

Predicting in-hospital death from derived EHR trajectory features

Rajeev BOPCHE^a, Lise Tuset GUSTAD^b, Jan Egil AFSET^a, Jan Kristian DAMÅS^a
and Øystein NYTRØ^a

^a*Norwegian University of Science and Technology, Trondheim, Norway.*

^b*Nord University, Levanger, Norway*

ORCID ID: Rajeev Bopche <https://orcid.org/0000-0003-4835-9488>, Lise Tuset Gustad <https://orcid.org/0000-0003-2709-3991>, Jan Egil Afset <https://orcid.org/0000-0003-2489-4179>, Jna Kristian Damås <https://orcid.org/0000-0003-4268-671X>, Øystein Nytrø <https://orcid.org/0000-0002-8163-2362>.

Abstract. Medical histories of patients can provide insight into the immediate future of a patient. While most studies propose to predict survival from vital signs and hospital tests within one episode of care, we carry out selective feature engineering from longitudinal historical medical records in this study to develop a dataset with derived features. We then train multiple machine learning models for the binary prediction whether an episode of care will culminate in death among patients suspected of bloodstream infections. The machine learning classifier performance is evaluated and compared and the feature importance impacting the model output is explored. The findings indicated that the logistic regression model achieved the best performance for predicting death in the next hospital episode with an accuracy of 98% and an almost perfect area under the receiver operating characteristic curve. Exploring the feature importance reveals that time to and severity of the last episode and previous history of sepsis episodes were the most critical features.

Keywords. Electronic Health Records, Health Trajectory Analysis, Machine Learning, Bloodstream Infection

1. Introduction

Electronic Health Records (EHR) data finds secondary use for applications like disease progression modelling [1], patient trajectory modelling [2], disease inference [3], risk stratification, and survival prediction [4]. These practice data driven analyses are increasingly needed in all kinds of health services or research. The problem with EHRs is mostly the sparseness and context dependent interpretation, which appears as incompleteness and to a lesser extent inconsistency and inaccuracy [5]. A series of patient discharge summaries provide a longitudinal perspective of patients' interactions with a hospital service. In Norway, with a predominantly public specialist healthcare, this means that patients often have long and continuous histories within one hospital's records. Thus, we can do retrospective medical history analysis for patients with poor outcomes.

In this study instead of fitting the model to data, we attempted to fit the data to the model. Bypassing all the complexity of machine learning models that handle episode data about state, exposure and intervals, we selectively derived the most relevant count and temporal features and used them to train classifiers for predicting in-hospital mortality for the next episode. We then studied the features that led to improvement in

model performance as predictors of in-hospital deaths. Similar to this work on prediction models using longitudinal medical data Chicco et al. (2020) showed that traditional machine learning models applied to minimal clinical records of patients diagnosed with sepsis predicted the survival of the patients [4]. In other work on using medical history, studies mainly focused on visualizing the history or building patient disease trajectories [6]. Some studies work on tackling the problems with the representation of medical data and codes. For example, Tran et. al. (2015) worked on building a low-dimensional representation of medical events using a modified Restricted Boltzmann Machine (RBM). Thereafter they trained a logistic regression classifier for suicide risk stratification [7]. Similarly, some disease-specific applications have treated the medical history as a sequence of events and then trained machine learning (ML) models to predict an outcome [8]. While Jia et al. (2020) used patient similarity-based frameworks to group similar patient histories together [9]. The aim of this study was to investigate patterns of events leading to death in the hospital, and to find if these patterns in prior episodes can help predict an impending medical episode with risk of death.

2. Methods

2.1. Data and Data Mining

The data comprises individual episodes of care with St. Olav’s university hospital, not including primary care or visits to other specialist care, of 35,594 patients that had at least one episode of suspected bloodstream infection (BSI) at between 2015-2020. The episodes range from the introduction of EHR in 1999 until 2020. The suspected BSI was identified through a physician-ordered blood culture test. The mean age of the complete cohort is 63.6 years, and the gender distribution is 52.5% males to 47.4% females. The data contains information on a total of 1.2 million medical episodes. Diagnosis is coded using ICD-10 (International Classification of Diseases 10th Revision). Python’s pandas library was used to preprocess and clean the dataset. Patient histories with only one episode were removed, leaving the final cohort with 32,313 patients. Patients with death date more than 30 days from the last episode were also excluded. New features were calculated by aggregating time under each ICD-10 code group and counting the number of respective episodes. To capture the repetitiveness of episodes, the time between episodes was calculated and normalized by total number of episodes. Other features such as age at the time of first visit, total length of medical history and total number of hospital visits were also included.

Table 1. Description of derived features.

Features	Unit	Range	mean
Age	Years	[0...107]	60.076
Age at first visit	Years	[0...100]	52.578
Time to last episode	Hours	[0...70489]	3287.287
Type of last episode	Category	[1, 3, 5, 8]	2.996
Time between episodes	Hours	[0... 28399]	2356.617
Time Hospitalized for implicit sepsis	Hours	[0...12712]	65.771
Implicit sepsis episodes	Count	[0...27]	0.227
Time Hospitalized for cancer disease	Hours	[0...10587]	119.825
Number of cancer episodes	Count	[0...365]	9.467
Time Hospitalized for explicit sepsis	Hours	[0...17299]	64.723
Number of explicit sepsis episodes	Count	[0...14]	0.189
Time Hospitalized for cardiovascular disease	Hours	[0...8894]	178.455

Number of cardiovascular episodes	Count	[0..66]	4.826
Number of infection episodes	Count	[0..93]	2.730
Total time under acute urgency	Hours	[0..22874]	717.139
Number of hospital visits	Count	[2..680]	31.107
Total length of stay	Hours	[0..30873]	1105.631

2.2. Prediction Modelling

In this study we derive temporal features like exposure and repetitiveness not otherwise available to non-temporal or process-blind ML methods. Initially, around 50 derived features were calculated, and a Random Forest classifier was trained, and the importance of each feature was examined. Then finally, the 20 most important features were selected to form the final dataset. All features are continuous values. All empty cell values were imputed to zero. The values were then standardized by removing the mean and scaling to unit variance. The target feature was labelled as 0, if death within 30 days of final episodes and 1, if the patient was alive. There were 21,266 patients labelled 1 and 7,041 patients labelled 0 making the data imbalanced 3:1. The dataset was further divided into a training set (80%) and a testing set (20%).

For our binary classification problem, we used Logistic Regression (LR) as the linear model, Gaussian Naïve Bayes (GNB) as the probabilistic model, K-Nearest Neighbors (KNN) as the non-parametric model, Random Forest (RF), Bagging and Boosting decision tree classifiers (BG and ADB), Voting Classifier as the ensemble model, multi-layer perceptron (MLP) as neural network based model, and eXtreme Gradient Boosting (XGBoost). Additionally, we explored interpretable machine learning by using SHapley Additive exPlanations (SHAP) values to explain the output of a machine learning model.

3. Results

3.1. Model performance

This study compared eleven different machine learning models on our dataset. Performance metrics for each of these models is given in Table 1. The results indicate that all models achieved high accuracy score. The logistic regression model and the XGBoost model gave overall best performance metrics. The receiver operating characteristic (ROC) curve for all models is given in Figure 1a. The confusion matrices on testing data for the logistic regression model and the XGBoost model are given as Figure 1b.

Table 2. Prediction performance metrics

Model	Accuracy	Sensitivity	Specificity	F1 Score	AUROC
Logistic Regression	0.987	0.964	0.994	0.982	0.999
Naïve Bayes	0.822	0.578	0.904	0.752	0.881
K-Nearest Neighbors	0.901	0.762	0.949	0.865	0.932
Support Vector Machine	0.938	0.844	0.970	0.916	0.983
Decision Trees	0.897	0.781	0.935	0.862	0.859
Bagging Decision Trees	0.932	0.848	0.960	0.908	0.970
Boosting Decision Trees	0.934	0.848	0.963	0.912	0.974
Random Forest	0.915	0.791	0.957	0.885	0.969
Ensemble	0.945	0.861	0.973	0.926	0.986

Neural Network	0.940	0.857	0.968	0.919	0.981
XGBoost	0.980	0.953	0.989	0.973	0.997

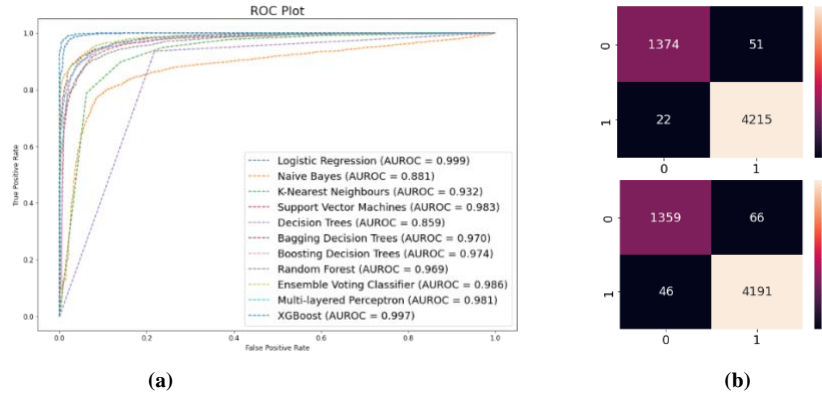


Figure 1. (a) ROC curve for all models (b) Confusion matrices logistic regression model (top right) and XGBoost model (bottom right)

3.2. Feature Importance

To study the features, SHAP values were estimated for each feature. SHAP values indicate the impact on model output. Figure 2a gives the feature importance as the impact on model prediction. In the figure Class 1 stands for alive patients and class 0 for patients who died. Age at the first visit and the information related to the last visit were the top features for the model shortly followed by time hospitalized for various disease episodes. Figure 2b summarizes the SHAP value by combining feature importance with feature effects. The plot depicts the features' overall influence on the model prediction. Each point represents an individual case. Data points shifted to the right indicates the features that contribute to high risk of death and vice versa.

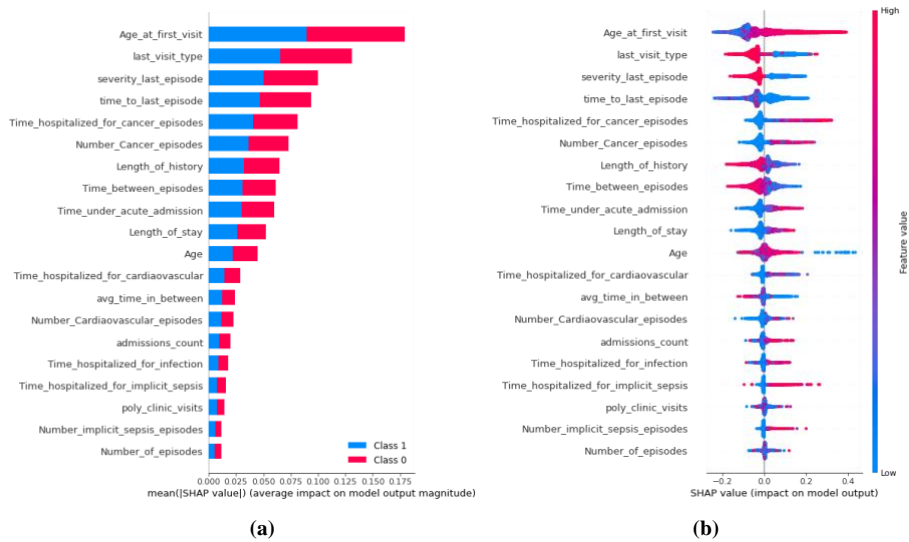


Figure 2. SHAP values for feature importance bar plot and dot plot.

4. Discussion

It was noticed during a thorough analysis of the patient histories that the last part of life in patients who died usually consisted of not one single episode but a series of episodes. On the other hand, such frequent episodes may be indicators of “swing door patients”, aged patients being discharged to primary care or home in a frail condition. This made our prediction model for in hospital mortality achieve high performance metrics as the derived features successfully capture these key indicators. The age at first visit being the most important feature indicated that patients arriving at the hospital for the first time very late in life are at a very high risk. Major limitations in this study are that alive patient’s histories were taken as complete instead of selecting only histories up to some critical episodes and the sensitivity to individual trajectory length and to rapid succession of episodes leading to a fatal episode. This can be avoided by normalizing patient histories by windowing, only using a fixed number of prior episodes or a fixed span of time.

5. Conclusions

We explored a new approach towards predicting in-hospital mortality based on minimal medical history data. We circumnavigated all the complexity of medical concept representation and derived indicators of severity, exposure and disease progression. We were able to predict if an impending disease episode entails risk of death. The derived features can be further well formulated with coordination and discussions with the hospital staff and stakeholders. These models can be easily implemented in the current and developing digital health platforms to predict adverse outcomes.

References

- [1] Mould DR. Models for disease progression: new approaches and uses. *Clin Pharmacol Ther.* 2012 Jul;92(1):125-31. doi: 10.1038/clpt.2012.53. Epub 2012 May 23. Erratum in: *Clin Pharmacol Ther.* 2012 Nov;92(5):666.
- [2] Ebadollahi S, Sun J, Gotz D, Hu J, Sow D, Neti C. Predicting Patient’s Trajectory of Physiological Data using Temporal Trends in Similar Patients: A System for Near-Term Prognostics. *AMIA Annu Symp Proc.* 2010 Nov 13;2010:192-6.
- [3] Zhao D, Weng C. Combining PubMed knowledge and EHR data to develop a weighted bayesian network for pancreatic cancer prediction. *J Biomed Inform.* 2011 Oct;44(5):859-68. doi: 10.1016/j.jbi.2011.05.004. Epub 2011 May 27.
- [4] Chicco, D., Jurman, G. Survival prediction of patients with sepsis from age, sex, and septic episode number alone. *Sci Rep* 10, 17156 (2020). <https://doi.org/10.1038/s41598-020-73558-3>.
- [5] Botsis T, Hartvigsen G, Chen F, Weng C. Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. *Summit Transl Bioinform.* 2010 Mar 1;2010:1-5.
- [6] Beck MK, Jensen AB, Nielsen AB, Perner A, Moseley PL, Brunak S. Diagnosis trajectories of prior multi-morbidity predict sepsis mortality. *Sci Rep.* 2016 Nov 4;6:36624. doi: 10.1038/srep36624.
- [7] Tran T, Nguyen TD, Phung D, Venkatesh S. Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBM). *J Biomed Inform.* 2015 Apr;54:96-105. doi: 10.1016/j.jbi.2015.01.012. Epub 2015 Feb 3.
- [8] Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F., & Sun, J. (2015). Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. *arXiv*. <https://doi.org/10.48550/arXiv.1511.05942>
- [9] Zheng Jia, Xian Zeng, Huilong Duan, Xudong Lu, Haomin Li, A patient-similarity-based model for diagnostic prediction, *International Journal of Medical Informatics*, Volume 135, 2020, 104073, ISSN 1386-5056, <https://doi.org/10.1016/j.ijmedinf.2019.104073>.