Sepsis Early Detection

Asif Mahdin amahdin@ucsd.edu
Colin Tran ctt005@ucsd.edu
Diego Zavalza dzavalza@ucsd.edu
Zhuji Zhang zhz044@ucsd.edu

Mentor: Professor Kyle M. Shannon
kshannon@ucsd.edu

Abstract

Our team aims to build on the bridge between advanced data science methods the health care industry by revolutionizing the way doctors, nurses, and hospitals identify sepsis in a timely and effective manner. Our project uses a Long Short Term Memory (LSTM) model to forecast the probability of a patient developing sepsis up to 50 hours before current methods diagnosis patients. By leveraging advanced machine learning algorithms and data science methods, we are able to take multiple features such as vital signs, patient backgrounds, and more to design a predictive model allowing health care professionals diagnosis and treat patients at an earlier stage resulting in prompt intervention and lowering patient's chances of death.

Website:
https://sepsis-early-detection.github.io/sepsis-early-detection.io/

Code: https://github.com/DZavalza2/sepsis_early_detection

Contents

1 Introduction

Sepsis is a life-threatening conditions in which the body is failing to responds to an infection, causing the body's organs to stop working. Furthermore, it stands as the leading cause of mortality for in house deaths in the ICU. Sepsis is a very complex illness in which it typically stems from pre-existing comorbidities and underlying health conditions.

For our model, we are using a dataset called MIMIC-III (Johnson et al. 2018) from PhysioNet (PhysioBank 2000). It is a relational database consisting of 26 tables that can be linked by identifiers, comprising deidentified health-related data associated with over forty thousand patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012. It is an extremely useful dataset when it comes to analysis since it includes information such as demographics, vital sign measurements made at the bedside, laboratory test results, procedures, medications, caregiver notes, imaging reports, diagnoses (represented by ICD-9 codes) and mortality (including post-hospital discharge). We will be using some of these variable in our analysis (Johnson, Pollard and Mark III 2016)

2 Methods

2.1 Data Cleaning

For our model, we are using the 4 major vital signs outlined by UCSD sepsis diagnosis protocol, it includes body temperature, white blood cell count, heart rate, and respiration rate. The dataset also included unique `subject_id` to identify patients and `hourly_bin` for the time which the measurement was taken.

We begin by looking at the types of each column, and the data types `hourly_bin` and `WBC` is off. The `hourly_bin` column has the data type of a object, and in order for us to better utilize the information time data provides, we decided to convert the data type to a more standard Python datetime object.

After getting the time stamps into the desired format, we moved on to the cleaning of vital signs. White Blood Cell count (WBC) has several invalid string values that was mixed with valid float values, causing the column type to be object instead of float. We replaced invalid values such as "no data" and "error" with `np.nan` and converted the column type to float for easier visualization and processing. Extremely high values of WBC are kept since there are diseases and conditions that can cause the WBC counts to shoot up.

For the temperature column, there is a peak around the value of 100, and a more careful examination suggests that they are temperatures in Fahrenheit. The Fahrenheit values were then converted to Celsius and invalid values such as negatives and 0s are replaced with `np.nan`.

2.2 Cohort Selection

For our cohort, we decided to work with patients above the age of sixteen (this is because symptoms are different for younger kids as opposed to adults). We then extracted all vital signs for patients we had available and merged them together. In some cases, we noticed that only certain vital signs were recorded therefore, we imputed the data with the average value for each vital sign. Furthermore, in order to ensure that our model was not uneven (and predicting one label because it was the majority), we balanced our dataset in terms of having equal patients with sepsis and without sepsis.

2.3 LSTM Model

2.3.1 Model Description

Our model utilizes an LSTM-based architecture, chosen for its proficiency in handling sequential data and its ability to capture long-term dependencies. LSTMs are particularly suited for medical time-series data, making them ideal for our objective of sepsis prediction.

2.3.2 Model Architecture

The core of our predictive model comprises two LSTM layers followed by dropout layers to mitigate overfitting. The first LSTM layer is configured to return sequences, allowing it to pass temporal information to the subsequent LSTM layer. This stacking of LSTM layers helps in learning complex patterns in the data. The final layer is a dense layer with a sigmoid activation function, designed to output the probability of sepsis occurrence.

- **Input Layer**: Accepts sequences with a predefined length, corresponding to the number of time steps considered for each patient's data.
- **LSTM Layers**: Two LSTM layers with 125 units each, chosen to balance model complexity and computational efficiency. The return sequences option is enabled for the first LSTM layer to maintain temporal information flow.
- **Dropout Layers**: Implemented with a dropout rate of 0.2 after each LSTM layer to prevent overfitting by randomly omitting a fraction of the neurons during training.
- **Output Layer**: A dense layer with a single neuron and a sigmoid activation function, providing the probability of sepsis occurrence.

2.3.3 Implementation

The model was implemented using TensorFlow and Keras, with the Adam optimizer and binary cross-entropy loss function. The choice of optimizer and loss function is standard for binary classification problems, given their effectiveness in handling such tasks.

2.3.4 Training and Validation

Training was conducted on a split of the MIMIC dataset, with separate sets for training, validation, and testing. We employed a batch size of 64 and ran the training process for 5 epochs, a decision made to balance between model performance and training time.

2.3.5 Hyperparameter Tuning

We explored several configurations of LSTM units (50, 75, 100, 125, 150) and dropout rates (0.2, 0.3, 0.4) to identify the optimal setup for our model. The final model architecture was selected based on its performance on the validation set, with a focus on maximizing recall to ensure the identification of as many true sepsis cases as possible, considering the critical nature of the condition.

3 Results

3.1 Model Performance

The model's performance was evaluated on the results of predicting 5 hours ahead of time, and used several metrics, including accuracy, recall, and AUROC to assess the model's ability to distinguish between sepsis and non-sepsis cases effectively.

- Accuracy and Recall: Correctly identifying true labels and underscoring its ability to capture the majority of actual sepsis cases with an 81% accuracy and recall. (Figure 1)

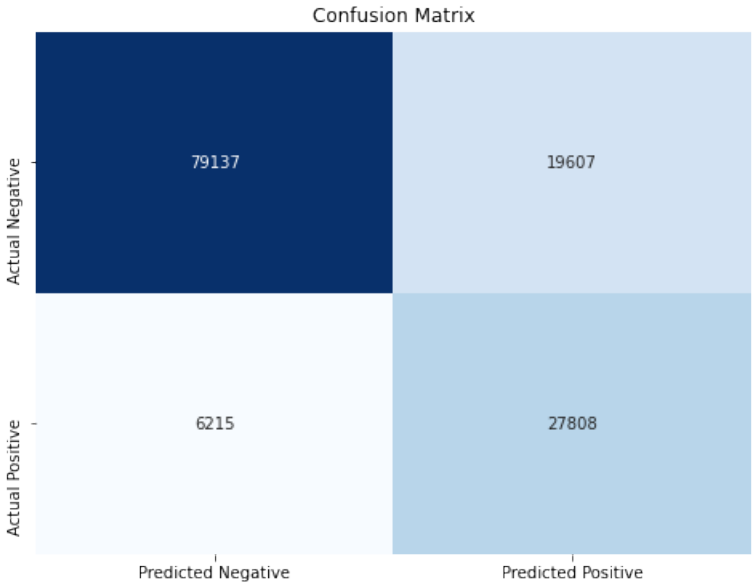

Figure 1: Confusion Matrix for predicting 5 hours ahead of time

- The area under the curve (AUC): The closer the curve is to the top left corner (closer to 1), the higher the accuracy of the test since it means high true positive rate and low false positive rate (Nahm 2022). In our case, Area Under the Receiver Operating Characteristic (AUROC) curve reached 0.89, which can be considered as a good model. (Figure 2)

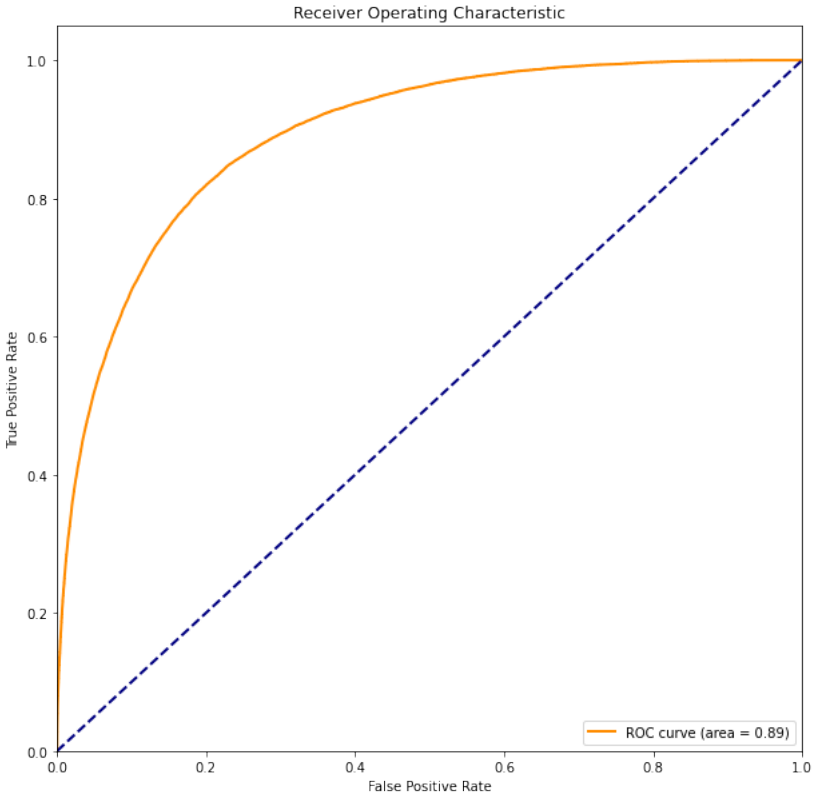

Figure 2: ROC Curve

We also tried predictions for multiple hours ahead of time, and we found that the accuracy drops as the hours ahead of time we are trying to predict increases (Table 1), and this is to be expected due to the variability in between the times.

Table 1: Results for predictions N hours ahead of time

	5 Hrs	10 Hrs	15 Hrs	20 Hrs	25 Hrs	50 Hrs
Accuracy	80.9%	79.6%	77.9%	77.6%	77.6%	75.1%
Recall	81.4%	78.5%	78.1%	75.1%	74.4%	67.9%
F1 Score	68.4%	71.2%	71.8%	72.4%	73.0%	70.2%
Precision	58.9%	65.2%	66.5%	69.9%	71.7%	72.6%

3.2 Conclusion

Using LSTM model with multiple layers, we were able to achieve an accuracy of 80.9% on our test dataset, and we were also able to find a balance between the true positive rate and false positive rate using the ROC Curve. Overall, we believe there is a future for the healthcare workers to utilize the model as a secondary opinion, and along with their experiences, the development of sepsis can be contained from the very beginning and the mortality rate can be reduced significantly through early intervention.

4 Discussion

4.1 Data Balancing for Algorithm Fairness

To ensure the fairness and reliability of our predictive model, a critical step undertaken was the balancing of the dataset. Given the inherent imbalance in medical datasets—where instances of sepsis are considerably rarer compared to non-sepsis cases—unadjusted, this skew could lead to a model biased towards predicting the majority class, thus undermining its clinical utility. For every training epoch, we ensured an equal distribution of sepsis and non-sepsis labels presented to the model, thereby maintaining balance in the learning process. Upon the completion of each epoch, we shuffled the predominant non-sepsis data with fresh instances. This approach guaranteed that our model was exposed to a varied set of data points, while still preserving balanced label representation throughout the training phase.

4.2 Limitations

When choosing our features, we decided to go with the four vital signs listed on the UCSD inpatient protocol, or SBAR (Situation, Background, Assessment, Recommendation), and blood pressure, heart rate, respiration rate, and body temperature are criteria for suspected infection. Even though these vital signs are crucial to determining sepsis, other measurements could be utilized to better train the model, including blood pressure (Mean Arterial Pressure), platelets, and lactate levels. We could add these measurements to our data for future improvements on the model.

Another concern is that MIMIC-III dataset is from Beth Israel Deaconess Medical Center between 2001 and 2012, and the protocol we used to extract features from the dataset in from present day UC San Diego Medical Center - Hillcrest Campus. This is a potential issue when it comes to assign labels (with sepsis and without sepsis). In the future, it would be beneficial to switch to UCSD in-house electronic medical records for a match between data and protocol.

4.3 Ethical Considerations/Tradeoffs

We identified three main stakeholders for our project: the users, including patients, healthcare professional, researchers, and data scientists; regulatory bodies, including government and industry entities responsible for ensuring the compliance to existing laws and regulation; funding entities, including organizations or individuals who provide financial resources to the project and the insurance companies which concerns the costs related with sepsis. With these stakeholders in mind, we proposed a few potential conflicts/tradeoffs that could arise in the future deployment of the model:

1. **Privacy vs model accuracy**: Patients values the privacy while data scientist/researcher values the details of the data in order to make accurate predictions
2. **Model Complexity vs. Explainability**: As a model get more complex, the results tend to get more accurate. However, a complex model is often hard to explain to people with non-technical backgrounds, and this could lead to hesitation during the implementation of the model.
3. **Cost vs. Benefit**: There is a cost that's related with integrating the model into the workflow of healthcare workers. The cost of using the model, along with the learning curve of the system could lead to confusion.
4. **New technology vs Regulation**: New technology will be under scrutiny, especially when it comes to the lives of millions of patients, and the lengthy process could potentially slow down the development of the model.

References

Johnson, Alistair E W, David J Stone, Leo A Celi, and Tom J Pollard. 2018. "The MIMIC Code Repository: enabling reproducibility in critical care research." *Journal of the American Medical Informatics Association* 25 (1): 32–39

Nahm, Francis Sahngun. 2022. "Receiver operating characteristic curve: overview and practical use for clinicians."

PhysioBank, PhysioToolkit. 2000. "Physionet: components of a new research resource for complex physiologic signals." *Circulation* 101 (23): e215–e220