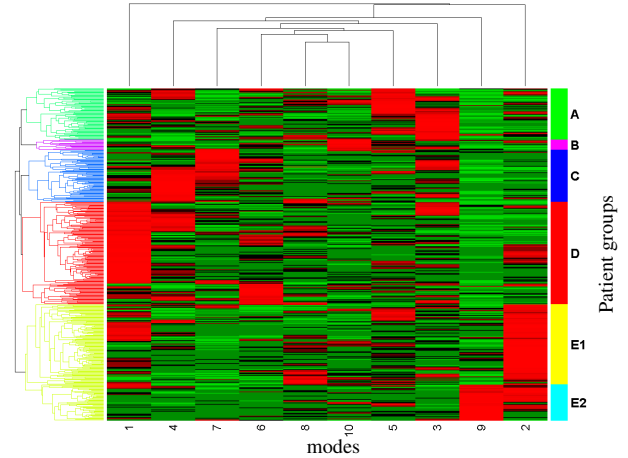

Fig. 1. Hierarchical clustering based on mode proportions. The horizontal axis represents the top 10 modes, and the vertical axis includes the 6 discovered subgroups. The color intensity at the intersection of a group and a mode is indicative of the proportion of time patients spent within the given mode: Light Red (very often), Dark Red (often), Black (sometime), Dark Green (rarely), Light Green (almost never).

Group	N	Mortality
Group A	74	15%
Group B	15	7%
Group C	75	9%
Group D	148	7%
Group E1	116	27%
Group E2	52	35%

TABLE V

HOSPITAL MORTALITY OF SUBGROUPS DEFINED BASED ON HIERARCHICAL CLUSTERING ON INFERRED MODE PROPORTIONS.

IV. DISCUSSION AND CONCLUSIONS

The goal of current work was to determine whether the dynamics of time series when combined with traditional indices of patient acuity scores can provide a more accurate assessment of patient survival/mortality (or more generally
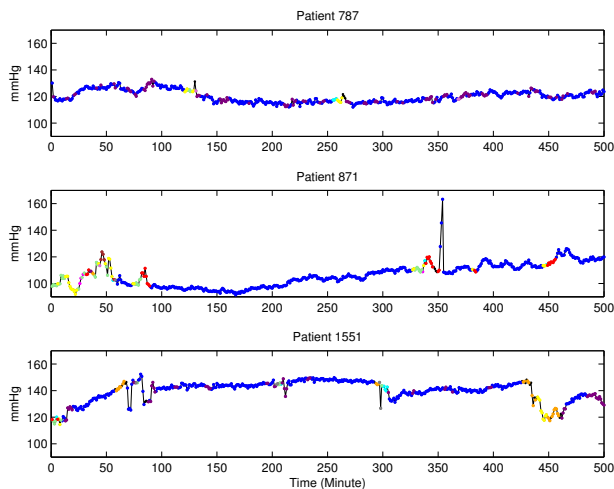
Fig. 2. Example BP time series (sample color coded by their mode assignment) from three patients with high proportions of "unhealthy" modes. Patient 871 expired at the end of the hospital stay. Legend for selected modes: Mode 1 (Purple), Mode 2 (Blue), Mode 4 (Green), Mode 5 (Red), Mode 6 (Light Green), Mode 9 (Magenta), Mode 10 (Cyan).
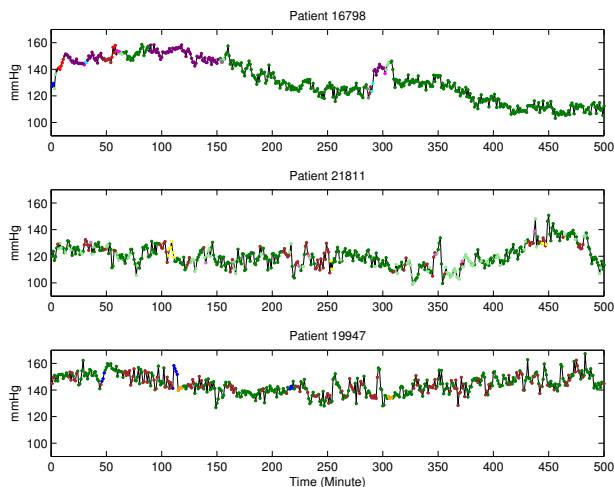

Fig. 3. Example BP time series (sample color coded by their mode assignment) from patients with high proportions of "healthy" modes. See Fig. 2 for color legends.

patient health state). We described a framework for extracting dynamics from time series of BP that learns sharing of dynamical behavior across a patient cohort. Using multivariate logistic regression and forward search techniques we demonstrated that dynamics of time series (mode proportions) may have independent prognostic value, beyond that of SAPS-I score and other co-morbidity. In particular, some of the dynamical modes are associated with improved patient survival and therefore provide complementary information to the SAPS-I severity score. Notably, combining SAPS-I with the mode proportions resulted in significantly improved prediction performance, primarily driven by an enhanced specificity of prognosis.

Our future work will involve assessing the utility of the discovered dynamical behaviors after including all available clinical data (lab tests, medication records, nursing notes, etc). Moreover, recent works by Celi et al. [13] have indicated that customization of mortality prediction techniques to specific patient populations (e.g., those with acute kidney injury) may result in significant improvement in prognostic performance. Although in this work we looked at a single vital sign, the utilized framework allows for modeling of multivariate time series, which may prove more informative in certain patient populations. For instance, patients with cardiovascular related complications may better be represented by the dynamics of the interaction of their heart rate and BP.

Although here we framed the problem in terms of a binary survival/mortality outcome, our ultimate goal is to construct an index of patient health state. To this end, it will be interesting to track a patient's progress during each passing hour of ICU stay.

References

[1] M. Saeed, M. Villarroel, A. Reisner, G. Clifford, L. Lehman, G. Moody, T. Heldt, T.H. Kyaw, B. Moody, and R.G. Mark, "Multiparameter intelligent monitoring in intensive care II (MIMIC-II): A public-access ICU database," Crit Care Med, 39(5), 2011.

[2] J.R. Le Gall, P. Loirat, A. Alperovitch, et al. "A simplified acute physiology score for ICU patients," Critical Care Medicine, 12(11), 975-977, 1984.

[3] J.R. Le Gall, S. Lemeshow, F. Saulnier, "A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study," JAMA, 270(24):2957-2963, 1993.

[4] J.E. Zimmerman, A.A. Kramer, D.S. McNair, F.M. Malila, "Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients," Crit Care Med, 34(5):1297-1310, 2006.

[5] A.L. Goldberger, L.A. Amaral, J.M. Hausdorff, P. Ivanov, C.K. Peng, and H.E. Stanley, "Fractal dynamics in physiology: alterations with disease and aging," Proc Natl Acad Sci U S A, vol. 99 Suppl 1, pp. 2466-72, Feb 19 2002.

[6] 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, pp. 55-67, Jul 2011.

[7] M. Costa, A.L. Goldberger, and C.K. Peng, "Multiscale entropy analysis of complex physiologic time series," Phys Rev Lett, vol. 89, p. 068102, Aug 5 2002.

[8] 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. 461-5, Jun 3 1999.

[9] 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," Am J Physiol, vol. 276, pp. R905-12, Mar 1999.

[10] E.B. Fox, E.B. Sudderth, M.I. Jordan, A.S. Willsky, "Sharing Features among Dynamical Systems with Beta Processes," Proceeding of Neural Information Processing Systems, Vancouver, Canada, December 2009.

[11] http://www.physionet.org/

[12] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.K. Peng, H.E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation, 101(23):e215-e220, June 13 2000.

[13] L.A. Celi, R.J. Tang, M. C. Villarroel, G.A. Davidzon, W.T. Lester, H.C. Chueh, "A Clinical Database-Driven Approach to Decision Support: Predicting Mortality Among Patients with Acute Kidney Injury," Journal of Healthcare Engineering, vol. 2(1), pp. 97-109, 2011.