Proceedings of Machine Learning Research 1–12

Methods for Clinical Time Series Analysis in Pediatrics

Kyle Liu KYLELIU@MIT.EDU
Massachusetts Institute of Technology

William Luo WQCLUO@MIT.EDU
Massachusetts Institute of Technology

Neha Hulkund NHULKUND@MIT.EDU
Massachusetts Institute of Technology

Abstract

The continual growth of medical data stored in electronic health records has motivated the development of techniques to extract information from clinical time-series data. Due to their high-dimensionality and irregularity in sampling, time-series data has been difficult to extract information from. Techniques from Natural Language Processing have been gradually adopted to the space due to their strengths in handling varying-length sequences, and models such as LSTM's have have been shown to outperform traditional methods involving hand-engineered features in a variety of clinical tasks(7). However, in order to handle irregularity or missing data, such techniques have often involved bucketing or imputing missing data points. To combat this issue, attention-based models have been applied to the space, as they've been shown to better encapsulate dependencies in long sequences, and have been used to achieve state-of-the-art performance in all MIMIC-III benchmark tasks(6). Despite this progress in the general medical field, there has been limited work to show that the same gains in performance hold in pediatrics. We propose applying a transformer-based model to time-series data from the PIC (Pediatric Intensive Care) dataset(5). We wish to evaluate it on predicting in-hospital mortality, ICU length of stay, Sepsis risk, and lab test results.

1. Introduction

PIC is a relatively new pediatric-specific dataset containing deidentified patient information from those admitted to the Children's hospital of Zhejiang University School of Medicine. Like MIMIC, PIC contains significant time-stamped data and measurements on ICU patients, which can be used for inference. Although structured similarly, PIC differs drastically from MIMIC in that its patient population resides almost entirely in the age-range that MIMIC lacks (0-18). This is significant in that pediatric care is relatively understudied in comparison to the general medical field. As a result, we develop reasonable baselines for standard clinical machine learning tasks for pediatrics, and specifically apply recent approaches from NLP to aid in early diagnosis of Sepsis.

Sepsis is a severe condition affecting both children and adults, characterized by symptoms such as fever, low blood pressure, or difficulty breathing. In particular, neonates are among the groups that are most at-risk (14), with pediatric sepsis causing 6500 deaths annually in the United States alone (9). Studies(15) have shown that early and accurate diagnosis paired with aggressive treatment can allay this risk. However, compared to sepsis

in adults, pediatric sepsis suffers from inconsistency between consensus definition (12) and clinical practice (16).

This may be due to the very dynamic clinical definition of Sepsis in children – due to the natural physiological changes of children, many criteria for Sepsis in adults must be delineated into age sub-cohorts in order to be clinically relevant in pediatrics (12). The difficulty of diagnosing Sepsis through rule-based methods motivates a machine-learning approach. Keeping the large physiological variations between children of different age groups in mind, we also develop models trained in aggregate and for age-specific subcohorts in order to capture and analyze the differences in risk factors between age groups.

Lab tests in the ICU can often be time and cost intensive, delaying treatment for the patient. Furthermore, diagnostic blood tests are often overused and redundant, estimated to waste over $5 billion in the United States (21), meaning that a rule based algorithm predicting lab test results can be helpful for clinician decision making in crucial hours before the diagnostic lab tests return. In particular, diagnostic blood tests can act as early indicators for infection in the patient, meaning that a machine learning approach to prediction of these values can aid clinicians in prescribing treatment for severe and time-sensitive cases (20).

In this paper, we develop several models and architectures for both representing and inferring from the time series data available in PIC for the tasks of Lab Test prediction and Sepsis risk prediction. We also study the MIMIC-III benchmark tasks of in-hospital mortality prediction and ICU length of stay prediction as gauges for the models we develop. Specifically, we train Random Forest and LSTM models for all tasks, reading the time series data as a sequence of hourly measurement averages, with forward imputation. We further experiment with a transformer-based model(1) in an attempt to represent the series of measurements for each patient in full resolution for the tasks of in-hospital mortality and ICU length of stay. We find that with effective imputation, dynamic NLP architectures such as LSTM models outperform static machine learning methods like Random Forest. However, our proposed Transformer-based architecture achieved sub-par results on the tasks of in-hospital mortality and ICU length of stay. We suggest possible improvements and modifications to be studied in the future.

2. Background

2.1. Sepsis Prediction

There have been several studies on Sepsis prediction with machine learning methods. Le et. al (2019)(9) uses boosted decision trees to predict severe pediatric sepsis from time series data on vital signs. LiSep(8) is an LSTM-based model for sepsis prediction in adults. Both utilize rule-based approaches for finding silver labels on sepsis diagnoses. Similar to them, we will use the consensus definition for Sepsis by Goldman et. al (12).

However, both of the groups above impute or bucket their data heavily (i.e. per hour bucketing). We propose a method to expand on their work by training a self-attention based model that doesn't require imputation or bucketing for temporal continuity. In addition, contrary to previous works, we cast the Sepsis prediction task as a rolling prediction task, in which we ingest a window of clinical measurements, and predict if a patient will develop Sepsis-like conditions in a predetermined prediction window. This is motivated by the fact

that Sepsis is a very dynamic condition, and is often not caused directly by the conditions which caused the patient to originally enter the hospital (3). Thus, it may be more clinically relevant to develop a model that is built to ingest data from a dynamic window, rather than a fixed window oriented at a patient's admit time.

2.2. Lab Test Prediction

The presence of white blood cell count, hemoglobin, and procalcitonin in the bloodstreams are early indicators of infection in the patient (20)(17). These has been significant previous work for diagnostic criteria for blood diagnostic lab criteria. De Baets, et. al (17) used a bidirectional LSTM to predict positive blood culture on features of vital signs time-series data. (19) used standard machine learning regression algorithms on demographic patient criteria for the prediction of ferratin in patient data from clinical laboratories, as a proof of concept of other diagnostic lab test predictors. However, imputation is central to both of these studies. Again, we additionally frame these lab test prediction tasks as dynamic-window prediction tasks, in order to better capture the clinical relevance of lab test prediction, compared to fixed-window prediction. We aim to predict if the lab tests fall in a safe range of clinically significant values for these diagnostic lab tests. We believe that these predictions of diagnostic lab tests based on vital signs will be able to aid clinicians in making decisions about treatment regarding time-sensitive cases for younger patients.

2.3. RNN's for Clinical Time Series

Lipton et. al(7) develop an LSTM-based model for phenotyping on clinical time series. They notably train on raw time series, via incorporating mask and time-delta terms, but only incorporate 13 clinical features in their model. Che et. al (11) develop the GRU-D model which is designed specifically to deal with time series with missing values, while also incorporating temporal structure, through introducing a decay term in the time series.

2.4. Transformers for Electronic Health Records

Song et. al(6) develop a transformer-based self attention model for general cinical tasks, evaluating it on the MIMIC-III benchmarks of mortality, ICU stay, phenotyping, and decompensation. They use positional encodings extracted from a single convolution layer to incorporate temporal order. Notably, they discretize the time series, and hence do not access the data in full resolution. We experiment with applying sinusoidal embeddings to incorporate temporal information in our Transformer-based model, as in the original Transformer paper(1).

3. Methods

For all of the above prediction tasks, we develop benchmarks with the standard models of random forest and logistic regression, as well as LSTM's. We then plan to experiment with self-attention based models in order to represent the data in full resolution.

The purpose of using attention-based transformers is due to the missingness of data that we often encounter in clinical settings. LSTMs traditionally depend on regularity of data, so some sort of imputation method is necessary in order to apply them. For our models, we

utilize forward imputation, in which a missing value is assumed to be the same as its most recent measurement, but a future extension could be to experiment with other methods, such as simple imputation. (11).

This albeit simple method has some clinical relevance. As noted by previous authors (7), since many measurements are taken by hand by nurses, a measurement's absence over a time interval typically indicates that clinicians expect the measurement to be relatively unchanged over that interval. Due to this, however, our models' ability to infer is in a sense limited by existing clinical practice, as we take these constant-value assumptions as fact.

Due to the irregularity and missingness of data, some work needs to be done in order for conventional NLP techniques, which depend on regularly sampled data, to be effectively applied.

For our baselines, we follow from work done previously on the MIMIC-III dataset (10), and induce regular sampling intervals by bucketing measurements by hour. Specifically, we let T be the number of hours that we collect data, and X_t^d is the average measurement of feature d over the t^{th} hour that a patient is in the hospital.

Crucially, for all the benchmarks and models we develop, we collect data in a fixed window, and employ a large gap size to prevent label leakage (10). In particular, we evaluate the models on data collection windows and 24 hours, with gap sizes of 6 hours. Any patients whose label was revealed prior to the end of the gap time is removed from the data cohort.

We normalized such that each feature had zero mean and unit variance over the training set, and applied the transformation to the held out test set.

3.1. Random Forest Models

Here, since these baseline models are not meant to handle time-series data, we will simply flatten the time series: $X_f = [x_1, x_2, ..., x_T]$. We will in addition append some static features to the vector – the patient's age at admission and gender (one-hot encoded). We used SciKit Learn's `RandomForestClassifier` class with hyperparameters selected through random search.

3.2. LSTM

Here, we can use whatever time series data we imputed above directly. Here, $T = 24$ is our window size, and the inputs x_i are the imputed hourly averages for each feature in hour i of an individual's visit (or in the case of the rolling tasks, hour i of the ingested window). In order to account for our relatively limited data, we apply certain regularization methods, utilizing dropout with $p = 0.3$, and l_2 regularization.

3.3. Self-Attention

We propose a self-attention based model inspired by the Transformer architecture for machine translation (1). We modify the above sequence representations in order to take advantage of the clinical time series data in full resolution. Our architecture takes similar structure to that proposed in Song et. al(6), with specific design changes.

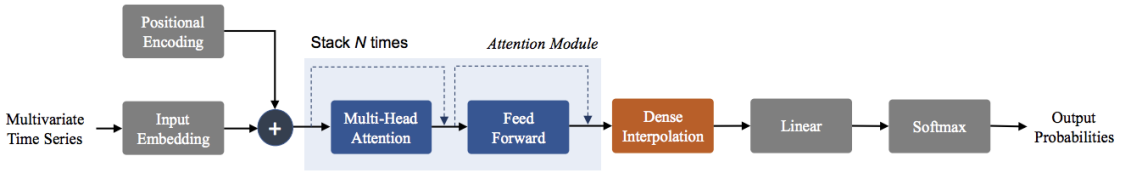

Figure 1: Non-recurrent Transformer Architecture

Concretely, each patient i is given measurements at time points $(t_1, ..., t_T)$, and these measurements are captured by feature vectors $(x_1, ..., x_T)$. Each feature vector x_i resides in $\mathbf{R}^d$, where d is the number of distinct features.

Each sequence of feature vectors is embedded and passed into N Transformer Encoder layers. Our embedding is twofold:

- **Feature Embedding** We embed each feature vector x_i through taking the output of a single feed forward neural network. Intuitively, our feature vectors x_i are very sparse, as measurements are often not taken at the same time. Using this feed forward layer gives us a dense embedding more suitable for learning.

- **Positional Embedding** Since our architecture uses no recurrence, a positional embedding is required to incorporate temporal information. We modify the sinusoidal positional embeddings of the original Transformer paper to support continous time steps. More specifically, the positional embedding at position pos is:

$$PE_{(pos,2i)} = sin(t_{pos}/t_{max}^{2i/d})$$

$$PE_{(pos,2i+1)} = cos(t_{pos}/t_{max}^{2i/d})$$

  The original authors motivate this embedding by the argument that this fixed positional embedding allows the model to easily extrapolate to unseen sequence lengths. This issue of extrapolation is compounded in the case of medical time series data, as our models need to generalize to arbitrary unseen time-step distributions. We conjecture that this fixed sinusoidal embedding allays that need.

The final input embedding of each x_i is the sum of its feature and positional embeddings.

The output of the positional embedding is a sequence of hidden states $(H_1, ..., H_T)$, from which we need to acquire a single logit. We need to reduce dimensions, while also incorporating temporal relationships (as this sequence of hidden states corresponds to a sequence of measurements). Following from the work of Song et. al (6), we apply a Dense Interpolated Embedding(23), which has been used in the realm of NLP to encapsulate syntactic features, and project the result onto a single logit. Specifically, we fix a parameter M, and compute a matrix $u \in \mathbb{R}^{M \times d}$, where u is defined as:

$$u_m = \sum_{i=1}^{T} \left(1 - \left|\frac{i}{T} - \frac{m}{M}\right|\right)^2 H_i$$

For regularization, we applied dropout with $p = 0.5$ and weight decay. The code for this architecture is based heavily on The Annotated Transformer (22).

3.4. Sepsis Prediction

Due to the lack of centralized gold labels for sepsis diagnosis, we utilize standard definition – a SIRS score of 2 or more and suspicion of infection (12). Here, we utilize a SIRS score modified for pediatrics, with varying thresholds for various age groups(12).

Moreover, we defined this as a dynamic prediction task. Here, we ingested a random 24-hour window of data, and predicted if the patient will develop sepsis-like conditions in the 6-hour prediction window following a 6-hour gap time to prevent label leakage. Windows of patients who developed sepsis-like conditions before the beginning of the prediction time were thrown out.

3.5. Lab Test Prediction

We utilized standard clinical definitions of infection for a binary classification scheme of prediction if the lab values are predicted to be within a normal range. We predicted if the diagnostic lab values fell below the clinically safe threshold of 4000 μ g/L for white blood cell count, below 14 g/dL for hemoglobin (17), and higher than 0.5 ng/mL for procalcitonin (20).

Similarly to sepsis, we defined this a dynamic prediction task, using a random 24-hour window of data, a 6-hour gap time, and a 6-hour prediction window. To reduce confounding factors, we removed lab tests that are closely linked to the lab test prediction in question, such as leukocytes, basophils, and neutrophils for white blood cell count.

3.6. Mortality Prediction & Length of Stay (3, 7 Days)

For all three of these tasks, we framed the question as a binary classification problem. We used a static window, using a data from the first 24 hours and leaving a gap size of 6 hours after that, so eliminating all patients that were discharged within the first 30 hours.

These tasks all used a static window, so unlike the previous two prediction tasks, we did not use a rolling window.

3.7. Feature Importance Analysis

To analyze the importance of features for our LSTMs, we utilized *feature perturbation*. Once we have a trained model, we analyze the importance of each feature by perturbing each feature slightly, and then analyzing how our LSTM's output changed as a result of the perturbation.

More specifically, we perturbed each feature by a random normal distribution with mean 0 and standard deviation 0.2 (after normalizing the features), and then computed the LSTM's output with the perturbed features.

Then to calculate the feature effect, we computed the Euclidean Distance between the two output vectors. See figures 4 through 7 below for our results on our feature importance analysis.

Table Name	Feature Name
Chart Events (14 Variables)	1 Temperature 2. Pulse 3. Heart Rate 4. Oxygen Saturation 5. Excrement 6. Urine Output 7. Input 8. Output 9. Blood Glucose 10. Pain Score 11. Height 12. Weight 13. Diastolic Pressure 14. Systolic Pressure
Lab Events (20 Variables)	15. Hematocrit 16. WBC Count 17. Platelet Count 18. Hemogloba 19. MCHC 20. MCH 21. MCV 21. Red Blood Cells 22. RDW 23. Potassium 24. Sodium, Whole Blood 25. Chloride, Whole Blood 26. Calculated Bicarbonate, Whole 27. Anion Gap 28. Urea 29. Creatinine 30. Glucose 31. Calcium, Total 32. INR(PT) 33. Lymphocytes, Percent
Surgery Vital Signs (7 Variables)	34. Heart Rate 35. O2 Saturation 36. Pulse 37. Systolic Pressure 38. Diastolic Pressure 39. Respiratory Rate 40. Central Venous Pressure
Demographics (2 Variables)	41. Gender 42. Age

Figure 2: Feature Set (All Models)

4. Data and Experiment Setup

Due to the presence of a few patients with multiple admissions, we decided to remove all but the first admission for each patient. Furthermore, due to the fact that less than 2% of patients were between the ages of 12 and 18, we removed patients in that age group prior to performing any analysis.

Along those lines, due to the large physiological differences between children of different ages, we decided to train our models individually on the age cohorts of 0-2 months, 2-24 months, 2-5 years, and 5-12 years. For the sake of comparison, we also trained models in aggregate as well.

For the rolling tasks, we limited the data-collection windows to the first 100 hours, as measurements past 100 hours were too sparse for significant inference. Similarly, for the LOS task, we removed those with LOS greater than the 99^{th} percentile (66 days).

Figure 2 shows the features that for each patient were extracted from the corresponding data tables in PIC. These feature measurements (with the exception of gender) were centered and scaled.

5. Results

Model(AUC)	total	0-2 months	2-24 months	2-5 years	5-12 years
Mortality					
RF	0.82	0.77	0.82	0.79	0.75
LSTM	0.83	0.75	0.81	0.82	0.83
Transformer	0.76	-	-	-	-
ICU LOS (3 day)					
RF	0.77	0.79	0.74	0.72	0.73
LSTM	0.76	0.82	0.68	0.79	0.75
Transformer	0.72	-	-	-	-
ICU LOS (7 day)					
RF	0.79	0.82	0.78	0.76	0.74
LSTM	0.76	0.78	0.71	0.79	0.65
Transformer	0.74	-	-	-	-
Sepsis					
RF	0.76	0.72	0.75	0.77	0.79
LSTM	0.78	0.80	0.79	0.80	0.81
WBC Count					
RF	0.82	0.76	0.72	0.86	0.79
LSTM	0.81	0.82	0.77	0.75	0.80
Hemoglobin					
RF	0.80	0.79	0.82	0.80	0.78
LSTM	0.82	0.80	0.81	0.83	0.81
Procalcitonin					
RF	0.77	0.76	0.77	0.75	0.78
LSTM	0.80	0.81	0.79	0.79	0.81

Figure 3: Validation AUC

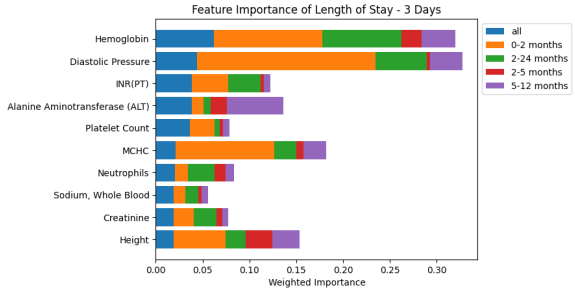

Figure 4: Feature ImportanceLOS (3 Days)

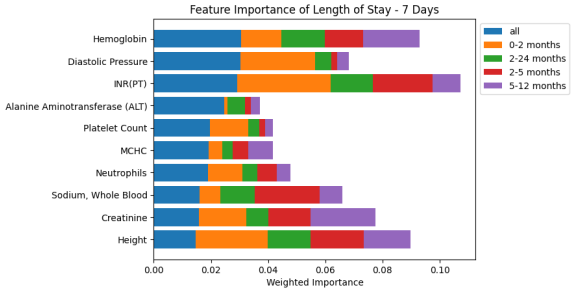

Figure 5: Feature ImportanceLOS(7 Days)

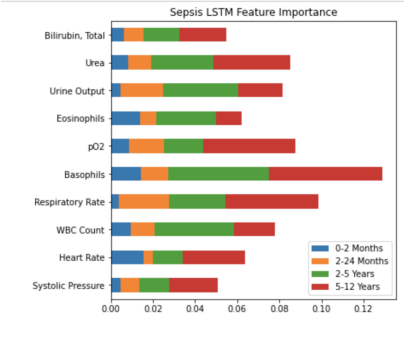

Figure 6: Feature Importance Sepsis Prediction

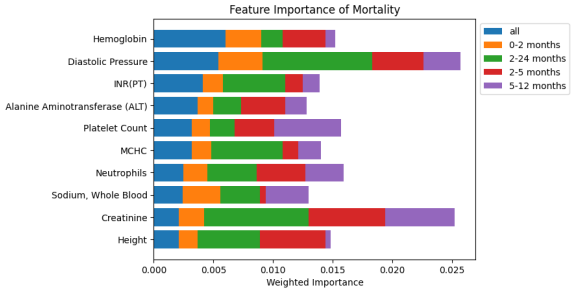

Figure 7: Feature ImportanceMortality

6. Discussion

Particularly of note, Figure 6 shows that the most important feature for Sepsis prediction in our LSTMs was the quantity of Basophils. Recent studies have shown that despite their relative small numbers, Basophils may play a key role in the prevention of Sepsis(24), so a large variation in the quantity of Basophils may indeed be a strong indicator for Sepsis risk. Furthermore, the feature importance magnitudes in the neonate cohort are relatively small, indicating that the causes of Sepsis in neonates indeed vary more for that age cohort.

Within the lab test prediction, it is evident that the relative performances of the random forest model and LSTM differed per age cohort. As a whole, neonates had lower performance than other measures, possibly due to greater variability in their vital signs than other age cohorts. White Blood Cell count had the best performance, possibly due to being closely linked to other markers of infection in vital signs. Future lab tests predictions we are looking into performing are C-reactive protein and leukocyte counts.

For mortality prediction, we have the highest performance from the baselines than any other task. However, the cohort performances are quite variable, often being lower for the

0-2 months age cohort. This makes sense, as 0-2 month old children are the most vulnerable, and often have the most varying of causes of severe conditions (such as Sepsis)(3).

Looking at length-of-stay prediction tasks, we see that the 2-24 month age cohort was fairly challenging for LSTMs on both tasks. Figure 6 shows that the most important features for length-of-stay predictions are hemoglobin and diastolic pressure. The former is an indicator of anemia, which is present in 97% of ICU patients, while the latter is an indicator of sepsis infection, heart attacks, and mechanical ventilation, all which would affect the length of stay of a patient (3). Additionally, as a whole, the 7 day length-of-stay prediction had a greater performance than the 3 day length-of-stay prediction, possibly because patients with a length of stay less than 3 days could also have readmission, which we did not account for in our analysis, and more volatile vitals. Furthermore by 7 days, patients are more likely to have had treatment and their vitals are more stabilized to indicate if they need long-term care.

Overall, as seen from the results section, the LSTMs and random forests (RFs) consistently out-perform the Transformer model on all of our tasks. RFs were competitive with LSTMs on most tasks, even outperforming the LSTMs on some cohorts. However, overall we end up with the baseline models outperforming the Transformer model on all tasks, the opposite of our initial hypothesis. This may be due to a lack of data in the study necessary for the Transformer model. In the future, we are looking to perform further hyperparameter tuning on the Transformer and see if results improve.

This may be due to a variety of reasons, in particular a lack of comprehensive data. The PIC database only contains data for about 10,000 patients, which is fairly limited for a Transformer model that needs to learn such complex associations with healthcare data. These are also preliminary results: further hyperparameter tuning and training is necessary for the Transformer. We may also experiment with other types of embeddings, such as relative positional embeddings.

Early warning systems are vital to a functioning and effective ICU and can be particularly important in a PICU, where it is important that we note that neonates are particularly volatile, and the health and status of a given patient can change drastically in a very short time span. Therefore, it becomes important that we have early warning systems in place to given clinicians and doctors an early heads-up about potential complications.

Some limitations of our study include the fact that PIC data is collected at a single hospital in China, so the relationships between risk factors and different cohorts may vary in other situations. Furthermore, our strongest models were developed by bucketing and imputing clinical time series, which results in some degree of information loss. It would be interesting to pursue the full-resolution self-attention model further, in order to capture more complex relationships in the data.

7. Acknowledgements

We would like to thank Dr. Chase Parsons for his clinical insights in the project. We would also like to thank Matthew McDermott and Professor Peter Szolovits for their valuable feedback and help in framing the project.

8. Member Contributions

1. **Neha Hulkund:** Coded baseline lab prediction task. Research thresholds and performed experiments on lab test prediction task. Performed feature analysis.

2. **Kyle Liu:** Coded framework for LSTM and RF models used on all tasks, set up relevant data for future experiments. Set up data for rolling window prediction tasks. Researched Sepsis prediction task, as well as Transformer architectures that would be useful for our task. Implemented and experimented with Transformer-based model. Performed experiments on Sepsis Prediction.

3. **William Luo:** Performed data exploration on preliminary data, researched and performed experiments on mortality and length of stay prediction tasks. Coded and performed feature analysis.

Code avaiable at https://github.com/kyleliu11111/MIT-MLHC-2020-FinalProject

9. References

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., Polosukhin, I. (2017). Attention Is All You Need.

Li, Y., Rao, S., Solares, J., Hassaine, A. (2018). BEHRT: Transformer for Electronic Health Records.

Singh, D., Mclean, C., Erdman, L., Radebe, L. (2017). Emergency Department Sepsis Alert System for Children.

Hager, H., Pronovost, D., Saria, S. (2015). A targeted real-time early warning score (trewscore) for septic shock.

Xian Zeng, Gang Yu, Yang Lu, Linhua Tan, Xiujing Wu, Shanshan Shi, Huilong Duan, Qiang Shu Haomin Li (2020) PIC, a paediatric-specific intensive care database.

Song, H., Jayarman, J. T., Spanias, A. (2017). Attend and Diagnose: Clinical Time Series Analysis using Attention Models.

Lipton, Z., Kale, D., Elkan, C., Wetzel, R. (2015). Learning to Diagnose with LSTM Recurrent Neural Networks.

Fagerström, J., Bang, M., Wilhemlms, D., Chew, M. M(2019). LiSep LSTM: A Machine Learning Algorithm for Early Detection of Septic Shock

Le, S., Hoffman, J., Barton, C., Fitzgerald, J., Allen, A., Pellegrini, E., Calvert, J., Das, R. (2019). Pediatric Severe Sepsis Prediction Using Machine Learning. Retrieved from

Wang, S., McDermott, M., Chauhan, G., Hughes, M., Naumann, T., Ghassemi, M. (2019). MIMIC-Extract: A Data Extraction, Preprocessing, and Representation Pipeline for MIMIC-III.

Che, Z., Purushotham, S., Cho, K., Sontag, D., Liu, Y. (2018). Recurrent Neural Networks for Multivariate Time Series with Missing Values.

Goldstein, B., Giroir, B., Randolph, A. International pediatric sepsis consensus conference: Definitions for sepsis and organ dysfunction in pediatrics.

Blagus, R., Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data.

Bobelyte, O., Gailute, I., Zubka, V., Zilinskaitel, V. (2017). Sepsis epidemiology and outcome in the paediatric intensive care unit of Vilnius University Children's Hospital. Retrieved from

Brierly, C., Choong, K., Cornell, T., DeCaen, A., Deymann A et al. . Clinical practice parameters for hemodynamic support of pediatric and neonatal septic shock: 2007 update from the American College of Critical Care Medicine. Crit Care Med. (2009)

Weiss F., Maffei K., Rodriguez-Nunez A., Hsing DD, et al.Discordant identification of pediatric severe sepsis by research and clinical definitions in the SPROUT international point prevalence study. Crit Care. (2015)

De Baets L., Ruyssinck J., Peiffer T., De Turck F., Ongenae F., Dhaene T. (2016). Positive blood culture detection in time series data using a BiLSTM network.

Dauvin A., Donado C., Bachtiger P., Huang K., Sauer C., Ramazzotti D., Bonvini M., Celi L. (2019). Machine learning can accurately predict pre-admission baseline hemoglobin and creatinine in intensive care patients.

Luo Y., Szolovits P., Dighe A., Baron J. (2016). Using Machine Learning to Predict Laboratory Test Results.

Ugarte, H., Silva, E., Mercan, D., De Mendonca, A., Jean-Louis V. (1999). Procalcitonin used as a marker of infection in the intensive care unit. Retrieved from

Jha AK, Chan DC, Ridgway AB, et al. (2009). Improving safety andeliminating redundant tests: cutting costs in U.S. hospitals. Retrieved from: Health Aff. 2009;28:1475-1484.

Klein, G., Kim, Y., Deng, Y., Senellart, J., Rush, A. (2017). OpenNMT: Open-Source Toolkit for Neural Machine Translation.

Trask, A., Gilmore, D., Russell, M. (2015). Modeling Order in Neural Word Embeddings at Scale.

Piliponsky, A., Shubin, N., Lahiri, A., Phuong, T., Clauson, M., Niino, K., Tsuha, A., Nedospasov, S., Karasuyama, H., Reber, L., Tsai, M., Mukai, K., Galli, S. Basophil-derived tumor necrosis factor can enhance survival in a sepsis model in mice. Nature Immunology, 2019.