

Interpretable Clustering for Patient Phenotyping using Advanced Machine Learning Models

Roy S. Zawadzki, Saman Parvaneh

Background: Heterogeneity in the characteristics of transcatheter aortic valve replacement (TAVR) recipients motivates the need to develop clinically meaningful patient phenotypes. In Zawadzki, Johnson, and Parvaneh (2022) (ZJP2022), we developed an interpretable clustering framework using decision trees after k-means to profile each cluster. Yet, while decision trees are more interpretable, they may be inferior compared to more sophisticated algorithms that can better capture complex, non-linear relationships between covariates and phenotypes. In this work, we close the gap between model complexity and explainability by using SHapley Additive exPlanations (SHAP) values to interpret patient phenotypes. SHAP values provide a model-free measure of how much each predictor influences the likelihood of cluster membership, allowing the use for a wider variety of models to interpret clusters.

Methods: We utilize open-source data from a single-center TAVR study in Germany (n=581). Data extraction and cleaning are detailed in ZJP2022, where we found six distinct clusters using k-means on continuous demographic, medical history, and pre-operational variables. We tested several models and selected the model that best classified cluster membership. Then, SHAP values were computed and used to interpret the meaning behind each cluster.

Results: The best algorithm was the light gradient boosting machine (LGBM), with an F1-score of 0.871, while the decision tree had an F1-score of 0.700. The SHAP values (Figure 1) show feature importances and which cluster a feature had the most impact on delineating. Annulus area was the most important feature, particularly for predicting clusters 1 and 2. We previously found that cluster 5 had a higher rate of myocardial injury post-procedure, and SHAP values reveal that high pre-procedure creatinine level was impactful in delineating this group.

Discussion: By extending our previous framework to incorporate SHAP values, we allow analysts to incorporate a wide range of ML models to identify interpretable phenotypes.

