Sepsis Prediction

CSE 6250 Project

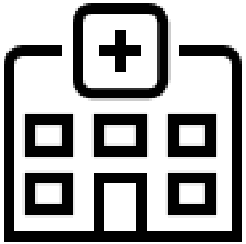
Team 11: Hao Lee, Michael Hur, Qasim Nazir, David Wu

Outline

- Introduction and Background
- Problem Formulation
- Approach and Implementation
- Experimental Evaluation
- Discussion and Challenges
- Conclusion

Introduction and Background

Introduction



- Sepsis is a medical condition where the immune system damages the body as a result of fighting infection
- If not treated, sepsis leads to septic shock, damaged organs, and death
- The CDC finds that each year, **1.7 million** adults develop sepsis, and nearly **270,000** die as a result of sepsis

Background

- There has been significant research in this field:
  - TREWScore (Henry et. al. 2015)
  - Insight (Calvert et. al. 2019)
  - LiSep LSTM (Fagerstorm et. al. 2019)

- Currently, there lacks a highly sensitive prediction system unique to acute sepsis
- Reliable sepsis identification and prediction for early treatment can save lives

Problem Formulation

Problem Formulation

Problem:

Early identification and prediction of sepsis, as defined by sepsis-3

Solution:

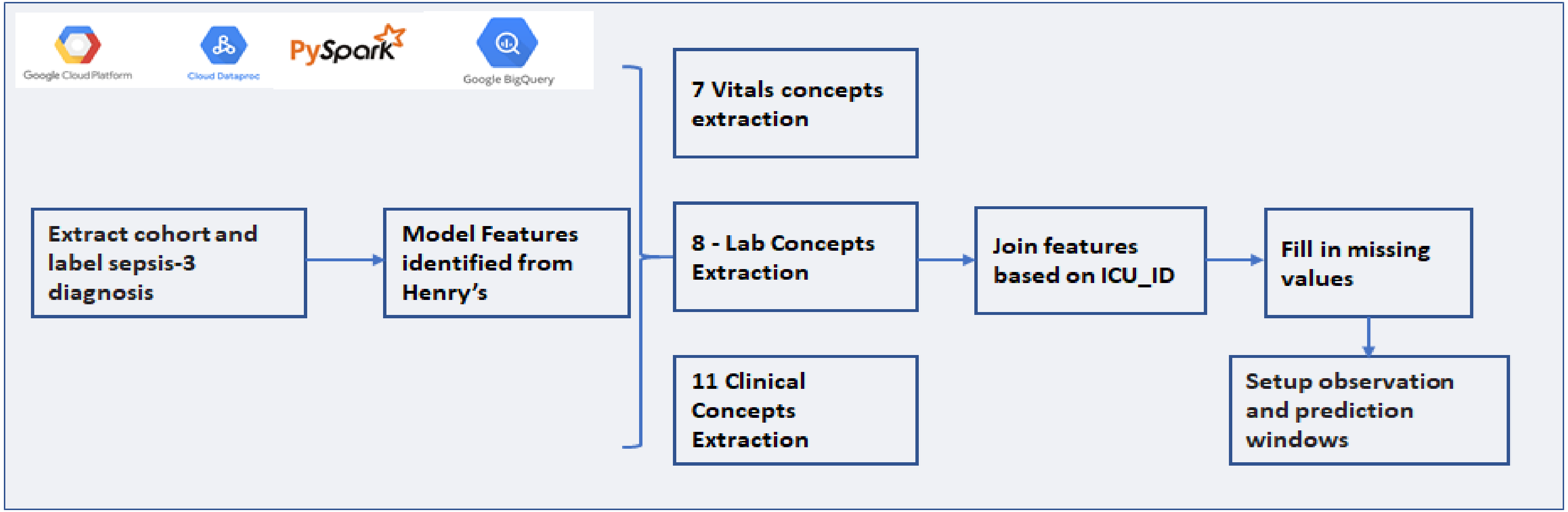
Introduce and replicate a Long Short-Term Memory (LSTM) neural network model

Data:

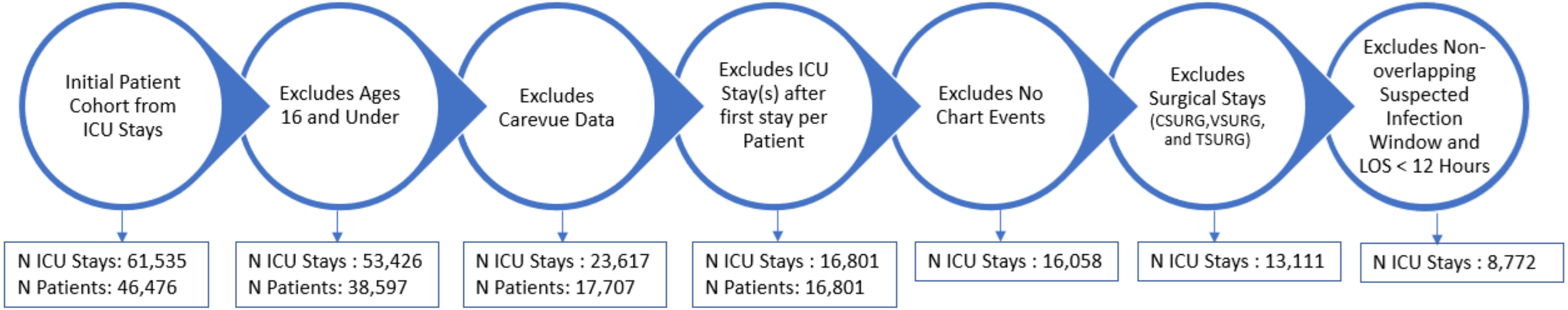
Patient features from the Medical Information Mart for Intensive Care (MIMIC)-III dataset

Approach & Implementation

ETL Process


Google Cloud Platform
Cloud Dataproc
PySpark
Google BigQuery
Extract cohort and label sepsis-3 diagnosis
Model Features identified from Henry's
7 Vitals concepts extraction
8 - Lab Concepts Extraction
11 Clinical Concepts Extraction
Join features based on ICU_ID
Fill in missing values
Setup observation and prediction windows

Cohort


Initial Patient Cohort from ICU Stays
N ICU Stays: 61,535
N Patients: 46,476
Excludes Ages 16 and Under
N ICU Stays : 53,426
N Patients: 38,597
Excludes Carevue Data
N ICU Stays : 23,617
N Patients: 17,707
Excludes ICU Stay(s) after first stay per Patient
N ICU Stays : 16,801
N Patients: 16,801
Excludes No Chart Events
N ICU Stays : 16,058
Excludes Surgical Stays (CSURG,VSURG, and TSURG)
N ICU Stays : 13,111
Excludes Non-overlapping Suspected Infection Window and LOS < 12 Hours
N ICU Stays : 8,772

Sepsis-3 Diagnosis

- We define Sepsis-3 Diagnosis by the “Gold Standard” (DeSautels et al, 2016) which includes two components:
  - Suspected Infection *and*
  - SOFA score increase of greater than or equal 2 ***within a 48 hour pre and 24 hour post window*** of the first suspected infection time
- We excluded ICU Stays whereby the the overlap between the Suspected Infection Window did not include at least 12 hours worth of ICU Stay data for SOFA evaluation

Prediction and Observation Windows

- Prediction Windows included (3,6,12) hours prior to the index date
    - Index date for Case is Sepsis-3 Onset Hour
    - Index date for Control included both last hour and average case Sepsis-3 Onset Hour (.25 LOS)
- Observation Windows included (Unlimited, 7, 12) hours
- Balanced labels case and control via downsampling our control to match number of cases

Model Architecture

```
MyLipSepFC(
  (fc1): Linear(in_features=31, out_features=1028, bias=True)
  (rnn): LSTM(1028, 100, num_layers=4, batch_first=True, dropout=0.2)
  (fcpost): Linear(in_features=100, out_features=32, bias=True)
  (out): Linear(in_features=32, out_features=1, bias=True)
)
```

Criterion	*Optimizer*	*Batch size*	*Cont. Learning rate*	*No of Epochs*
Binary Cross Entropy Loss	ADAM	3	0.0001	100

Experimental Evaluation

Experimental Evaluation

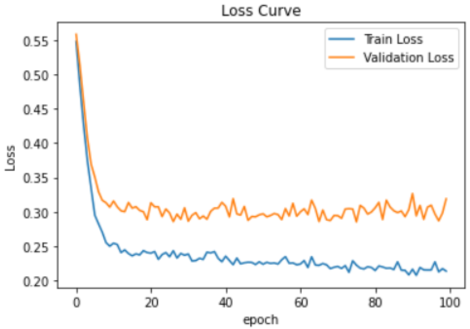
Trial #	Observation Window (Case)	Prediction Window (Case)	Index Date (Case)	Observation Window (Control)	Prediction Window (Control)	Index Date (Control)
1	Unlimited to Index Time	3	Sepsis Diagnosis Hour	Unlimited to Index Time	None	Last Hour of Stay
2	0 < Window Length <= 7 Hrs	3	Sepsis Diagnosis Hour	0 < Window Length <= 7 Hrs	None	Last Hour of Stay
3	0 < Window Length <= 12 Hrs	3	Sepsis Diagnosis Hour	0 < Window Length <= 12 Hrs	None	Last Hour of Stay
4	0 < Window Length <= 12 Hrs	3	Sepsis Diagnosis Hour	0 < Window Length <= 12 Hrs	None	Average Sepsis Diagnosis Hour
5	0 < Window Length <= 12 Hrs	6	Sepsis Diagnosis Hour	0 < Window Length <= 12 Hrs	None	Last Hour of Stay
6	0 < Window Length <= 12 Hrs	12	Sepsis Diagnosis Hour	0 < Window Length <= 12 Hrs	None	Last Hour of Stay

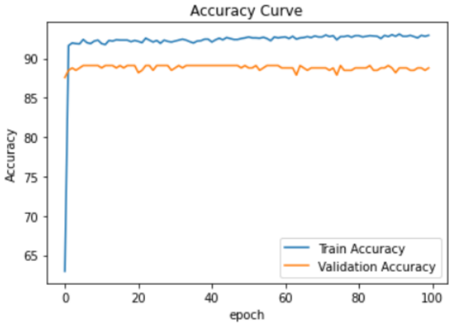
a) Trial Desriptions

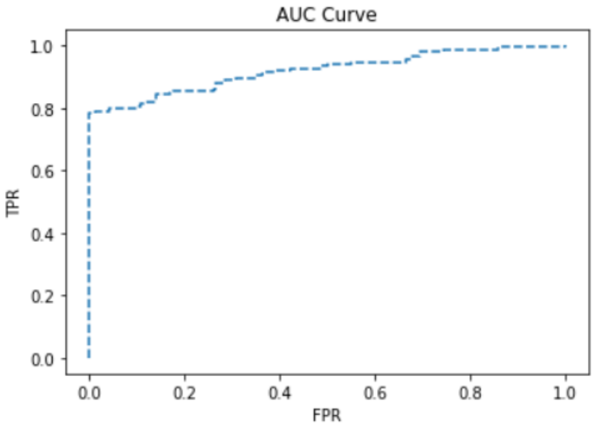
Trial #	Accuracy	AOC	Sensitivity	Specificity	Precision
1	0.71	0.75	0.93	0.51	0.64
2	0.87	0.92	0.75	0.98	0.97
3	0.89	0.92	0.78	0.99	0.99
4	0.75	0.75	0.85	0.62	0.73
5	0.90	0.90	0.73	1.00	1.00
6	0.74	0.83	0.55	0.94	0.89

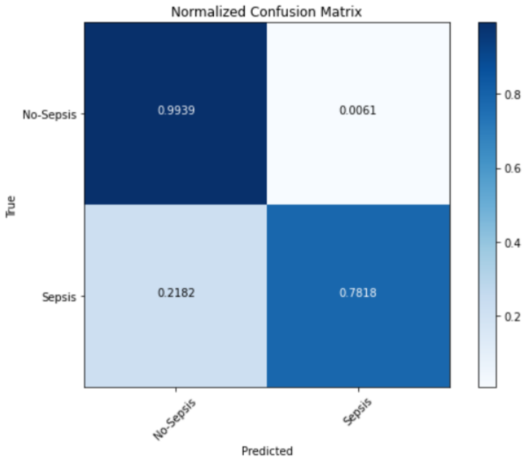
b) Predictive Metrics

Experimental Evaluation









Discussion & Challenges

Key Learnings

- In-Memory Dataset for PyTorch training
- Balanced Dataset
- Small Batch Size
- Utilization of Prediction and Observation Windows
- Results have high Accuracy AUC, Sensitivity and Specificity

Conclusion

Conclusion

- Use of Sepsis-3 diagnosis with our LSTM Deep Learning Architecture, we were able to develop a predictive model to predict Sepsis-3 onset
- Our model showed strong performance results including high sensitivity and specificity
- Further refinement of choice of observation window and prediction window from acceptable clinical standpoint can lead to better application/generalizability to real world use cases.