2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340317

Risk Estimation for ICU Patients with Personalized Anomaly-Encoded Bedside Patient Data

Kai Wu[1,3], Ee Heng Chen[2,3],
Felix Wirth[1], Keti Vitanova[1], Rüdiger Lange[1] and Darius Burschka[2,3]

***Abstract*— We propose a novel framework to estimate intensive care unit patients' health risk continuously with anomaly-encoded patient data. This framework consists of two modules. In the first module, we use Gaussian process models to learn change trend and day-night circulation in temporal patient data and annotate abnormal data. Such models provide dynamically adaptable bedside patient monitoring instead of conventional threshold-based monitoring. In the second module, we use the abnormal data together with the learned Gaussian models to estimate patients' risk level by predicting their in-hospital mortality and remaining length of stay in ICU ward. We show that prediction models with anomaly-encoded data have better performance than those with raw patient measurements, and they are comparable with state-of-art prediction models.**

I. INTRODUCTION

Intensive care unit (ICU) patients face an increased risk of clinical deterioration [1], [2]. A continuous and close monitoring of inpatients is therefore crucial. Currently, bedside patient monitoring devices are widely used in ICU wards to monitor patients' vital signs and trigger an alarm when a vital sign exceeds the pre-set thresholds. As stated in our previous research [3], vital signs have varying normal ranges from person to person and response differently to ongoing activities. Such rule-based monitoring method yields high rate of false alarms that interrupts nursing staff [4]. Beside raw data monitoring, multi-parameter scoring methods, e.g., National Early Warning System (NEWS), Modified Early Warning System (MEWS), and Acute Physiology And Chronic Health Evaluation (APACHE), are widely applied to estimate patient's deterioration risk according to a set of criteria chosen by medical experts. These methods are typically criticized by their lack of patient-specificity, empirically chosen threshold values, and disregarding temporal trends and the history of vital-sign values preceding the current set of observed values [5].

In recent decades, machine learning (ML) methods have been proposed to develop personalized patient monitoring systems that comprehensively estimate patients' health condition with demographics information and physiological measurements. However, characteristics of medical data, e.g., irregularly-spaced time-series measurements and high missing-value proportion, are still problematic for many ML models. Data interpolation, imputation or feature extraction are usually required before applying any learning model. Gaussian process (GP) is able to handle these problems due to its probabilistic framework and imputes the missing values with uncertainty estimation [6], [7]. Another advantage of GP is that prior knowledge and constraints can be added to the model by selecting, combining and tuning the kernels [8]. This will allow medical staff to feed their expert knowledge to the model and have more control of the learning process. Works are done to apply GP models to detect anomaly in vital signs [5], [6], predict clinical deterioration [9], [10], and combine them with neural networks for further classification tasks [11], [12].

Current works on anomaly detection for bedside patient data mainly focus on learning personalized trends and finding abnormal sudden changes in the signals. On this basis, we investigate day-night circulation as another criteria to model normal ranges of physiological data and use detected abnormal data as input for patient risk estimation. The following are the main contributions of this work:

- We demonstrate an interpretable personalized anomaly detection for physiological measurements by learning long-term trend and day-night seasonality with GP model.
- We design an anomaly-encoding method using results from the learned Gaussian models for further analysis tasks.
- We show that length-of-stay (LOS) and mortality prediction models using anomaly-encoded data as input have better performance than those using raw data as input.

II. METHOD

A. Gaussian Process Modelling

Gaussian process describes the distribution of an arbitrary function, defined as

$$f(x) \sim \mathcal{GP}(\mu(x), k(x, x')), \tag{1}$$

where $\mu(x)$ is the mean function,

$$\mu(x) = \mathbb{E}[f(x)] \tag{2}$$

and $k(x, x')$ is the kernel, i.e., the covariance function.

$$k(x, x') = \mathbb{E}[(f(x) - \mu(x))(f(x') - \mu(x'))]. \tag{3}$$

[1]Kai Wu, Felix Wirth, Keti Vitanova, Rüdiger Lange are with the German Heart Center Munich, Lazarettstrasse 36, 80636 Munich, Germany, {wuk,wirth,vitanova}@dhm.mhn.de

[2]Ee Heng Chen and Darius Burschka are with MIRMI - Munich Institute of Robotics and Machine Intelligence, Technical University of Munich, Georg-Brauchle-Ring 60-62, 80992 Munich, Germany, eeheng.chen@tum.de,burschka@cs.tum.de

[3]Kai Wu, Ee Heng Chen and Darius Burschka are with the Machine Vision and Perception Group, Department of Computer Engineering, TUM School of Computation, Information and Technology, Technical University of Munich, Parkring 13, 85748 Garching, Germany

979-8-3503-2447-1/23/$31.00 ©2023 IEEE

There are different expressions of kernels to model characteristics of a function, e.g., linearity, smoothness and periodicity. In this work, we choose a radial basis function (RBF) kernel and a periodic kernel to model the long-term trend and day-night seasonality in patient physiological data. The RBF kernel is defined as

$$k_{RBF}(x, x') = \sigma^2 \exp\Big(- \frac{(x - x')^2}{2l^2} \Big), \tag{4}$$

where σ^2 is the variance, and l is the length scale. The larger the l, the smoother the approximated function. The periodic kernel is defined as

$$k_{periodic}(x, x') = \sigma^2 \exp\Big(- \frac{\sin(\frac{\pi}{T}(x - x'))^2}{2l^2} \Big), \tag{5}$$

where T is the period parameter.

B. Anomaly Encoding

After fitting the GP model to patient data, the mean and covariance functions are optimized to describe distribution of each individual physiological signal. We define outliers as measurements outside the 95% confidence interval (CI) of the Gaussian model. The raw physiological time-series data is then encoded according to Alg. 1, where $x_{i,t}$ is a measurement at time t of the i-th signal, $\mu_{i,t}$ is the distribution mean, $bu_{i,t}$ and $bl_{i,t}$ are the upper and lower boundary of the model's 95% CI, and $trend_{i,t}$ is the change trend of $\mu_{i,t}$:

$$trend_{i,t} = \mu_{i,t+1} - \mu_{i,t}. \tag{6}$$

This anomaly encoding contains information about the expected normal interval (mean and confidence interval) of individual physiological signal. Measurements within in this normal interval are removed from the data representation, where as abnormal measurements are described as their difference to the expected values ($\mu_{i,t}$).

Algorithm 1: Anomaly encoding

Input: $x_{i,t}$, $\mu_{i,t}$, $trend_{i,t}$, $bu_{i,t}$, $bl_{i,t} \in \mathbb{R}^1$
Output: $\mathbf{e}_{i,t} \in \mathbb{R}^6$

$\mathbf{e}_{i,t}$ = [$\mu_{i,t}$, $bu_{i,t} - bl_{i,t}$, 0, 0, 0, 0]
if $x_{i,t} < bl_{i,t}$ **then**
  if $trend_{i,t} < 0$ **then**
    $\mathbf{e}_{2,t} = x_{i,t} - \mu_{i,t}$
  else
    $\mathbf{e}_{3,t} = x_{i,t} - \mu_{i,t}$
else if $x_{i,t} > bu_{i,t}$ **then**
  if $trend_{i,t} < 0$ **then**
    $\mathbf{e}_{4,t} = x_{i,t} - \mu_{i,t}$
  else
    $\mathbf{e}_{5,t} = x_{i,t} - \mu_{i,t}$

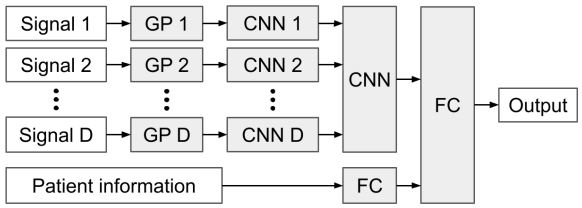


Fig. 1: Patient risk estimation framework.

C. Patient Risk Estimation Model

In this work, we estimate patients' risk by predicting their length of stay in ICU ward and in-hospital mortality. Random Forest (RF) and XGBoost are widely applied in clinical prediction tasks and achieve satisfying performance. However, instead of using raw time-series data, these two models usually require feature engineering to obtain low-dimensional input [13], [14]. Therefore, we implemented a two-layer Convolutional Neural Network (CNN) that works with raw and encoded data (see Fig. 1). Model configurations for RF, XGBoost and CNN are summarized in Tab. I. The model input consists of static patient information of dimension D_S and temporally aligned physiological signals of dimension $D_T \times T$, where D_T is the data dimension and T is the length of temporal sequences. In the first CNN layer, we extract feature for each time-series signal separately and compress it in the temporal dimension by applying a convolutional layer to each input signal. In the second layer, we learn correlation between all time-series signals by convolving the output of the first layer with one single convolutional layer. The output of the second CNN layer is concatenated with static patient information and fed into a fully connected layer to output the prediction result.

III. EXPERIMENTS

A. Data Preprocessing

Experiments in this paper are conducted with MIMIC IV dataset [15]. According to our cohort selection pipeline (see Fig. 2), we end up with 33778 unique ICU admissions and 27952 unique patient subjects. The number of ICU survivors is 24483 (87.6%). We use the entire cohort for mortality prediction models and the survivor cohort for LOS prediction models. For each prediction task, the corresponding cohort is split into a training cohort (80%) and a test cohort (20%), where a 5-fold cross validation is applied in the former for model training (input data statistics see Tab. II).

Each data sample contains patient information and time-series measurements of patient's physiology (see Tab. III). Time-series data are temporally aligned to a sampling frequency of 1 measurement per hour. Then we slice the aligned data sample into non-overlapping 48-hour segments as input of prediction models. Numeric data is normalized according to Eq. 7, whereas categorical data is represented by one-hot encoding.

$$x_{normalized}(i, t) = \frac{x(i, t) - \bar{x}_i}{\sigma_i} \tag{7}$$

TABLE I: Model Configuration.

Random Forest and XGBoost models are implemented with XGBoost library. CNN models are implemented with Pytorch library. Input dimension D_T of CNN is 38 for raw data and 228 for encoded data.

Model	Parameter
Random Forest	learning_rate = 1 max_depth = 6 number_of_trees = 5000 L_2_regularization = 10^{-5} objective_function = 'binary:logistic' tree_method = 'gpu_hist' subsample = 0.8 colsample_bynode =0.8
XGBoost	learning_rate = 0.3 max_depth = 6 number_of_trees = 5000 objective_function = 'binary:logistic' tree_method = 'gpu_hist'
CNN	input_dimension = D_T cnn_num_layer = 2 cnn_num_hidden_neuron = [380, 256] cnn_kernel_size = [3, 3] cnn_stride_size = [1, 1] cnn_group_size = [38, 1] num_fc_layer = 1 num_hidden_neuron_fc = 512 droup_out_fc = 0.5 learning_rate = 10^{-5} batch_size = 64

TABLE II: Input data statistics.

Mortality prediction			**LOS prediction**		
Number of subjects: 27952			Number of subjects: 24483		
train	validation	test	train	validation	test
17884	4472	5596	15648	3912	4923
Number of samples: 95868			Number of samples: 78127		
train	validation	test	train	validation	test
61371	15546	18951	49704	12723	15700

where $\bar{x}_i$ and σ_i are the mean and standard deviation of the i-th signal of train cohort. After normalization, we keep the raw data and compute anomaly-encoded representation for numeric time-series data.

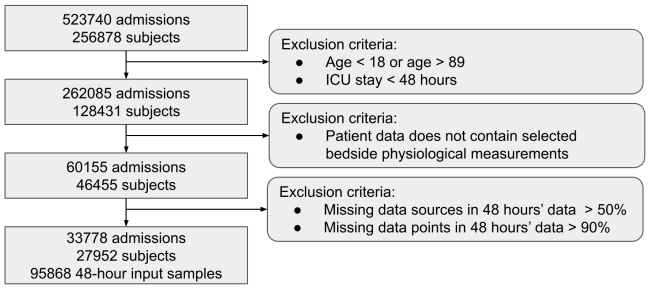


Fig. 2: Cohort selection pipeline.

B. GP-based Anomaly Detection

In this section, we demonstrate the GP modelling for patient physiological data. We first remove sudden change peaks in the raw data, which can interfere with fitting process of GP model (see red triangle markers in Fig. 3). The kernel of our model is a combination of a RBF kernel (see Eq. 4) and a periodic kernel (see Eq. 5). To extract the long-term trend and day-night circulation in the raw data, we heuristically assume that: 1) the trend component should be smooth, to avoid over-fitting to noises, and 2) the period length of the day-night seasonality component should be roughly around 24 hours. We insert these assumptions here by constraining kernel parameters:

$$24 < l_{RBF} < 48 \tag{8}$$

$$0.5 < l_{periodic} < 10 \tag{9}$$

$$22 < T_{periodic} < 26 \tag{10}$$

We constrain RBF kernel to have larger length scale to learn the long-term trend, and periodic kernel to have smaller length scale to avoid averaging out local patterns. Period length of the periodic kernel is constraint between 22 hours to 26 hours.

Fig. 3 shows an example GP modeling of blood pressure signals. Black cross markers are normalized raw data. The trend kernel learns a smooth change trend in the signal, while periodic kernel captures local repetitive patterns. Data points outside model's confidence region (light blue area) are considered as abnormal points.

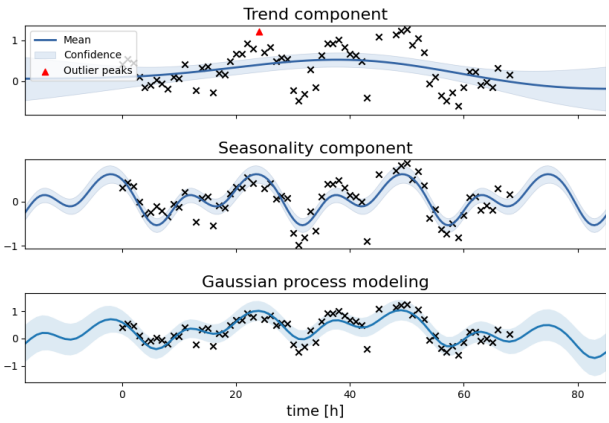


Fig. 3: GP modelling of normalized systolic blood pressure.

C. LOS and Mortality Prediction

To validate whether our anomaly detection and encoding can capture abnormal information in patients' physiology, we compare encoded data with raw data regarding several patient risk estimation tasks: mortality prediction and LOS prediction. If the anomaly-encoded data contains critical information in the raw data, it is expected to achieve similar or better performance in the prediction tasks. We use data samples from the entire patient cohort for mortality prediction task, and only the survivor cohort to predict if the patient can be discharged within 7, 14, or 21 days.

We consider RF and XGBoost as our baseline models, where time-series signals are represented by their average or

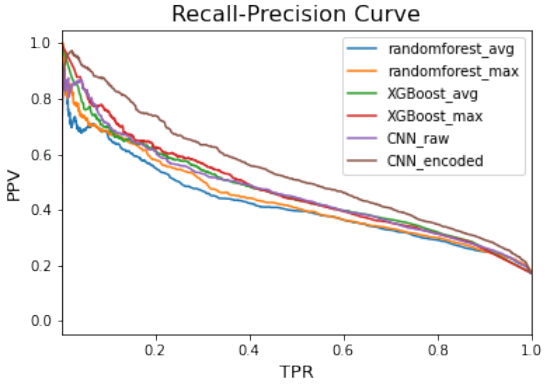
TABLE III: Input features.

Data Category	Data Type	Input Features
Patient Information	numeric categorical	age, admission weight gender, admission type
Vital Sign	numeric categorical	heart rate, oxygen saturation, respiration rate, systolic/mean/diastolic blood pressures, temperature, O2 flow, FiO2 temperature site, heart rhythm, ectopy type, ectopy frequency
Lab Test	numeric	Chloride (serum), Sodium (serum), Potassium (serum), Creatinine (serum), HCO3 (serum), BUN, Anion gap, Hematocrit (serum), Glucose (serum), Hemoglobin, Platelet Count, WBC, Magnesium, Phosphorous, Calcium non-ionized, PT, INR, PTT, Glucose, FS (range 70 -100), Lactic Acid, ALT, AST, Total Bilirubin, Alkaline Phosphate, PH (Arterial), TCO2 (calc) Arterial, PO2 (Arterial), PCO2 (Arterial), Arterial Base Excess

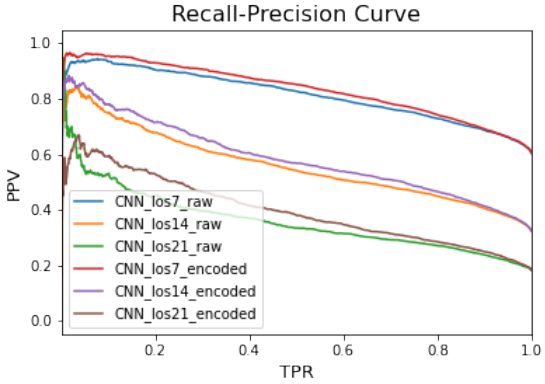
TABLE IV: Model performance scores.

Precision and specificity scores are compared at a recall score of 80%. XGB: XGBoost, enc.: encoded, mort.: mortality.

				AUROC	AUPRC	accuracy	precision80	specificity80
RF	avg	mort.	0	0.773	0.421	0.837	0.291	0.595
		los	7	0.729	0.803	0.674	0.701	0.477
		los	14	0.708	0.538	0.707	0.420	0.471
		los	21	0.700	0.354	0.820	0.253	0.472
	max	mort.	0	0.780	0.438	0.838	0.299	0.613
		los	7	0.731	0.803	0.679	0.708	0.495
		los	14	0.711	0.536	0.707	0.424	0.478
		los	21	0.708	0.359	0.821	0.260	0.493
XGB.	avg	mort.	0	0.801	0.467	0.839	0.320	0.649
		los	7	0.760	0.826	0.704	0.734	0.556
		los	14	0.732	0.571	0.720	0.437	0.505
		los	21	0.717	0.370	0.819	0.254	0.457
	max	mort.	0	0.797	0.475	0.841	0.311	0.631
		los	7	0.746	0.815	0.690	0.718	0.519
		los	14	0.720	0.555	0.712	0.428	0.487
		los	21	0.703	0.357	0.817	0.253	0.476
CNN	raw	mort.	0	0.795	0.464	0.757	0.312	0.636
		los	7	0.755	0.815	0.684	0.729	0.546
		los	14	0.733	0.560	0.682	0.448	0.528
		los	21	0.720	0.361	0.686	0.271	0.518
	enc.	mort.	0	**0.828**	0.536	0.757	0.348	0.690
		los	7	**0.774**	0.835	0.698	0.742	0.574
		los	14	**0.751**	0.586	0.689	0.466	0.560
		los	21	**0.747**	0.396	0.688	0.285	0.551



(a) Precision-Recall curve for mortality prediction.



(b) Precision-Recall curve for LOS prediction.

Fig. 4: Precision-Recall curves.

maximal value in the 48-hour window. Our CNN model is trained with the raw and encoded time-series data.

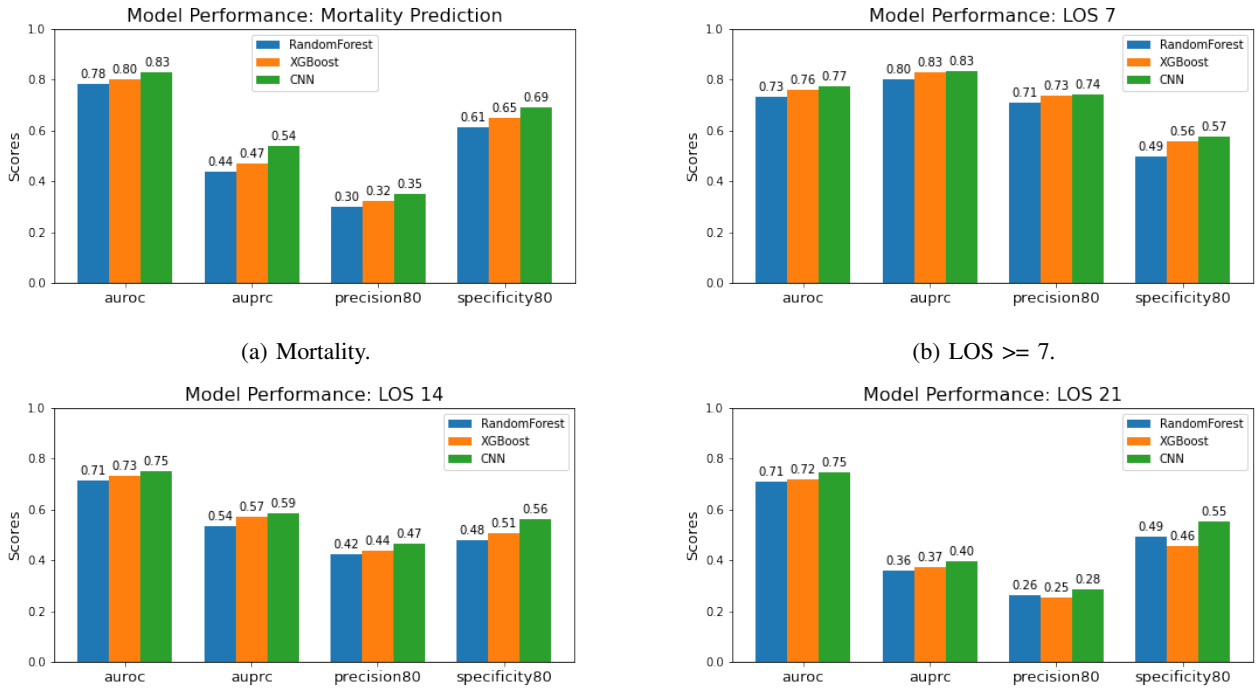
We train each model in a 5-fold cross validation scheme and ensemble the resulted models for prediction in the test cohort. Model performance scores are summarized in Tab. IV. The overall model performance is evaluated by Area Under Receiver Operating Characteristic Curve (AUROC). We observe that RF and XGBoost models have better performance with maximal and average data representation respectively. CNN trained with encoded data achieves an AUROC of 0.83 and an Area Under Precision Recall Curve (AUPRC) of 0.54 for mortality prediction task. It effectively increases AUPRC of CNN trained with raw data (see Fig. 4a), and it has overall better performance than RF and XGBoost models (see Fig. 5a). The same applies to LOS prediction models (see Fig. 4b, 5b, 5c, 5d).

IV. CONCLUSION AND FUTURE WORKS

In this work, we use GP model to estimate the expected physiological ranges for individual patient. Such model, which takes personalized change trends and 24-hour circulation in physiological signals into account, is promising to reduce false alarm rate in bedside patient monitoring compared with current threshold-based monitoring methods.

Based on the GP model, we introduced an anomaly-encoding method. It proved to effectively improve model performance in mortality and LOS prediction tasks compared with the raw data. With commonly available patient information, bedside physiological measurements and lab test, our CNN model trained with encoded data achieves an AUROC score of 0.83 and an AUPRC score of 0.54 for mortality prediction. Tuning the classification threshold to detect 80% mortality cases, the model has a false alarm rate of 65% and correctly recognizes 69% survivor cases.

Understanding more criteria that influence patient physiological signals can be an effective way to improve current



(a) Mortality.

(b) LOS >= 7.

(c) LOS >= 14.

(d) LOS >= 21.

Fig. 5: Comparison of Model performance scores.
Precision and specificity scores are compared at a recall score of 80%.

bedside patient monitoring and reduce false alarms. With more densely recorded patient data, it is also possible to use Gaussian models to learn the transient patterns, e.g., how patients' vital signs responses to certain medicines, treatments and physical activities. These transient models can be integrated with our current models to provide a comprehensive personalized event-aware monitoring for patient bedside measurements.

ACKNOWLEDGMENT

The project was funded by the Bavarian State Ministry of Science and Arts within the framework of the "Digitaler Herz-OP" project under the grant number 1530/891 02.

REFERENCES

[1] L. G. Donowitz, R. P. Wenzel, and J. W. Hoyt, "High risk of hospital-acquired infection in the icu patient." *Critical care medicine*, vol. 10, no. 6, pp. 355–357, 1982.

[2] K. Singbartl and J. A. Kellum, "Aki in the icu: definition, epidemiology, risk stratification, and outcomes," *Kidney international*, vol. 81, no. 9, pp. 819–825, 2012.

[3] K. Wu, E. H. Chen, X. Hao, F. Wirth, K. Vitanova, R. Lange, and D. Burschka, "Adaptable action-aware vital models for personalized intelligent patient monitoring," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 826–832.

[4] M.-C. Chambrin, "Alarms in the intensive care unit: how can the number of false alarms be reduced?" *Critical Care*, vol. 5, no. 4, pp. 1–5, 2001.

[5] G. W. Colopy, S. J. Roberts, and D. A. Clifton, "Gaussian processes for personalized interpretable volatility metrics in the step-down ward," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 949–959, 2019.

[6] D. Wong, D. A. Clifton, and L. Tarassenko, "Probabilistic detection of vital sign abnormality with gaussian process regression," in *2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE)*. IEEE, 2012, pp. 187–192.

[7] M. A. Pimentel, D. A. Clifton, and L. Tarassenko, "Gaussian process clustering for the functional characterisation of vital-sign trajectories," in *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2013, pp. 1–6.

[8] D. Duvenaud, "Automatic model construction with gaussian processes," Ph.D. dissertation, University of Cambridge, 2014.

[9] A. M. Alaa, J. Yoon, S. Hu, and M. Van der Schaar, "Personalized risk scoring for critical care prognosis using mixtures of gaussian processes," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 1, pp. 207–218, 2017.

[10] G. W. Colopy, M. A. Pimentel, S. J. Roberts, and D. A. Clifton, "Bayesian optimisation of gaussian processes for identifying the deteriorating patient," in *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE, 2017, pp. 85–88.

[11] J. Futoma, S. Hariharan, and K. Heller, "Learning to detect sepsis with a multitask gaussian process rnn classifier," in *International conference on machine learning*. PMLR, 2017, pp. 1174–1182.

[12] K. Zhang, S. Karanth, B. Patel, R. Murphy, and X. Jiang, "A multi-task gaussian process self-attention neural network for real-time prediction of the need for mechanical ventilators in covid-19 patients," *Journal of biomedical informatics*, vol. 130, p. 104079, 2022.

[13] S. Iwase, T.-a. Nakada, T. Shimada, T. Oami, T. Shimazui, N. Takahashi, J. Yamabe, Y. Yamao, and E. Kawakami, "Prediction algorithm for icu mortality and length of stay using machine learning," *Scientific Reports*, vol. 12, no. 1, pp. 1–9, 2022.

[14] M. Zabihi, S. Kiranyaz, and M. Gabbouj, "Sepsis prediction in intensive care unit using ensemble of xgboost models," in *2019 Computing in Cardiology (CinC)*. IEEE, 2019, pp. Page–1.

[15] A. Johnson, L. Bulgarelli, T. Pollard, S. Horng, L. A. Celi, and R. Mark, "Mimic-iv," *PhysioNet. Available online at: https://physionet.org/content/mimiciv/1.0/(accessed August 23, 2021)*, 2020.