Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220408

Phenotyping of Heart Failure with Preserved Ejection Faction Using Health Electronic Records and Echocardiography

Morgane PIERRE-JEAN[a,b,1], Erwan DONAL[a,b], Marc CUGGIA[a,b] and Guillaume BOUZILLÉ[a,b]
[a] *CHU de Rennes, Cardiologie, France*
[b] *LTSI, Inserm, 1099, Rennes, France*

Abstract. Patients suffering from heart failure (HF) symptoms and a normal left ventricular ejection fraction (LVEF 50%) present very different clinical phenotypes that could influence their survival. This study aims to identify phenotypes of this type of HF by using the medical information database from Rennes University Hospital Center. We present a preliminary work, where we explore the use of clinical variables from health electronic records (HER) in addition to echocardiography to identify several phenotypes of patients suffering from heart failure with preserved ejection fraction. The proposed methodology identifies 4 clusters with various characteristics (both clinical and echocardiographic) that are linked to survival (death, surgery, hospitalization). In the future, this work could be deployed as a tool for the physician to assess risks and contribute to support better care for patients.

Keywords. machine learning, phenotyping, Heart failure, Health electronic records, echocardiography

1. Introduction

The number of patients who suffer from heart failure (HF) symptoms and a normal left ventricular ejection faction (LVEF $\geq$50%) is growing up [1,2]. This pathology is referred to HF with preserved ejection faction (HFpEF). However, it represents an heterogeneous syndrome with very different clinical phenotypes [1,3]. This study aims to use both the clinical variables available in health electronic records and echocardiographic parameters to classify patients suffering from HFpEF into groups who likely share similar physiological profiles. Here, we proposed an algorithm to classify the patients into phenogroups. Finally, we attempted to links these phenogroups to their follow-up. We study three survival outputs: death, surgery and admission in cardiology. Heading

[1] Corresponding Author, Morgane Pierre-Jean; E-mail: morgane.pierre-jean@chu-rennes.fr.

2. Methods

The number of patients who suffer from heart failure (HF) symptoms and a normal left ventricular ejection faction (LVEF ≥50%) is growing up [1,2]. This pathology is referred to HF with preserved ejection faction (HFpEF). However, it represents an heterogeneous syndrome with very different clinical phenotypes [1,3]. This study aims to use both the clinical variables available in health electronic records and echocardiographic parameters to classify patients suffering from HFpEF into groups who likely share similar physiological profiles. Here, we proposed an algorithm to classify the patients into phenogroups. Finally, we attempted to links these phenogroups to their follow-up. We study three survival outputs: death, surgery and admission in cardiology.

We performed a two-step cluster analysis to identify common characteristics among patients. As a first step, we performed Principal Component Analysis (PCA) following by a spectral clustering to the 10 first coordinates of the PCA. The follow-up information about the vital status of the patients was obtain from both data available from the data warehouse and from the National Institute of Statistics and Economic Studies. We matched firstname, surname, date of birth and town of birth when it was available. We collected stays into cardiology department from the data warehouse and CIM-10 codes were used to extract the surgery information. We conducted a survival analysis on death, surgery, and the admission to the cardiology department.

The final aim is to classify new patients, then we used supervised machine learning algorithms to predict the phenogroup that are defined by the clustering algorithm. We optimized two algorithms: SVM and Random forest. We split the cohort into two sets (train and test), the algorithms are trained on the first set and we evaluate there performance on the second one. Performance was evaluated by computing AUC ROC and accuracy.

3. Results

The final sample eligible included around 2500 patients with echocardiography and suffering from HFpEF. The variables included in the two-step cluster analysis were both clinical (13 variables) and echographic (17 variables). Four clusters were identified from the clustering algorithm. The groups were relatively well balanced with respectively: 753, 744, 519 and 545 patients.

From these clusters, we extract significant distinct variables between clusters. After performing survival analysis, we observed that clusters have significative different survival curves in particular for death.

Optimization of machine learning models has been conducted on training sets (75%) in addition to bootstrap. Then, we evaluated the models on the test sets. The performances of these two models are satisfying with an AUC upper than 0.97 and an accuracy upper than 0.92. We retain the SVM model that provides better performance on the two criteria to classify patients.

4. Discussion

The clinical variable extraction from health records is not perfect because of automatic extraction, and some variables are of poor quality or missing for a non-negligible number of patients. Even if, the cardiologists help us to perform quality control of data, we can improve this part by using NLP techniques.

Imputation of missing values was done using KNN- algorithm. Our algorithm could be tested with the dataset imputed by the mean of the median of the variables for examples. Then, we could determine which imputation method gives the better results.

The final step could be to build a score to predict the phenotype of patients or even their survival (death, hospitalization, surgery).

5. Conclusion

We develop a POC of machine learning model to predict the phenotype of patients suffering from HFpEF from both clinical and echocardiography data from data warehouse of Rennes University Hospital Center. The phenotyping of HFpEF could improve the characterization of patients, the definition of the most appropriate treatments, and the care pathways.

References

[1] Fayol A, Wack M, Livrozet M, Carves JB, Domengé O, Vermersch E, Mirabel M, Karras A, Le Guen J, Blanchard A, Azizi M, Amar L, Bories MC, Mousseaux E, Carette C, Puymirat E, Hagège A, Jannot AS, Hulot JS. Aetiological classification and prognosis in patients with heart failure with preserved ejection fraction. ESC Heart Fail. 2022 Feb;9(1):519-530. doi: 10.1002/ehf2.13717. Epub 2021 Nov 29. PMID: 34841727; PMCID: PMC8788026.

[2] van Riet EE, Hoes AW, Wagenaar KP, Limburg A, Landman MA, Rutten FH. Epidemiology of heart failure: the prevalence of heart failure and ventricular dysfunction in older adults over time. A systematic review. Eur J Heart Fail. 2016 Mar;18(3):242-52. doi: 10.1002/ejhf.483. Epub 2016 Jan 4. PMID: 26727047.

[3] Galli E, Bourg C, Kosmala W, Oger E, Donal E. Phenomapping Heart Failure with Preserved Ejection Fraction Using Machine Learning Cluster Analysis: Prognostic and Therapeutic Implications. Heart Fail Clin. 2021 Jul;17(3):499-518. doi: 10.1016/j.hfc.2021.02.010. PMID: 34051979.