2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Unsupervised learning approach for predicting sepsis onset in ICU patients*

Guilherme Ramos [1], Erida Gjini [2], Luis Coelho [3] and Margarida Silveira [1]

Abstract— **Sepsis is a life-threatening condition caused by a deregulated host response to infection. If not diagnosed at an early stage, septic patients can go into a septic shock, associated with aggravated patient outcomes. Research has been mostly focused on predicting sepsis onset using supervised models that require big labeled datasets to train. In this work we propose two fully unsupervised learning approaches to predict septic shock onset in the Intensive Care Unit (ICU). Our approach includes learning representations from patient multivariate timeseries using Recurrent Autoencoders. Then, we apply an anomaly detection framework, using clustering-based algorithms, on the representation space learned by the models. When evaluating the performance of the proposed approaches in the septic shock onset prediction task, the Variational Autoencoder (VAE) using Gaussian Mixture Models in the anomaly detection framework was competitive with a supervised LSTM network. Results led to an AUC of 0.82 and F1-score of 0.65 using the unsupervised approach in comparison with 0.80, 0.66 for the supervised model.**

Clinical relevance— **This work proposes an unsupervised septic shock onset prediction framework which can improve current procedure for monitoring infection progression in the ICU.**

I. INTRODUCTION

An infection is the invasion of the human body by microorganisms, such as virus or bacteria, that rapidly spread and affect the well-being of the host. A deregulated host response to infection leading to damaging of tissues and organs corresponds to a life-threatening condition defined as *Sepsis*. According to a global report published by the World Health Organization (WHO), sepsis was responsible for 11 million deaths in 2017, 20% of all-cause global deaths, and affected around 49 million individuals [1].

If not diagnosed at an early stage, septic patients can go into a *Septic Shock*, where underlying circulatory, cellular and metabolic abnormalities strongly increase 1) the aggressiveness of treatments, 2) the overall costs for health units, and 3) mortality up to 38% [2]. This way, experts agree that an early sepsis diagnosis is essential, as a delay in antibiotic treatment has been documented to result in increased in-hospital mortality [3].

Researchers have been achieving promising results in predicting septic shock onset ahead of time [4]. These data-driven studies use Electronic Health Records (EHRs) to train innovative ML-based predictive frameworks. In [5], a novel deep learning architecture - composed of an LSTM to capture temporal structures, a CNN that detects time-invariant features and a fully connected neural network to process static information; achieved an AUC of 0.8 and F1-score of 0.75 in predicting septic shock 4 hours in advance. A Temporal Sequential Pattern-based approach, specifically focused on mining EHRs and extracting temporal dependencies among features, using an SVM classifier outperformed 6 classic machine learning models and an LSTM network in the same shock prediction task, consistently achieving an AUC above 0.85 for a prediction up to 20 hours before shock onset [6].

Despite the performance achieved by previous studies, the vast majority of the machine learning models proposed are trained in a supervised fashion, which is highly dependent on accurate data labeling, not always guaranteed in EHRs, and do not take advantage of available unlabeled data.

Recently some unsupervised models have been proposed [8], [7] that use generative models to learn representations of data without the need for big labeled data sets. Motivated by their success, we propose an anomaly detection framework to predict septic shock onset in the ICU which is not only fully unsupervised but also less susceptible to class imbalance. We first learn representations of time series using Recurrent Autoencoders trained only with non-septic patients. Afterwards, based on those representations, we detect septic-shock patients by performing anomaly detection with unsupervised clustering methods.

II. METHODOLOGY

A. Data

This study used data from Medical Information Mart for Intensive Care (MIMIC) III [9] - a public dataset composed of anonymized information on over 40.000 patients, 58.000 hospital admissions and more than 60.000 ICU stays.

Following the guidelines of Sepsis-3 criteria [10], a patient in septic shock can be clinically identified by presenting hypotension only reversed by sustained need of vasopressor therapy and an elevated serum lactate level, despite adequate fluid resuscitation. In other words, a septic shock occurs if, during the 48 hours before and up to 24 hours after suspected infection, the patient presents any:

- vasopressor initiation

* This work was supported by Portuguese funds through FCT (Fundação para a Ciência e Tecnologia) under LARSyS - FCT Project UIDB/50009/2020

[1] Institute for Systems and Robotics, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal guilherme.b.ramos@tecnico.ulisboa.pt, msilveira@isr.tecnico.ulisboa.pt

[2] Center for Computational and Stochastic Mathematics, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal erida.gjini@tecnico.ulisboa.pt

[3] Hospital de São Francisco Xavier, Nova Medical School, Lisbon, Portugal luismiguelcoelho16@gmail.com

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

- mean arterial pressure (MAP) < 65 mm Hg
- lactate > 2 mmol/L (18 mg/dL)

Septic shock onset is defined as the moment the patient initiates vasopressor treatment to maintain a MAP above 65 mm Hg with elevated lactate levels, despite adequate fluid resuscitation as defined by the Surviving Sepsis Campaign guidelines [11].

B. Cohort selection

The following rules were applied: 1) remove patients under 18 years old, 2) only include each subject's first ICU visit, 3) exclude admissions to the Coronary Care Unit, 4) duration of ICU stay of at least 24 hours and a max of 10 days and 5) remove patients with shock onset before the first 14 hours. After applying these exclusions, we were left with a group of 18814 eligible patients, of which 7164 developed sepsis. Furthermore, after applying Sepsis-3 criteria, we identified 1177 patients that progressed to a septic shock.

C. Feature Set

In this work, an ICU patient is represented by time series for each of the clinical variables extracted. For a certain variable, the data is grouped into hourly-bins, meaning the number of timesteps corresponds to the size of the window analyzed. To monitor sepsis progression, we selected 7 vital signs, 18 lab values, 4 clinical interventions (mechanical ventilation, vasopressor administration, crystalloid and colloid bolus) and 3 demographic variables (age, gender and ethnicity). Static variables were categorized and included in the time series using one-hot encoding.

D. Missing Data Strategy

To replace missing data, we start by applying forward-filling. If there are no previous values, we fill the missing data with the individual-specific mean. Finally, if that patient has no observations for that variable, the missing values are replaced with the global mean.

One additional technique when dealing with missing values is to use a missing indicator (MI), proposed by Lipton et al. (2016) in [12], which sees missing data itself as a feature for clinical prediction. In fact, lack of observations for a certain variable might indirectly tell us important information about that patient condition. Hence, for each vital and lab result sequence, x^i, a MI vector m^i is concatenated where:

- $m_t^i = 1$, if x_t^i is an observation
- $m_t^i = 0$, otherwise

Using MIs allows the RNN to learn missingness patterns that may relate with the progression of sepsis in the ICU.

E. Window extraction

To approach the event-level early prediction task, patient time series were right-aligned, as shown in Figure 1, using feature windows of 10 hours (10 timesteps) and prediction windows of 3 hours.

For a shock patient, the shock onset is located and both windows are computed with $P = t_{shock}$, to ensure the feature window is the same for all shock patients, i.e. we extract the same period before shock onset. For patients with sepsis that did not progress to shock we use $P = 24h$, considering at the end of the first ICU day the patient has already received clinical interventions and has both symptoms of infection and organ dysfunction. Finally, for patients that were not diagnosed with sepsis, we use $P = t_{end}$.

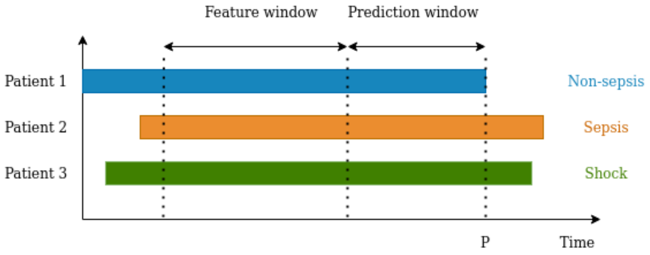

Fig. 1. Feature and prediction window extraction by patient type.

F. Data-split Design

Data was separated into 80% training (X_{train}) and the remaining 20% in testing (X_{test}). Both train and test set have 16.4% of shock patients. Additionally, a validation set (X_{val}) was created by splitting 12.5% of the train set, with the same shock distribution of the previous sets.

III. PROPOSED APPROACH

Our model is composed of two stages: *representation learning*, conducted by recurrent autoencoders and *anomaly detection*. We take advantage of the ability of recurrent neural networks to learn the characteristics of time series data, combined with the representation learning capabilities of autoencoders. Then, we predict septic shock onset by performing anomaly detection on the learned representations. Both stages are unsupervised.

A. Representation Learning

The first step, corresponds to extracting meaningful representations from high-dimensional data in an unsupervised manner, using autoencoders. An autoencoder is a neural network composed of two parts: the *encoder* and the *decoder*. The former maps the original input data $\mathbf{x} \in \mathbb{R}^{d_\mathbf{x}}$ to the latent space $\mathbf{z} \in \mathbb{R}^{d_\mathbf{z}}$ while the latter maps this vector back to the input dimension, creating a reconstruction of the original sample, $\hat{\mathbf{x}} \in \mathbb{R}^{d_\mathbf{x}}$. Autoencoders learn how to reconstruct the input data by minimizing a loss function that measures the dissimilarity between the output $\hat{\mathbf{x}}$ and the input $\mathbf{x}$. Frequently a regularization term is added.

In the autoencoders used in this work both encoder and decoder are parametrized by a Long Short-Term Memory (LSTM). This is a variant of recurrent neural network known for achieving remarkable results with sequential data. LSTMs have cell states that work as a memory controlled by three gates: the forget gate, the input gate and the output gate. Adding extra interactions with these gates solves the vanishing gradient problem and allow LSTMs to learn long term dependencies and relations on sequential data. Furthermore, to force the models to learn the most important features, we propose undercomplete autoencoder structures, in which the code dimension is less than the input dimension.

1) Standard Autoencoder (AE): This autoencoder learns to extract meaningful features by minimizing the mean squared error between the output sequences and the original input data, computed as follows:

$$MSE = \frac{1}{N}\sum_{n=1}^{N} \|x_n - \hat{x}_n\|^2 \tag{1}$$

Using the encoder of a trained AE we are able to reduce our high-dimensional patient time series into fixed-size vectors, defined by the code size parameter.

2) Variational Autoencoder (VAE): The variational autoencoder learns the parameters of a probability distribution - the encoder maps the patient time series to a vector of mean and covariance matrix, $\Sigma = \sigma_n^2 I$. In order to feed the decoder, the model takes a sample from the latent distribution, picking a random variable z from a continuous space, and reconstructs the original sequences.

Considering the true posterior for the random variable is intractable, we approximate the distribution to a normally distributed Gaussian, $\mathcal{N}(\mathbf{0},\mathbf{I})$. This optimization problem is called variational inference. In order to approximate the two distributions, a second term is added to the training objective to express the similarity between the true posterior $p_\phi(\mathbf{z} \mid \mathbf{x})$ and the approximation $q_\phi(\mathbf{z} \mid \mathbf{x})$, where ϕ corresponds to the parameters of the encoder. Training a VAE corresponds to maximizing the *evidence lower bound* defined as,

$$\mathscr{L}_{\text{ELBO}} = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x} \mid \mathbf{z})\right] - \mathscr{D}_{\text{KL}}\left(q_\phi(\mathbf{z} \mid \mathbf{x}) \| p_\theta(\mathbf{z})\right) \tag{2}$$

where the first term corresponds to the reconstruction of the original time series and the KL-divergence term measures the similarity between the two probability distributions, and is always non-negative.

Additionally, we apply the reparametrization trick proposed by Kingma & Welling (2013) [13], which defines a random variable z as,

$$\mathbf{z} = \mu_{\mathbf{z}} + \sigma_{\mathbf{z}} \odot \varepsilon \tag{3}$$

where ε is an external noise that follows a normal distribution $\mathcal{N}(\mathbf{0},\mathbf{I})$ and $\odot$ is an element wise multiplication.

Let $\mathbf{x}^{(n)} = \left(\mathbf{x}_1^{(n)}, \mathbf{x}_2^{(n)}, \ldots, \mathbf{x}_T^{(n)}\right)$ be a patient timeseries, with T corresponding to the size of the feature window. The training objective of the VAE is

$$\begin{aligned}\mathscr{L}\left(\theta,\phi;\mathbf{x}^{(n)}\right) = & \,\mathbb{E}_{q_\phi\left(\mathbf{z}^{(n)}|\mathbf{x}^{(n)}\right)}\left[\log p_\theta\left(\mathbf{x}^{(n)} \mid \mathbf{z}^{(n)}\right)\right] \\ & - \beta\mathscr{D}_{\text{KL}}\left(q_\phi\left(\mathbf{z}^{(n)} \mid \mathbf{x}^{(n)}\right) \| p_\theta\left(\mathbf{z}^{(n)}\right)\right).\end{aligned} \tag{4}$$

where β is a trade-off parameter between the two loss terms.

B. Anomaly detection

Anomaly Detection (AD) is the identification of rare events, i.e. anomalies, in an environment where most data is considered normal. Looking at a shock patient as an anomaly among ICU patients, one can apply an AD framework on the representation space learned by the autoencoders and perform septic shock onset prediction. To do this, we use clustering algorithms that group data points into 2 clusters: shock and non-shock. AD is performed under the assumption that the cluster with the most patients corresponds to the normal group (non-shock).

To predict a septic shock onset we apply this framework on the representation space learned by the AE and on the mean posterior space (μ_z) of the VAE. The unsupervised clustering techniques applied in this framework are: k-Means, spectral clustering, Gaussian Mixture Models (GMM) and one-class (OC) SVM.

IV. RESULTS

In this chapter we present the main results obtained using the proposed approach. Both models were trained using normal samples (patients from the non-septic group) and applied to the same test set - with non-septic, septic and shock patients.

The models were developed using Keras deep learning library with TensorFlow backend. Implementation and training were performed in Google Colab's virtual machine.

A. Representation Learning

The first step was to train both autoencoders to learn compressed representations of patient time series.

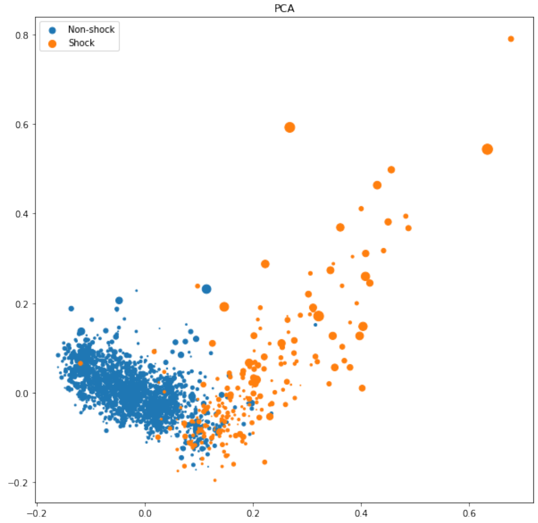

Fig. 2. Visualization of μ_z space via PCA. The size of each data point corresponds to sum of the log-variance of the distribution.

To analyze the representations learned by the VAE we visualized the mean posterior space, depicted in Figure 2, where the area of each data point corresponds to the sum of the log-variances of that patient distribution. When looking at the obtained result one can see that most orange bubbles are bigger than blue ones. In fact, non-shock patients are mapped to distributions with variances close to 1, whereas shock time series have variances above this value. This is particularly visible in the transition region, where both classes have similar means and orange bubbles higher areas. This is because the VAE did not learn how to represent shock patients as a normal distribution.

B. *Septic Shock onset prediction*

After learning representations using autoencoders, we applied the AD framework. Additionally, a supervised LSTM network was trained to compare with our unsupervised approaches. In Table I we summarize the obtained results, where both unsupervised models used a code size equal to 80 and the VAE had $\beta = 250$.

TABLE I

PERFORMANCE OF THE PROPOSED MODELS IN EVENT-LEVEL EARLY PREDICTION. BEST RESULTS ARE HIGHLIGHTED IN BOLD.

		ACC	F1-score	Recall	Precision
AE	**k-Means**	0.7527	0.3885	0.8596	0.2509
	Spectral	0.7547	0.3892	0.8553	0.2519
VAE	**k-Means**	0.8915	0.6246	0.9276	0.4708
	Spectral	0.5474	0.2815	**0.9702**	0.1646
	GMM	**0.9342**	**0.6529**	0.6766	**0.6309**
	OC-SVM	0.9265	0.6103	0.6298	0.5920
Supervised LSTM		0.9343	0.6606	0.7000	0.6255

Looking at the performance of the different models we see that the VAE and the supervised LSTM achieved the best results. Furthermore, the VAE outperformed the AE on every metric, reaching the best accuracy, F1-score and precision using GMM. Despite having achieved the best recall using k-Means and spectral clustering, both autoencoders showed lower precision, meaning a large number of false positives, consequently leading to a reduced F1-score. The VAE with GMM outperformed the LSTM in precision while achieving competitive results in the remaining metrics.

In order to test the impact of the prediction window in early shock prediction we applied the same methodology for different hours before shock onset. As depicted in Figure 3, the models' performance decreases as the distance to shock increases. In fact, predicting a septic shock becomes more challenging the further away the feature window gets from onset. For the AUC, the VAE with k-Means consistently outperforms the other models. At the same time, VAE with GMM outperforms the LSTM network, especially for bigger prediction windows. For the F1-score, we see that the LSTM obtains the best results when close to onset (under 3 hours). For the remaining window sizes, the VAE with GMM becomes very competitive and ends up outperforming the supervised network when predicting a septic shock 6 and 7 hours before onset.

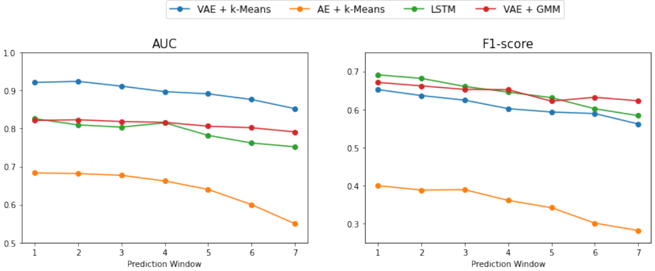

Fig. 3. AUC and F1-score for different prediction windows (in hours).

V. CONCLUSIONS

This work achieved a motivating result for applying an anomaly detection unsupervised approach for septic shock onset prediction by showing it can compete with a supervised network. The proposed framework can be extended to perform a continuous risk assessment of ICU patients and provide an additional indicator to support clinical physicians in their decision-making, mainly regarding treatment initiation. Future work will focus on using the variance of the normal distributions obtained with VAE in the clustering stage.

REFERENCES

[1] World Health Organization. (2020). Global report on the epidemiology and burden of sepsis: current evidence, identifying gaps and future directions.

[2] Vincent, J. L., Jones, G., David, S., Olariu, E., & Cadwell, K. K. (2019). Frequency and mortality of septic shock in Europe and North America: a systematic review and meta-analysis. Critical care, 23(1), 1-11.

[3] Kumar, A., Roberts, D., Wood, K. E., Light, B., Parrillo, J. E., Sharma, S., ... & Cheang, M. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Critical care medicine, 34(6), 1589-1596.

[4] Fleuren, L. M., Klausch, T. L., Zwager, C. L., Schoonmade, L. J., Guo, T., Roggeveen, L. F., ... & Elbers, P. W. (2020). Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive care medicine, 46(3), 383-400.

[5] Lin, C., Zhang, Y., Ivy, J., Capan, M., Arnold, R., Huddleston, J. M., & Chi, M. (2018). Early diagnosis and prediction of sepsis shock by combining static and dynamic information using convolutional-LSTM. 2018 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 219-228). IEEE.

[6] Khoshnevisan, F., Ivy, J., Capan, M., Arnold, R., Huddleston, J., & Chi, M. (2018). Recent temporal pattern mining for septic shock early prediction. 2018 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 229-240).

[7] J. Yao, M. L. Ong, K. K. Mun, S. Liu and M. Motani, (2019). Hybrid Feature Learning Using Autoencoders for Early Prediction of Sepsis, 2019 Computing in Cardiology (CinC), Singapore, (pp. 1-4).

[8] P. Javia, A. Rana, N. Shapiro and P. Shah, Machine Learning Algorithms for Classification of Microcirculation Images from Septic and Non-septic Patients, 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 607-611.

[9] Johnson, A. E., Pollard, T. J., Shen, L., Li-Wei, H. L., Feng, M., Ghassemi, M., ... & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific data, 3(1), 1-9.

[10] Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., ... & Angus, D. C. (2016). The third international consensus definitions for sepsis and septic shock (Sepsis-3). Jama, 315(8), 801-810.

[11] Surviving Sepsis Campaign Guidelines Committee including the Pediatric Subgroup. (2013). Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. Intensive care medicine, 39(2), 165-228.

[12] Lipton, Z. C., Kale, D. C., & Wetzel, R. (2016). Modeling missing data in clinical time series with RNNs. Machine Learning for Healthcare, 56.

[13] Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114.