Disaentangling the heterogeneity of dilated cardiomyopathy: use of unsupervised clustering for mixed-data type in the search of novel clinical sub-phenotypes

Gandin Ilaria [1], Baj Giovanni[2], Paldino Alessia[1,3], Zaffalon Denise[3], Merlo Marco[1,3], Dal Ferro Matteo[3], Barbati Giulia[1]

1 Department of Medical Sciences, University of Trieste, Trieste, Italy
2 Department of Mathematics and Geosciences, University of Trieste, Italy
3 Cardiovascular Department, Azienda Sanitaria Universitaria Giuliano Isontina, Italy

Introduction

Dilated cardiomyopathy (DCM) is a condition affecting the heart muscle that is associate with approximately 20% of 5-year mortality. Despite this, there is currently a lack of accurate prognostic risk models for DCM. The disease has a significant genetic component, with around 30% of cases associated with pathogenic mutations. However, there is also considerable variability and heterogeneity in the phenotypic expression and progression [1]. In particular, DCM and arrhythmogenic right ventricular cardiomyopathy (ARVC), another type of cardiomyopathy, frequently present overlapping features that challenge the clinical conventional classification.

Aims

The aim of the present study is to identify novel DCM sub-phenotypes based on the information collected at the first cardiological visit using unsupervised clustering methods.

Methods

The study included detailed information of a longitudinal cohort of 409 DCM patients with a median follow up of 100 months (IQR [51, 185]). The variables under analysis ranged from clinical characteristics, to ECG parameters, Holter and imaging results. The analysis presented two major challenges: the large number of variables (p=102) and their mixed-data type. An innovative two-step approach was applied: 1) a recent extension of principal component analysis for mixed data [2] was employed to perform dimensionality reduction on the entire dataset; 2) agglomerative hierarchical clustering was applied to the first 11 principal components, after the identification of k=2 as optimal number of clusters using the average silhouette width criteria.

Results

Using a post-hoc cluster representation technique, we obtained a clinical characterization for the smaller group (Gr2, n=75) that appeared quite homogeneous. Gr2 included mostly thickened heart muscles with hypertrophic features and high prevalence of left bundle branch block. Moreover, using a multivariate cause-specific Cox model, incorporating 9 well-known risk factors, we demonstrate that Gr2 is at lower risk for life-threatening arrhythmic events (HR=0.21, 95% CI [0.08,0.54]). Coherent results were obtained considering genetic information: Gr2 showed a lower yield of causative mutations compared to Gr1 (15% vs 47%, p<0.001) and none of those pathogenic variants were found in what are considered arrythmogenic DCM-genes (DSP, PKP2, LMNA). These findings are concordant with a recent work [3] in which an association between the presence of causative mutations and the absence of left bundle branch block was identified.

Conclusions

Using a combination of dimensionality reduction and unsupervised clustering approach suitable for mixed-type data, two novel DCM subphenotypes were identified and characterized. The two groups differ in

terms of progression of the disease, as well as in the genetic etiology. An independent and external validation cohort has been recently identified, and there are plans to conduct a validation process of the obtained results in the near future. This findings could be relevnt in the prospective of a refining in the conventional phenotype-based classification of cardiomyopathies.

Bibliography:

[1] A Paldino, M Dal Ferro et al. Prognostic Prediction of Genotype vs Phenotype in Genetic Cardiomyopathies. J Am Coll Cardiol. 2022 Nov, 80 (21) 1981–1994

[2] Marie Chavent, Vanessa Kuentz-Simonet, Amaury Labenne et al. Multivariate Analysis of Mixed Data: The R Package PCAmixdata. PREPRINT (v5) available at arXiv: 1411.4911

[3] L Escobar-Lopez, JP Ochia, A Royuela et al. Clinical Risk Score to Predict Pathogenic Genotypes in Patients With Dilated Cardiomyopathy. J Am Coll Cardiol. 2022 Sep 20,80(12) 1115-1126