Vol.64: e21210142, 2021
https://doi.org/10.1590/1678-4324-75years-2021210142
ISSN 1678-4324 Online Edition

*Article - 75 years - Special Edition*

# Early Identification of Patients at Risk of Sepsis in a Hospital Environment

**Everton Osnei Cesario[1]***
https://orcid.org/0000-0003-3599-7528

**Viviane Maria de Carvalho Hessel Dias[2]**
https://orcid.org/0000-0001-6465-8213

**Yohan Bonescki Gumiel[1]**
https://orcid.org/0000-0001-8239-2930

**Claudia Moro[1]**
https://orcid.org/0000-0003-2637-3086

**Marcia Cristina Marins Martins[1]**
https://orcid.org/0000-0002-9684-1183

**Deborah Ribeiro Carvalho[1]**
https://orcid.org/0000-0002-9735-650X

[1]Pontifícia Universidade Católica do Paraná, Graduate Program on Health Technology, Curitiba, Paraná, Brazil; [2]Hospital Nossa Senhora das Graças, Curitiba, Paraná, Brasil.

Editor-in-Chief: Alexandre Rasi Aoki
Associate Editor: Daniel Fernandes

Received: 2021.03.18; Accepted: 2021.07.21.

*Correspondence: evertoncesario@hotmail.com; Tel.: +55-41-9958-3361 (E.O.C.)

**HIGHLIGHTS**

- Early identification of sepsis despite low-quality and sparse data.
- New strategy to identify cases due to the underreporting with ICD-10 codes.
- Recommendation of more frequent data collection and with better quality.
- Identification of variables that most contribute to the system prediction.

**Abstract:** Sepsis is a systematic response to an infectious disease, being a concerning factor because of the increase in the mortality ratio for every delayed hour in the identification and start of patient's treatment. Studies that aim to identify sepsis early are valuable for the healthcare domain. Further, studies that propose machine learning-based models to identify sepsis risk are scarce for the Brazilian scenario. Hence, we propose the early identification of sepsis considering data from a Brazilian hospital. We developed a temporal series based on LSTM to predict sepsis in patients considering a three-day timestep. The patients were selected using both criteria, ICD-10, and qSOFA, where we supplemented qSOFA with the additional identification of words referring to infections in the clinical texts. Additionally, we tested a Random Forest classifier to classify patients with sepsis with a single timestep before the sepsis event, evaluating the most relevant features. We achieved an accuracy of 0.907, a sensitivity of 0.912, and a specificity of 0.971 when considering a three-day timestep with LSTM. The Random Forest classifier achieved an accuracy of 0.971, a sensitivity of 0.611, and a specificity of 0.998. The features age, blood glucose, systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate, and admission days had the most influence over the algorithm classification, with age being the most relevant feature. We achieved satisfactory results compared with the literature considering a scenario of spaced measures and a high amount of missing data.

**Keywords:** sepsis; machine learning; healthcare.

## INTRODUCTION

Healthcare-associated infection (HAI) is a frequent event adverse among hospitalized patients [1]. A high infection rate is related to the use of invasive devices, especially central lines, urinary catheters, and ventilators [2]. Sepsis is a life-threatening organ dysfunction generated by a dysregulated host response to infection [2]. Infectious disease, whether caused by bacteria, fungi, viruses, or protozoa, manifesting itself in different clinical stages of the same pathophysiological process [3], can be triggered by HAI, especially in people who are already at risk. It is characterized by the presence of inflammatory mechanisms, which leads to cellular and circulatory alterations like vasodilation and increased capillary permeability, promoting hypovolemia, hypotension, capillary density reduction, disseminated intravascular coagulation, leading to the reduction of tissue oxygen supply, causing increased anaerobic metabolism and hyperlactatemia [4].

As reported in the study of [5], there was an estimation of 48.9 million sepsis cases and 11 million sepsis-related deaths worldwide in 2017, around 19.7% of the global deaths in that year. The sepsis scenario in Brazil is represented in the study of [6], in which 30% of the beds in adult intensive care units (ICUs) are occupied by patients that acquired sepsis during their stay. It was evidenced a progressive growth in sepsis cases in the ICUs, from 19.4% in 2010 to 25.2% in 2016. Additionally, there was a decrease in the death ratio from 39% in 2010 to 30% in 2016. However, the death ratio is still a concerning factor. Thus, it is essential to provide tools to facilitate the early identification of sepsis.

One of the main reasons for such a higher number of deaths relies on the limited knowledge and comprehension about the complex inflammatory response mechanism, resulting in late sepsis recognition [7]. Its manifestations can be confused with other non-infectious processes and can even go unnoticed [3]. Initial interventions rely on early recognition, a continuous and manual process performed by health professionals, being one of the most significant difficulties in clinical practice, as it depends directly on the professional's ability to identify patients at risk [8].

The patient identification relies on specific scores that measure the severity and are used to recognize cases but fail to identify around one in every eight patients with severe sepsis [9]. Among those specific scores, we highlight Sequential [Sepsis-related] Organ Failure Assessment (SOFA) and quickSOFA (qSOFA). The first has the objective of quantitatively describing the organ dysfunction or failure resulting from the septic clinical picture, aiming to describe a sequence of complications in the critically ill [10]. The score is based on available clinical data and laboratory parameters, not considering other associated factors, such as age or other comorbidities [11]. The second is less robust than SOFA, not requiring laboratory exams, which facilities the bedside analysis [2]. The overall recommendation in clinical suspicion of infection is to use the SOFA score of two or more points for patients in ICUs and qSOFA for patients outside the ICUs [12,13].

Early sepsis recognition is essential since the survival rate reduces by up to 8% for each treatment delayed hour [14]. In addition, an early start of therapy, administering antimicrobials, preferably within 60 minutes after the hypotension recognition, causes a positive impact in reducing mortality [14]. [15] reported an early recognition reduced mortality from 46.5% to 30.5% in their study. Additionally, [14] observed in their study that the efficient antibiotic introduction during the first hour of a documented hypotension increases the patient's with a septic shock survival rate.

Machine learning-based models can be used to support sepsis prediction, being an active research topic demonstrated by studies such as [16-20]. Besides predictor models, other studies have mapped several progressions related to sepsis, such as mortality in [21], systematic organ failure in [22], and therapies for the treatment of patients in [7].

The challenges for this study rely on working with data from the health domain, where the data tend to be noisy and to have a lot of missing information. Also, the inconsistency in labeling the International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) [23] code impacts the predictive power of the developed model. Additionally, none of the evaluated studies dealt with Brazilian hospitals' records, verifying the performance of machine learning-based techniques for identifying sepsis.

Henceforth, in this study, we propose to develop a model to contribute to the early identification of sepsis, verifying its efficacy on data from Brazilian hospitals, focusing on data gathered from the patient and not on data related to the ambient condition.

## MATERIAL AND METHODS

The research scenario involved admitted patients between 2017 and 2018. Our dataset is related to a hospital from Brazil, restricting the data to patients admitted to the hospital to both infirmary and ICU. The dataset contained 15,189 patients with 55,590 records, with a minimum age of 18 years old and a maximum age of 106 years old, and a median age of 50. We verified a lack of vital signs' measures and Glasgow coma scale values during our initial analysis. Consequently, we adopted the following criterion to eliminate incomplete data from vital signs' measures: records with at least 50% of the vital signs' documentation. We chose the threshold of 50% because we still maintained 87% of the total data with this value. After these procedures, our dataset was composed of 4,331 patients with 4,810 admissions.

To identify sepsis occurrences, we tested two different strategies: (i) searching for the ICD-10 codes belonging to the ICD-10 A41 group in primary, secondary, and death ICD-10 fields over the records and (ii) applying the criteria from qSOFA proposed by [2] as an alternative way to identify patients with sepsis, comparing the results with ICD-10 afterward. We chose qSOFA over SOFA because of the non-disponibility of laboratory data in a systematic form, while qSOFA considers only measures taken from physical examination and the occurrence of infections. However, due to the missing values Glasgow coma scale, we did not consider it in our study for the qSOFA criteria. To search for occurrences of infections, we adopted a strategy of searching for words that referred to infections in the patient's records. In summary, for this project, the criteria for identifying patients with sepsis were: (i) respiratory rate higher than 22 per minute, (ii) systolic blood pressure lower than 100 Hg, and (iii) had terms related to infections over the free text.

We calculated the Kappa coefficient [24] to evaluate the concordance between the sepsis identification by ICD-10 and qSOFA and used the criteria from Landis [25] to score the concordance.

### Preprocessing

One of the significant challenges of this study was the preprocessing, in which we had the objectives of standardizing the data, removing the noise, and collecting information to choose the strategies to deal with the missing information [18,26]. We applied the following preprocessing steps: (i) merging of clinical and laboratory data, (ii) identification of clinical and surgical procedures, (iii) identification of medication prescription containing antibiotics, (iv) identification of terms over the free-text that are related to infections and (v) filling the missing data. All these steps will be detailed below.

#### *Merging of clinical and laboratory data*

In datasets such as Medical Information Mart for Intensive Care (MIMIC) [27], widely used in sepsis-related studies [18,28-34], the patient data collection is more frequent, allowing clustering the data in intervals varying from minutes to hours. However, in this study, we clustered data by date following [21]. For every admitted day, the ideal scenario would be to have at least one record of the patient's vital signs and an exam result for the same date. In the scenario of having an exam result without a vital sign result for a specific date, it would be associated with the closest vital sign record within a one-day range. However, if there were no vital sign records within the one-day range, we would exclude the results.

#### *Identification of clinical and surgical procedures*

Every record was related to a specific procedure, and we distinguished those in clinical and surgical procedures. In patients with SIRS secondary to polytrauma or surgery, the correct diagnosis is hampered by inflammatory response signs due to the previous aggression [3]. The procedure code can be changed during the hospitalization, so we used the last registered procedure code. We defined a list of 5,892 terms related to clinical and surgical procedures specific to the hospital. The terms *acompanhamento de doença hepática* (monitoring hepatic disease in English*)*, *consulta neurológica* (neurology consultation in English*)*, and *sessão de auriculoterapia* (auriculotherapy session in English) are examples of clinical procedures. Further, the terms *angioplastia coronariana* (coronary angioplasty in English*)*, *redução cirúrgica de fratura de costela* (surgical reduction of ribbon fracture in English*)*, and *laringorrafia* (laryngography in English*)* are examples of surgical procedures.

#### *Identification of medication prescription containing antibiotics*

The incorrect initial approach to the infectious agent through antibiotics directly relates to sepsis mortality, with clear evidence that delaying antibiotic therapy increases death risk [3]. The antibiotic treatment should be started as soon as possible to control the infection's focus as a prerequisite for eliminating the aggressor

to enable the patient to recover [35]. Thus, we gathered this information to evaluate the relationship between antibiotics administration and the patient outcome. We defined a list of 175 antibiotic-related terms specific to the hospital. The terms *amoxicilina* (amoxicillin in English*)*, *doxiciclina* (doxycycline in English*)*, and *levofloxacina* (*levofloxacin* in English) are examples of antibiotics.

*Identification of terms over the free-text that are related to infections*

We identified words that referred to infections because the identification of infection focus is a part of the process of recognizing patients at risk of sepsis, which is difficult because the presence of infectious focus is not always clear [3]. The utilization of free-text to complement structured data was also used by [34], improving around 19% their identification of infections. Therefore, we also chose to develop a strategy to deal with free-text identifying words related to infections. We verified the fields: (i) History of Current Disease, filled by the screening professional at the patient's arrival, (ii) Evolution, which includes the patient's daily evolution, filled by the physician, and (iii) Physical Examination, filled daily by the nursing professionals.

The keywords were selected by a professional specialized in infection in hospitals. The keywords in Portuguese, with their respective translation to English in parenthesis, were: *bronquite* (bronchitis), *nefrite* (nephritis), *peritonite* (peritonitis), *meningite* (meningitis), *apendicite* (appendicitis), *cistite* (cystitis), *celulite* (cellulite), *pneumonia* (pneumonia), *sepsis* (sepsis), *septicemia* (sepsis), *osteomielite* (osteomyelitis), *conjuntivite* (conjunctivitis), *sinusite* (sinusitis), *otite* (otitis), *gengivite* (gingivitis), *laringite* (laryngitis), *faringite* (pharyngitis), *endocardite* (endocarditis), *gastroenterite* (gastroenteritis), *erisipela* (erysipelas), *amigdalite* (tonsillitis), *pielonefrite* (pyelonephritis), *mastoidite* (mastoiditis), *abscesso* (abscess), *pericardite* (pericarditis), *endometrite* (endometritis), *colecistite* (cholecystitis), *pancreatite* (pancreatitis), *diverticulite* (diverticulitis), *colite* (colitis), *mastite* (mastitis) e *anexite* (salpingitis).

To identify the keywords in the free-texts, we utilized the function *stringdist* with the method of Damerau-Levenshtein [36] to calculate the distance between the vectors of characters or between vectors that represent generic sequences. A degree of distance equals to two was applied to turn the identification process more flexible; that is, two changes would be needed for the identified word to become a searched word. With this process, words written with a maximum of two wrong characters were still identified.

*Filling the missing data*

We filled the missing data from the following vital signs: temperature, blood glucose, systolic blood pressure, diastolic blood pressure, heart rate, and respiratory rate. We followed [17,21,37,38] to fill the missing data, considering nearby data or typical clinical values according to gender and age. We only utilized records that had at least 50% filled data completing the missing data with the following criteria: (i) in cases where there were up to two daily readings, the median between the two closest values was used, the day before and the day after; (ii) in the absence of two to three readings, typical clinical values were used. For the average temperature values, we used [39], obtaining the temperature by the variables sex, age, and time of the day. For blood glucose data, we used [40]; for heart rate, we used [41]; for respiratory rate, we used [42]; and for systolic and diastolic blood pressure, we used the 7th Brazilian Guideline on Hypertension [43]. However, it is known that these procedures are not appropriate for all situations [17] and are just a way to bypass the lack of daily measures.

## Experiments

For our experiments, we tested two different approaches, first a temporal series developed with LSTM because it is resilient to missing data, being a widely used architecture in the field of medical diagnostic [17]. Additionally, as [17] has shown, it can achieve superior results than techniques such as multilayer perceptron (MLP) and linear traditional machine learning classifiers. Besides, using this strategy can consider patient longitudinal data, understanding the patient history over the admitted days. The experiment structured is presented in Figure 1, where longitudinal data from day 1 (the day before) to day *n*, where *n* is calculated considering the timesteps. Additionally, we tested obtaining the sepsis indication from both qSOFA and ICD-10. In the second experiment, a Random Forest classifier used the information from the previous day to classify the patient as a positive or negative to sepsis.

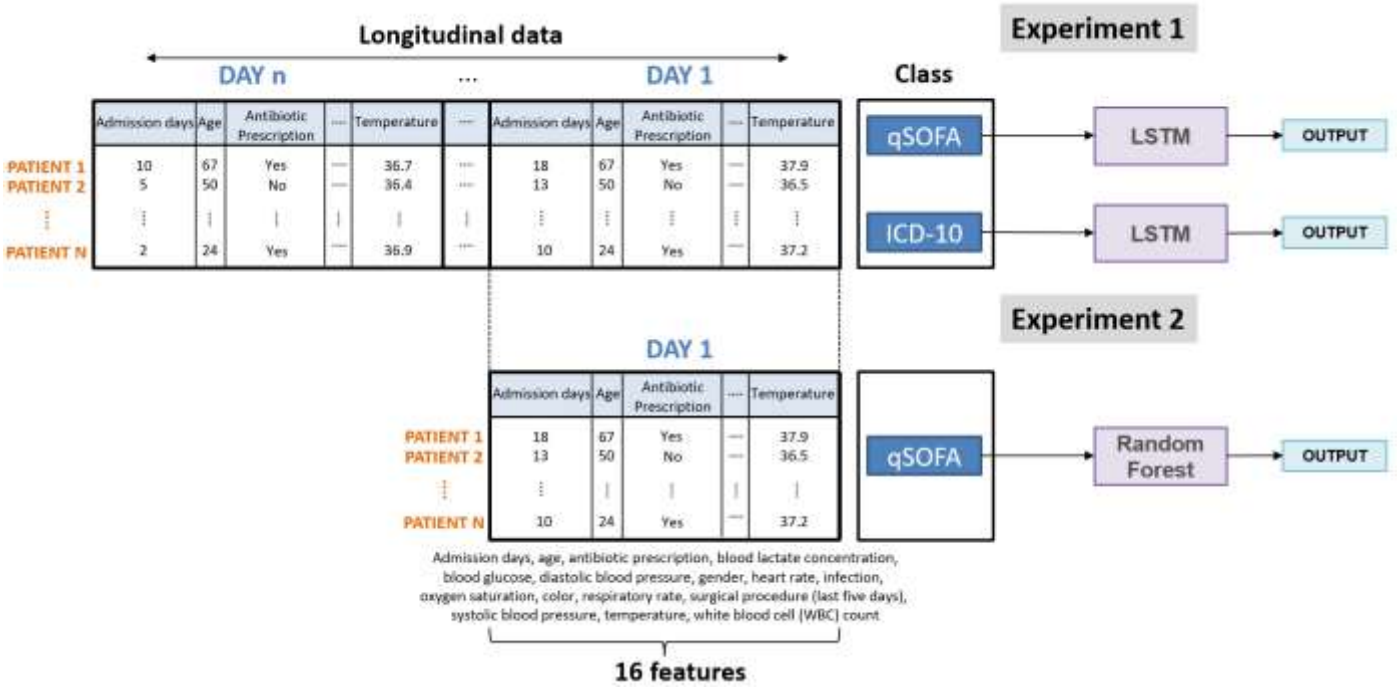



**Figure 1.** Model development for both experiments.

We used the same features for both experiments, which are detailed in Table 1, where the features *antibiotic prescription*, *infection,* and *surgical procedure (last five days*) are obtained from the preprocessing section. Additionally, the feature *admission days* is a feature that reflects how many days from the admission until that specific day.

**Table 1**. Features used in the experiments with their respective units.

| Feature | Unit |
|---|---|
| Admission days | Days |
| Age | Years |
| Antibiotic prescription | Yes, No |
| Blood lactate concentration | mg/dL |
| Blood glucose | mg/dL |
| Diastolic blood pressure | mmHg |
| Gender | Male, Female |
| Heart rate | Bpm |
| Infection | Yes, No |
| Oxygen saturation | % |
| Color | Black, White, Pardo, Yellow, Indigene, Not informed |
| Respiratory rate | Bpm |
| Surgical procedure (last five days) | Yes, No |
| Systolic blood pressure | mmHg |
| Temperature | ◦C |
| White blood cell (WBC) count | $10^3/mm^3$ |

*Experiment 1 – LSTM*

For the initial network parameters, we followed [17] because we have a similar environment, with temporal series prediction and severe unbalanced data. The initial parameters are detailed in Table 2. Additionally, we searched for the best parameters and the optimal network configuration with a grid-search approach; the value ranges and the optimized values are also in Table 2. We tested several configurations varying amounts of neurons and hidden layers, achieving the best results with a hidden layer of 180 neurons followed by two hidden layers of 64 neurons. We tested both Stochastic Gradient Descent (SGD) and ADAM [44] for the optimizer, achieving superior results with SGD. Additionally, we tested the log loss, Mean Square Error (MSE), and Root Mean Square Error (RMSE) loss functions, achieving superior results with MSE.

**Table 2**. Initial parameters for the LSTM model, the value range for the grid-search, and the optimized values.

| Parameter | Initial | Value range | Optimized |
|---|---|---|---|
| Epochs | 100 | [100, 10.000] | 200 |
| Batch size | 32 | [32, 64] | 32 |
| Dropout | 0.5 | [0, 0.8] | 0.5 |
| Neurons | 64 | [64, 512] | 128 (1), 64 (2) |
| Hidden Layers | 2 | [2, 3] | 3 |
| Momentum | - | [0, 0.5] | 0.5 |
| Optimizer | SGD | SGD, ADAM | SGD |
| Loss function | Log loss | Log loss, MSE, RMSE | MSE |
| Activation | Sigmoide | - | Sigmoide |

We tested both ICD-10 and qSOFA criteria considering a timestep of eight days, meaning that the classifier considered data from the eight days before. The timestamp of eight days has the advantage of allowing the monitoring of cases that evolved gradually; following the time series concept, the maximum historical value is obtained for each value to be classified [45]. Accordingly, we chose eight days to define which of the models better performs with maximum historical value.

In the following experiments, we considered only the model related to qSOFA, as it achieved the best results in the previous evaluation. To improve the system's performance, we modified the timesteps, testing values of one, three, five, and eight. Intuitively, the closer to the sepsis event, the more accurate the predictions will be.

*Experiment 2 – Random Forest*

In our last experiment, we aimed to evaluate the performance on a classification task, where we classified in the moment if the patient had sepsis or not. Additionally, and more importantly, to verify the feature importance by utilizing mean decrease GINI and mean decrease accuracy. We reported the values of the 14 most relevant features in relevance order. The Random Forest classifier was trained with the package *randomForest* [46] with parameters *ntree* equal to 100 e *mtry* equal to 2.

## Evaluation Criteria

For the experiment involving LSTM, we divided the dataset into 70% for training, 20% for validating, and 10% for testing. In addition, for the experiment involving Random Forest, we used 10-fold cross-validation. We used the accuracy, sensitivity, specificity for evaluating our binary classification results, as presented in [47]. Below we show the formulas where false positive (FP), false negative (FN), true positive (TP), and true-negative (TN).

- Sensitivity:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{1}$$

- Specificity:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{2}$$

- Accuracy:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{3}$$

## RESULTS

Using the ICD-10 strategy, we identified 164 admissions (3.52% of the total admissions) from 163 patients (3.72% of the total amount of patients) related to sepsis. With the qSOFA strategy, we identified 216 (4.49% of the total admissions) due to sepsis. We achieved a kappa value of 0.68 for the concordance between ICD-10 and qSOFA, which can be considered a substantial concordance by Landis, showing a positive concordance between ICD-10 and qSOFA sepsis detection. However, the concordance value also showed that there were differences between identified sepsis cases.

Regarding the experiments, the LSTM model results considering both ICD-10 and qSOFA criteria in a timestep of eight days are shown in Table 3. There was a significant difference in the LSTM model's accuracy based on ICD-10 from the model based on qSOFA. However, the difference was not that significant for the sensitivity, meaning that both models were able to determine true positives correctly. In the health domain, having a higher sensitivity is crucial as the objective is to correctly predict all patients with sepsis, even if several patients that do not have sepsis are misclassified. So, the best model was the one related to qSOFA.

**Table 3**. Accuracy, sensitivity, and specificity for the LSTM model considering both ICD-10 and qSOFA.

| Model | Accuracy | Sensitivity | Specificity |
|---|---|---|---|
| ICD-10 | 0.316 | 0.084 | 0.75 |
| qSOFA | 0.872 | 0.877 | 0.869 |

[1] Confidence interval of 95%.

Afterward, our experiments relied only on the qSOFA-based model, testing different timesteps. The first experiment results are shown in Table 4, where timesteps are the number of days since the suspected sepsis occurrence. We had better predictive power with small timesteps, especially using timesteps of one and three days. Thus, utilizing data from days closer to the sepsis events benefits the model's predictive power. In this scenario, timesteps of three days are adequate because of considering information from few days before the sepsis event.

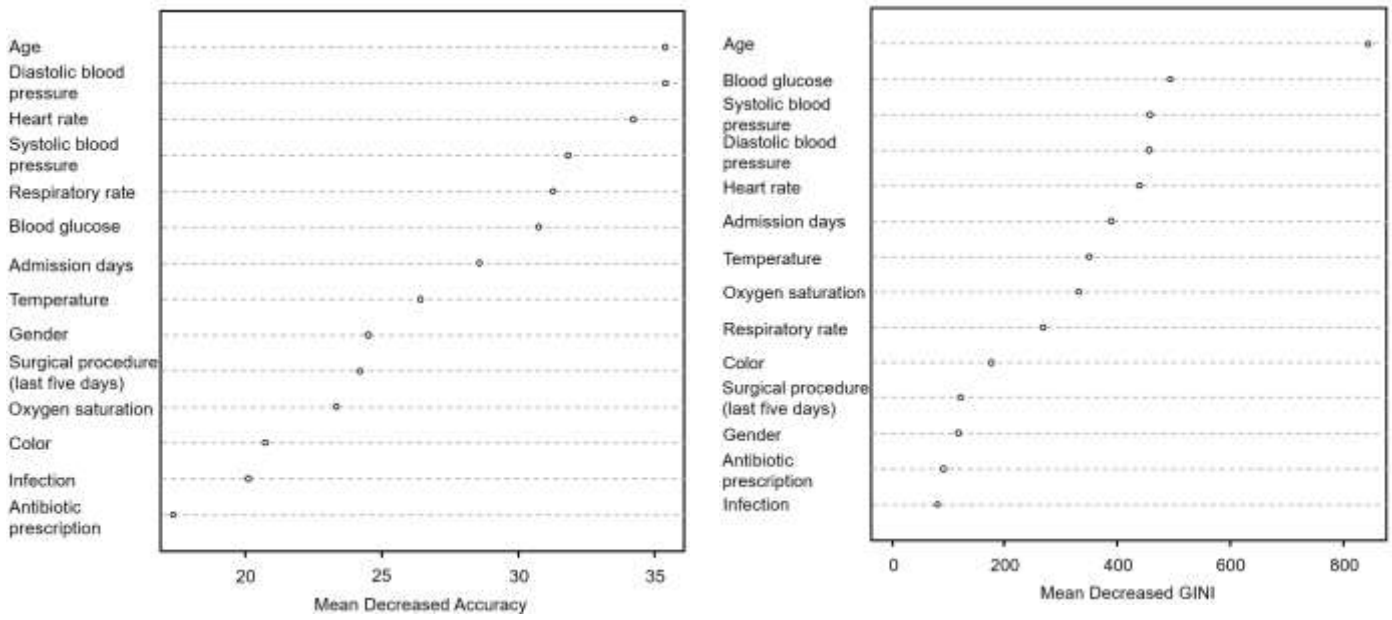
**Table 4**. Accuracy, sensitivity, and specificity for the LSTM model considering qSOFA with different timesteps.

| Timesteps | Accuracy | Sensitivity | Specificity |
|---|---|---|---|
| 8 | 0.872 | 0.877 | 0.869 |
| 5 | 0.896 | 0.899 | 0.936 |
| **3** | **0.907** | **0.912** | **0.971** |
| **1** | **0.913** | **0.922** | **0.989** |

[1] Confidence interval of 95%.

Afterward, we evaluated the performance of the Random Forest classifier. We achieved 0.971 for accuracy, 0.611 for sensitivity, 0.998 for specificity. The obtained results were expected since close to 95% of the records had missing data and were filled with pre-determined criteria, such as the median of their relative values or typical clinical reference values. However, as we noticed, the model presents a low sensitivity, misclassifying patients that have sepsis.

To understand which features were more valuable to the Random Forest classifier, we used the mean decrease accuracy and GINI. The relevance of each feature for both evaluation methods is shown in Figure 2. In both methods, the features *age*, *blood glucose*, *systolic blood pressure*, *diastolic blood pressure*, *heart rate*, *respiratory rate,* and *admission days* had the most contribution to the classifier's prediction. The feature *age* was the most relevant for both methods, showing a higher impact on classifier prediction when considering mean decrease GINI.




(a) (b)

**Figure 2**. (a) Feature importance with mean decreased accuracy for the Random Forest model; (b) Feature importance with mean decreased GINI for the Random Forest model.

## DISCUSSION

The percentual number of admissions related to sepsis identified by the ICD-10 and qSOFA, 3.52% and 4.49% respectively, were below the value reported in [3], where around 230 ICU units were analyzed, and patients with sepsis or septic shock occupied 30% of the hospital beds. The identification of patients with sepsis with ICD-9 codes was pointed out by [48] as a limitation of their study. Identifying patients with sepsis may be underestimated with ICD codes, as the cause of death can often be attributed to the underlying pathology and not to sepsis [3]. Moreover, analyzing the qSOFA criteria, we verified that the low percentage of occurrence detection might be related to the process of filling missing data with standard values for heart and respiratory rate, pressure, and temperature. Hence, abnormal values that could change the qSOFA score are hidden over traditional values. Thus, both ICD-10 and qSOFA criteria are underreporting sepsis cases, which could negatively impact our model performance.

Regarding the model developed for early identification of patients at risk of sepsis, the results showed that the model trained based on qSOFA selection criteria had superior performance than the ICD-10 criteria. These results were obtained using only seven vital signs and some information about the patient's hospitalization, a simple set of features. Our results were positive, especially when we achieved only around 4% lower than the study of [20], a reference for developing a commercialized tool specialized in the early recognition of sepsis.

The model's predictive power was tested by reducing the timesteps, with the proposed model showing promising results, reaching 0.913 for accuracy, 0.922 for sensitivity, and 0.989 for specificity in predicting sepsis on the day that the sepsis manifested. The models from [18,20,30] had predictions with timesteps based on hours, considering the hours preceding the sepsis event. Besides, their experiments were based on the MIMIC datasets, freely available data sources containing health and demographic information, where the data collection frequency varies from minutes to hours. In our study, we had a limitation due to the quality of data, in which the data collection frequency was low (one measure per day), and there was a large amount of missing data. For instance, we identified that 94.79% of the vital signs' documentation was incomplete (with at least one missing measure), especially for respiratory rate, temperature, and blood glucose, where the amount of missing data was 78.78%, 85.78%, and 86.21%, respectively.

Further, the Glasgow coma scale was rarely registered on our dataset, being registered in less than 1% of the records. Unfortunately, this scenario of having missing values in electronic health records (EHRs) is a reality for hospitals [49], directly impacting the dataset size and adding noise into the predictive model. One

aspect contributing to missing data is that several measures are often only recorded proportionally to the way they change over time, meaning that clinicians generally considered it normal and did not add it into the EHRs [17].

Our models had the ability to predict patients with sepsis in a scenario of noisy and scarce data. Comparing the Random forest results with the LSTM results with the timestep of three days, the first presented better accuracy and specificity values. However, the sensitivity dropped by 30.1%, which indicates that the use of time series in this context is still the best alternative since the sensitivity represents the algorithm's ability to identify patients at risk of sepsis.

The features that were the most important to the Random Forest classifier were *age*, *blood glucose*, *systolic blood pressure*, *diastolic blood pressure*, *heart rate*, *respiratory rate,* and *admission days*. The features *respiratory rate* and *systolic blood pressure* are part of the qSOFA criteria; so its positive to verify that these features had a high contribution to the classifier's prediction. The feature *blood glucose* was relevant to the classifier because it highlights patients with possible comorbidities, which are more vulnerable. Besides, sepsis alone can also affect blood glucose levels [50]. The feature *age* was the most relevant feature, similar to the one obtained by [21]. Moreover, in the study of [21], *heart rate* was also a feature with high impact, similar to the result obtained in this study. We highlight that antibiotic prescription and infection features had a low influence over the classifier's prediction because there is a delay in obtaining information about infection and starting the antibiotic. When this information is documented over the records, the patient is generally already with sepsis.

Unfortunately, we could not add the Glasgow coma scale into our feature set due to the high amount of missing data, with less than 1% of filled data. In the study of [51], they identified that the most relevant feature to the qSOFA score was the Glasgow coma scale. Hence, further experiments with this feature would be beneficial.

## CONCLUSION

This study aimed to develop a predictive model using LSTM and multivariate time series to recognize patients at risk of sepsis. The LSTM was chosen due to memorizing the temporal dependencies of long periods, making it possible to capture even the most subtle progressions of sepsis. It was possible to analyze sepsis cases in the studied hospital, a relevant study since studies correlating sepsis and machine learning are scarce for Brazilian hospitals.

We achieved an accuracy of 0.907, a sensitivity of 0.912, and a specificity of 0.971 when considering a three-day timestep with LSTM. These results were positive compared to the literature, especially in a scenario of spaced measures and a high amount of missing data. Also, we analyzed the features that had the most impact over the Random Forest classifier, verifying the high relevance of the features *age*, *blood glucose*, *systolic blood pressure*, *diastolic blood pressure*, *heart rate*, *respiratory rate*, and *admission days*. The feature *age* was the most relevant feature; thus, we indicate future sepsis prediction studies adding this feature to their predictive model.

We proposed an initial study aiming to provide tools to support the clinical staff in reducing the time between the onset of symptoms and the first medical care, as this is the only way to optimize the planning of interventions regarding the implementation of care protocols. However, this was a preliminary study, and the developed tool still needs to be validated by health professionals in a real scenario.

Besides, to deploy a tool that could be used in a clinical scenario, we highlight the need for more complete data to train and test the algorithms. The scenario of only a single measure per day and a high amount of missing data is not ideal when training a classifier. For sepsis, in which the patient survival rate reduces by up to 8% for each treatment delayed hour [14], more frequent patient data collection (several measures per day) is desired. Hence, we signalize that the minimum would be three vital signal measures per day, one measure in the morning, another in the afternoon, and another at night. Moreover, there is a need for more significant personnel and infrastructure investments to obtain higher quality data.

For future work, we highlight the possibility of modifying some of the definitions used in this study. For instance, we utilized the definition for hypotension from [2], which was created considering the data availability and the necessity of an objective criterion to compose the risk score at the bedside. However, instead, we could evaluate other criteria, such as switching to the hypotension definition defined by the 6th Ambulatory Blood Pressure Monitoring Guidelines and the 4th Residential Blood Pressure Monitoring Guidelines [52]. Further, both ICD-10 and qSOFA indicate sepsis, yet the ideal scenario would be a sepsis confirmation by specialists, validating the gold standard. Additionally, we could use more elaborated natural language processing tools to extract some of the missing information, such as substituting the Glasgow

comma scale missing values with conclusions from mentions in the free text representing the patient's neurological condition.

**Acknowledgments:** The Brazilian Government Agency Coordination for the Improvement of Higher Education Personnel (CAPES) supported this work.
**Conflicts of Interest:** The authors declare no conflict of interest.

## REFERENCES

1. WHO, World Health Organization. Report on the burden of endemic health care-associated infection worldwide, 2011.
2. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016 Fev;315(8):801-810.
3. ILAS, Instituto Latino Americano De Sepse. Sepse: um problema de saúde pública. 2017.
4. Zonta, FNS, Velasquez PGA, Velasquez LG, Demetrio LS, Miranda D, Silva, MCBD. Características epidemiológicas e clínicas da sepse em um hospital público do Paraná. Revista de Epidemiologia e Controle de Infecção. 2018;8(3):224-231.
5. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. The Lancet. 2020;395 (10219):200-2011.
6. Lobo SM, Rezende E, Mendes CL, de Oliveira MC. Mortalidade por sepse no Brasil em um cenário real: projeto UTIs brasileiras. Rev Bras Ter Intensiva. 2019;31(1):1-4.
7. Ghalwash M, Radosavljevic V, Obradovic Z. Early Diagnosis and Its Benefits in Sepsis Blood Purification Treatment. Proceedings of the 2013 IEEE International Conference on Healthcare Informatics (ICHI);2013. p. 523-528.
8. Westphal GA, Lino AS. Rastreamento sistemático é a base do diagnóstico precoce da sepse grave e choque séptico. Rev Bras Ter Intensiva. 2015;27(2):96-101.
9. Kaukonen KM, Bailey M, Pilcher D, Cooper DJ, Bellomo R. Systemic Inflammatory Response Syndrome Criteria in Defining Severe Sepsis. N Engl. J. Med. 2015;372(17):1629-38.
10. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive Care Med. 1996;22(7):707-10.
11. Macdonald SP, Williams JM, Shetty A, Bellomo R, Finfer S, Shapiro N, et al. Review article: Sepsis in the emergency department - Part 1: Definitions and outcomes. Emerg Med Australas.;29(6):619-25.
12. Kushimoto S, Gando S, Ogura H, Umemura Y, Saitoh D, Mayumi T, et al. Complementary Role of Hypothermia Identification to the Quick Sequential Organ Failure Assessment Score in Predicting Patients With Sepsis at High Risk of Mortality: A Retrospective Analysis From a Multicenter, Observational Study. J Intensive Care Med. 2020;35(5):502-10.
13. Serafim R, Gomes JA, Salluh J, Póvoa P. A comparison of the Quick-SOFA and systemic inflammatory response syndrome criteria for the diagnosis of sepsis and prediction of mortality a systematic review and meta-analysis. Chest. 2018;153(3):646-55.
14. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Critical Care Med. 2006; 34(6):1589-96.
15. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. New Engl J Méd. 2001;345(19):1368-77.
16. Gultepe E, Nguyen H, Albertson T, Tagkopoulos I. A Bayesian network for early diagnosis of sepsis patients: a basis for a clinical decision support system. Proceedings of the 2012 IEEE 2nd International Conference on Computational Advances in Bio and Medical sciences (ICCABS); 2012. p. 1-5.
17. Lipton ZC, Kale DC, Elkan C, Wetzel R. Learning to diagnose with LSTM recurrent neural networks. Proceedings of the 2016 International Conference on Learning Representations (ICLR); 2016.
18. Kam, HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. Comput Biol Med. 2017 Oct;89:248-55.
19. Zhang Y, Lin C, Chi M, Ivy J, Capan M, Huddleston JM. LSTM for septic shock: adding unreliable labels to reliable predictions. Proceedings of the 2017 IEEE International Conference on Big Data (Big data); 2017. p. 1233-1242.
20. Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open. 2018;8(1):e017833.
21. García-Gallo JE, Fonseca-Ruiz NJ, Celi LA, Duitama-Munõz JF. A machine learning-based model for 1-year mortality prediction in patients admitted to an intensive care unit with a diagnosis of sepsis. Med Intensiv. 2018;44(3):160-70.

22. Gultepe E, Green JP, Nguyen H, Adams J, Albertson T, Tagkopoulos I. From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. J Am Med Inform Assoc. 2014;21(2):315-25.
23. WHO, World Health Organization. International statistical classification of diseases and related health problems: Tabular list, 2004.
24. Landis RJ, Kosh G. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.
25. Kuritz SJ, Landis JR. Attributable risk estimation from matched case-control data. Biometrics. 1988;44(2):355-67.
26. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. AI Magazine. 1996 Mar;17(3):37.
27. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Mengling F, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Scientific data. 2016 May;3(1):1-9.
28. Guillén J, Liu J, Furr M, Wang T, Strong S, Moore CC, et al. Predictive models for severe sepsis in adult ICU patients. Proceedings of the 2015 IEEE Systems and Information Engineering Design Symposium (SIEDS); 2015. p. 182-187.
29. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Science Translational Medicine. 2015; 7(299):299ra122.
30. Calvert JS, Price DA, Chettipally UK, Barton CW, Feldman MD, Hoffman JL. et al. A computational approach to early sepsis detection. Comput Biol Med. 2016;74:69-73.
31. Ghosh S, Li J, Cao L, Ramamohanarao K. Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. J Biomed Inform. 2017 Feb;66:19-31.
32. Jiang Y, Tan P, Song H, Wan B, Hosseini M, Sha L. A self-adaptively evolutionary screening approach for sepsis patient. Proceedings of the 2016 International Symposium on Computer-based Medical Systems (CBMS); 2016. p. 60-65.
33. Mitchell S, Schinkel K, Song Y, Wang Y, Ainsworth J, Halbert T, et al. Optimization of sepsis risk assessment for ward patients. Proceedings of the 2016 IEEE Systems and Information Engineering Design Symposium (SIEDS); 2016. p. 107-112.
34. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PLoS One. 2017 Apr;12(4):e0174708.
35. Salomão R, Diament D, Rigatto O, Gomes B, Silva E, Carvalho NB, et al. Diretrizes para tratamento da sepse grave/choque séptico: abordagem do agente infeccioso – controle do foco infeccioso e tratamento antimicrobiano. Rev Bras Ter Intensiv. 2011 Jun;23(2):145-57.
36. Van Der Loo MPJ. The stringdist package for approximate string matching. The R Journal 2014;6(1):111-22.
37. Saqib M, Sha Y, Wang MD. Early prediction of sepsis in EMR records using traditional ml techniques and deep learning LSTM networks. Proceedings of the 2018 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018. p. 4038-4041.
38. Kale DC, Gong D, Che Z, Liu Y, Medioni G, Wetzel R, et al. An Examination of Multivariate Time Series Hashing with Applications to Health Care. Proceedings of the 2014 IEEE International Conference on Data Mining (ICDM); 2014. p. 260-269.
39. Kelly G. Body temperature variability (part 1): a review of the history of body temperature and its variability due to site selection, biological rhythms, fitness, and aging. Alternat Medi Rev. 2006 Dec; 11(4):278-93.
40. SBD, Sociedade Brasileira de Diabetes. Diagnóstico e classificação do diabetes mellitus e tratamento do diabetes mellitus tipo 2 [Internet]. 2000 May [cited 2021 Mar 9]: 71 p. Available from: http://bvsms.saude.gov.br/bvs/publicacoes/consenso_bras_diabetes.pdf.
41. Paschoal MA, Volanti VM, Pires CS, Fernandes FC. Variabilidade da freqüência cardíaca em diferentes. Rev Bras Fisioter. 2006;10:413-419.
42. Parreira VF, Bueno CJ, França DC, Vieira DSR, Pereira DR, Britto RR. Padrão respiratório e movimento toracoabdominal em indivíduos saudáveis: influência da idade e do sexo. Rev Bras Fisioter. 2010 Oct; 14(5):411-6.
43. Malachias MVB, Plavnik FL, Machado CA, Malta D, Scala LCN, Fuchs S. 7ª Diretriz brasileira de hipertensão arterial. Arquiv Brasil Cardiol. 2016;107(3):1-6.
44. Kingma PD, Ba LJ. Adam: A Method for Stochastic Optimization. Proceedings of the 2015 International Conference on Learning Representations (ICLR); 2015.
45. 45. Bontempi G, Ben Taieb S, Le Borgne YA. Machine Learning Strategies for Time Series Forecasting. In: Aufaure MA, Zimányi E, editors. Business Intelligence. eBISS 2012; 2012; Berlin, Heidelberg: Springer; 2013. p. 62-77.
46. Liaw A, Wiener M. Classification and regression by randomForest. R news. 2002;2(3):18-22.
47. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag. 2009; 45(4):427-37.
48. Van Wyk F, Khojandi A, Kamaleswaran R, Akbilgic O, Nemati S, Davis RL. How much data should we collect: a case study in sepsis detection using deep learning. Proceedings of the 2017 IEE Healthcare Innovations and Point of Care Technologies (HI-POCT); 2017. p. 109-12.
49. Beaulieu-Jones BK, Moore, JH. Missing data imputation in the electronic health record using deeply learned autoencoders. Proceedings of the 2017 Pacific Symposium on Biocomputing; 2017. p. 207-218.

50. Aleman L, Guerrero J. Hiperglicemia por sepsis: del mecanismo a la clínica. Rev Méd Chile. 2018;146(4):502-10.
51. Gupta A, Liu T, Shepherd S, Paiva W. Using statistical and machine learning methods to evaluate the prognostic accuracy of SIRS and qSOFA. Healthcare Informatics Research. 2018 Apr; 24(2):139-47.
52. Brandão AA, Alessi A, Feitosa AM, Machado CA, de Figueiredo CAP, Amodeo C, et al. 6ª Diretrizes De Monitorização Ambulatorial Da Pressão Arterial E 4ª Diretrizes De Monitorização Residencial Da Pressão Arterial. Arquiv Brasil Cardiol. 2018; 110(5):1-29.