Local Explanation-Based Method for Healthcare Risk Stratification

Jean-Baptiste EXCOFFIER[a,1], Elodie ESCRIVA[a,b], Julien ALIGON[b] and Matthieu
ORTALA[a]

[a] *Kaduceo, Toulouse, France*
[b] *Université de Toulouse-Capitole, IRIT, (CNRS/UMR 5505), Toulouse, France*

Abstract. Decision support tools in healthcare require a strong confidence in the developed Machine Learning (ML) models both in terms of performances and in their ability to provide users a deeper understanding of the underlying situation. This study presents a novel method to construct a risk stratification based on ML and local explanations. An open-source dataset was used to demonstrate the efficiency of this method that well identified the main subgroups of patients. Therefore, this method could help practitioners adjust and build protocols to improve care deliveries that would better reflect patient's risk level and profile.

Keywords. Machine Learning, Explainability, Clustering, Risk Stratification

1. Risk Stratification Method

Clinical decision support tools based on predictive Machine Learning (ML) modeling often provide the prediction alone, raising concerns about their reliability and acceptance. Several methods have thus been perfected to provide local explanations, also called influences, for a individual prediction. Moreover, a recent trend in ML is that predictive tools should not only provide local explanations but also enable the user to contextualize the observation [1]. This need is particularly high in healthcare, where physicians need to link a patient to a more global context so as to deliver the most relevant care. Care management through risk stratification (RS) is thus commonly used in healthcare [2]. It consists in the identification of subgroups, where patients in the same group have similar condition and risk level. Nevertheless, no method has yet been perfected that would take into account both the predictive power of ML models and the identification of such homogeneous subgroups. This study proposes a novel RS procedure based on ML modeling, local explainability methods and clustering that would provide physicians a clearer view of the different profiles of patients, also called subphenotypes or typologies. This method is divided into the three following steps.

Step 1: *Modeling of the underlying dataset using a ML predictive model*, that provides for each patient a risk level, for example a probability of having a disease.

Step 2: *Identification of personal protective and risk factors using a local explainability technique* that quantifies for each patient the influence of each feature.

Step 3: *Identification of subgroups of patients* with respect to their condition and risk level, using a clustering technique on the local explanations computed in the

---

previous step. Then a instance selection method can be applied in order to provide clear view of characteristics and specificity of each subgroups.

2. Use-Case

The open-source *Acute Inflamations* dataset was used [3]. It contains information about 120 patients and whether they have an *Acute Inflammation of Urinary Bladder* (AIUB). The techniques used for each step are the following.

Step 1: An ensemble of boosted trees based on the XGBoost implementation was used as the ML predictive risk model.
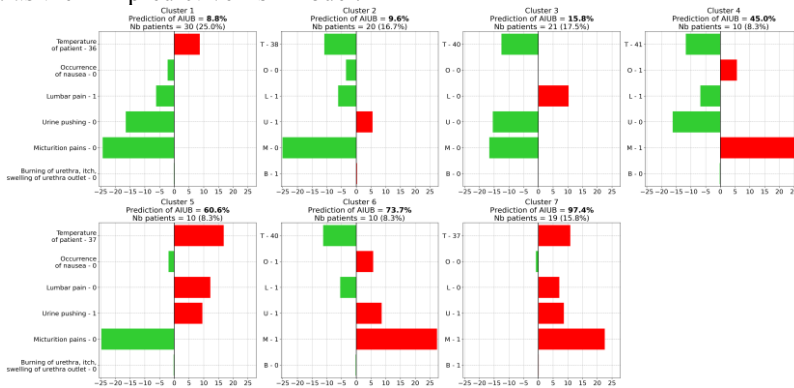
Figure 1. Influences of patients corresponding to the medoid of each identified cluster. Feature names are represented by their initials for subgraph not located on the far-left. Initial values of features are indicated after the hyphen. A positive influence (represented in red) increased the AIUB risk

Step 2: The SHapley Additive exPlanation (SHAP) technique was applied to compute local explanations for each patient [4].

Step 3: The K-Medoid algorithm was used to perform the clustering task on the local explanation values. The optimal number of groups was selected with the Silhouette coefficient. The K-medoid also had the advantage of being both a clustering and instance selection technique, since it can provide for each identified subgroup its most representative patient (*i.e.* medoid), as shown in Figure 1.

Identified subgroups of patients were coherent with medical indications about AIUB given in the original paper [13]. Therefore, it indicates that the novel proposed method is a strong option to consider when the final objective of the data analysis is the construction of a RS, for the purpose of guiding physicians to adjust or create medical protocols that best answer the specific needs of each patient.

References

[1]   Lundberg S, Erion G, Chen H, DeGrave A, Prutkin J, Nair B, et al.. From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell., 2, 56–67; 2020.

[2]   Dera JD. Risk stratification: A two-step process for identifying your sickest patients. Family practice management. 2019;26(3):21-6.

[3]   Czerniak J, Zarzycki H. Application of rough sets in the presumptive diagnosis of urinary system diseases. In: Artificial intelligence and security in computing systems. Springer; 2003. p. 41-51.

[4]   Lundberg S, Lee S. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 2017, vol. 30.