Latent Modeling for Early Prediction of Sepsis from Clinical Data

Kenneth Ezukwoke
Mines Saint-Étienne *
Saint-Etienne, France

Abstract

Sepsis, a life-threatening systemic inflammatory response to infection, remains a global health concern with substantial morbidity and mortality rates. Despite recent medical advancements, sepsis diagnosis remains challenging due to its complex presentation and overlap with other conditions. The inability to accurately and promptly identify sepsis hinders timely treatment initiation, contributing to increased mortality.

To address this challenge, we propose a latent model approach using a generalized controllable variational autoencoder (GCVAE) to project high-dimensional sepsis data into a low-dimensional, well-structured, and disentangled latent space. This latent space effectively captures the underlying patterns and relationships within the data, enabling accurate sepsis prediction one hour prior to clinical diagnosis. Our approach holds the potential to improve sepsis detection and treatment outcomes, ultimately saving lives.

1 Introduction

Sepsis is a life-threatening dysfunction of organs resulting from an unregulated host response to infection (Singer et al., 2016), represents a significant global health issue (Rudd et al., 2020). Despite promising medical progress in recent decades, sepsis continues to be a prevalent cause of in-hospital fatalities. It is linked to exceptionally high rates of mortality and morbidity, placing substantial burdens on healthcare systems worldwide (Rudd et al., 2020). This can be partly attributed to the difficulties associated with promptly recognizing sepsis and initiating timely and appropriate treatment (Ferrer et al., 2014).

An increasing body of research indicates that mortality rates rise with each hour of delayed antimicrobial intervention, emphasizing the critical importance of timely recognition and treatment initiation (Pruinelli et al., 2018; Ferrer et al., 2014). A major obstacle to early recognition lies in differentiating sepsis from other medical conditions (such as inflammation) that share similar clinical signs (e.g., changes in vitals), symptoms (e.g., fever), and molecular manifestations (e.g., dysregulated host response) (Lever and Mackenzie, 2007). Given the systemic nature of sepsis, efforts have been made to propose biological and molecular indicators, referred to as biomarkers, to enhance the precision of sepsis diagnosis and detection. However, despite substantial endeavors to identify suitable biomarkers, there is currently no universally accepted single biomarker or combination thereof for sepsis diagnosis and treatment, primarily due to issues related to sensitivity and specificity.

The early prediction of sepsis holds the potential to save lives. Consequently, the challenge lies in predicting sepsis one hour prior to the clinical diagnosis of sepsis. On the flip side, delayed sepsis prediction poses a potential threat to life, and predicting sepsis in individuals without sepsis (or predicting sepsis very early in those already diagnosed with sepsis) places a strain on limited hospital resources. To address this challenge, we propose the use of a generalized controllable variational autoencoder loss to project the original data into a low-dimensional, well-structured and disentangled space. The restructured space is then used with a classification algorithm to predict early sepsis.

2 Related works

Recent works on Sepsis prediction have successfully forecasted the onset of sepsis 17 hours before the initial administration of antibiotics and 36 hours before meeting the criteria for a sepsis diagnosis (Futoma et al., 2017). This approach was inspired by Li and Marlin (2016), who initially introduced the Gaussian process adapter. This adapter integrates single-task Gaussian processes imputation

*Institut Fayol- Mathematical and Industrial Engineering

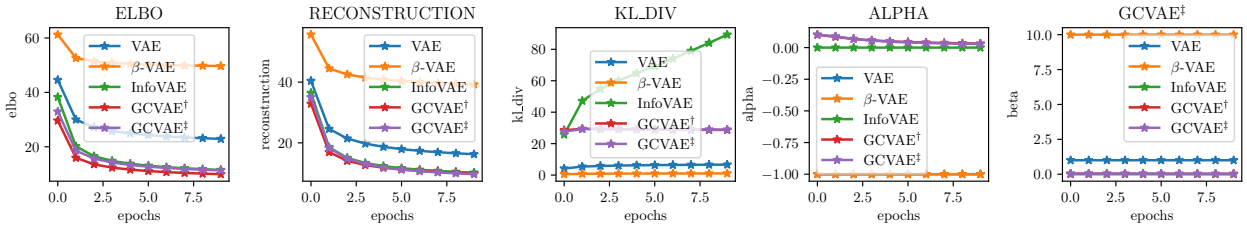

Figure 1: Latent-15: Behavior of the different losses for 10 training steps.

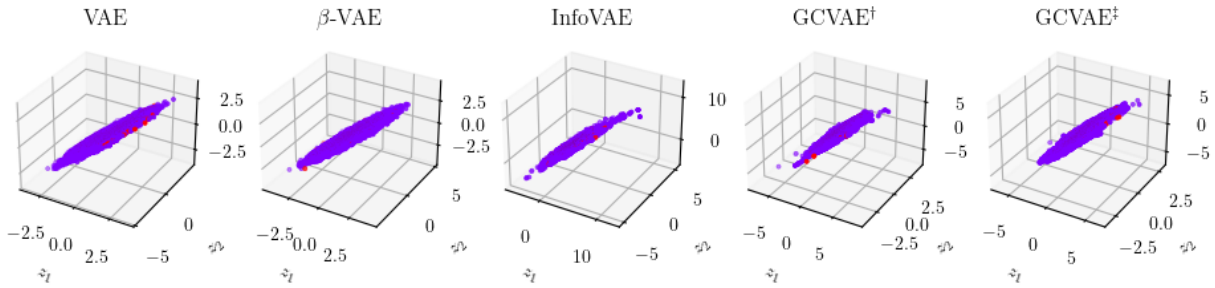

Figure 2: 3D-Projection of the original data into the Latent space. Notice the overlapping coloring of Sepsis data-points over non-Sepsis class

with neural networks in an end-to-end learning environment. A more recent study enhanced predictive accuracy by integrating the Gaussian process adapter framework with temporal convolutional networks (MGP-TCN).

This study represents the first application of a variational autoencoding model to the early prediction of Sepsis from clinical data. We use Generalized-Controllable Variational Autoencoder (Ezukwoke et al., 2022) to acquire a re-representation of the original data space, along with its disentanglement properties where clusters exhibit distinct separation of overlapping clusters. The disentangled space is subsequently passed to a classification algorithm for the early prediction of Sepsis.

3 Methodology

Given a d-dimensional input space $\{x_i\}_{i=1}^N \in \mathcal{X}$ consisting of N-independently and identically distributed (**i.i.d**) samples; k-dimensional latent space $\{v_i\}_{i=1}^N \in \mathcal{V}$ (where $k \ll d$) over which a generative model is defined. We assume an empirical prior distribution $p_\theta(v) \sim \mathcal{N}(0, I)$ to infer an approximate posterior distribution $q_\phi(v|x) \sim \mathcal{N}(v|\mu_\phi(x), \sigma_\phi^2(x)I)$, with mean $\mu_\phi(x)$ and variance $\sigma_\phi^2(x)I$ used for re-parameterization sampling of the latent space v (Kingma and Welling, 2014). We model the data using conditional distribution $p_\theta(x|v) \sim \mathcal{N}(x|\mu_\theta(x), \sigma_\theta^2(x)I)$. Let us suppose that the underlying distribution of the input space $p(x)$ follows a normal distribution, and its empirical distribution is denoted by $p_\mathcal{D}(x)$. $I_p(x', v)$ is the joint mutual information space between x' and v generated from the posterior $p_\theta(x|v)$ after obtaining an inference posterior $q_\phi(v|x)$.
The optimization framework proposed in (Ezukwoke et al., 2022) is given as follows:

$$
\begin{aligned}
&\max_{\theta,\phi,\xi^+,\xi^-,\xi_p\in\mathbb{R}} \quad I_p(x', v) \\
&s.t \quad \mathop{\mathbb{E}}_{p_\mathcal{D}} D_{KL}(q_\phi(v|x) \parallel p_\theta(v)) + I_p(x', v) \leq \xi^- \\
&s.t \quad -\mathop{\mathbb{E}}_{p_\mathcal{D}} D_{KL}(q_\phi(v)||p_\theta(v)) \leq \xi^+ \\
&s.t \quad I_p(x', v) \leq \xi_p \\
&s.t \quad \xi_i^+, \xi_i^-, \xi_{ip} \geq 0, \quad \forall i = 1, \ldots, n
\end{aligned} \tag{1}
$$

The expansion of the above equations using sets of Lagrangian multipliers is as follows,

$$
\begin{aligned}
&\mathcal{L}(\theta, \phi, \boldsymbol{\xi^+}, \boldsymbol{\xi^-}, \boldsymbol{\xi_p}, \alpha, \beta, \gamma, \boldsymbol{\eta}, \boldsymbol{\tau}, \boldsymbol{\nu}) \\
&= I_p(x', v) - \beta(\mathop{\mathbb{E}}_{p_\mathcal{D}} D_{KL}(q_\phi(v|x)||p_\theta(v)) + \\
&\quad I_p(x', v) - \xi^+) + \gamma(\mathop{\mathbb{E}}_{p_\mathcal{D}} D_{KL}(q_\phi(v)||p_\theta(v)) + \xi^-) \\
&\quad - \alpha(I(x'; v) - \xi_p) - \boldsymbol{\eta}\xi^+ - \boldsymbol{\tau}\xi^- - \boldsymbol{\nu}\xi_p
\end{aligned} \tag{2}
$$

The negative Lagrangian multipliers pose no challenge since they only exist to eliminate the error terms ξ^+, ξ^-, ξ_p,

$$
\begin{aligned}
&\mathcal{L}(\theta, \phi, \boldsymbol{\xi^+}, \boldsymbol{\xi^-}, \boldsymbol{\xi_p}, \alpha, \beta, \gamma, \boldsymbol{\eta}, \boldsymbol{\tau}, \boldsymbol{\nu}) \\
&= (1 - \alpha - \beta) I_p(x', v) - \beta \mathop{\mathbb{E}}_{p_\mathcal{D}} D_{KL}(q_\phi(v|x)||p_\theta(v)) \\
&\quad + \gamma D_{KL}(q_\phi(v)||p_\theta(v)) + (\beta - \boldsymbol{\eta})\xi^+ + (\gamma - \boldsymbol{\tau})\xi^- \\
&\quad + (\alpha - \boldsymbol{\nu})\xi_p
\end{aligned} \tag{3}
$$

We take the gradient over the loss, $\nabla\mathcal{L}$ for ξ^-, ξ^+, ξ_p and apply KKT optimality conditions to obtain,

$$\begin{aligned}\mathcal{L}(\theta,\phi,\alpha,\beta,\gamma) &= (1-\alpha-\beta)I_p(x',v) \\ &\quad - \beta \mathop{\mathbb{E}}_{p_\mathcal{D}} D_{KL}(q_\phi(v|x)||p_\theta(v)) \\ &\quad + \gamma D_{KL}(q_\phi(v)||p_\theta(v)) \qquad (4) \\ &= (1-\alpha-\beta) \mathop{\mathbb{E}}_{v\sim q_\phi(v|x)}[\ln p_\theta(x|v)] \\ &\quad - \beta \mathop{\mathbb{E}}_{p_\mathcal{D}} D_{KL}(q_\phi(v|x)||p_\theta(v)) \\ &\quad + \gamma D_{KL}(q_\phi(v)||p_\theta(v)) \qquad (5)\end{aligned}$$

We set the Lagrangian adaptive hyperparameters as follows,

$$\begin{aligned}\mathcal{L}(\theta,\phi,\alpha,\beta,\gamma) =& (1-\alpha_t-\beta_t) \mathop{\mathbb{E}}_{v\sim q_\phi(v|x)}[\ln p_\theta(x|v)] \\ &- \beta_t \mathop{\mathbb{E}}_{p_\mathcal{D}} D_{KL}(q_\phi(v|x)||p_\theta(v)) \\ &+ \gamma_t D_{KL}(q_\phi(v)||p_\theta(v)) \qquad (6)\end{aligned}$$

The adaptive weight α_t controls the reconstruction error while β_t ensures the posterior latent factor $q_\phi(v|x)$ does not deviate significantly from its prior $p_\theta(v)$. Varying both terms gives us better control of the degree of disentanglement and helps us to understand the parameters affecting density disentanglement. The first term of the loss in Equation 6 with weight $(1-\alpha_t-\beta_t)$ is the reconstruction loss, the second term with weight β_t is the Kullback-Leibler divergence, and the third term with weight γ_t is a distance measure. α_t, β_t and γ_t are controllable optimizable parameters based on reconstruction loss, KL-divergence and the distance measure respectively.

The adaptive weight α_t controls the reconstruction error while β_t ensures the posterior latent factor $q_\phi(v|x)$ does not deviate significantly from its prior $p_\theta(v)$. Both prior, $p_\theta(v)$ and posterior, $q_\phi(v|x)$ are typically modeled by a Gaussian distribution. Note that the resulting latent vector obtained on inference is re-parameterized following VAE style (Kingma and Welling, 2014).

The resulting latent-v_k obtained from GCVAE is passed into the classification algorithm to improve prediction.

4 Experimentation

4.1 General Setup

Experimental Setup. The experimentation is conducted on a High-Performance Computing (HPC) cluster comprising 80 cores, 2 × Intel Xeon E5-2698 v4 2.20GHz CPUs (80 cores), 512GB RAM, and 8 × Nvidia V100 32GB GPUs.

4.2 Dataset

Data used in for this research is sourced from ICU patients in three separate hospital systems (source: https://physionet.org/content/challenge-2019/1.0.0/). The data for each patient will be contained within a single pipe-delimited text file. Each file will have the same header and each row will represent a single hour's worth of data. The dataset is characterized by imbalanced data, with $x \in \mathbb{R}^{1552210\times 42}$, and includes labels for Sepsis/NoSepsis. It comprises $1,524,294$ observations with label 0's and $27,916$ with label 1's. The majority of variables in the dataset follow a Gaussian distribution (See Fig. 3).

Due to the presence of correlating variables (as seen on Fig. 4), we filter the variables accordingly, reducing the dimension of the data to $\mathbb{R}^{n\times 33}$, where n is the number of observations in the dataset. Finally, for modeling, we subsample the data into training and testing data with dimensions, $\mathbb{R}^{155315\times 33}$ and $\mathbb{R}^{84279\times 33}$ respectively; with the labels, l (training and testing), following the size of the aforementioned.

4.3 Evaluation metric

We evaluate the performances of the variational models using the original variables of x as latent factors (y), and we employ metrics commonly used for evaluating disentanglement to compare with the latent space (v), as outlined below:

Model	z-score		
	Log loss↓	Acc.↑	AUC↑
MLP	**0.08**	0.98	0.50
AdaBoost	0.65	0.98	0.50
GradientBoost	0.08	0.98	0.50
LDA	**0.08**	0.98	0.50
QDA	0.55	0.90	0.57
Gaussian-NB	0.61	0.88	0.58
LSTM	0.54	-	**0.76**

Table 1: Performance metric of the different models on the z-score (standardized) data. LSTM is the best performing model with highest AUC score.

Factor-VAE Score (Kim and Mnih, 2019) or **Z-min variance score**. Measures the accuracy of majority vote classifier that predicts the index of a fixed ground-truth factor (y_k) using latent code dimensions (v_j's). It is computed by normalizing the latent factor by their standard deviation (σ). A random sample of the y_k is selected and compared

with v_j and their variance computed. The code with the minimum variance best predict the fixed factor over a majority vote classifier for several subsets of the fixed factors.

Latent-$v \in \mathbb{R}^3$			
Model	Loss↓	Recons↓	KL-div↗
VAE	23.684	18.194	5.4896
β-VAE	49.568	38.921	1.0647
InfoVAE	22.185	18.161	15.6632
GCVAE†	15.286	16.941	25.2697
GCVAE‡	**15.150**	**16.492**	25.2169
Latent-$v \in \mathbb{R}^{10}$			
VAE	22.785	16.091	6.6943
β-VAE	49.576	38.874	1.0703
InfoVAE	13.743	11.967	52.1380
GCVAE†	**10.443**	**10.749**	28.1314
GCVAE‡	11.487	10.797	28.1542
Latent-$v \in \mathbb{R}^{15}$			
VAE	22.899	16.251	6.6482
β-VAE	49.711	39.237	1.0474
InfoVAE	11.494	10.175	89.3161
GCVAE†	**9.917**	9.881	28.7155
GCVAE‡	11.313	**9.672**	28.9077

Table 2: Comparison of Loss (total loss), Recons (Reconstruction loss) and KL-div (KL divergence) for the different latent sizes. GCVAE loss obtains the lowest reconstructions and total loss across all latent sizes.

Latent-$v \in \mathbb{R}^3$				
Model	Factor-VAE↑	MIG↑	Mod↑	JMIG↑
VAE	3387.45 ± 23.84	**0.22**	0.88	0.30
β-VAE	3379.80 ± 24.94	0.19	0.61	0.25
InfoVAE	3378.60 ± 26.60	0.09	0.82	0.27
GCVAE†	**3388.10 ± 32.26**	0.10	**0.89**	**0.43**
GCVAE‡	3383.20 ± 16.59	0.07	0.84	0.42
Latent-$v \in \mathbb{R}^{10}$				
VAE	**1053.20 ± 13.72**	**0.21**	0.77	0.28
β-VAE	1051.70 ± 19.12	0.19	0.53	0.25
InfoVAE	1051.20 ± 15.29	0.16	0.86	0.28
GCVAE†	1048.80 ± 16.37	0.11	**0.92**	**0.29**
GCVAE‡	1050.15 ± 14.82	0.09	0.89	0.25
Latent-$v \in \mathbb{R}^{15}$				
VAE	712.25 ± 8.75	**0.23**	0.65	0.29
β-VAE	707.55 ± 8.74	0.18	0.54	0.30
InfoVAE	711.90 ± 12.86	0.09	**0.90**	0.25
GCVAE†	714.35 ± 12.11	0.12	0.89	0.27
GCVAE‡	**720.65 ± 16.25**	0.15	0.89	**0.31**

Table 3: Performance comparison of different models on **Sepsis** data. Comparison metrics MIG (Chen et al., 2019), Mod (Ridgeway and Mozer, 2018) and JEMMIG (Do and Tran, 2021) for different sizes of Latent space. The direction of the arrow indicates the best performing model. Higher is better for MIG, Modularity and JEMMIG($1-JEMMIG$). **VAE** performs best on MIG disentanglement metric across al latents while **GCVAE**† performs the best on robustness and interpretability of latent; plus having the lowest reconstruction error.

Mutual Information Gap (MIG) score (Chen et al., 2019). Normalized mutual information gap between the top two latent factors $(I(y_k, v_I) - I(y_k, v_{II}))/(\sum_{j=1}^{d} I(y_i, v_j))$. The average MIG score is taken by normalizing the total mutual information. The score reports the compactness of the latent code by ensuring that the information contained in a fixed ground truth y_k is expressed by only one latent factor v_j at a time. High values imply a high level of disentanglement in the latent space.

Joint Entropy Minus Mutual Information Gap (JEMMIG) score (Do and Tran, 2021). This indirectly measures the modularity in latent space, since a single latent factor may explain more than one ground truth factor. It is expressed as $H(y_k, v_I) - I(y_k, v_I) + I(y_k, v_{II})$. A lower JEMMIG score is preferred, or a high $(1 -$ JEMMMIG$)$ score. The average JEMMIG score $(\frac{1}{K}\sum_{k=0}^{K-1} JEMMIG(y_k))$ quantifies the interpretability of the latent variables, measuring both its compactness and its explicitness.

Modularity score (Ridgeway and Mozer, 2018) expresses the number of latent factors v_j with high mutual information and explains the ground truth factors.

5 Quantitative evaluation

5.1 z-score prediction

The first approach to addressing this challenge is predicting on the z-score transformed data. The z-score is based on zero mean normalization of the data by centering as follows, $\frac{x-x_0}{\sigma}$. We adopt seven classification algorithms including, Multi Layer Perceptron (MLP), Ada Boosting (AdaBoost), Gradient Boosting (GradientBoost), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Gaussian Naive Bayes (Gaussian-NB) and Long Short Term Memory (LSTM).

Model	VAE			β-VAE			InfoVAE			GCVAE		
	Log loss	Acc	AUC	Log loss	Acc	AUC	Log loss	Acc	AUC	Log loss	Acc	AUC
MLP	0.09	0.98	0.50	0.09	0.98	0.50	0.09	0.98	0.50	0.09	0.98	0.50
AdaBoost	0.65	0.98	0.50	0.65	0.98	0.50	0.65	0.98	0.50	0.09	0.98	0.50
GradientBoost	0.09	0.98	0.50	0.09	0.98	0.50	0.25	0.98	0.50	0.09	0.98	0.50
LDA	0.09	0.98	0.50	0.09	0.98	0.50	0.09	0.98	0.50	0.09	0.98	0.50
QDA	0.98	0.98	0.50	0.09	0.98	0.50	0.27	0.93	**0.54**	0.09	0.98	0.50
Gaussian-NB	0.98	0.98	0.50	0.09	0.98	0.50	0.25	0.94	**0.53**	0.09	0.98	0.50
LSTM*												

Table 4: Performance metric of the different classification models on the generated variational latent spaces $v \in \mathbb{R}^{15}$.

Our emphasis is on enhancing the Receiver Operating Characteristic Area Under the Curve (ROC AUC), which gauges the tradeoff between the true positive rate (TPR) and false positive rate (FPR) considering the substantial data imbalance. LSTM surpasses other benchmark models, outperforming the second-best model with an increase in the AUC score of over 30% (See Table 1).

5.2 Latent-v_k prediction

We examine the performance of the different variational autoencoding models on the original data space for latent sizes 3, 10 and 15. Observe the the low reconstruction and total loss of GCVAE† (MMD) and GCVAE‡ (Mahalanobis) across all latent sizes (See Fig. 2). However, low reconstruction and moderately high KL-divergence does not directly imply high disentanglement of the latent space. This is the case, when we examine the results on Table 3.

Unexpectedly, the VAE demonstrates superior performance compared to other metrics in terms of disentanglement quality. We hypothesize that this may be attributed to the assumption that the original data x, used as y_k, does not represent a true latent factors. This assumption is merely made to affirm the presence of existing factors of x in v_k. However, the GCVAE surpasses other models in terms of measured Modularity (robustness) and interpretability of the latent space. This implies that more factors of x exist in v despite the low disentanglement.

We note that the robustness of the model has a causal implication for the classification AUC. To confirm this hypothesis, we employ the $\mathbb{R}^{15}$ latent space for classification. It is noteworthy that the top-performing models, namely QDA and Gaussian-NB, demonstrate superior performance on the InfoVAE latent space, as indicated by its high Modularity score. This observation further affirms that substantial information is preserved in the latent space generated by InfoVAE, leading to a higher AUC score compared to other models (See Table 4).

6 Conclusion

This research present a Latent model approach to addressing early prediction of Sepsis from Clinical data 1h ahead. We presented two approaches, using classification of z-normalized score and classification of latent space generated by Generalized-Controllable Variational AutoEncoder and its derivatives. Both approaches showed promising results, with LSTM outperforming other benchmark models in the z-score normalization approach and VAE demonstrating superior performance in the latent-v_k prediction approach. Note the absence of LSTM* scores on latent. This experiment was omitted due to lack of time and busy teaching schedules. The authors also found that the robustness of the model has a causal implication for the classification AUC.

7 Limitation

Exhaustive experimentation needed. Secondly, identify factors responsible for high AUC in InfoVAE.

Acknowledgments

None

Competing Interest

The author(s) declare no competing interest.

Funding

Not applicable.

Data Availability Statement

Available on Physionet.

References

Ricky T. Q. Chen, Xuechen Li, Roger Grosse, and David Duvenaud. 2019. Isolating sources of disentanglement in variational autoencoders.

Kien Do and Truyen Tran. 2021. Theory and evaluation metrics for learning disentangled representations.

Kenneth Ezukwoke, Anis Hoayek, Mireille Batton-Hubert, and Xavier Boucher. 2022. Gcvae: Generalized-controllable variational autoencoder.

Ricard Ferrer, Ignacio Martin-Loeches, Gary Phillips, Tiffany Osborn, Sean Townsend, Richard dellinger, Antonio Artigas, Christa Schorr, and Mitchell Levy. 2014. Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour. *Critical care medicine*, 42.

Joseph Futoma, Sanjay Hariharan, Mark Sendak, Nathan Brajer, Meredith Clement, Armando Bedoya, Cara O'Brien, and Katherine Heller. 2017. An improved multi-output gaussian process rnn with real-time validation for early sepsis detection.

Hyunjik Kim and Andriy Mnih. 2019. Disentangling by factorising.

Diederik P Kingma and Max Welling. 2014. Auto-encoding variational bayes.

Andrew Lever and Iain Mackenzie. 2007. Sepsis: Definition, epidemiology, and diagnosis. *BMJ (Clinical research ed.)*, 335:879–83.

Steven Cheng-Xian Li and Benjamin Marlin. 2016. A scalable end-to-end gaussian process adapter for irregularly sampled time series classification.

Lisiane Pruinelli, Bonnie Westra, Pranjul Yadav, Alexander Hoff, Michael Steinbach, Vipin Kumar, Connie Delaney, and Gyorgy Simon. 2018. Delay within the 3-hour surviving sepsis campaign guideline on mortality for patients with severe sepsis and septic shock. *Critical Care Medicine*, 46:1.

Karl Ridgeway and Michael C. Mozer. 2018. Learning deep disentangled embeddings with the f-statistic loss.

Kristina Rudd, Sarah Johnson, Kareha Agesa, Katya Shackelford, Derrick Tsoi, Daniel Kievlan, Danny Colombara, Kevin Ikuta, Niranjan Kissoon, Simon Finfer, Carolin Fleischmann, Flavia Machado, Konrad Reinhart, Kathryn Rowan, Christopher Seymour, Scott Watson, Eoin West, Maria De Fatima Marinho de Souza, Simon Hay, and Mohsen Naghavi. 2020. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the global burden of disease study. *The Lancet*, 395:200–211.

Mervyn Singer, Clifford S. Deutschman, Christopher Warren Seymour, Manu Shankar-Hari, Djillali Annane, Michael Bauer, Rinaldo Bellomo, Gordon R. Bernard, Jean-Daniel Chiche, Craig M. Coopersmith, Richard S. Hotchkiss, Mitchell M. Levy, John C. Marshall, Greg S. Martin, Steven M. Opal, Gordon D. Rubenfeld, Tom van der Poll, Jean-Louis Vincent, and Derek C. Angus. 2016. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8):801–810.

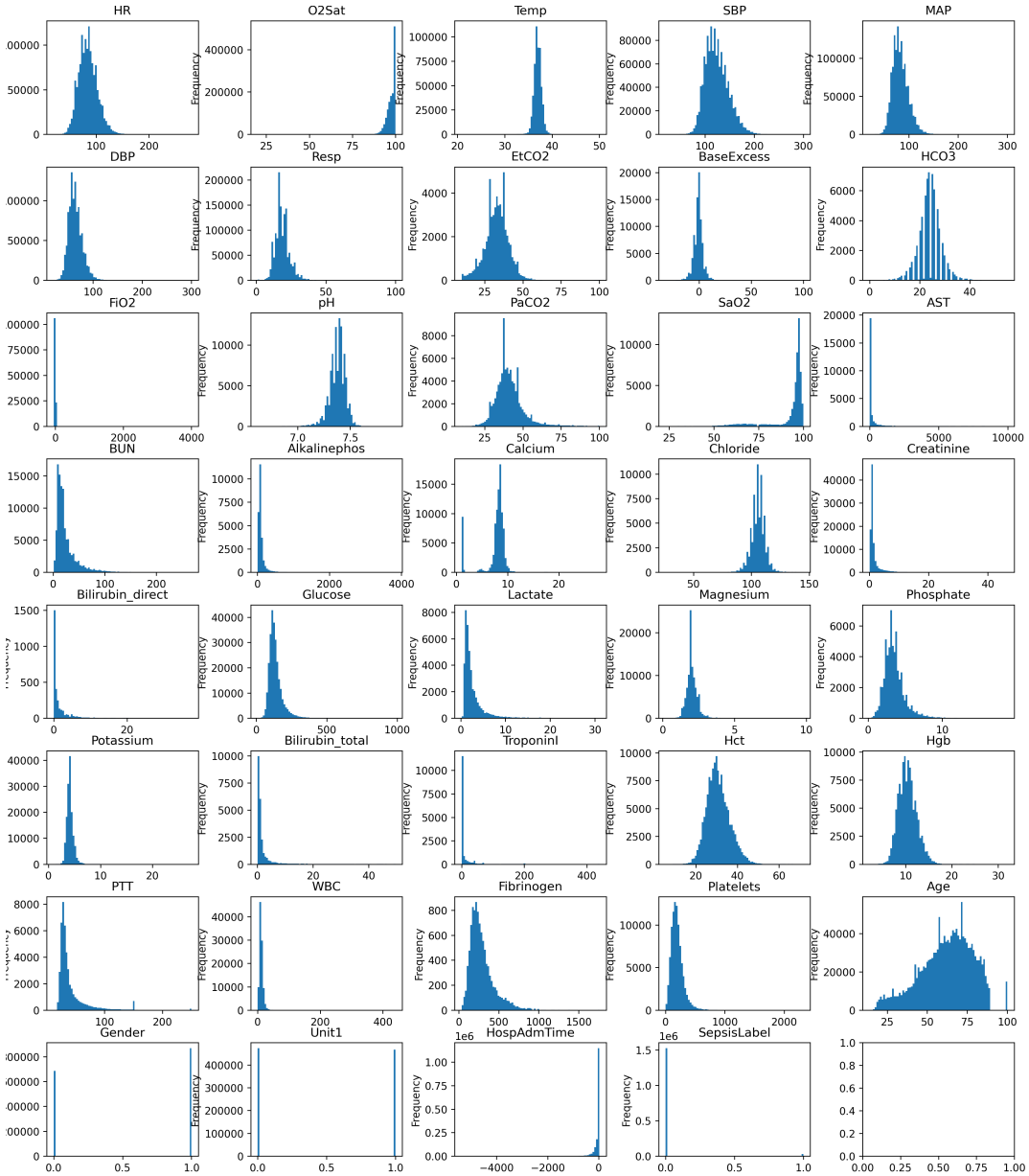

Figure 3: Distribution of variables in the dataset.

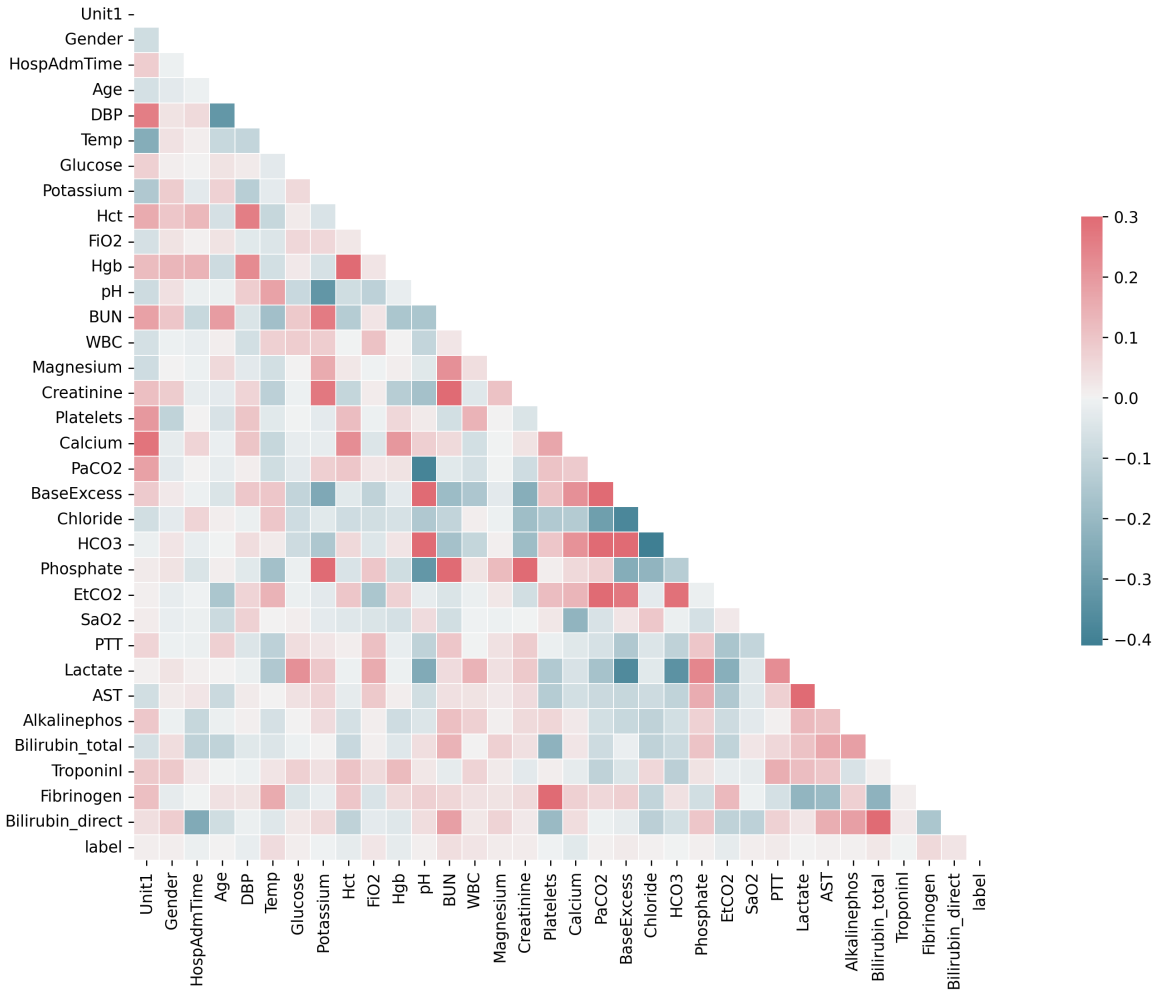

Figure 4: Correlation heatmap of the variables in the dataset