An Embedding Approach for Biomarker Identification in Hypertrophic Cardiomyopathy

Arash Kazemi-Díaz[1]*, Luis Bote-Curiel, María Sabater-Molina, Juan Ramón Gimeno-Blanes, Salvador Sala-Pla, Francisco Javier Gimeno-Blanes, Sergio Muñoz-Romero, José Luis Rojo-Álvarez

[1]Teoría de la Señal y Comunicaciones y Sistemas Telemáticos y Computación. Universidad Rey Juan Carlos,
Madrid, Spain.

Introduction: Hypertrophic Cardiomyopathy (HCM) consists of a thickening of the cardiac muscle, causing fatigue, changes in the cardioelectric system, arrhythmias, and even sudden deaths. Variants in gene MYBPC3 are a well-known cause of this illness. Our objective was to find variants in other genes that can cause this pathology.

Experiments and results: For that purpose, genetic data from a group of patients (affected and not affected) were analyzed using Machine Learning techniques. More precisely, we propose embedding methods that allow a lower dimensional representation, which is very helpful for visualization, diagnosis, and therapy personalization. Our results, applying different methods. The example of PCA is shown in Figure 1. This visualization allowed us to identify 10 variants that affected 11 different genes that cause that separability. Once the causes of the separability were identified, we applied again the same methods in order to check the new data distribution. The separability in the new space was measured applying a machine learning classifier (Support Vector Machines) and checking how good it fitted to the data. In the case of PCA it gave a 0.64 of accuracy, meaning that the separability was low.

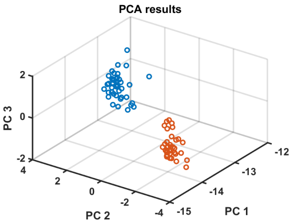

Figure 1. Results of the implementation of PCA to the matrix with all the variants. Blue points represent HCM patients and orange points represent controls.

Conclusion: This study explored the differences between controls and HCM patients embedding the original data onto lower-dimensional latent spaces. Thanks to that, we were able to identify 10 variants that where potential causes of the disease. Although this information is not conclusive enough to determine whether those variants are a cause of HCM or not, it may help clinicians in the task of identifying new HCM cases.