Utilisation of the Signature Method to Identify the Early Onset of Sepsis from Multivariate Physiological Time Series in Critical Care Monitoring

Authors

James H Morrill, MMath[1 4]

Andrey Kormilitzin, PhD[1 2]

Alejo J Nevado-Holgado, PhD[2]

Sumanth Swaminathan, PhD[3 4]

Samuel D Howison, DPhil[1]

Terry J Lyons, DPhil[1]

This work was performed at: Department of Mathematics, University of Oxford, UK.

Reprints

---

[1] Department of Mathematics, University of Oxford, UK.

[2] Department of Psychiatry, University of Oxford, UK.

[3] Department of Mathematics, University of Delaware, Newark, DE, USA.

[4] Iterex Therapeutics, 500 7th Ave, NY 10018 New York, US.

James Morrill, Andrew Wiles Building, Woodstock Rd, Oxford OX2 6GG.

Reprints will not be ordered.

Funding

James Morrill is supported by the ESPRC under the program grant EP/L015803/1 in collaboration with Iterex Therapeutics.

Andrey Kormilitzin and Alejo Nevado-Holgado are supported by the MRC under the Pathfinder programme grant MC/PC/17215.

Terry Lyons is supported by the EPSRC under the grant EP/S026347/1 and by the Alan Turing Institute under the EPSRC grant EP/N510129/1.

Terry Lyons, James Morrill and Andrey Kormilitzin are members of the DATASIG programme.

Corresponding author

Name: James H Morrill

Address: 34 West St, Oxford, UK, OX2 0BQ

Phone: +44 7772 663314

Email: james.morrill.6@gmail.com

Key words

Sepsis; critical care; PhysioNet; early detection and treatment; Competition

Abstract

Objective Patients in an intensive care unit (ICU) are particularly vulnerable to sepsis. It is therefore important to detect its onset as early as possible. This study focuses on the development and validation of a new signature-based regression model, augmented with a particular choice of the hand-crafted features, to identify a patient's risk of sepsis based on physiological data streams. The model makes a positive or negative prediction of sepsis for every time interval since admission to the ICU.

Design The data was sourced from the PhysioNet/Computing in Cardiology Challenge 2019 on the “Early Prediction of Sepsis from Clinical Data”. It consisted of ICU patient data from three separate hospital systems. Algorithms were scored against a specially designed utility function that rewards early predictions in the most clinically relevant region around sepsis onset and penalises late predictions and false positives.

Setting The work was completed as part of the PhysioNet 2019 Challenge alongside 104 other teams.

Patients PhysioNet sourced over 60,000 ICU patients with up to 40 clinical variables for each hour of a patient’s ICU stay. The Sepsis-3 criteria was used to define the onset of sepsis.

Interventions None

Measurements and Main Results The algorithm yielded a utility function score which was the 1st placed entry in the official phase of the challenge.

Introduction

Sepsis is a life-threatening organ dysfunction caused by a dysregulated host response to infection. Sepsis management represents an extremely challenging problem throughout the world. Recent estimates suggest around 1.7 million adult Americans develop sepsis each year, of which nearly 270,000 dies, this accounting for one in three hospital deaths [1]. A recent study [2, 3, 4] evaluated the cost to Medicare of sepsis inpatient admission and subsequent patient care to be in excess of $41.5 billion, and found that both the numbers of sepsis-related claims and the aggregate dollar costs are rising. Additionally, it has been reported that mortality rates increase significantly with delays in receiving antibiotics. For example, Kumar et al. [5] report that in cases of septic shock, the risk of death increases by ~10% for every hour of delay in antibiotic treatment. Early detection of sepsis events is thus essential to improving sepsis management and mortality rates in the ICU.

Machine-learning algorithms have been investigated as a method for early prediction of sepsis [6, 7, 8]. However, it is difficult to compare the different algorithms since each addresses a slightly different problem, utilises different data sets, maintains a different definition of sepsis onset, and uses different metrics. Furthermore, none use the new Sepsis-3 definition [9] to denote the onset of sepsis.

In this study, we develop a machine-learning model for early detection of sepsis based on readily attainable ICU data. The model was developed during the 2019 PhysioNet challenge entitled ‘Early Prediction of Sepsis from Clinical Data’. Participants were invited to submit algorithms that were trained on the same set of data and validated under a common performance metric on unseen test data. Sepsis onset was defined using the Sepsis-3 criteria. This paper is an extension of the Computing in Cardiology conference proceedings paper [10] with the inclusion of further method details, additional results, and discussion.

Methods

Dataset

Data for this study was taken from the PhysioNet/Computing in Cardiology Challenge 2019. Full details are found in the official challenge paper [11]. In brief, the data was sourced from ICU patients in three separate hospital systems. Data from two hospital systems was split into a publicly available training set and an undisclosed test set, both to be used for model development and testing. The test set and third hospital system data were kept private, with submitted models being scored against these unseen data from all three systems. A total of 40,336 patients were used in model training with 40 features per patient, consisting of demographic, vital sign, and laboratory data. The data for each patient was indexed with time at one-hour increments and predictions were to be made sequentially at each hour in a patient's time-series with any future information being censored. The models were scored against the custom utility function defined in [11].

Data Imputation

Missing values during a patient's hospital stay were imputed by using a forward-fill method. When a value was missing, it was taken to be the most recent measurement of that particular variable. If no previous value existed the value was left as ‘NaN’, since decision tree-based algorithms can handle such missing data effectively.

The Path Signature

Sequential time-series data exhibits a natural path-like structure. A path X of finite length in d dimensions can be described by the mapping $X\colon [a,b] \rightarrow \mathbb{R}^d$, or in terms of co-ordinates $X = (X_t^1, X_t^2, \dots, X_t^d)$, where each coordinate X_t^i is real-valued and parameterised by $t \in [a,b]$. For example, the path of a patient's systolic blood pressure changing in time can be thought of as a path in $\mathbb{R}^2$ with the first dimension being time and the second being the value of the systolic blood pressure at that point in time.

The signature transformation S of a path X is defined as an infinite sequence of terms:

$$S(X)_{a,b} = (1, S(X)_{a,b}^1, S(X)_{a,b}^2, \dots, S(X)_{a,b}^d, S(X)_{a,b}^{1,1}, S(X)_{a,b}^{1,2}, \dots)$$

where each term is a k-fold iterated integral of X with multi-index $i_1, \dots, i_k$:

$$S(X)_{a,b}^{i_1,\dots,i_k} = \int_{a<t_k<b} \int_{a<t_1<t_2} dX_{i_1}^{t_1} \dots dX_{i_k}^{t_k}.$$

This defines a graded sequence of numbers associated with a path which is known to characterize it up to a generalized form of reparameterization [12]. One can think of the signature as a collection of summary statistics that determine a path (almost) uniquely. Furthermore, any continuous function on the path X can be approximated arbitrarily well as a linear function on its signature [13]; the signature unravels the non-linearities on functions on the space of unparameterized paths.

Despite its mathematical abstraction, the signature is simply working as a data compression tool, and is particularly well suited to data streams of the kind so prevalent in healthcare. In Figure 1 we give a more concrete illustration of how the signature enables us to discriminate between different paths. Here we plot two examples of how heart rate might change with blood pressure, we are assuming there exists an implicit time dimension such that the values in each plot are moving from left to right along the blue lines in each figure. The first order of the signature is simply the change in each variable over the interval (this is $(\Delta HR, \Delta SBP)$) giving us the first level of description of the path, however, we can see this is not yet sufficient to discriminate between the two cases. The second level gives us the signed areas (the shaded orange regions), where the orientation of the left most plot is such that the negatively signed area is produced whereas the second gives the positive value, and thus, at order 2 in the signature we now have sufficient information to discriminate between these two situations where in the first rise in heart rate occurs before (or at least, initially faster than) the rise in blood pressure, and vice versa. Here we are hinting at something fundamental about the signature: It structurally picks out the information concerning the ordering in which events occur. If the ordering of events in any stream-like data is thought to be important, then the signature is a natural tool with which we can utilise. We refer to [14] for a more thorough introduction to the subject of signatures and their applications in machine learning.

[Insert Figure 1]

The usefulness of signatures as features of sequential data has been demonstrated theoretically for non-parametric hypothesis testing [15] and algebraic geometry [16] as well as in numerous machine learning applications. For example: in healthcare [17, 18, 19, 20, 21], finance [22], computer vision [23], topological data analysis [24] and deep signature learning [25, 26].

Non-Signature Feature Extraction

We augmented the data sets by including a number of additional features typically thought to be useful for discerning the onset of sepsis. These features, based on literature review and “expert knowledge”, included a partial construction of the Sequential [Sepsis-related] Organ Failure Assessment (SOFA) score [9], which we call PartialSOFA. The term “partial” was used because the data set did not include all variables that comprise the SOFA. Hence, the score was calculated based on the available required information which consisted of threshold conditions on each of the platelet count, bilirubin, mean arterial pressure, and creatinine variables. Given that one of the requirements of the Sepsis-3 definition is a deterioration of the SOFA score, any approximation of this quantity would be expected to aid in prediction of sepsis onset. Other added features included the ratio of heart rate to systolic blood pressure (the “shock index”) and the ratio of bilirubin to creatinine as suggested in [6].

We hypothesised that measurement frequency would provide an indication of patient health, given the anticipated increase in sampling when physicians are concerned about patients. For this reason, we also included a counter variable for the temperature variable and the laboratory values, which denotes the number of times a given variable had been measured in some look-back window. Finally, we included the maximum and minimum of the vital signs over a look-back window. The size the look-back windows were treated as hyper-parameters and optimised in model training.

Signature Feature Extraction

To extract additional longitudinal information from the time series, we turned to the signature transformation (as outlined above). A simple sliding window approach was used so that signature features were computed for each time-point over a window of some given look-back size. We computed the signatures of the ‘PartialSOFA’, ‘MAP’ and ‘BUN/CR’ variables where the input paths contained an augmented time dimension. We also computed the signatures of all non-stationary features after first applying the cumulative sum followed by the lead-lag transformation on the input paths (refer to [14] for information on the cumulative sum and lead-lag transformation). The signature truncation levels and the lengths of the look-back windows were treated as hyper-parameters and optimised in model training.

Sepsis Labels

The data was pre-labelled with the value 1 at any location of sepsis occurrence or pre-defined window around sepsis onset, and zero otherwise. Given that the aim was to optimize the utility score (as defined in [11]), not simply the percentage of correct binary predictions, we created an alternative labelling that accounted for information about the utility score and enabled the classifier to place higher weights on points that lead to a larger score if predicted correctly. Let $U_y(s,t)$ denote the utility score of predicting $y \in \{0,1\}$ for patient s at time t. The “modified utility score” is defined as $U_{MUS}(s,t) = U_1(s,t) - U_0(s,t)$ and constitutes the case labelling against which the regressor was trained. This allowed for increased weighting on cases that yield a larger utility difference if predicted correctly.

Model Training and Validation

We used a stratified 5-fold cross validation with a uniform distribution of time-points and sepsis labels in each fold for hyper-parameter optimisation of the predictive algorithm. No patient had data in more than one fold. We used the LightGBM implementation [27] of the gradient boosting machine algorithm with the mean square error loss function for the regression task. Hyper-parameters were chosen to maximise the mean utility score over the 5-folds. The final values of the parameters can be found in [10].

Results

The submitted model was the 1st-place official entry to the challenge with an average utility score of 0.360. Table 1 shows the scores for each hospital both for the public training data and the unseen test data, the scores for the public training data are taken to be the average utility over the 5-folds used in cross-validation. One can see that the algorithm scores on hospitals A and B in the training and test sets are similar, which suggests that our model was not over-fitted when restricted to testing on the same hospital system as that on which the model was trained. Conversely, the scores on hospital C were significantly worse, which highlights the potential danger of using an algorithm that was trained on one hospital system to make predictions on a different one.

The Usefulness of Signatures

Table 2 shows the cross-validated and averaged utility score predictions on the training data for different subsets of features. Note that the score for each row is from the algorithm containing the features listed for that row, along with any features that are listed above it. The data indicates that the time since admission feature alone can be used to achieve a utility score of 0.282; this feature was, indeed, found to be the most important. Of particular further interest is the additional benefit to be gained from supplementing the dataset with the signature features. We see that the addition of the signature terms increases the utility score from 0.422 to 0.434 and this, while not a dramatic improvement, does show that the representation of the information fed into the algorithm after the signature transformation is beneficial to learning, and thus worthy of inclusion in comparable modelling approaches for similar problems.

[Insert Table 2]

Predictions in an In-Hospital Environment

While the goal of the challenge was to optimise the score achieved on the pre-defined utility function, it is useful to consider how the model could be used in an in-hospital environment to provide clinically actionable information. At each time-point the model outputs a value with increasing values associated with higher risk of sepsis. When the model output exceeds a specified operating-point threshold, the subject is designated as a *sepsis-risk patient*, thus indicating to a doctor that closer monitoring or further tests may be warranted. This operating-point threshold can be chosen to achieve the most clinically meaningful sensitivity and specificity. The area under the receiver operating characteristic (ROC) curve of the optimised model was 0.868. A closer analysis is displayed in Figure 2. In the leftmost plot we have a confusion matrix, where a true negative represents that no call was made and the person didn't develop sepsis, and a true positive represents that the patient being flagged and developing sepsis *at some point* after this call. The operating-point threshold is set so that 33% specificity (the proportion of correct positives that were correctly identified) is achieved. On the right of the figure we give a performance analysis in increasing time-windows before onset; for example, the first bar gives the number of correct predictions made one or more hours in advance at 20% precision.

[INSERT FIGURE 2]

Discussion

The signature-based model that we have presented for the early prediction of sepsis constitutes a competitive approach to discerning early onset sepsis from health data streams. The model achieved the highest score in the official phase of the challenge [11]. We believe the strength of our model lies primarily in the smart collection of hand-crafted features derived from the data, such as 'PartialSOFA', along with effective utilisation of the signature method to extract additional features from the time-series. The non-signature features were not unique to our submission, however, the signature provided us with an extra piece of technology that allowed us to represent the longitudinal multidimensional clinical data as a collection of low-dimensional characteristic features that were not considered by other teams. In addition to this, training against the utility score as opposed to the provided binary labels for sepsis, and choice of classifier paired with an extensive hyperparameter search were two other things that we found made significant improvements to the score. Despite the success of this method with comparison to other teams, there are a few limitations and caveats to be addressed.

First, one of the primary aims of the PhysioNet/CinC challenge was to develop a model to make accurate predictions in a window around 6 hours prior to the onset of sepsis, as this was determined to be the most clinically relevant region for plausible intervention. However, we see from Figure 2 (right) that the majority of predictions were not made in this desired time window. Similar long-time predictions have been found in other studies such as [6], in which the governing model predicted sepsis shock with a median time between prediction and onset of 28.2 hours, and more recently [8] demonstrated a median time-to-event prediction of 48.0 hours. Even when training against the utility function, which is biased towards making predictions in the desired 6-hour time-window, our model makes predictions with a median time of 25.0 hours in advance (at 33% specificity). In [3] Buchman et al. provide evidence suggesting that sepsis may often occur as a result of a manifestation of some other chronic conditions, as opposed to a 'bad luck' event from a more ordinary infection. This hypothesis would make sense in the context of these model predictions, with the conclusion being that the model may be identifying characteristic features of chronic conditions from which sepsis is likely to manifest. A simple example being high-temperature correspondent to a fever from which sepsis can develop. This would explain why predictions are being made in advance of where the symptoms of sepsis are expected to have developed.

In Figure 3 we show the number of people that have not yet developed sepsis as a fraction of the subset of people who do eventually develop it. This is plotted for both hospitals in the training set (blue solid lines) along with a best fit exponential (orange dashed lines). The figure shows that the fitted trend line matches the true data very closely. Given that the population level curve shows a clean exponential decay, it could be that the likelihood of any individual patient developing sepsis is dependent on a variable or variables that are governed by a Poisson process. Alternatively stated, there could be random events at play that are difficult to resolve with our modelling formalism. One could consider, for example, that sepsis arose out of the accidental use of a dirty needle. The point at which a dirty needle is used is almost impossible to predict (assuming it is a true Poisson process). If these events happened far enough in advance of onset, one could feasibly expect to build a model to make a positive prediction in a sensible time-window prior to the emergence of sepsis if the feature data could be collected; alternatively, the time-scale between such events and sepsis onset could be very short making predictive models difficult. An increase in the data sampling rate could be beneficial as 1-hour bins would likely not contain the required resolution to accurately resolve all of the relevant features.

[Insert Figure 3]

Finally, as previously noted, the model's inability to maintain high prediction performance on hospital C indicates a limitation in generalising predictions to hospital data-sets on which the algorithm was not trained. There are a number of potential reasons as to why this might be the case: For example, the models were trained on variables that were highly dependent on physician decision-making processes and thus local hospital policies. A measurement or assessment that is encouraged to be taken by a doctor in one hospital may not comport with the practices of another hospital. As such, any model that is trained of data-sets from healthcare systems in which a gold standard for assessment procedures and measurements does not exist, will inherently adopt some of the biases of the underlying training set. One remedy to this limitation would be to only train on variables that are sampled at pre-determined times of a patients stay and are independent of a doctor decision-making process. One could then expect to more safely transfer an algorithm from across multiple hospital systems. Any remaining uncontrolled variables between hospitals such as patient demographics, socio-economic factors, etc would, of course, persist as sources of error.

Conclusion

We have presented a signature-based model for early prediction of sepsis. The signature representation produced a useful summary of the longitudinal physiological measurements that were used to effectively discriminate septic from non-septic cases. The addition of the signature terms improved the algorithms prediction performance significantly as demonstrated in Table 2. The method proposed has achieved the highest official score on the utility function on the full test data set from 104 submissions. We have further shown that model predictions can be turned into clinically-actionable information for use by doctors. Finally, we explored: 1) why the model struggles to predict more cases in the desired time-window before onset, 2) the model's limited performance on new hospital systems and 3) some potential methodological enhancements for better generalising the model predictions.

Acknowledgements

We are thankful to Scott Ganis for his stimulating discussions during this project.

References

1. National Center for Emerging and Zoonotic Infectious Diseases (NCEZID), Centers for Disease Control and Prevention, Division of Healthcare Quality Promotion (DHQP). https://www.cdc.gov/sepsis/datareports/index.html (2016)
2. Buchman, T.G., Simpson, S.Q., Sciarretta, K.L., et al: Sepsis Among Medicare Beneficiaries: 1. The Burdens of Sepsis, 2012–2018*. Critical Care Medicine 48(3), 276–288 (2020). doi:10.1097/CCM.0000000000004224
3. Buchman, T.G., Simpson, S.Q., Sciarretta, K.L., et al: Sepsis Among Medicare Beneficiaries: 2. The Trajectories of Sepsis, 2012–2018*. Critical Care Medicine 48(3), 289–301 (2020). doi:10.1097/CCM.0000000000004226
4. Buchman, T.G., Simpson, S.Q., Sciarretta, K.L., et al: Sepsis Among Medicare Beneficiaries: 3. The Methods, Models, and Forecasts of Sepsis, 2012–2018*. Critical Care Medicine 48(3), 302–318 (2020). doi:10.1097/CCM.0000000000004225
5. Kumar, A., Roberts, D., Wood, K.E., et al: Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*:. Critical Care Medicine 34(6), 1589–1596 (2006). doi:10.1097/01.CCM.0000217961.75225.E9

6. Henry, K.E., Hager, D.N., Pronovost, P.J., et al: A targeted real-time early warning score (TREWScore) for septic shock. Science Translational Medicine 7(299), 299–122299122 (2015). doi:10.1126/scitranslmed.aab3719
7. Calvert, J.S., Price, D.A., Chettipally, U.K., et al: A computational approach to early sepsis detection. Computers in biology and medicine 74, 69–73 (2016) Fagerstro ̈m, J., B ̊ang, M., Wilhelms, D., Chew, M.S.: LiSep LSTM: A machine learning algorithm for early detection of septic shock. Scientific reports 9(1), 1–8 (2019)
8. Fagerstro ̈m, J., B ̊ang, M., Wilhelms, D., et al.: LiSep LSTM: A machine learning algorithm for early detection of septic shock. Scientific reports 9(1), 1–8 (2019)
9. Singer, M., Deutschman, C.S., Seymour, C.W., et al: The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 315(8), 801 (2016). doi:10.1001/jama.2016.0287
10. Morrill, J., Kormilitzin, A., Nevado-Holgado, et al: The signature-based model for early detection of sepsis from electronic health records in the intensive care unit. In: 2019 Computing in Cardiology Conference (CinC). IEEE (2019
11. Reyna, M., Josef, C., Jeter, R., et al: Early Prediction of Sepsis from Clinical Data – the PhysioNet Computing in Cardiology Challenge 2019. Critical Care Medicine 2019.
12. Hambly, B., Lyons, T.: Uniqueness for the signature of a path of bounded variation and the reduced path group. Annals of Mathematics 171(1), 109–167 (2010). doi:10.4007/annals.2010.171.10
13. Lyons, T.J.: Differential equations driven by rough signals. Revista Matem ́atica Iberoamericana 14(2), 215–310 (1998)
14. Chevyrev, I., Kormilitzin, A.: A Primer on the Signature Method in Machine Learning. arXiv:1603.03788 [cs, stat] (2016). 1603.03788
15. Chevyrev, I., Oberhauser, H.: Signature moments to characterize laws of stochastic processes. arXiv preprint arXiv:1810.10971 (2018)
16. Pfeffer, M., Seigal, A., Sturmfels, B.: Learning paths from signature tensors. SIAM Journal on Matrix Analysis and Applications 40(2), 394–416 (2019)
17. Kormilitzin, A., Saunders, K., Harrison, P., et al: Application of the signature method to pattern recognition in the cequel clinical trial. arXiv preprint arXiv:1606.02074 (2016)
18. Kormilitzin, A., Saunders, K.E., Harrison, P.J., et al: Detecting early signs of depressive and manic episodes in patients with bipolar disorder using the signature-based model. arXiv preprint arXiv:1708.01206 (2017)
19. Arribas, I.P., Saunders, K., Goodwin, G., et al: A signature-based machine learning model for bipolar disorder and borderline personality disorder. Translational Psychiatry 8 (2018)
20. Moore, P., Lyons, T., Gallacher, J., Initiative, A.D.N., et al.: Random forest prediction of Alzheimer's disease using pairwise selection from time series data. PloS one 14(2), 0211558 (2019)
21. Kormilitzin, A., Walsh, A., Matini, L., et al: P283 Patient-reported symptoms over a period of 14 days reliably predict endoscopic and histological disease activity in ulcerative colitis (UC). Journal of Crohn's and Colitis 14(Supplement 1), 293–294 (2020)
22. Arribas, I.P.: Derivatives pricing using signature payoffs. arXiv preprint arXiv:1809.09466 (2018)
23. Xie, Z., Sun, Z., Jin, L., et al: Learning spatial-semantic context with fully convolutional recurrent network for online handwritten Chinese text recognition. IEEE transactions on pattern analysis and machine intelligence 40(8), 1903–1917 (2017)
24. Chevyrev, I., Nanda, V., Oberhauser, H.: Persistence paths and signature features in topological data analysis. IEEE transactions on pattern analysis and machine intelligence, 1–1 (2018)
25. Bonnier, P., Kidger, P., Perez Arribas, I., et al.: Deep signatures. arXiv preprint arXiv:1905.08494 (2019)
26. Kidger, P., Lyons, T.: Signatory: Differentiable computations of the signature and logsignature transforms, on both CPU and GPU. arXiv preprint arXiv:2001.00706 (2020)
27. Ke, G., Meng, Q., Finley, T., et al: LightGBM: A Highly Efficient Gradient Boosting Decision Tree, 9

Figure Legends

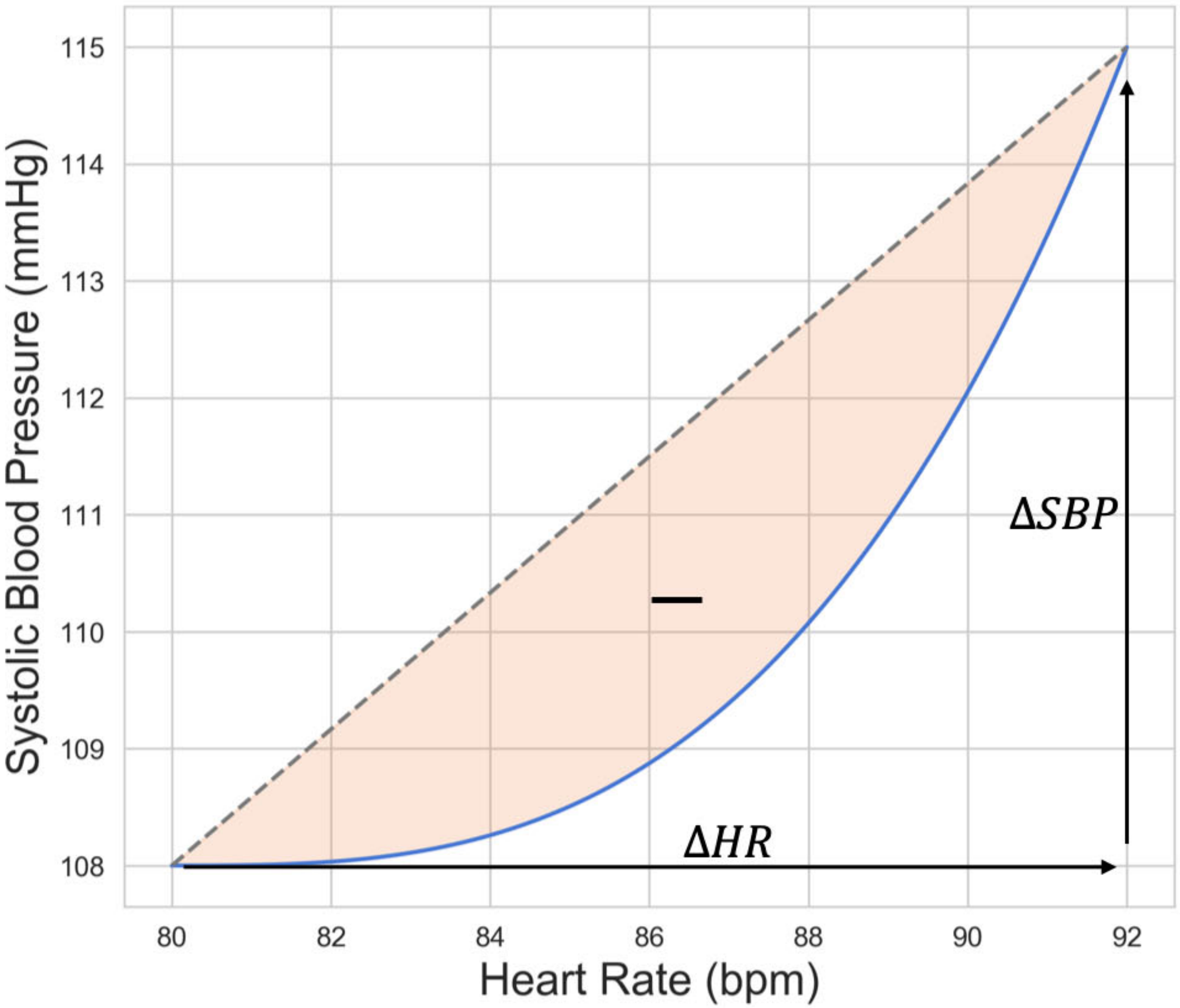
Figure 1: *Paths of heart rate and systolic blood pressure to visualize the signature terms up to order 2. The first order terms are the changes in the variables over the interval,* $(\Delta HR, \Delta SBP)$*, and the second order terms contain the signed values of the areas (shaded orange region) with the left plot having a negative sign and the right plot a positive sign.*

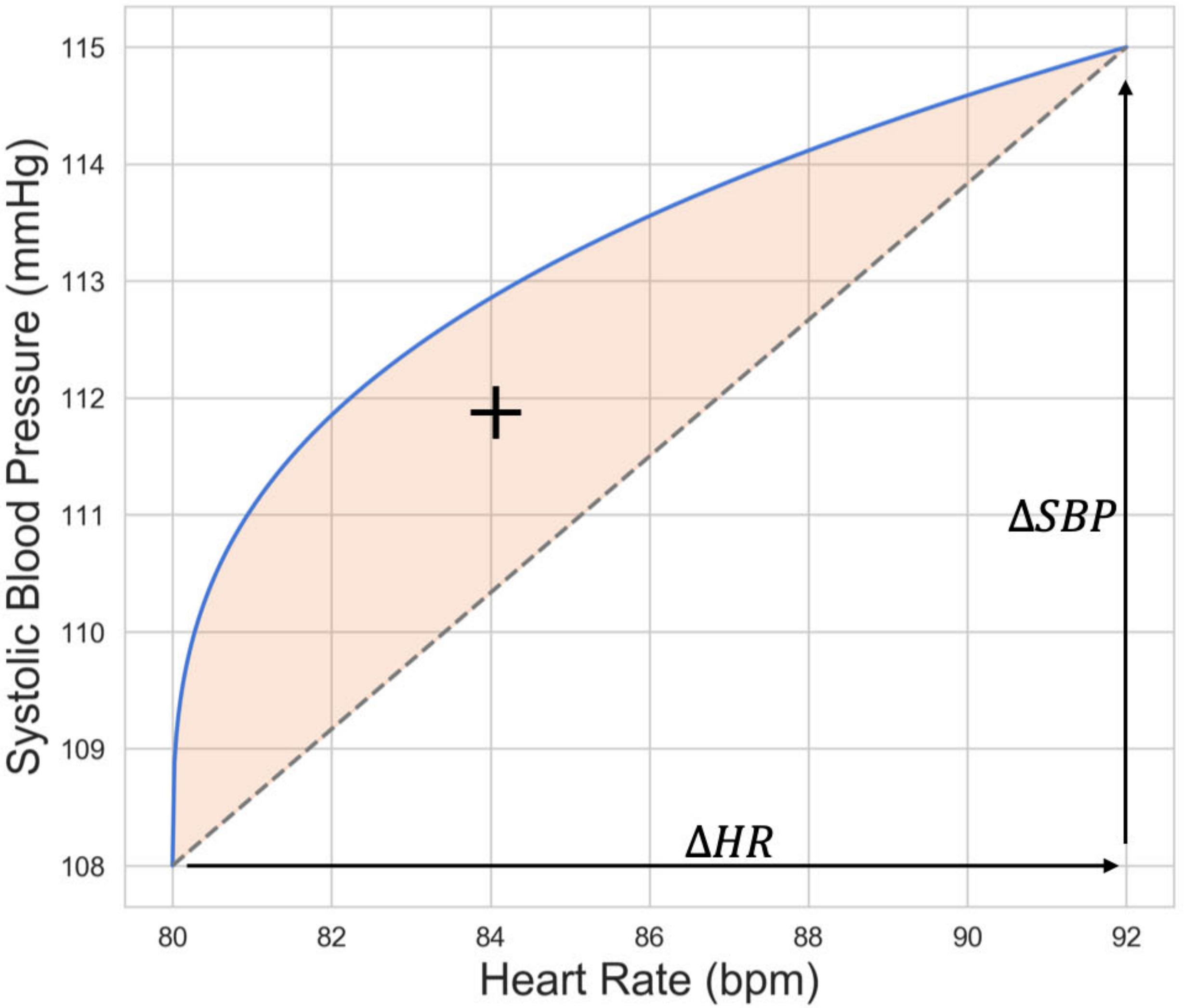
Figure 2: *Confusion matrix displaying the number of people predicted as likely to get sepsis compared with those who actually end up with sepsis with the threshold tuned to 33% specificity (left). Proportion of sepsis cases predicted correctly in different time windows tuned to different precision levels (right).*

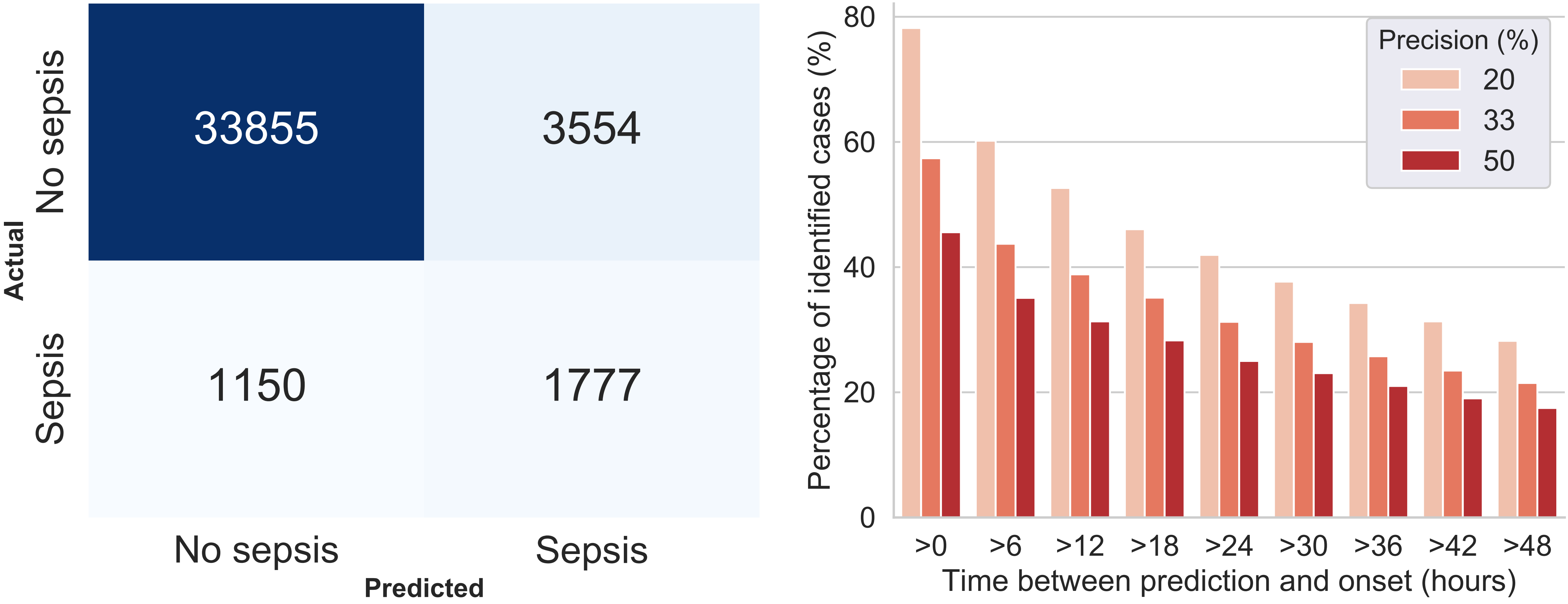
Figure 3: *The number of people who have not yet developed sepsis at each time, as a fraction of those who develop sepsis eventually, for each of the hospitals in the training set.*

Tables

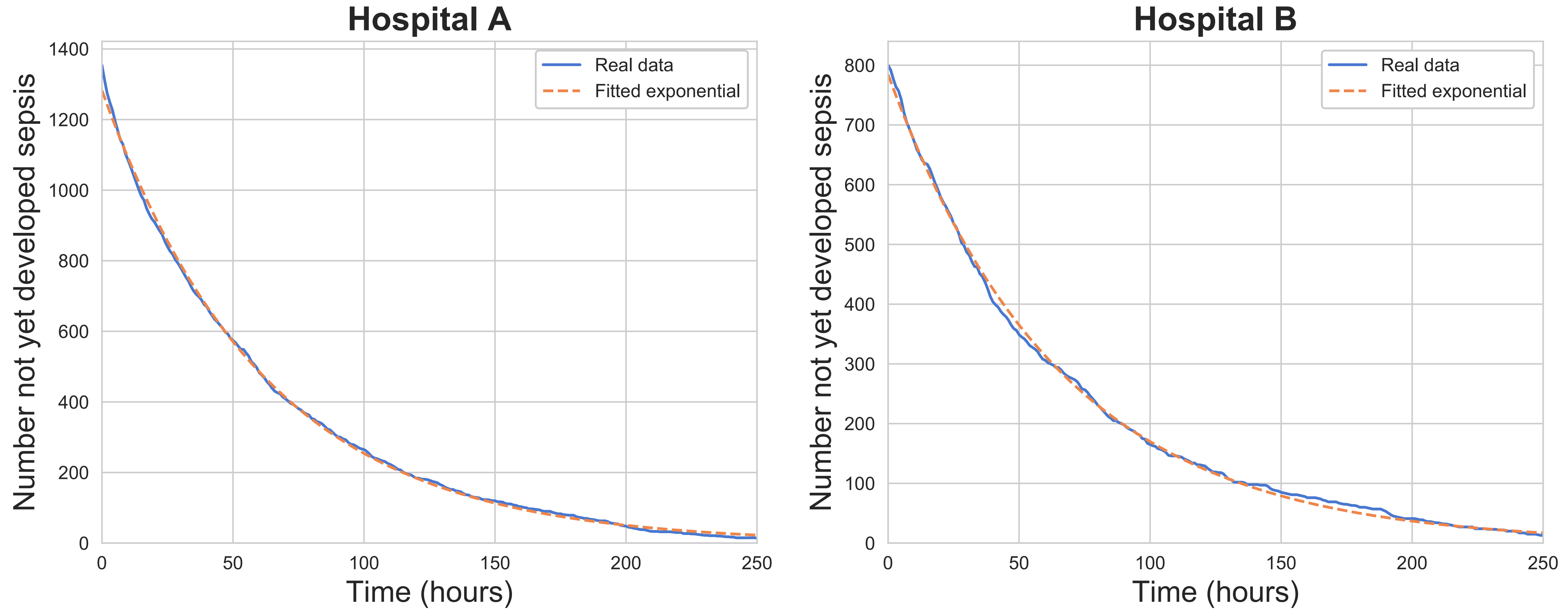
Provided as supplementary word files.


Systolic Blood Pressure (mmHg)
115
114
113
112
111
110
109
108
−
ΔSBP
ΔHR
80
82
84
86
88
90
92
Heart Rate (bpm)


Systolic Blood Pressure (mmHg)
115
114
113
112
111
110
109
108
+
ΔSBP
ΔHR
80
82
84
86
88
90
92
Heart Rate (bpm)



Actual \ Predicted	No sepsis	Sepsis
No sepsis	33855	3554
Sepsis	1150	1777


Hospital A
Hospital B
Number not yet developed sepsis
Time (hours)
Real data
Fitted exponential
0
200
400
600
800
1000
1200
1400
0
50
100
150
200
250
0
100
200
300
400
500
600
700
800

Hospital				
	A	B	C	**Average (std)**
Training	0.442	0.421	--	0.434 (0.016)
Test	0.433	0.434	-0.123	0.360

Table 1: *The utility scores achieved for each hospital from the public training set (top row) and the scores returned by PhysioNet on the hidden test data (bottom row). The scores on the public training set are taken to be the average utility over the 5-folds used in cross-validation.*

Features	**Averaged Utility Score**
Time only	0.282
+ Original 40 features	0.389
+ Non-Signature features	0.422
+ Signature features	0.434

Table 2: *Cross-validation results on the public training set from models trained on different subsets of features. Each row also contains the features listed in any rows above it.*