**Introduction:**

Sepsis occurs when bacteria enter the bloodstream and elicit a systemic inflammatory response that injures tissues and organs around the body, leading to systemic organ failure and eventually death. In the United States, sepsis is the most common cause of death among patients in non-coronary intensive care units (ICUs) ([Angus *et al.*, 2001](https://insights.ovid.com/pubmed%3Fpmid%3D11445675)) and in Canada, Sepsis was reported to contribute to one in every eighteen deaths in 2011, making it the twelfth leading cause of death ([statcan](https://www150.statcan.gc.ca/n1/pub/82-624-x/2016001/article/14308-eng.htm)). Globally, it is estimated that 6 million people die from Sepsis annually. In addition to its negative impact on health, Sepsis cost US hospitals an estimated $24 billion USD annually. Much of these costs are associated with patients that were not detected with Sepsis early in their admission.

Identification of organ dysfunction associated with sepsis is done using the Sequential [Sepsis-related] Organ Failure Assessment (SOFA) score, a scoring system that evaluates oxygen levels, blood pressure, liver function, platelet levels, mental state, and kidney function. To detect early sepsis, physicians use the quick SOFA (qSOFA) score which consists of three components: respiratory rate ≥ 22 breaths per minute, altered mental state, and systolic blood pressure ≤100 mmHg, for which the presence of at least two components is associated with poor outcomes due to sepsis. Although the qSOFA score may facilitate earlier detection and management of sepsis, an analysis of 184,875 ICU patients showed that it was limited in its ability to predict in-hospital mortality (AUROC 0.61 for qSOFA vs AUROC 0.75 for SOFA) ([Raith *et al.,* 2017](https://jamanetwork-com.myaccess.library.utoronto.ca/journals/jama/fullarticle/2598267)). As each hour of delayed treatment for sepsis can increase mortality by approximately 4-8% ([Kumar *et al.*, 2006](https://www.ncbi.nlm.nih.gov/pubmed/16625125); [Seymour *et al.*, 2017](https://www.ncbi.nlm.nih.gov/pubmed/28528569)), there is thus a critical need for novel approaches that enable both early and accurate detection of sepsis in critically ill patients.

As the physiological signs and laboratory features of sepsis extend beyond those considered by the SOFA score, incorporation of more features may allow for more accurate early detection of sepsis. This, together with ever-increasing capacity to collect and monitor physiological markers in patients over the course of their stay in the ICU, provides an opportunity to employ deep learning approaches to utilize this data and develop better models to detect sepsis early. We are thus proposing to build a deep learning model towards predicting sepsis earlier than currently possible by using physiological data collected from ICUs.

**Data:**

Data for this project will be sourced from an ongoing Computing in Cardiology Challenge (<https://physionet.org/challenge/2019/>). Each patient’s data from the challenge dataset is contained in a pipe-delimited file, of which there are 40 336. Each file consists of 41 columns, corresponding to the 41 clinical measurements collected (40 of which are time-dependent; 1 is the sepsis label, which has been shifted six hours ahead), and the number of rows is dependent on the time in which the patient was in the ICU (one row corresponds to one hour). In total, the data can be represented by a 1 522 210 x 41 dimensional matrix. Missing data will be imputed either through forward propagation of the most recent observation or using a random forest regressor.

**Approach:**

Given that we will be working with time-series data and that our aim is to predict sepsis hours in advance, we believe that variants of the Recurrent Neural Network (RNN) architecture would be best suited for this task. We will be employing Long Short-Term Memory (LSTM) networks to model the data and predict sepsis. LSTM is a variation of the RNN that includes gates to modulate the ‘memory’ of the RNN (i.e. what past outputs will be included and to what extent) and attenuates the vanishing and exploding gradient problems, which thus allows for longer-term ‘memory’ of past outputs. In this way, LSTMs are ideal for time-series data where many observations are collected over time. We may also experiment with other variations of RNNs that have been used for time series data, including Gated Recurrent Units (GRUs). GRUs are similar to LSTMs but utilize fewer gates, which reduces model complexity and may potentially lead to increased training efficiency as well as reduced overfitting.

We will be employing a 50-25-25 Train-Validation-Test split to train the model. To build the base architecture of the model, we will gradually increase the complexity of the model by adding layers and increasing the number of units in each layer until the model begins to overfit the data. At this point, we will add regularization and begin tuning model hyperparameters such as dropout rate, learning rate, activation functions, and optimizers to maximize the performance of the model.

**Evaluation Plan:**

The evaluation for our proposed neural network model will be based on the established

machine learning classification task performance criteria. The accuracy of the trained neural

network in classifying sepsis patients will be tested using a testing data set and compared with

actual patient outcomes. The results will be summarized using multiple confusion confusion matrices (by varying the diagnosis threshold for the prediction) and its various derived measures. Each confusion matrix as a 2x2 contingency table is structured as follows:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Actual Sepsis Diagnosis** | |
|  |  | **Yes** | **No** |
| **Predicted Sepsis Diagnosis** | **Yes** | True Positive (TP) | False Positive (FP) |
| **No** | False Negative (FN) | True Negative (TN) |

From the confusion matrices, several derived model performance measures are to be estimated for assessing model performance. True positive rate (sensitivity or recall; defined as TP/(TP + FN)) and false positive rate (1 – specificity; defined as 1 - TN/(TN + FP)) measure the probability of producing true positives and false positives when conditioned on individuals be actually sepsis patients or not. The receiver operating curve (ROC) will be plotted that shows the true positive and false positive rates at each diagnosis prediction criteria threshold and its area (AUROC) will be quantified as an evaluation metric.

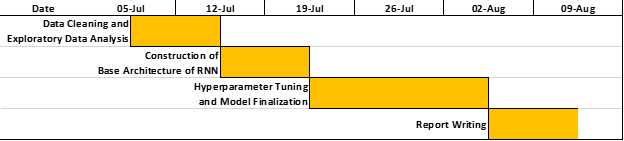
As the ROC does not account for the incidence of Sepsis in the ICU patient population, we will also evaluate precision (or positive predictive value; defined as TP / (TP + FP)). Accordingly, a precision-recall curve across different thresholds will be plotted and its area (AUPRC) will also be quantified to evaluate the performance of our model.

Finally, a lift score which compares the neural network predicted sepsis diagnosis to predictions made in truly random fashion is also used in this project. This score will assess the improvement in our neural network over random guess. The score is defined as follows.



A lift chart will also be constructed that plots the number of true positive and the proportion of sepsis-positive predictions at each diagnosis threshold.

**Project Timeline:**

****

**Group Responsibilities:**

Data cleaning, which will include 1) formatting the 40 336 pipe-delimited files for each participant into one comma-separated file and 2) imputing missing data, will be done primarily by 1) Steven and Faizan and 2) Yidi and Andy. The initial exploratory data analysis will be done by all members of the group to get an initial visual representation of potential relationships between variables. Building of the base architecture will be done together as a group, while hyperparameter tuning will be done in parallel between the group members and the hyperparameters of the best performing model will be used for in our final model. All members will contribute to various portions of the final written report and presentation.