Patient Phenotyping using Interpretable Clustering to Study Clinical Outcome in TAVR Patients

Roy S. Zawadzki, Terri Johnson, Saman Parvaneh

Background: Discovering and understanding subpopulations (i.e., phenotypes) of Transcatheter Aortic Valve Replacement (TAVR) recipients and their differing clinical outcomes (e.g., adverse events) is valuable for delivering safe and effective care. Due to physiological complexity, vulnerable subpopulations are more likely to be identified by clustering across many patient characteristics rather than univariately (e.g., only age). This study explores an interpretable clustering framework that allows a clinical team to understand the clustering rules generated by a decision tree and investigate patient characteristics across different clusters. Furthermore, the team can study the associations between phenotypes and clinical outcomes via post-clustering statistical tests.

Method: We utilized public data from a single-center study on TAVR in Germany (n= 581). We clustered the data on continuous demographic, medical history, and pre-operational variables using K-means clustering. In the preprocessing, pairwise correlations were computed, and if there was a pairwise correlation over 0.6, the more clinically meaningful variable was kept. Missing values for the selected columns were imputed using the MissForest algorithm. Then, the number of clusters was selected using the Silhouette score. Next, we examined differences between clusters using a one-way analysis of variance (ANOVA). Furthermore, we fit a decision tree to predict cluster membership using the clustering variables, extracting and visualizing the prediction rules for each cluster for interpretation. Lastly, the clusters were analyzed for differences in rates of adverse events using a chi-squared test of independence.

Results: Six clusters were found using silhouette score. Cluster 2 had a statistically significant higher incidence of myocardial injury during operation. Decision tree rules for cluster 2 were extracted for interpreting this cluster (Figure 1).

Discussion: Our proof-of-concept analysis highlights the potential of interpretable clustering to understand patient phenotypes. This methodology can be used to find clinically meaningful associations between phenotypes and adverse events.
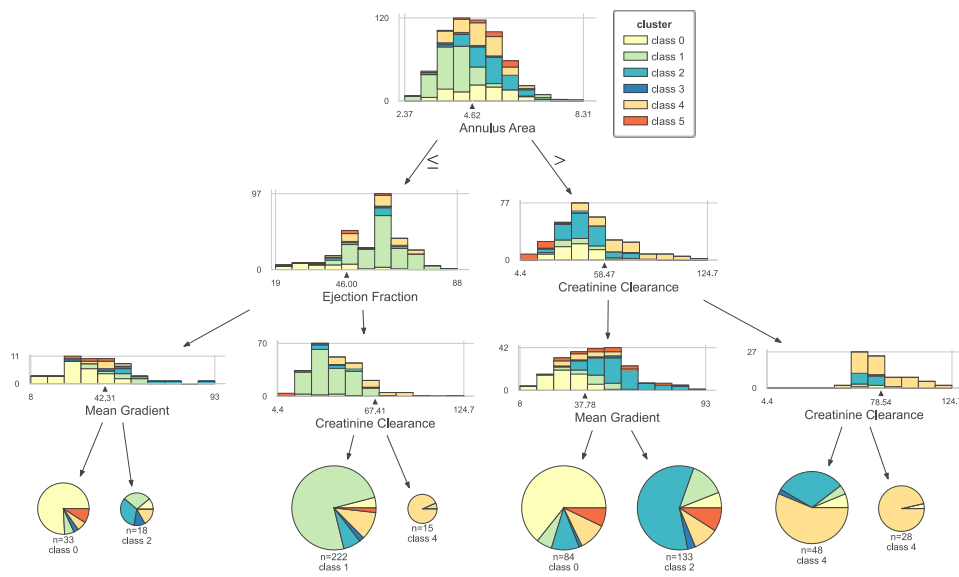

Figure 1: Decision Tree Rules for Cluster 2