Machine Learning Solutions for early sepsis detection

Jonas Grabbe
grabbe.jonas@gmail.com

Sepsis is one of the leading causes of morbidity and mortality in hospitals. Early diagnosis could substantially improve the patient outcomes and reduce the mortality rate. However, while professional critical care societies have proposed new clinical criteria that aid sepsis recognition, the fundamental need for early detection and treatment remains unmet.
This resume is the first draft to outline how one could proceed in developing an MLA solution for questions like: early detections, survival or length of stay. The data we have access to at the moment the the PhysioNet 2019 Challenge dataset which consists of 40k UCI patients with 40-time dependent features.

1. Current state of sepsis

Sepsis is a major health crisis. It is worldwide one of the leading causes of hospital admissions and hospitalized patient' s mortality. It is thought of to play a role in over 50% of hospital deaths and accounts for between 6-30% of ICU admissions. It is also a major contributor to Hospital length of stay.

In figures that translates into 970.000 sepsis patients each year in the US alone with over 270.00 fallacies. Worldwide there are over 30 million people (4.2 million new-borns and children) who develop sepsis and 6 Million people who are estimated to die from it.

Besides the personal tragedies, with 24 Billion each year in the US, sepsis is a major burden on hospital budgets (13%) [1].

2. What is sepsis

Sepsis is a life-threatening condition caused by your body' s response to an infection (bacterial, viral or fungal) and damages its own tissues [2]. This can lead organs to function poorly or abnormally. Septic shock which is a dramatic drop in blood pressure can itself can cause severe organ problems and death.

Early treatment with antibiotics and intravenous fluids improves chances for survival, while age, in particular being senior or infant, having a compromised immune system or diabetes are among other increase risk factors.

3. Why would a ML solution be helpful

Machine Learning solutions for early sepsis predictions is a perfect task for AI in medicine. Since each hour of delay in treatment can roughly increase mortality by 4–8 % combined with the fact that few electronically monitoring methods of patients provide predictive capabilities to enable early intervention, an MLA could be a mayor help in sepsis detection.

Furthermore note, that in case of septic shock, the risk of dying increases by approximately 10% for every hour of delay in receiving antibiotics and by the time sepsis is diagnosed (SOFA) mortality is already by 10%.

4. Objective

The goal of this work is to work out a ML solution for the early detection of sepsis using physiological data. The early prediction of sepsis is potentially life-saving, and we aim to predict sepsis at least 6 hours before the clinical prediction of sepsis.

Furthermore, we would like to find ML prediction algorithms on patient admission, seveners, survival and hospital length of stay.

5. Sepsis risk indicators

Various rule-based disease-severity scoring systems are widely used in hospitals in an attempt to identify patients with sepsis. These scores are such as the Modified Early Warning Score (MEWS), the Systemic Inflammatory Response Syndrome (SIRS) and the Sequential Organ Failure Assessment (SOFA) and the quick SOFA (qSOFA) [3].

Although these disease severities scoring systems were designed to predict patient risk, rather than specifically to identify sepsis, they are commonly used in severe sepsis diagnostic criteria due to their designed purposes of identifying systemic inflammation as a sign of possible infection and detecting possible organ dysfunction [4].

6. DataSet

The data obtained is a 40k patient strong dataset which was made public for the 2019 Physionet Challenge [5]. The data was obtained from two geographically distinct U.S. hospital systems with two different electronic medical record systems: Beth Israel Deaconess Medical Center (hospital system A) and Emory University Hospital (hospital system B), collected over the last decade.

The data has 40-time dependent features which can broadly be classified into:

- → Demographics (26)
- → Vital Signs (8)
- → Laboratory values (6)

Furthermore, the de-identified data was given a SepsisLabel for each patient according to the Sepsis-3 guidelines, i.e., a two-point change in the patient's SOFA score and clinical suspicion of infection (as defined by the ordering of blood cultures or IV antibiotics).

The SepsisLebel is zero until 6 hours before Sepsis-3 guidelines would diagnose sepsis, when it is set to one. For patients who never develop sepsis the data is labelled zero everywhere.

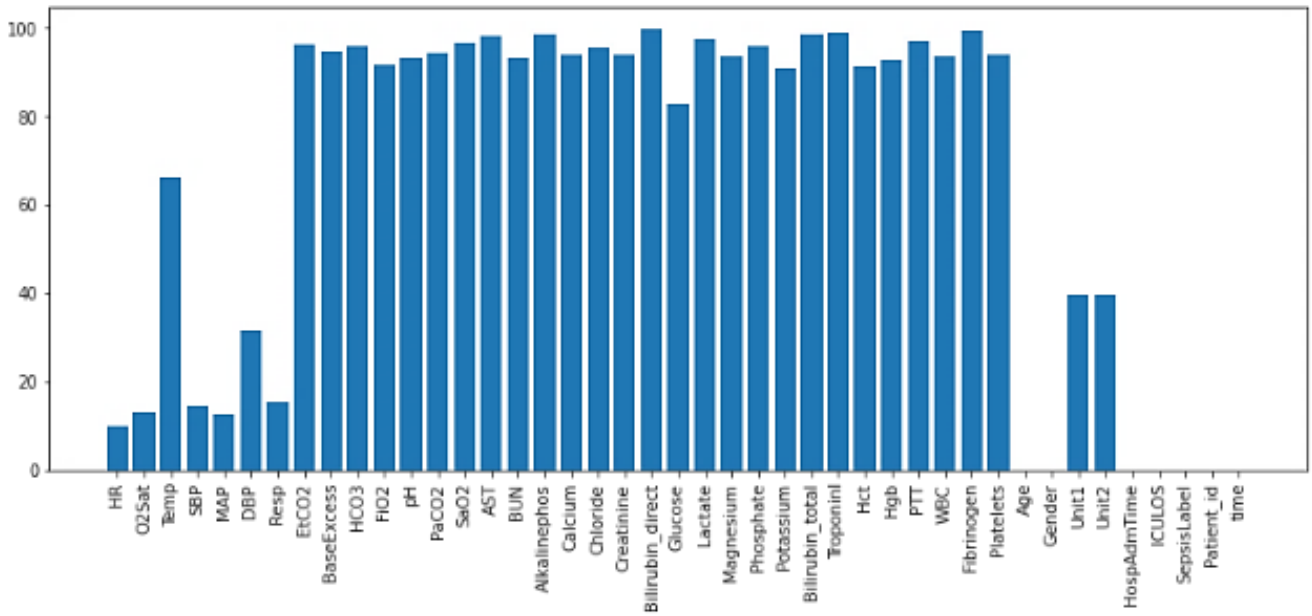
There are two ways in which one can approach this problem:

1. Temporal Approach: Take into the account the time component for the data. Sepsis is diagnosed for each patient at each hour using the past data.
2. Non-temporal Approach: Ignore the time component and treat record as independently and identically distributed. This approach would help in predicting Sepsis at each hour for any patient (with or without patient past data).

Although the Temporal Approach seems more promising, we will start first with the Non-temporal approach.

In the dataset of 40.366 individual patients with 2932 individual sepsis patients in the Non-temporal approach, we can consider 1.524.294 non-sepsis patients and 27.916 sepsis patients, when merged, which is about a 1% vs 99% ratio.

A part from the high ration the data set has also a lot of missing data. In particular the laboratory values have missing rates between 82% and 99%.



7. Data processing pipeline and feature selection

We eliminate variables which have greater than 82% missing values, except Bilirubin direct, Lactate, Partial thromboplastin time (PTT), Creatinine, Leukocyte count (WBC) and Glucose which are known as significant variables for detecting sepsis.

In addition, the variables Unit1 and Unit2 are equally distributed across the whole dataset, and are uncorrelated with the SepsisLabel and is therefore removed.

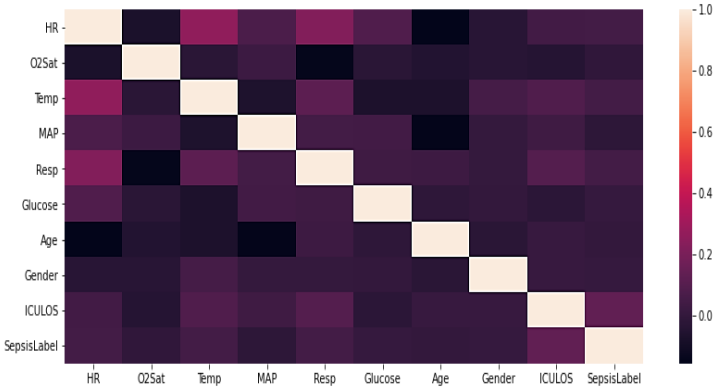
The vitals Mean arterial pressure (MAP), Systolic BP and Diastolic BP are highly correlative since MAP = (SBP + 2*DBP) / 3, thus we eliminate SBP and DBP.

Imputation

For the laboratory values remaining variables, we impute any missing values with the mean obtained from the healthy group for this variable if the patient is healthy and conversely if the patient is sick. (For testing on the unseen dataset, we impute the mean between both groups if there are missing values.)

For the vital signs' variables, if a patient did not have a measurement in a given hour, the missing measurement was filled in using carry-forward imputation (Forward backfill).

The variable O2Sat corresponds to the pulse oximetry percentage, and would be controlled for ventilated patients and our correlation method finds that O2Sat is not appearing as a predictive variable and is thus eliminated.



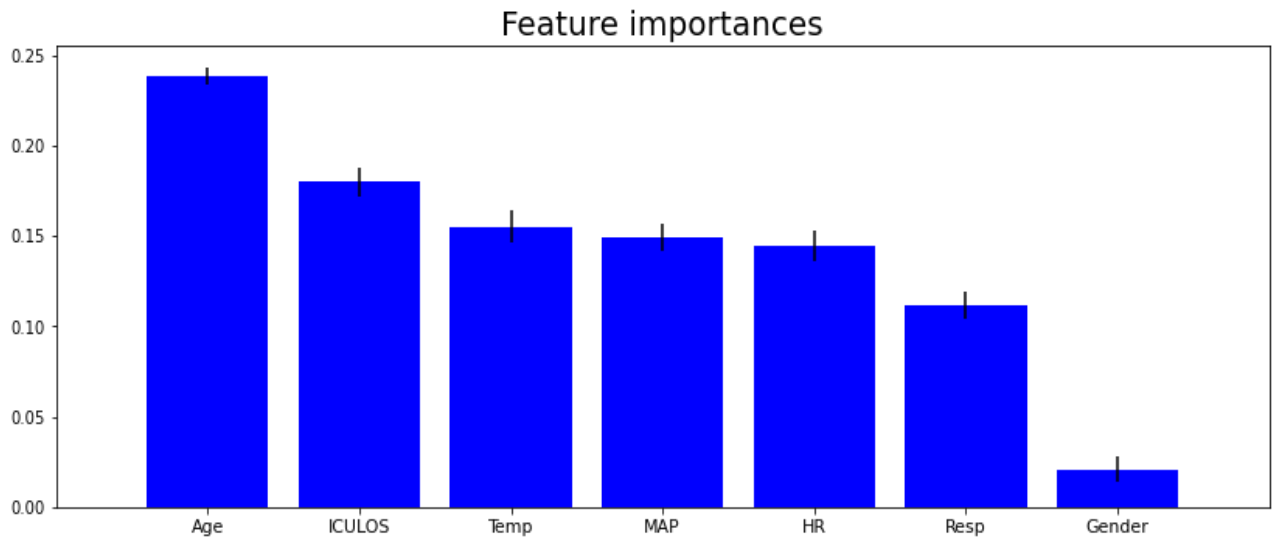
For the demographic variables, we select Age, Gender and Intensive care unit length of stay (ICULOS).

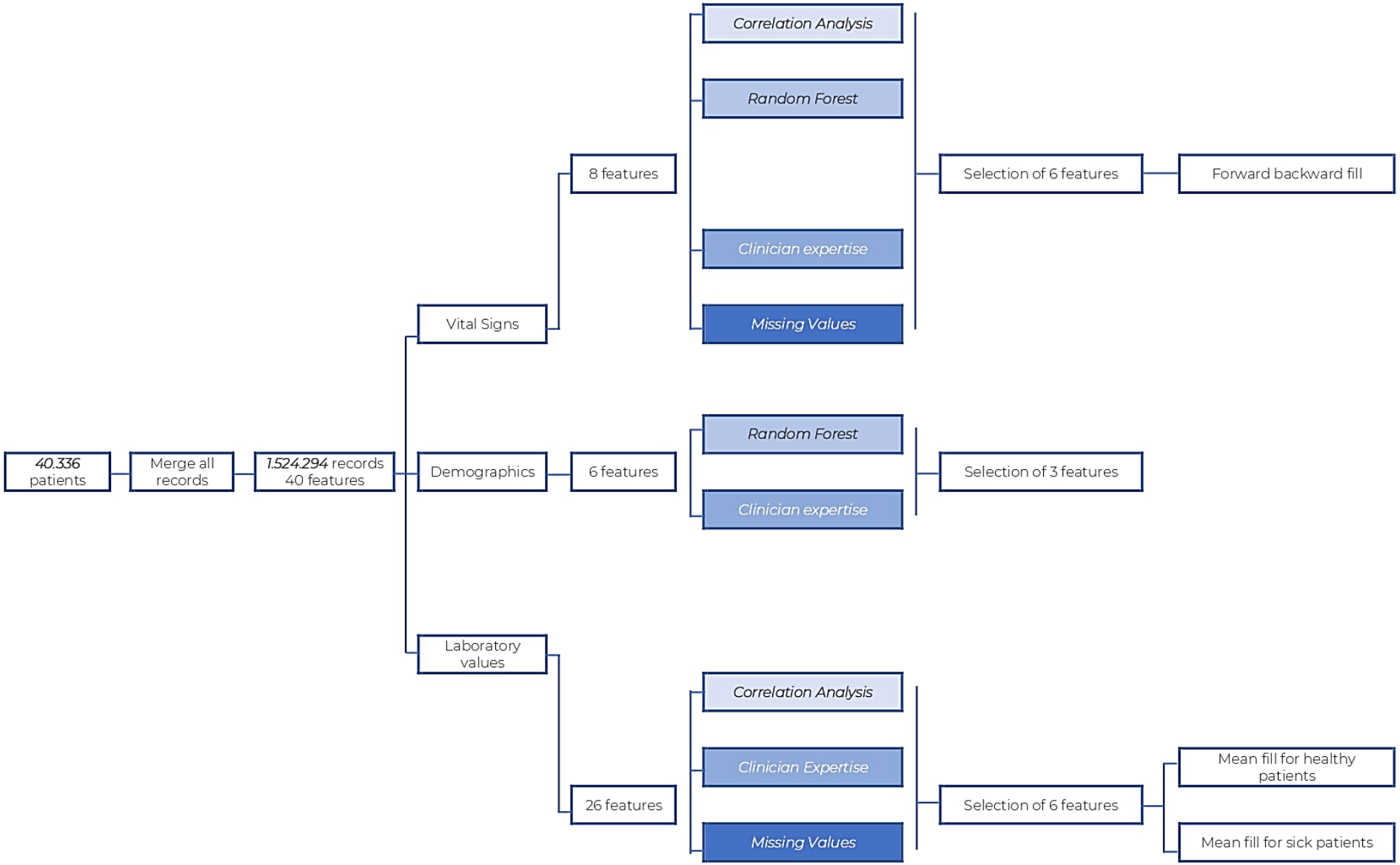
Random forest importance features computation

We want to evaluate the importance of the remaining features in the prediction of sepsis. We split the dataset into training and test sets in proportion 70:30.

As might be expected, the variable Age is the most significant. Sepsis is equally prevalent in both genders, and so as expected the Gender variable is judged to be of low importance.

(Davide Chicco and Giuseppe Jurman: Survival prediction of patients with sepsis from age, sex, and septic episode number alone) [6].




40.336 patients
Merge all records
1.524.294 records 40 features
Vital Signs
8 features
Correlation Analysis
Random Forest
Clinician expertise
Missing Values
Selection of 6 features
Forward backward fill
Demographics
6 features
Random Forest
Clinician expertise
Selection of 3 features
Laboratory values
26 features
Correlation Analysis
Clinician Expertise
Missing Values
Selection of 6 features
Mean fill for healthy patients
Mean fill for sick patients

8. Models

To get a Baseline one can quickly run a mix of standard ML algorithms to get familiarized the dataset and its limits, such as:

- → Logistic Regression
- → AdaBoost
- → Gradient Boosting
- → RandomForest

Autoencoders / anomalies detection

Auto-encoders have proven to be useful for anomaly detection use-cases which involves high class imbalance.

Auto-encoders would be expected to perform better than traditional Machine Learning models as they are modelling the behaviour of positive class and treating the negative class as an anomaly.

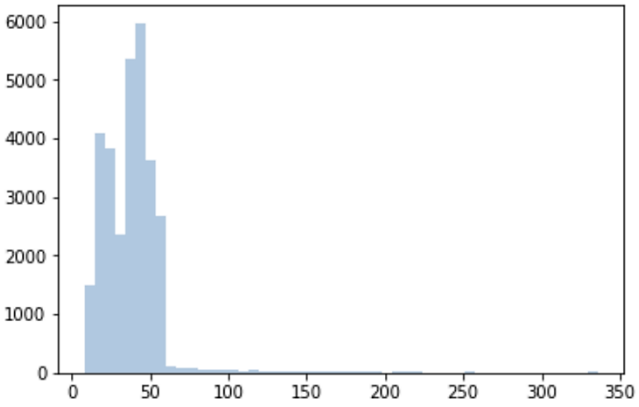
That's why one could see sepsis detection as an anomaly detection problem where you may think about an individual who developed sepsis as a patient with specific abnormal representative features, i.e., for each patient, we aim to detect a change from their usual clinical measurements.

DNN or GXBOOST

With help of hyperparameter optimisation and 5-fold CVgrid search on can further apply it to the DNN or GXBOOST as it was done in (Predicting hospital admission at emergency department triage using machine learning by Woo Suk Hong) [7] for the prediction of hospital admission and UCILOS.

TimeSerie

When considering the Non-temporal Approach, the obvious choice would be to start to apply it to



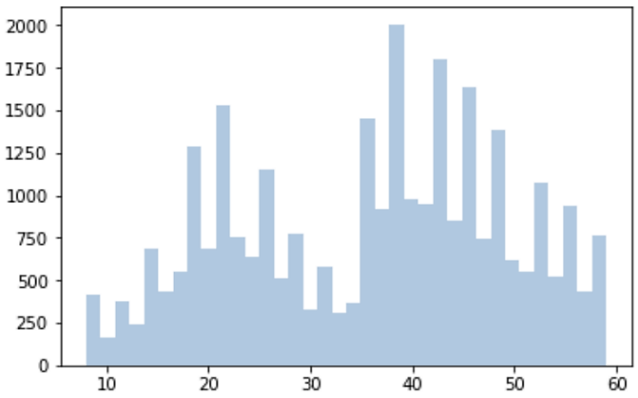
the Long Short-Term Memory (LSTM) algorithm.

Number of hours a patient spends in Hospital

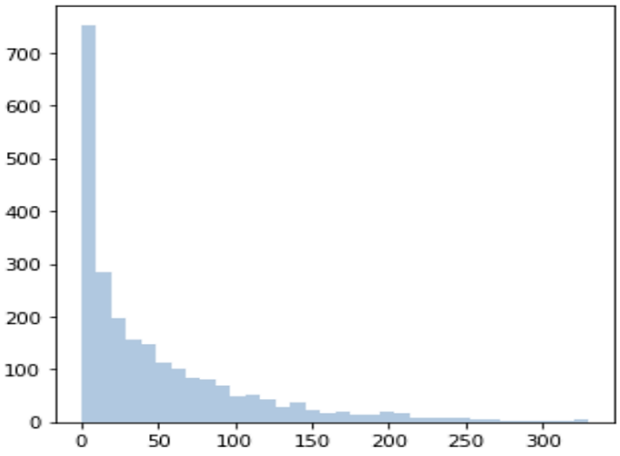
Looking back at the data, we could remove those patients with more than 60 hours in the hospital.

These patients are considered outliers. The reason behind this is that we will be using LSTM for modelling and window of size 350 won't be a great idea.

After removing patients with more than 60 hours, we see that we lose a considerable number of



patients from the positive class. Hence, they are not being removed for the moment



Time at which a patient contracts Sepsis:

Hence, we look for inspiration to the solution of James Morrill in "The Signature-Based Model for Early Detection of Sepsis From Electronic Health Records in the Intensive Care Unit" [8].

1. Hand-Crafted features

He engineered with the give data new features, like the 'ShockIndex', which is defined as the heart rate divided by the systolic blood pressure and 'BUN/CR' which is the ratio of levels of bilirubin to creatinine. It was introduced by Katharine E. Henry in "A targeted real-time early warning score (TREWScore) for septic shock" [9], where the authors showed their importance in sepsis detection.

As sepsis is labelled only if there has been a 2-point deterioration in 'SOFA' score within a 24-hour period, an auxiliary 'PartialSOFA' score is additionally introduced. The partial 'SOFA' score is constructed from the variables that present both in the 'SOFA' score and the dataset. Additionally, the indicator variable SOFA deterioration is marked as '1' if 'PartialSOFA' saw this deterioration over the last 24 hours.

2. Signature Features

They used the theory of controlled differential equations to extract the information about a path needed to predict the behaviour of a system (the so-called Signature of a path).

Then a sliding window approach was used so that the signature features are computed for each time-point over a window of some given look-back size.

The signatures of 'PartialSOFA', 'MAP' and BUN/CR' were computed with a time dimension and the lead-lag transformation and then

signatures of all non-stationary columns were computed after first applying the cumulative sum followed by the lead-lag transformation.

3. Model starts with 5-fold CV

Then with about 40 hyperparameters (such as look-back windows), they trained an LSTM and use the output as a feature in the LGBM Regressor with a stratified 5-fold cross validation method.

At las one could like to try BayesianNet and Reinforced Learning algorithms or see if converting to date into categories would be helpful, e.g.:

→ Heart Rate: Normal, Abnomal, Missing
→ Temperature: Normal, Abnomal,Missing
→ Age: Old,Infant, Child/Adult
→ Etc.

9. Monitorization

A market ready MLA could be sold as electronic health record (EHR) agnostic and hence can be integrated with all major EHR systems.

The MLA can run in the background using the readily-available vitals and alert clinicians of sepsis cases. The usefulness of this should be determined by clinicians and a threshold can be set to achieve clinically meaningful sensitivity and specificity.

But before the MLA can be verified in both a randomized controlled trial and a real-world post-marketing study it has to surpass the Gold standards test SOFA, SIRS and MEWS which have an area under the receiver operating characteristic (AUROC) curve of 0.725, 0.609 and 0.803, respectively.

The MLA should as well surpass the already market ready MLA InSight by Dascena, which has an AUROC of over 0.90 and a sensitivity score of 93% and specificity of 91%.

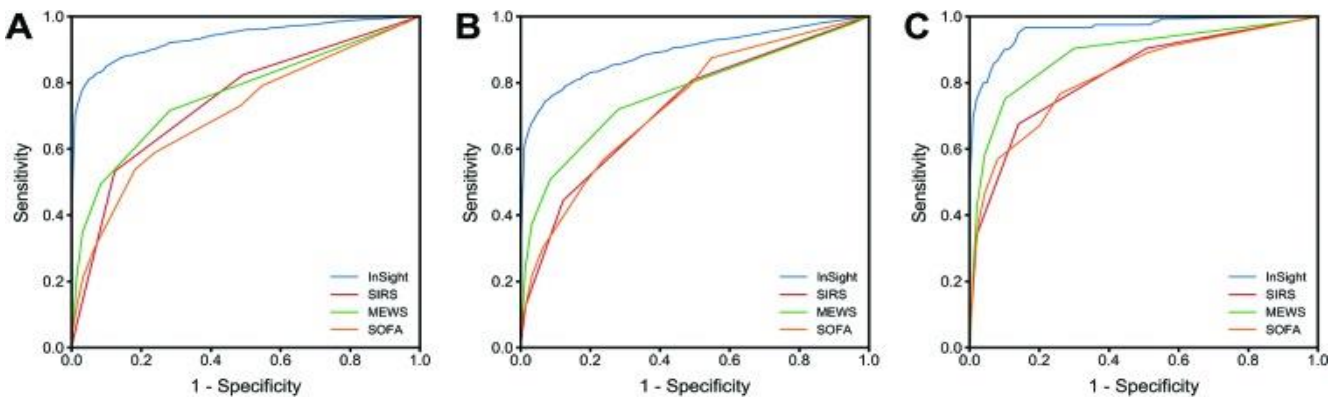
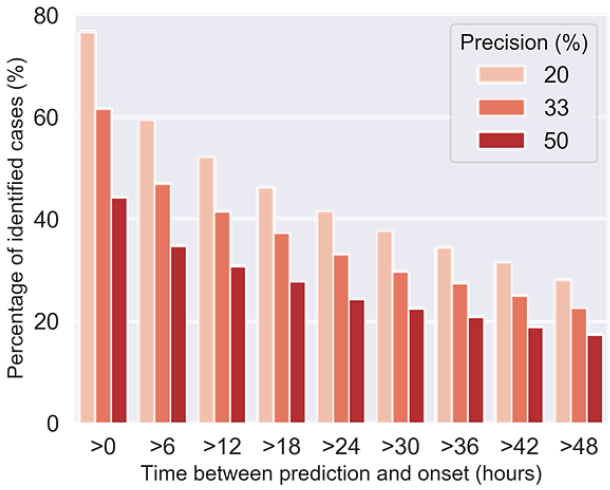


Fig: *ROC curves for InSight and common scoring systems at the time of (A) sepsis onset, (B) severe sepsis onset and (C) 4 hours before septic shock onset.*

Furthermore, it should hold stand to the 2019 PhysioNet Challenge winner, which model often detects sepsis much further in advance than the 6 hours, regularly over 24 hours.



References

[1] Carly J Paoli et al. *Epidemiology and Costs of Sepsis in the United States-An Analysis Based on Timing of Diagnosis and Severity Level.*

[2] Singer et al.: *The Third International Consensus Definitions for Sepsis and Septic Shock* (Sepsis-3).

[3] Qingqing Mao et al. *Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general Ward and ICU.*

[4] David W Shimabukuro: *Effect of a machine learning-based.*

[5] Matthew Reyna et al.: *Early Prediction of Sepsis from Clinical Data -- the PhysioNet Computing in Cardiology Challenge 2019 .*

[6] Davide Chicco and Giuseppe Jurman: *Survival prediction of patients with sepsis from age, sex, and septic episode number alone).*

[7] Woo Suk Hong et al.: *Predicting hospital admission at emergency department triage using machine learning.*

[8] James Morrill et al.: *The Signature-Based Model for Early Detection of Sepsis From Electronic Health Records in the Intensive Care Unit.*

[9] Katharine E. Henry et al.: *A targeted real-time early warning score (TREWScore) for septic shock.*