Data, Informatics and Technology: An Inspiration for Improved Healthcare
A. Hasman et al. (Eds.)
IOS Press, 2018

doi:10.3233/978-1-61499-880-8-97

On the Representation of Machine Learning Results for Delirium Prediction in a Hospital Information System in Routine Care

Sai VEERANKI[a,1], Dieter HAYN[a], Alphons EGGERTH[a], Stefanie JAUK[b], Diether KRAMER[c], Werner LEODOLTER[c] and Günter SCHREIER[a]
[a] *AIT Austrian Institute of Technology, Graz, Austria*
[b] *CBmed, Graz, Austria*
[c] *Steiermärkische Krankenanstaltengesellschaft m.b.H. (KAGes), Graz, Austria*

Abstract. Digitalisation of health care for the purpose of medical documentation lead to huge amounts of data, hence having an opportunity to derive knowledge and associations of different attributes recorded. Many health care events can be prevented when identified. Machine learning algorithms could identify such events but there is ambiguity in understanding the suggestions especially in clinical setup. In this paper we are presenting how we explain the decision based on random forest to health care professionals in the course of the project predicting delirium during hospitalisation on the day of admission.

Keywords. Electronic health records, machine learning, delirium, important features

1. Introduction

1.1. Background

Delirium is an acute neuropsychiatric syndrome which is common in elderly patients. It is generally under diagnosed and 40% of the cases can be prevented when identified earlier [1]. Therefore, early prediction of Delirium is a promising candidate for machine learning (ML) approaches. There are high expectations on the application of ML methods in various areas [2]. However, interpretation of results obtained from the ML algorithms is often difficult for health care professionals. Hence there is a necessity for a tool that shows why a certain prediction has been obtained.

1.2. State-of-the-art and motivation

Several attempts have been made to predict delirium using clinical assessment methods and risk factors during hospitalisation [3]. They have mainly used regression models

[1] Corresponding Author, Sai Pavan Kumar Veeranki, AIT Austrian Institute of Technology, Reininghausstraße 13, 8020, Graz, Austria; E-Mail: sai.veeranki@ait.ac.at.

with rather small patient groups and obtained an Area Under Receiver Operating Curve (AUROC) ranging between 70 % and 85 % [3]. In a previous publication, we have received an AUROC of approximately 90 % in a retrospective analysis of data from more than 8,500 patients [4] of “Steiermärkische Krankenanstaltengesellschaft m.b.H”(KAGes). KAGes as Styrian healthcare provider with about 90% market share in terms of acute care hospital beds can use more than 1 million longitudinal health records as basis for its analyses in its premises. One strategic initiative of KAGes Management was to prevent delir as good as possible and to provide its healthcare teams with tools for diagnosis and prevention. Lead by KAGes several research partners contribute to this project with the goal to provide clinical decision support for everyday work.

A systematic statistical analysis has been applied on data to exclude irrelevant attributes. In remaining features, some attributes were represented as binary values and others were derived from attributes such as number of diseases, Charlson comorbidity index [5], number of procedures etc., are represented as numerical values.

By applying RandomForest (RF) with one thousand trees and a10-fold cross validation, we achieved an AUROC of 89.8%. Among all ML approaches at hand, RF has performed the best. The results were presented in one of our papers [4]. In [6] and [7] authors proposed how to explain individual decisions obtained from different classifiers. There have been methods for visualising feature importance of a RF using tree structures, bar charts, graph networks, etc., to explain the predictions [8]. We have published our approach to derive feature importance for individual decisions from RF [9]. However, none of these approaches were able to reach sufficient acceptance when presented to the health care professionals.

1.3. Objective

The objective of the paper was to develop a tool to visualize results from a ML system that predicts delirium in hospitalised patients for health care professionals to check the plausibility and also to understand ML results and thus support their clinical reasoning.

2. Material, Technologies and Methods

2.1. Data

KAGes routinely records data about their patients related to diseases, procedures and laboratory measurements based on international standards, such as ICD-10, ICPM/ICHI and LOINC, respectively. Inclusion and exclusion criteria for extracting the cohort and control group are described in [4]. We derived a cohort of 4,596 delirium patients and randomly selected 25,000 patients for a control group.

2.2. Methods

2.2.1. Technologies

Data were retrieved from the hospital information system (HIS) through a data warehouse platform (HANA, SAP, Walldorf, Germany). R software (R Foundation for Statistical Computing, Vienna, Austria) was used for data pre-processing. the caret package has been used for modelling and online deployment and the shiny package [10] for presenting decisions to health care professionals. Additionally, further investigations and simulations were done using AIT's Predictive Analytics Toolset for Health and Care (PATH) based on Matlab (The MathWorks, Nattick, USA) [11].

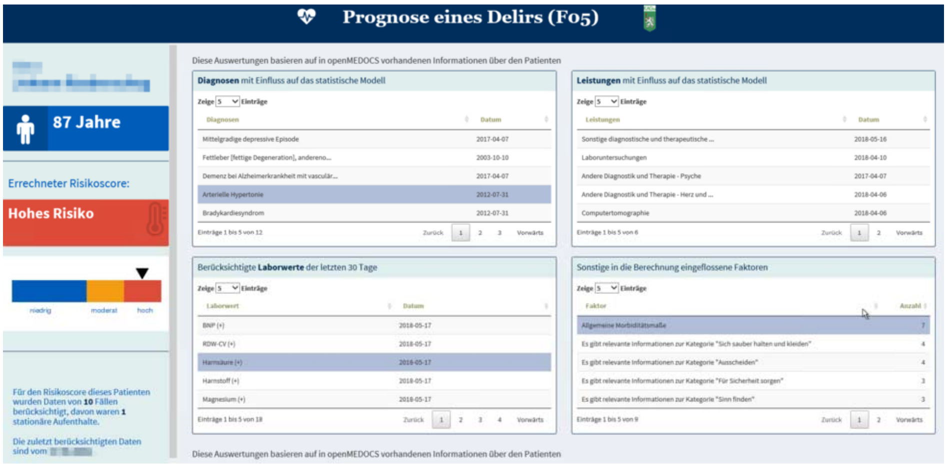

Figure 1. Presentation of delirium probability to health care professionals, along with features that influenced the particular decision

2.2.2. Feature selection requirements

The health care professionals need to understand the results derived from ML algorithms to conceive a decision. We have had discussions with health care professionals in focus groups as well as individual debates. The focus groups included doctors, management, nursing, IT staff along with academicians and statisticians.

3. Results

Probability of occurrence of delirium is derived for each patient from the model. For each feature, the importance was calculated from RF based on Out-of-Bag permuted predictor importance. We have selected features that were relevant for individual patient and sorted them based on their importance and classified into four different categories: diseases, procedures, laboratory results and others. Others included nursing assessment, transfers and demographical data. Diseases which were stated as high risk factors in the literature were given higher importance (E.g. Dementia). Each feature is represented with the description and date recorded. A screenshot of the visualization tool is provided in Figure 1. Thus, we are able to provide a context sensitive view on the longitudinal electronic health record. At any time, it is possible for the user to navigate to the full EHR of the patient in order to evaluate the situation in more detail.

4. Discussion

Our tool provides the features along with the associated information and date of that feature recorded for individual patient. This representation gives the personnel a quick glance of a patient's health history and complications.

Initial feedback for this representation was encouraging as the concept was accepted by the health care professionals. We are continuously updating the tool according to new health care professional's requirements (e.g., related diagnoses). Although the tool has been accepted for its usability, for some patients with few number of previous hospital admissions or none, the features that are presented may not be important for delirium. Presently, we work on a solution by analysing the results obtained for such cases.

5. Conclusion and Outlook

Our solution based on listing the most important features which are relevant for an individual patient and grouping the features in four categories is a promising approach, with the potential to increase acceptance of ML solutions significantly at hospital care, not just in the delirium case but in additional real-world applications as well.

Acknowledgements

This work has been carried out with the K1 COMET Competence Centre CBmed, which is funded by the Federal Ministry of Transport, Innovation and Technology (BMVIT); the Federal Ministry of Science, Research and Economy (BMWFW); Land Steiermark (Department 12, Business and Innovation); the Styrian Business Promotion Agency (SFG); and the Vienna Business Agency. The COMET program is executed by the FFG. KAGes and SAP provided significant resources, manpower and data as basis for research and innovation.

References

[1] S. Wass, et al., Delirium in the elderly: a review, *Oman Med J* **23**(3) (2018), 150-7.
[2] I.A. Scott, Machine Learning and Evidence-Based Medicine, *Ann Intern Med (*2018).
[3] M.W. Newman, L.C. O'Dwyer, L. Rosenthal, Predicting delirium: a review of risk-stratification models, *Gen Hosp Psychiatry* **37**(5) (2015), 408-13.
[4] D. Kramer *et al.*, Development and Validation of a Multivariable Prediction Model for the Occurrence of Delirium in Hospitalized Gerontopsychiatry and Internal Medicine Patients, *Health informatics meets eHealth*, Vienna, 2017.
[5] M. Charlson, T.P. Szatrowski, J. Peterson, J. Gold, Validation of a combined comorbidity index, *J Clin Epidemiol* **47**(11) (1994), 1245-51.
[6] D. Baehrens, S. Harmeling, M. Kawanabe, K. Hansen, K.R. Müller, How to explain individual classification decisions, *Journal of Machine Learning Research* **11** (2010), 1803-1831.
[7] M.T. Ribeiro, et al., Why Should I Trust You? Explaining the predictions of any Classifier.
[8] F.G. Hudson, M.A.G. Cristiano, Visualizing Random Forest's Prediction Results, *Psychology* **5** (2014), 2084-98.
[9] D. Hayn, H. Walch, J. Stieg, K. Kreiner, H. Ebner, G. Schreier, Plausibility of Individual Decisions from Random Forests in Clinical Predictive Modelling Applications, *Stud Health Technol Inform* **236** (2017), 328-335.
[10] C. Winston, et al., Shiny: Web application Framework for R, 1.0.5 ed, 2017.
[11] D. Hayn et al., Predictive Analytics for Data Driven Decision Support in Health and Care, *IT Information Technology* **57** (2018).