Implementation of LiSep LSTM Model for Early Detection of Septic Shock

https://youtu.be/kCu_SascO2E

Derek Chapman derek4@illinois.edu
Karan Gupta karang3@illinois.edu
Mark Mulcaire-Jones markm5@illinois.edu
Richa Gupta richag2@illinois.edu

ABSTRACT

Sepsis is a serious condition where the body has an unusually severe response to an infection. It is also known as septicemia. During sepsis the immune system releases chemicals into the blood which triggers widespread inflammation leading to organ damage. In severe cases, sepsis causes a dangerous drop in blood pressure, which in medical terms is called "septic shock". It can quickly lead to organ failure in major organs throughout the body, such as the lungs, kidneys, and liver. Because of this rapid and widespread damage, septic shock has a high rate of mortality [1]. According to CDC, each year at least 1.7 million adults in America develop sepsis. Nearly 270,000 Americans die annually as a result of sepsis. One in three patients who died in a hospital had sepsis [2]. The incidence of sepsis world-wide is difficult to determine, although a recent scientific publication estimated that in 2017 there were 48.9 million cases and 11 million sepsis-related deaths worldwide, which accounted for almost 20% of all global deaths.[3]. It remains a major global health concern.

In this project, our primary objective was to implement and recreate the predictions made by a state of the art LSTM (Long Short-Term Memory)[19] neural network model, designed for early detection of septic shock, called LiSep from the paper by Fagerström et al. titled: *LiSep LSTM: A Machine Learning Algorithm for Early Detection of Septic Shock* [4]. While the original paper developed the LiSep LSTM model using the machine learning framework Keras with a Google TensorFlow back end, we used PyTorch as our machine learning framework instead. Also we used Python Pandas for data analysis and manipulation. The model is trained with data from the Medical Information Mart for Intensive Care (MIMIC-III) database [5] which contains vital signs, laboratory data, and ICU stay entries from approximately 53,000 ICU patients.

In most cases, the LiSep model was able to predict septic shock a full day earlier than all other models. We have reproduced the results with similar results.

1. INTRODUCTION AND BACKGROUND

Our primary goal with this project was to replicate and validate the LiSep model. Model replication is becoming an increasingly common practice in the world of machine learning. The notion of model bias has recently garnered national and global attention due to faults in 'algorithmic decision making'. This has been highlighted in social media, the financial sector, and even medical systems [6]. As models become more complex and the data going into these models becomes ever larger and more complex there is a greater need for replication and validation of the final models. This has been the process in science for almost a century and now needs to be applied to machine learning and advanced methods in the data science field as well.

Septic shock is a well described medical condition with a high morbidity and mortality rate. However it is a condition whose onset is difficult to predict due to the fact that it has numerous underlying causes and is driven by complex interactions of many biochemical pathways, most of which are only partially understood. Predicting septic shock is a problem that is ripe for better solutions as it affects people in every country and at every age. This type of problem is a great one for the application of more advanced machine learning methods such as Neural Networks (NN). In particular, LSTMs are well suited for early detection of septic shock as they can model the sequential nature of disease progression.

Septic shock is just one of many complex syndromes that have been studied for inclusion in algorithmic decision making systems. The current tools available to healthcare providers (such as the early warning score – EWS [7]) are often not sensitive enough. Many of these systems inadvertently created 'alarm fatigue' with health practitioners since they were overly sensitive and gave many false positives. The LiSep model tries to resolve these problems by making more accurate predictions with fewer false positives.

There have been several attempts at using NNs to predict sepsis and septic shock. Predicting sepsis was the main competition on PhysioNet in 2019 [16]. Due to a wide range of hospital wards and definitions of sepsis, comparison across models is limited. However, Area Under Receiver Operator Curve (AUROC) for sepsis prediction usually ranges from 0.68 to 0.99 [15]. This provides a rough target for other sepsis prediction projects. In addition to AUROC, one of the other main factors to consider when looking at the models is the hours before onset (HBO) when the prediction is made. One group used a LSTM-CNN model to successfully predict sepsis 4, 8 and 12 hours out. They achieved an AUROC at 12 hours of 0.84 to 0.86 [14] Ultimately, integrating the model into normal clinical care would be the goal of any project. DeepAISE is a RNN to predict sepsis and has been successfully implemented in a clinical setting. It has a similar AUROC as other models for prediction at 12 hours before onset, however no clear clinical outcomes have been reported [17].

One key barrier to clinical implementation is the black box nature of many NN models. Clinicians will be reluctant to act on data if they do not understand what it is and how it was created. DeepAISE by Yang, Meicheng, et al. [18] dealt with this barrier by letting clinicians know which variables were driving a prediction of future sepsis. If a project can both have clinician acceptance and demonstrate a usefulness in improving mortality, time to discharge or cost of the hospitalization it would be a big step forward for machine learning in healthcare.

2. APPROACH AND IMPLEMENTATION

2.1 MIMIC Dataset and Septic Shock Definition

Medical Information Mart for Intensive Care (MIMIC) III is a database with Intensive Care Unit (ICU) admission data for over 53,000 adult ICU patients gathered between 2001 and 2012 [5]. It has extensive records for each admission including notes, vital signs, labs, and diagnosis. Patients 15 years and older who met septic shock criteria per the Systemic Inflammatory Response Syndrome (SIRS) criteria were identified. The features used in the LiSep model were extracted from this dataset and processed to pass to the neural network.

	Septic Shock	**No Shock**	**Total**
Patients	2208	19735	21943
Female Patients	1042	8642	9684
Age	78.0 years	75.6 years	75.9 years
Mean ICU LOS	10.2 days	2.8 days	3.5 days
Deaths	1376	5984	7360

Table 1: Patient Demographics used for model training and testing

While the most current and widely accepted definition of sepsis is Sepsis-3 [8], the SIRS criteria was widely used for many years and the LiSep paper also used this criteria. Therefore the SIRS criteria was used for this project as well. SIRS is a set of criteria that was developed to identify how sick a potentially infected patient is. A patient can be designated as having SIRS, sepsis, severe sepsis or septic shock. For the purposes of this model we focus only on whether a patient will get septic shock or not. Thus our target label is a binary classification problem and not a multi-classification problem.

Measurement	**Criteria**
Temperature	>38.0 or <36.0 ℃
Heart Rate	> 90 beats per minute
White Blood Cell (WBC) Count	>12,000 cells/mL or <4,000 cells/mL
Respiratory Rate (RR) or arterial pressure of carbon dioxide ($PaCO_2$)	RR > 20 breaths per min $PaCO_2$ < 32 mmHg

Table 2: SIRS Criteria - the patient has two or more of the above

Definitions and Typical Disease Progression:

1. Sepsis - the patient meets SIRS criteria and has a known or suspected infection.
2. Severe Sepsis - the patient meets sepsis criteria and has evidence of organ dysfunction.
3. Septic Shock - a patient who has severe sepsis and persistent hypotension (defined again as systolic blood pressure < 90 mmHg) despite adequate fluid resuscitation [9]. Adequate fluid resuscitation for this project is defined as either at least 20ml/kg of IV fluids in the last 24 hrs or at least a total fluid amount given of 1200 mL [10].

2.2 Data Cleaning

While MIMIC III offers rich opportunities for medical machine learning projects the data cleaning aspect is non-trivial.

There were two main barriers:

First, the size of the dataset meant that significant work needed to be performed to make it usable. MIMIC-III is made up of around 25 files linked via item numbers, patient IDs, and hospital admission IDs among others. Data was extracted from 12 different files for this project. Three of them were “dictionary” files which provided descriptions of various numerical codes. The other 9 files contained various data sources that were combined to create the input data [11]. In addition to the number of files needed, many of them were quite large and needed to be processed in chunks. The largest file used was around 35 GB covering 360 million records, an initial filtering process to remove bad or unneeded data it still needed to be split into ten 350 MB files to be processed separately. Even with the resources available to us via Google Colab we were only able to load about 10% of the original data at once. Unfortunately, the MIMIC data doesn't come presorted, so a patient may have records at the beginning and end of the 360 million records. For this project we needed all data records for a single hospital admission to be together.

Second, the nature of medical data does not lend itself to easy data cleaning. There is frequently bad, discordant, or missing data. These problems have many causes such as practice variation between providers, lab errors, imprecise and inaccurate measurements, and the unavoidable breakdown of complex reporting systems. Additionally, medical terms can have multiple names, numerous abbreviations and different unit types. This dataset reflected all of these problems. For example, arterial pressure of oxygen (PaO2) had four different names and five different item IDs associated with it in the database. Even something as basic as patient weight is reported in both kilograms and pounds.

2.3 Data Preparation

Python Panda dataframes were heavily used for the data cleaning process. All work was done through Google Colab to allow for more resources as well as easier collaboration.

There were two uses of features extracted from the dataset. One use of the extracted features was as direct inputs into the model. These include features such as heart rate, White Blood Count and a diagnosis of metastatic cancer. The second use was feature creation where the extracted features were used to calculate other model features. For example, bilirubin was a feature we obtained from the dataset but was not a direct input into our model. It was used to calculate the hepatic SOFA score which was a direct input into the model. SOFA scores are defined using industry standard values [12]. All model features are shown in Appendix A.

There was some challenge to understand what the LiSep input variables were. The paper stated “we used the same set of medical parameters as Henry *et al.*” and “roughly 30” features. Reviewing the supplemental material of LiSep and TREWScore [13] it seems TREWScore used 26 input variables. The wording of the LiSep paper was ambiguous but our team decided to use the same 26

input variables. In addition to the final 26 variables the model used [Appendix A - Table 1], another 8 variables were found that were needed to calculate either septic shock or another of the input variables [Appendix A - Table 2].

The data fed into the LiSep model needed to be coerced into a specific format. The final output from the data cleaning pipeline [Appendix B] is in the form of one entry per hour per admission. For example, if a patient is in the ICU for two days, they will have 48 records. Records are 'right-aligned' onto hour 0 which is when the patient either has septic shock, leaves the ICU or passes away. Thus for a patient with a stay of two days their first record will be at hour 47. The last record for all patients is hour 0. Some individual patients have multiple hospital admissions. These are treated as separate events and as such have different IDs and a different feature matrix and target label attached. If a patient had multiple ICU stays during one stay in the hospital only the first ICU trip was retained.

As expected due to the aforementioned problems with any massive dataset and the nature of medical data; patients were missing numerous data points. Outside of basic vital signs it is unusual to check labs every hour. For missing data the first preference was to forward-fill previously obtained data until a new value was found. For example, if patient A had a White Blood Count (WBC) count of 9 at hour 20 and a WBC count of 18 at hour 10; hours 19 through 11 would have a WBC count coded as 9. If no value was obtained yet in the ICU we used that feature's population mean. For example, not all patients get Arterial Blood Gases (ABGs) levels taken when in the ICU however our model requires pH and pO2 levels which can only be obtained from an ABG so not all patients will have those values in their chart. The population mean of the pH and pO2 would then be used.

Once all the features were extracted, each patient was identified as meeting septic shock criteria or not, during their ICU stay. After a patient was identified as having septic shock further data was removed. A patient who died or left the ICU without meeting that criteria was identified as not having septic shock. For the purpose of this model there was no consideration whether the final ICU outcome was discharge or the patient passing away.

During the time when the data was recorded, between 2001 and 2012, the system to track certain parameters changed from CareVue to MetaVision. After this switch, antibiotics were consistently charted at the actual time given. Prior to that change the actual time from first antibiotics being given could not be accurately calculated. As it was felt having accurate data was important and our model could still train and test off the remaining 21,943 patients, the patients without accurate antibiotic charting were dropped.

2.4 Model Building

Our project follows the LiSep paper as closely as possible. In cases of confusion related to the construction or implementation of the model, we used similar papers [14][16] and other sources [15] to make informed decisions.

After data cleaning and feature construction, our final data table is small enough that we are able to load the entire table into memory as a Pandas dataframe. We next standardized the numerical columns without changing the categorical and score columns. Although the score columns such as SOFA score are technically numerical columns, they signify a range and we wanted to retain the intuition that the ordering and differences between scores are significant. For example, in the data for 'platelets', the fact that two different readings are 10% above or below the mean is not important. But for a SOFA score, the fact that a patient is a two instead of a one is very significant [12]. Also we standardize instead of normalizing the data, due to the fact that there isn't a set minimum and maximum. Although we set a range of values for each numerical column and cut off what are considered input errors at the extremes, the maximum value isn't guaranteed to be a certain value unlike the range of intensities in a picture.

Next we set up a custom `collate_fn` to pad the data into equal length sequences. The structure of a LSTM network requires that our sequences be in equal lengths even though the individual ICU stays vary in length by a significant amount.

After prepping and loading our data into batches we construct the actual LSTM network. Our model consists of a four layer LSTM network with 100 hidden units and a dropout rate of 0.4 between the layers. In a multi-layer LSTM network, the inputs to subsequent layers are the hidden states of the previous layer. We end with a fully-connected linear layer that reduces our feature length to one and then a sigmoid activation layer to calculate our final probability of septic shock. The network architecture is shown in Fig. 1 below.

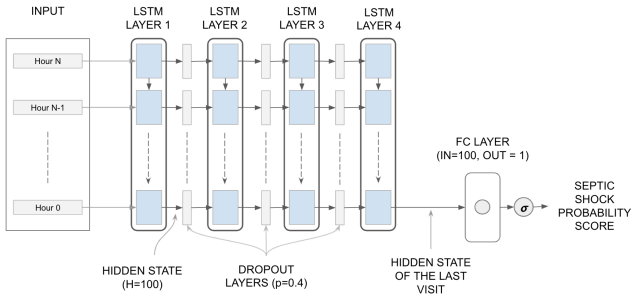

Figure 1: LiSep LSTM network architecture

For comparison, we also trained a less complex version of the LiSep model. This new model, called Lisep LSTM v2, is a two layer LSTM network with 120 hidden units. We wanted to check if we can get the same level of performance by reducing the layers and increasing the hidden units. Lisep LSTM v2 has two-thirds the number of parameters as the original Lisep LSTM model. Detailed comparison of the models is given in Fig. 3.

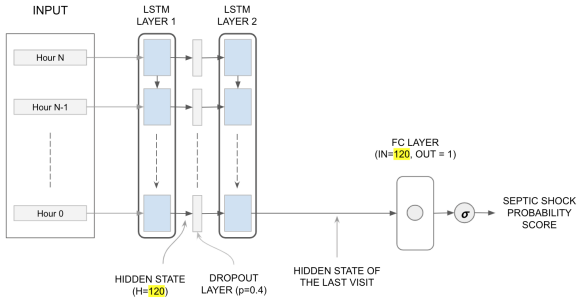

Figure 2: LiSep LSTM v2 network architecture

```
Lisep LSTM Model Summary

==========================================================================================
Layer (type (var_name):depth-idx)        Input Shape     Output Shape    Param #         Mult-Adds
==========================================================================================
LisepLSTM                                --              --              --              --
├─LSTM (lstm): 1-1                       [32, 24, 26]    [32, 24, 100]   293,600         223,027,200
├─Linear (lin): 1-2                      [32, 100]       [32, 1]         101             3,200
├─Sigmoid (sig): 1-3                     [32, 1]         [32, 1]         --              --
==========================================================================================
Total params: 293,701
Trainable params: 293,701
Non-trainable params: 0
Total mult-adds (M): 223.03
==========================================================================================

Lisep LSTM v2 Model Summary

==========================================================================================
Layer (type (var_name):depth-idx)        Input Shape     Output Shape    Param #         Mult-Adds
==========================================================================================
LisepLSTMv2                              --              --              --              --
├─LSTM (lstm): 1-1                       [32, 24, 26]    [32, 24, 120]   187,200         142,295,040
├─Linear (lin): 1-2                      [32, 120]       [32, 1]         121             3,840
├─Sigmoid (sig): 1-3                     [32, 1]         [32, 1]         --              --
==========================================================================================
Total params: 187,321
Trainable params: 187,321
Non-trainable params: 0
Total mult-adds (M): 142.30
==========================================================================================
```

Figure 3: Detailed summary of both the models

For training the model we use the Binary Cross Entropy Loss function since we are ultimately looking for a classification of "Sepsis Shock" or "Not Sepsis Shock". We use Adam [20] as the optimization algorithm with a learning rate of 0.001 and a batch size of 32 patients.

For the training of the model we investigate both a traditional 80/20 split and then a k-fold training regime. The k-fold method seems to produce more consistent results, as we would expect. To address possible class imbalances in our dataset, we use the method `StratifiedKfold` from the `SK-learns` package. This version of K-Fold allows us to create multiple folds without inadvertently creating a large class imbalance in one of the splits. Class imbalance when creating the folds is a possible concern in our final cleaned dataset as we have a 10-1 ratio of negative to positive cases.

3. EXPERIMENTAL EVALUATION

The model was trained and tested using 6 fold cross validation using the 21,943 patients. The training loss and AUROC across 5 epochs are shown in Fig. 4 and 5 respectively. From the graphs we can see that the model is optimised early on and that we do not gain much from training the model beyond 5 epochs.

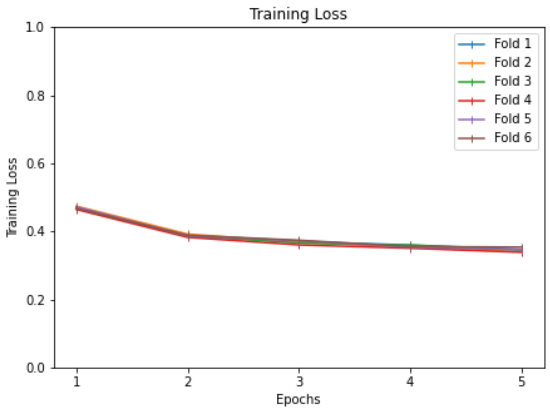

Figure 4: Training loss across 5 epochs

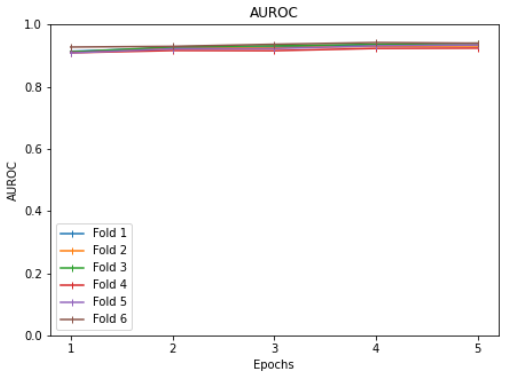

Figure 5: AUROC across 5 epochs

The evaluation was done on the test dataset of each fold. Results from each fold are shown in Table 3. These results are for a prediction window of 24 hours and use a probability threshold of 0.5 for predicting septic shock.

Fold	Precision	Recall	F1	AUROC
1	0.3995	0.5757	0.4717	0.8358
2	0.3841	0.5265	0.4441	0.8414
3	0.4300	0.5615	0.4870	0.8382
4	0.3813	0.5408	0.4473	0.8375
5	0.3890	0.5544	0.4572	0.8273
6	0.4424	0.5949	0.5074	0.8586
Overall	0.4044	0.5590	0.4691	0.8397

Table 3: Results from 6 fold CV at 24 hours before sepsis onset

Figure 6 shows the Receiver Operating Characteristic (ROC) curve. Compared to other models developed to predict septic shock, our model produced a similar Area Under the Receiver Operating Characteristic (AUROC) curve of 0.8397 (0.827, 0.858). This is similar to the LiSep paper of 0.83 (0.82, 0.84).

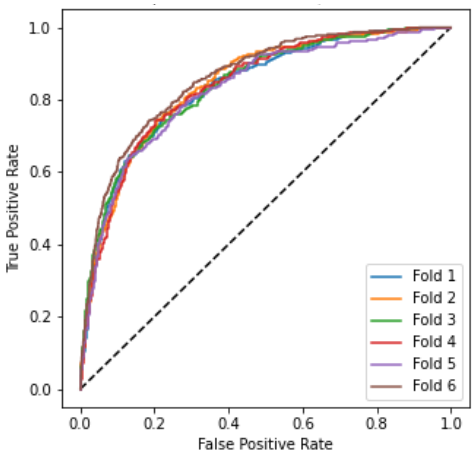

Figure 6: ROC curve of 6 fold cross validation training.

To see how the model performs at different prediction windows, we computed the AUROC for each of the 48 hours preceding the septic shock onset. This is shown in Figure 7. We can see that the performance of the less complex Lisep LSTM v2 model degrades relative to Lisep LSTM as the prediction window increases. This is likely because the v2 model uses less number of LSTM layers and hence cannot predict as accurately ahead of time. As a future work, it will be interesting to see if the accuracy improves when the LSTM layers are increased beyond four.

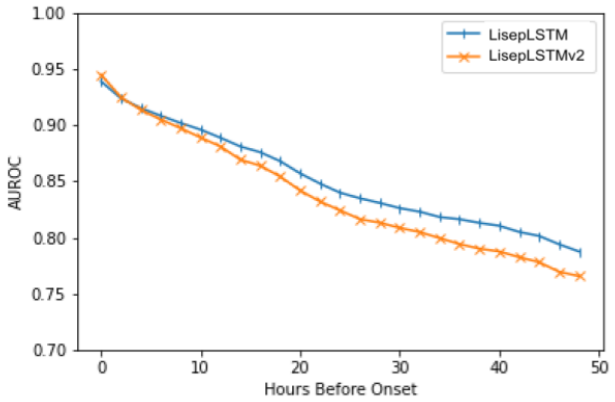

Figure 7: AUROC over the 48 hours directly preceding septic shock onset.

We also computed the HBO (time difference between the onset of septic shock and the model's first positive prediction) for all correctly predicted septic shock-positive patients. This is where the results of our model and the original paper differ. The original model had a HBO median of 40 hours (IQR 20-135) before onset and our model had a HBO median of 64 hours (IQR 30-131) before onset. There are two plausible reasons for this:

First, while predicting septic shock more than 40 hours out is impressive, there are some caveats. The definition of septic shock for LiSEP is based on the SIRs criteria which captures some patients who aren't actually that sick. This allows any model, not just ours, to achieve a higher true positive rate when predicting septic shock. The patient population being considered is only ICU patients which means patients that are at higher risk for developing complications have already been pre-selected for. HBO is a useful measurement for this model but any model that frequently predicts septic shock in ICU patients will look impressive by a metric that doesn't account for false positives.

Second, as discussed in the Data Cleaning section, the original paper used all of the MIMIC-III data while we dropped half of the patients due to poorly recorded data in the CareVue system. This means the LiSep authors included patients that probably had inaccurate 'time since first antibiotics' values.

4. CONCLUSION

The primary goal of this project was to replicate and validate the LiSep model. This goal has been achieved. The model that was produced during this project shows a very similar AUROC for its overall ability to predict septic shock. While this method of comparison has some limitations it is the one readily available. Without additional metrics from the original LiSep model further comparison is not possible. AUROC is often used to compare machine learning models. With the similar AUROC we can state that the LiSep model is reproducible and a valid way to predict if a patient from the MIMIC-III dataset will develop septic shock while in the ICU.

LiSep shows some promise in predicting septic shock but we do not feel it is ready for clinical implementation. We would suggest the following modifications and potential improvements:

First, retraining the model to predict septic shock onset as defined by the more current sepsis-3 criteria.

Second, the model only has 26 variables. While indiscriminately including every data point from the EMR would not be a wise approach, other promising variables could be included. These range from basic labs such as lactate, which is commonly elevated in sepsis patients, to antibiotic exposure prior to hospitalization.

Third, consider including patient populations in other areas of the hospital such as the medical and surgical wards. Identifying patients at high risk of septic shock in these wards would allow better utilization of limited ICU resources.

This project provided both challenges and learning opportunities. As discussed in the data cleaning section, the MIMIC III dataset was a challenge to clean and get into a format for the NN to process. As a result all of us are now much more comfortable handling large and complex datasets and are more proficient using pandas. Our team was made up of 4 people all in different time zones which made coordinating meetings difficult. However, it also provided a good opportunity to practice breaking up a complex project into smaller sections that individuals could work on separately. It was difficult to reproduce the specific architecture of a NN with the description provided in the LiSEP paper. However, this provided an opportunity to think critically on how to structure our own NN. Finally, the long data cleaning and training times made error checking difficult. Without the immediate feedback of success or failure that other programs often give, good time management was essential. Our group fortunately did not procrastinate and we had sufficient time to troubleshoot.

Machine Learning models show great promise in the medical field. Their ability to make meaningful predictions from large amounts of complex data could lend themselves to integration with the electronic medical record (EMR) system to improve patient care. However, there are numerous steps that are needed before this integration could happen, model validation is a major one that often goes under-recognized. If a model is to be used to make meaningful clinical decisions, especially if it is a "black box model", then the model needs to be independently validated. The best way to further validate the model would be to apply the trained model on test patients from another healthcare system to determine if the model remains accurate and robust. However, due to logistical and legal issues this is not a trivial step. One possible intermediate step is to do what this project has done and replicate the model on the same data from scratch. If it again shows significant clinical promise then consideration can be made to devoting the resources to looking at it with an entirely different patient dataset.

5. CONTRIBUTIONS

Derek Chapman (derek4@illinois.edu): Model training and evaluation, data cleaning, report and presentation.

Karan Gupta (karang3@illinois.edu): Model training and evaluation, report and presentation.

Mark Mulcaire-Jones (markm5@illinois.edu): Understanding the data, data cleaning and report and presentation.

Richa Gupta (richag2@illinois.edu): Understanding the data, data cleaning, report and presentation.

6. REFERENCES

[1] Sepsis as defined by WEBMD

[2] CDC mortality statistics

[3] WHO mortality statistics

[4] Fagerström, Josef, et al. "LiSep LSTM: a machine learning algorithm for early detection of septic shock." Scientific reports 9.1 (2019): 1-8.

[5] Johnson, A., T. Pollard, and R. Mark III. "MIMIC-III Clinical Database (version 1.4). PhysioNet." (2016). https://physionet.org/content/mimiciii/1.4/.

[6] Obermeyer, Ziad, et al. "Dissecting racial bias in an algorithm used to manage the health of populations." Science 366.6464 (2019): 447-453.

[7] Subbe, Christian P., et al. "Validation of a modified Early Warning Score in medical admissions." Qjm 94.10 (2001): 521-526.

[8] Singer, Mervyn, et al. "The third international consensus definitions for sepsis and septic shock (Sepsis-3)." Jama 315.8 (2016): 801-810.

[9] SIRS Definition: Rangel-Frausto, M. Sigfrido, et al. "The natural history of the systemic inflammatory response syndrome (SIRS): a prospective study." Jama 273.2 (1995): 117-123.

[10] Liu, Ran, et al. "Data-driven discovery of a novel sepsis pre-shock state predicts impending septic shock in the ICU." Scientific reports 9.1 (2019): 1-9.

[11] MIMIC-III description

[12] SOFA scores from MDCALC

[13] Henry, Katharine E., et al. "A targeted real-time early warning score (TREWScore) for septic shock." Science translational medicine 7.299 (2015): 299ra122-299ra122.

[14] Rafiei, Alireza, et al. "SSP: Early prediction of sepsis using fully connected LSTM-CNN model." Computers in biology and medicine 128 (2021): 104110.

[15] Fleuren, Lucas M., et al. "Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy." Intensive care medicine 46.3 (2020): 383-400.

[16] Reyna, Matthew A., et al. "Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019." 2019 Computing in Cardiology (CinC). IEEE, 2019.

[17] Shashikumar, Supreeth P., et al. "DeepAISE--An end-to-end development and deployment of a recurrent neural survival model for early prediction of sepsis." arXiv preprint arXiv:1908.04759 (2019).

[18] Yang, Meicheng, et al. "An Explainable Artificial Intelligence Predictor for Early Detection of Sepsis." Critical Care Medicine 48.11 (2020): e1091-e1096.

[19] Gers, Felix A., Jürgen Schmidhuber, and Fred Cummins. "Learning to forget: Continual prediction with LSTM." (1999): 850-855.

[20] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

APPENDIX A

Table 1: Model Input Variables

Variable Name	Range	Population mean
Antibiotics started hours ago	0,1	-
Blood Urea Nitrogen (BUN)	1-300	32
BUN/CREAT ratio	2-200	-
Fraction of Inspired Oxygen (FiO2) [1]	21-100	49
Glasgow Coma Scale (GCS)	3-15	
Heart Rate	1-320	87
ICD-9 code(s) for Chronic Liver Disease[2]	0,1	-
ICD-9 code(s) for Chronic Organ Dysfunction[2]	0,1	-
ICD-9 code(s) for Congestive Heart Failure[2]	0,1	-
ICD-9 code(s) for Diabetes[2]	0,1	-
ICD-9 code(s) for Hematological Cancer[2]	0,1	-
ICD-9 code(s) for Immunosuppression[2]	0,1	-
ICD-9 code(s) for Metastatic Cancer [2]	0,1	-
Meets SIRS criteria	0,1	-
Partial Pressure of Oxygen (PaO2)	1-700	132
Patient Originally admitted to Cardiac ICU	0,1	-
pH of arterial blood	1-10	7.39
Platelets	1-1500	217
Respiration Rate	1-150	20
Shock Index	0-11	-
SOFA Score - hepatic portion	0-4	-
SOFA Score - neuro portion	0-4	-
SOFA Score - renal portion	0-4	-
Systolic Blood Pressure (SBP)	1-300	121
Urine output in the last 6 hours (mL)	0-10000	-
White Blood Cell Count (WBC)	1-1000	12

(1) Room air's FiO2 is 21% so the lower limit was changed to reflect that.

(2) ICD9 codes taken from TREWSCORE paper.

Table 2: Variables not used to train the model

Name	Range	Use in Project
Age	15-110	Remove patients under 15 years old
Bilirubin	1-900	Calculate Hepatic Sofa
Creatinine	10-300	Calculate Renal SOFA and BUN to Creatinine ratio
Date and Time - Admit	date and time	Calculate various time related values
Date and Time - Discharge	date and time	Calculate various time related values
Fluid input in last 24 hours	25-110	Decide if patient has been resuscitated
Fluid Input total	0-50000	Decide if patient has been resuscitated
ICD-9 code(s) for Infection	0,1	Calculate if sepsis presence
Arterial Partial Pressure of carbon dioxide	1-200	Calculate SIRS
Temperature	0,1	Calculate SIRS
Weight	10-500	Decide if patient has been resuscitated

APPENDIX B - Data Cleaning Workflow