Unsupervised Cluster Analysis Enables Outcomes Prediction Among Patients with LVH

Applied M.Sc. research project conducted in collaboration with the Hadassah Heart Institute

By: Ariel Vishne, Avi Surit, and Ranel Loutati

Hadassah Heart Institute collaborators:
David Luria, Offer Amir, and Yitschak Biton

Academic guidance: Prof. David Zucker and Bitya Neuhof

The applied project is based on UK Biobank data, accessible subject to payment and ethical approvals.
UK Biobank application number: 96272.

Abstract:

Background: The traditional classification of left ventricular hypertrophy (LVH), which relies on left ventricular geometry, fails to encompass the accompanying comorbidities and clinical features. Objectives: To identify unique clinical phenotypes of LVH using unsupervised cluster analysis and explore their association with clinical outcomes. Methods: Among the UK Biobank participants with available left ventricular mass (LVM) estimation, we identified individuals with a CMR-based diagnosis of increased LVM using previously determined cutoffs within the UK Biobank: LVM index ≥ 72 g/m^2 for men, and LVM index ≥ 55 g/m^2 for women. Baseline demographic, clinical, laboratory, and imaging data were collected from the database. Using Ward's minimum variance method, patients were clustered based on 27 clinical variables and 30 MRI-based features. The primary outcome was a composite of all-cause mortality with heart failure (HF) admissions, ventricular arrhythmia, and atrial fibrillation (AF). Cox proportional hazard model and Kaplan-Meier survival analysis were applied. Results: Among the UK Biobank participants who underwent CMR, increased LVM was found in 4,255 individuals, with an average age of 64 ± 7 years. Of these patients, 2,447 (58%) were women. Through cluster analysis, four distinct subgroups were identified. Over a median follow-up period of 5 years (IQR: 4-6), 100 patients (2%) died, 118 (2.8%) were admissioned due to HF, 29 (0.7%) were admissioned due to VA, and 208 (5%) were admissioned due to AF. Univariate Cox analysis revealed that compared to the 1^{st} cluster (n=1,578), patients in the 2^{nd} (n=1,296), 3^{rd} (n=824) and 4^{th} (n=557) clusters had 60%, 104%, and 164% increased risk of mortality, respectively (95% CI 1.2-2.16, p<.001 for cluster 2; 95% CI 1.49-2.78, p<.001 for cluster 3; 95% CI 1.92-3.62, p<.001 for cluster 4). Further investigation of patients in each cluster revealed different clinical phenotypes. Cluster 2 patients were mostly overweight women, with the highest prevalence of chronic lung disease. Cluster 3 patients were mostly men, and had the highest burden of comorbidities, including diabetes, hypertension, hyperlipidemia, chronic kidney disease and elevated aminotransferases. Cluster 4 patients were mostly men. Those patients presented with the most abnormal cardiac measures, including left ventricular end-diastolic volume (LVEDV), left ventricular end-systolic volume (LVESV), and left ventricular ejection fraction (LVEF). Comparison of clustering using clinical data with image-based clusters did not show any correlation. Conclusions:

Unsupervised cluster analysis identified four subgroups among patients with increased LVM. These subgroups bear notable relevance to overall mortality risk and clinical outcomes. This phenotypic classification holds promise in offering valuable insights regarding clinical course and outcomes in LVH patients.

INTRODUCTION

Increased left ventricular mass (LVM), commonly referred as left ventricular hypertrophy (LVH), is a well-known independent predictor of cardiovascular events [1-4] including heart failure, atrial fibrillation, stroke, myocardial infraction, sudden cardiac death, and mortality [5,6]. Furthermore, it was intriguingly recognized as a potential treatable target with prognostic implications in clinical trials for hypertension treatment [7].

Cardiac magnetic resonance imaging (CMR) is an accurate and reproducible tool for the assessment of LVM. Compared to the most common imaging modality, which is echocardiography, CMR offers additional data regarding the shape, size, and possible pathologies of the left ventricle [8].

The conventional four-tiered classification of LVH patients was initially purposed by Khouri et al., based on the CMR images of patients from the Dallas Heart Study (DHS) [9]. This classification divides the two common types of LVH, which are concentric and eccentric hypertrophy, into purely eccentric, eccentric-concentric, purely concentric, and concentric-eccentric. Concentric LVH is characterized by thickening of the LV wall without an increase in chamber size, while eccentric LVH involves both chamber enlargement and wall thickening. This classification is founded on the geometry of the left ventricle, and exhibits a strong link to the pathophysiological mechanisms of

pressure and volume overload, which act as the driving forces behind LVH development [10]. However, this classification has not demonstrated consistent correlations with clinical outcomes in population studies [11,12].

Unsupervised cluster analysis plays a pivotal role in the identification of distinct disease phenotypes, proving particularly advantageous when exploring cohorts characterized by heterogeneity [13-15]. Therefore, the objectives of this study were to utilize cluster analysis to identify various disease phenotypes among patients with increased LVM within the heterogeneous, low-risk population of the UK Biobank participants. Additionally, we aimed to investigate potential correlations between the identified phenotypes and significant clinical outcomes, and explain our results by clusters exploration.

METHODS

Description of the data

Our dataset consists of information obtained from the UK Biobank, which is a population-based prospective cohort comprising 502,629 participants recruited in the United Kingdom between 2006 and 2010. The primary aim of the UK Biobank is to investigate the genetic and lifestyle factors that contribute to the development of various diseases. The recruitment process involved inviting approximately 9.2 million individuals aged 40-69 years residing within a 25-mile radius of the 22 assessment centers located in England, Wales, and Scotland. Ultimately, 5.4% of the invited individuals participated in the baseline assessment. During recruitment, extensive data were collected through questionnaires, physical measurements, and biological samples. Furthermore, ongoing data collection was conducted in subsets of the cohort, including

repeated assessments and multimodal imaging. To monitor health outcomes, all participants are followed up through linkage to national health-related databases.

For the purpose of our specific project, we analyzed the clinical, laboratory, and imaging data of patients who underwent CMR. The data for each patient encompassed core information such as previous diagnoses obtained from hospitals or family doctors, comorbidities, prescribed medications, allergies, habits, environmental exposures, past surgeries and procedures, and previous admissions with health outcome phenotypes. Additionally, assay data consisting of biochemical and hematological test results were available, including platform-based assays performed on patients with suspected cardiac diseases, such as BNP and troponin assays, as well as measurements of vital signs. Imaging data included various cardiac imaging modalities, such as CMR, echocardiography series, and Electrocardiogram (ECG) data. Our primary focus for the initial clustering analysis was on the clinical data, encompassing demographics, diagnoses, health questionnaires, physical assessments, and health outcomes. Subsequently, we utilized the CMR data for further clustering analysis.

For all analyses, we included individuals who underwent CMR during the UK Biobank imaging assessment and whose bulk CMR data were available for download as of January 01, 2023. The full CMR protocol of the UK Biobank has been described in detail previously [16]. Briefly, all CMR examinations were performed in the United Kingdom on a clinical wide-bore 1.5 Tesla scanner (MAGNETOM Aera, Syngo Platform VD13A; Siemens Healthineers, Erlangen, Germany). All acquisitions used balanced steady-state free precession with typical parameters.

The estimation of LVM was based on the contours from the InlineVF algorithm which are stored in each DICOM file's metadata. Those contours were further processed into

pixel masks that labeled myocardium, LV cavity, and background. The DICOM metadata stores the inner and outer contours of the myocardium as a 1-pixel-wide pixel mask; we converted these contours to polygons, which we processed into a segmentation pixel mask using morphologic image operators.

Data Preparation

The data preparation process for this project was carried out separately for the clinical data and the imaging data. Starting with the clinical data, we initially downloaded, converted, and decoded the entire UK Biobank dataset. The complete dataset consisted of 502,369 rows, each corresponding to a different patient, and comprised over 4000 features. To narrow down the features of interest, we conducted feature selection in consultation with cardiologists from the Hadassah Heart Institute. Eventually, we identified and selected 900 possibly-relevant features.

The diagnosis of LVH was based on the LVM cutoffs that were in use by researchers from Mass General Brigham (MGB) hospitals and the Broad Institute; LVM index ≥ 72 g/m^2 for men, and LVM index ≥ 55 g/m^2 for women.

Additionally, we timestamped the selected diagnoses and significant events to distinguish between clinical data that was known prior to the CMR and subsequent LVH diagnosis, which was used for clustering purposes, and clinical outcomes (including specific diagnoses, events, and mortality) that occurred after the assignment of LVH diagnosis. Clinical outcomes were utilized for evaluating the predictive ability of our clustering approach.

Regarding the imaging data, our primary focus was on CMR images. Each CMR test generates images in various planes, heights, and timeframes. For our analysis, we specifically considered the short-axis CMR images in the transverse plane. To

encompass the entire heart, we collected all available "height" series for each patient, which typically ranged from 5 to 13 series, and concatenated them. We then computed an average image over time to represent the heart cycle. As the images had different shapes, we applied padding to standardize them to the maximum shape of 256x256 pixels.

It is important to highlight that each image, with an average of 500 images per patient, has a size of approximately 40 Mb. Consequently, a significant portion of our pre-processing efforts was dedicated to developing methods for efficiently downloading, compressing, and loading the data onto the university clusters. The main table has a size of 4.5 Gb.

Unsupervised Clustering

In order to decide how many clusters we aim to find, we utilized a few commonly used methods for deciding the cutoff that dictates the number of clusters. First we use the SVD decomposition to observe the singular values of the data matrix. Based on these values 2 methods can be used. The former is the Elbow-method, in which the singular values are plotted as bars, and an "elbow" is manually identified as they descent. The point where the reduction in values starts to slow down creates an "elbow" shape on the plot, and it's often where the optimal number of clusters lies. The latter is to observe how many components of the singular values are needed in order to explain at least 90% of the explained variance. In the domain of tabular data, the common rule of thumb is trying to explain 90%, as opposed to other applications like in visual or auditory data, where higher percentages are pursued.

In order to add more certainty to our choosing, 3 more commonly used algorithms that inspect the wellness of the clustering were used. The first was Silhouette Coefficient,

that measures the quality of clustering by considering both the compactness of clusters and the separation between clusters. It ranges from -1 to 1, where higher values indicate better clustering. A value close to 1 suggests that the clusters are well-separated and compact, indicating a good clustering solution. The second method was Calinski-Harabasz Index. The index, also known as the variance ratio criterion, quantifies the ratio of between-cluster dispersion to within-cluster dispersion. Higher values of the index indicate better clustering quality. A higher value suggests that the clusters are more compact and well-separated from each other. The third algorithm was Davies-Bouldin Index, which evaluates both the compactness of clusters and the separation between clusters. Lower values of the index indicate better clustering quality. A lower value suggests that the clusters are more compact and well-separated, indicating good clustering.

For the unsupervised clustering analysis, we employed a widely used algorithm, agglomerative clustering. Agglomerative clustering is a hierarchical method that begins with individual data points and gradually merges them into clusters based on their similarity. Initially, each data point is treated as a separate cluster, and then, using a chosen distance metric, clusters are iteratively merged based on their proximity until a desired number of clusters or a specified stopping criterion is reached. Specifically, we applied Ward's minimum variance method that aims to minimize the variance within each cluster as new data points are added. The result of the hierarchical structuring is presented via a dendrogram, and is then colored in a fixed number of colors, according to the number of clusters chosen by the methods aforementioned above.

After specifying the number of clusters and creating them, we proceed with creating a heat map of all the features in the data, calculated per cluster. Since some of the columns in the data are the results of bloodwork, their values were converted to binary values,

where 1 indicating an abnormal value, and 0 is in the normal range. Continuous and discrete columns in the data then were processed differently. For the continuous features we calculated the aggregated means per cluster. In order to provide the same scale as the binary data, we reported for each mean its respective quantile from all the observations, providing a value between 0 and 1. For the binary columns, we calculated for each cluster the ratio of number of abnormalities in the cluster over all abnormalities, per feature. Thus we guarantee these numbers sum up to 1, which means that we can distinguish if any cluster has a more prominent percentages of abnormal diagnoses compared to the others. The values for both types of data were then presented in a 2-dimensional table, where the x-axis represents the clusters ($C_1, \dots C_{n_{clusters}}$), and the y-axis represents the features ($F_1, F_2, \dots$). The range 0-1 was divided to 5 even segments and a respective color, and each cell was colored according to its respective segment color, in order to visually express the qualities that identify each cluster uniquely.

Image processing and feature extraction

For training an Auto-Encoder based Neural Network on the images we used an Auto-Encoder architecture, in which training is conducted in order to optimize the reconstruction loss based on the original image, whilst propelling the image through an encoding stage which is – hopefully – a latent space in which the inner metrics are useful for understanding similarity between different images. Training was conducted with MSE loss between image and decoded image of the encoding. For all experiments we used three symmetric convolution layers, wrapped with strides and max pooling at each stage. The convolution layer size decreases exponentially approaching the

encoding layer. The latent space of the encoding stage was fixed at $24^2 = 576$. We used the Adam optimizer with learning rate $lr = 10^{-4}$ and weight decay 10^{-5}.

Each sample was defined as a single patient, whereas the initial input was a cropping of all images for that patient. Images were dealt with two different approaches: The first approach (denoted "mean") takes a mean of all images in a single series, and then several of those mean images was considered an input (i.e., every channel in the input is a mean of a series of images). The second approach (denoted "series") takes all 50 images from a single series as the basic unit of analysis, and in this case a channel represents a single MRI image. For both "mean" and "series" implementations, different thresholds were placed for minimum number of provided images, specific series numbers, and consistent dimensions for all images per patient.

Finally, we applied K-means clustering with $k = \#clusters$ from clinical data, in order to create a basis for comparison between the images and clinical data.

Clustering agreement

We sought to compare the results obtained from the two types of data. For this comparison, we intended to determine the agreement rate between the two methods and assess which method exhibited stronger correlations with clinical outcomes. For calculating the degree of agreement between different outputs of clustering algorithms (that either derive from different sources of data, i.e., clinical tabular data and images, or that derive from different data processing methods on a single source of data), we use metrics that account only for the level of agreement between the labels of the dataset. This is because the labelling in itself is arbitrary and any permutation of the labels on the data is essentially the same. Therefore, we only look at how different samples are categorized into the clusters. For this we use adjusted rand index, which

examines every possible pair in the dataset and sees whether the pair is in the same class in both clustering results. Another measure based on entropy is the adjusted mutual information index. Both metrics are adjusted in order to account for data imbalance inherent within clustering outputs.

Statistical analysis

Continuous variables were expressed as mean ± standard deviation if normally distributed or median with interquartile range if skewed. Categorical variables were presented as frequency (%). Differences between the four clusters were analyzed using one-way ANOVA for continuous variables that were normally distributed, while the Kruskal–Wallis test was used to compare continuous variables that did not adhere to a normal distribution. Multiple comparisons for continuous and categorical variables were tested using Bonferroni's correction. The primary outcome was a composite endpoint of all-cause mortality, heart failure admissions, and events of ventricular or atrial arrhythmias. For survival analysis patients were censored only in the case of death. The probability of outcome according to the study groups was graphically displayed according to the method of Kaplan–Meier, with a comparison of cumulative event across strata by the log-rank test. Univariate and multivariate Cox proportional hazards regression modeling were used to compare patients in each of the four clusters, with adjustment made to the strongest predictor of mortality in LVH patients which is LVM (as a continuous variable). Unsupervised cluster analyses and image processing and AE-based feature extraction were performed using Python software version 3.11 (packages in use: scikit-learn, matplotlib, pandas, numpy, pillow, pyTorch, and more – see in code). All statistical analyses were performed using R software version 3.4.4 (R Foundation for Statistical Computing). An association was considered statistically significant for a two-sided P value of less than 0.05.

RESULTS

Study population and baseline characteristics

Among the 60,650 UK-Biobank participants who underwent CMR, 4,255 individuals met the cutoff for LVH diagnosis. The final study had an average age of 64 ± 7 years, of whom 58% were women. The average body mass index (BMI) was 28 (5) which indicated that as an average, the study population was over-weight. The most common co-morbidity was hypertension, that was present in 47% of patients. Other common comorbidities included dyslipidemia (30%), peripheral vascular disease (26%), and diabetes mellitus (8%). Cardiac measures that were calculated based on the CMR images included left ventricular ejection fraction (LVEF), left ventricle end diastolic volume (LVEDV), left ventricular end systolic volume (LVESV), and left ventricle mass (LVM). All of those measures didn't appear to be abnormal in average, as expected in a relatively healthy population such as the UK-Biobank participants. All baseline characteristics are summarized in **Table 1**. SVD based methods suggested that a total of three or four clusters would be appropriate for this study population. We then employed different algorithms including the Silhouette Coefficient, the Calinski-Harabasz index, and the Davies-Bouldin index. Those algorithms, together with expert opinion, had directed us to make the final choice of 4 clusters (see **Figure 1a-b, Table 2**).

Unsupervised clustering results

The agglomerative clustering identified 4 distinct clusters, as depicted in the dendrogram (**Figure 2**). The 1st and the 2nd clusters were larger compared to the later two. Comparisons of baseline characteristics across clusters are provided in **Table 3.**

Cluster 1 consisted on 1,578 patients, almost all were females (93%). The average age was similar to the entire population, and the BMI was lower – with an average of 26 (4). Patients in this cluster had the lowest prevalence of comorbidities, including comorbidities that are known to be cardiovascular risk factors such as diabetes mellitus (6%), hypertension (39%), dyslipidemia (22%), and chronic kidney disease (0.3%). The median of all cardiac measures was the lowest, in comparison to all other clusters, and none of the laboratory tests was abnormal. Therefore, this cluster served as the baseline for all of our comparisons.

Cluster 2 (n=1,296) also included mainly females (63%), with a slightly higher age than the other clusters (average 65[7]). Patients in this cluster had more comorbidities compared to cluster 1, with the highest prevalence of diabetes mellitus (10.5%), and chronic lung disease (4%). In similar to the first cluster, cardiac measures were low, pointing the lack of overt structural heart disease among those patients. None of the laboratory tests was abnormal.

Cluster 3 (n=824) was the cluster with the most comorbidities and cardiovascular risk factors, including highest portion of males (89%), obesity (average BMI 30 [5]), diabetes mellitus (10.3%), hypertension (54.5%), dyslipidemia (41%), peripheral vascular disease (29%), and chronic kidney disease (1.8%). Not surprisingly, this was also the cluster with the most abnormal laboratory tests including liver function tests (AST, ALT), kidney function tests (Urea, Creatinine), Urate, and lipid profile (lowest

HDL and the highest LDL). In terms of cardiac measures, this cluster showed lower values than the first two, but much better measures compared to the 4th cluster.

Cluster 4 (n=557) was the most interesting one. It was also consisted mainly of males (86%), with an average age of 63 (8). This cluster had higher prevalence of comorbidities comparing to the 1st cluster, but less than the 2nd and 3rd clusters. The cardiac measures of this cluster were the most abnormal – indicating pathological changes in the structure of the heart which are expressed by the left ventricular volumes (LVEDV median 209 [193-231], LVESV median 99 [86-117]), the left ventricle mass (LVM index median 79 [74-86]), and function (LVEF 51[10]). In conjunction with those findings, this cluster exhibited the highest prevalence of prior myocardial infraction (8.3%), and prior diagnosis of congestive heart failure (3.4%).

A variable heatmap that underscores the differences between clusters is presented in **Figure 3.**

LVH clusters and outcomes during follow-up

During a median follow-up period of 5 years (IQR: 4-6), 100 patients (2%) died, 118 (2.8%) were admissioned due to heart failure, 29 (0.7%) were admissioned due to ventricular arrhythmia, and 208 (5%) were admissioned due to atrial fibrillation. The crude number of events in each cluster is specified in **Table 4a/b**. Kaplan-Meier survival analysis revealed that the cumulative probability of the composite endpoint at 5 years of follow-up was 5.2%±0.6%, 8%±0.9%, 1.2%±1.3%, and 15.5%±1.7% for each cluster, respectively (**Figure 4**; p Log rank <.001 for the overall difference during follow-up). Univariate Cox analysis revealed that compared to the 1st cluster, patients in the 2nd, 3rd, and 4th clusters had 60%, 104%, and 164% increased risk of a major event, respectively (95% CI 1.2-2.16, p<.001 for cluster 2;

95% CI 1.49-2.78, p<.001 for cluster 3; 95% CI 1.92-3.62, p<.001 for cluster 4). Multivariate Cox analysis adjusted for LVM index, which is the most contributing factor for mortality among LVH patients, consistently demonstrated that compared to the 1st cluster, patients in the 2nd, 3rd, and 4th clusters had 38%, 49%, and 94% increased risk of a major event, respectively (95% CI 1.02-1.88, p=0.039 for cluster 2; 95% CI 1.06-2.10, p=0.022 for cluster 3; 95% CI 1.36-2.78, p<.001 for cluster 4). Every 10 g/m^2 increase in LVM index was associated with 33% increased risk of a major event (95% CI 1.23-1.44, p<.001).

Breaking down the composite outcome to each endpoint, the most common outcomes were admissions due to heart failure (n=118 [2.8%]) and atrial fibrillation (n=208 [5%]). The association of all clusters with each endpoint was consistent with the composite endpoint as seen in **Figure 5**. However, because of the relatively small rate of events, certain clusters were underpowered, especially cluster 2 which did not demonstrated statistical significance with any endpoint. A summary of crude number of events for each outcome, as well as their association with all clusters are presented in **Table 5**.

Image-Based clustering results

The AE training was conducted for either 10, 20, 50 or 100 epochs. Batch size was 4 patients per batch, determined randomly (and changing at each epoch). For each model trained, the cohort of patients was divided into a train set (90% of the data) and a test set (10% of the data), in order to determine where convergence has reached. As the dataset provided was moderate in size, it is evident that training even for a large number of epochs had a decreasing effect also on the test set, so these (rather large) numbers of epochs were used. **Figure 6a** demonstrates the learning process by plotting the

reconstruction loss for training and test sets. **Figure 6b** provides examples of reconstruction across all clusters.

After training the neural network, all data was passed again in order to create full encodings and decoded images for all data. The clustering was applied to the latent space (the encoding), either directly, or to the 30 top principal components applied on the latent space. **Figure 7** depicts the t-SNE results on the clusters of embeddings.

Surprisingly, level of agreement between the clustering algorithms remained low, between different models and also between the model clustering outputs and the clinical clustering outputs. Levels of agreement for both adjusted rand score and adjusted mutual information score remained close to 0, which suggests a level of agreement to that of random labelling. This is despite the reconstruction of images from the decoded stage which proved satisfactory. An illustrative example is provided in **Figure 8**, where we take T-SNE-5 projection from the encodings of a "series" model and color the clusters by the clinical data.

DISCUSSION

Our study offers several important findings. Firstly, it uses large population-based data to assess the presence of different clusters among LVH patients with respect to mortality and cardiovascular outcomes, therefore expanding previous observation on the heterogeneity of this population. Secondly, this study demonstrates that the association between LVH and cardiovascular events and mortality is mainly influenced from the presence of high burden of comorbidities (i.e. cluster 2 and 3) or very abnormal cardiac

measures (i.3 cluster 4). Finally, this analysis provides original comparison between clinical-based clustering to imaging-based clustering.

Previous unsupervised clustering studies have been done in the field of cardiology, and especially in heart failure patients, in an attempt to predict certain classes of patients that will respond to therapy and/or are at risk for cardiovascular events. Segar et al. [15] performed a study aimed to identify distinct phenotypic subgroups within a complex dataset of heart failure patients with preserved ejection fraction (HFpEF) using unsupervised clustering analysis. The researchers were able to characterize three exclusive phenogroups within the subset. These phenogroups demonstrated varying co-morbidity burdens, natriuretic peptide levels, and profiles of left ventricular function. Phenogroup 1 had a higher risk of adverse clinical events, including all-cause mortality and HF hospitalization, compared to phenogroup 3. Phenogroup 2 had a distinct risk profile, being associated with higher HF hospitalization risk but lower atherosclerotic event risk, and comparable mortality risk. In conjunction with our analysis, a higher rate of events was seen in the group of patients with high burden of comorbidities. Another study [13] which utilized unsupervised clustering on patients with heart failure attempted to find phenotypes of patients who are considered responders for cardiac resynchronization therapy (CRT). Analyzing 1106 HF patients from the MADIT-CRT trial, the researchers employed an unsupervised machine learning algorithm to group patients based on clinical and baseline left ventricular data. Four phenogroups were identified, differing significantly in baseline characteristics and cardiac measures. Two phenogroups exhibited a higher proportion of predictors for CRT response and demonstrated markedly better CRT treatment effects on the primary outcome (all-cause death or HF event) compared to the other groups. From our point of view, this study was intriguing. Primarily since it highlighted the role of unsupervised machine learning

not only in identifying phenogroups that correlate with outcome, but also in providing evidence-based insights regarding life-saving therapies. Additionally, in this study the data that was used for clustering consisted not only clinical and echocardiographic measures but also time-dependent data that was interpolated from the echocardiography. Therefore, their results underscore the possible synergism that can be exist between clinical and imaging data.

The apparent mismatch between the clustering algorithms provided by the clinical data and the neural network models is intriguing, especially in the light of the previous study. However, it may be less astounding at second thought. The MRI images themselves provide only a narrow understanding of the clinical issues associated with LVH. The images are general in scope and the spatial characteristics of the imaging in itself (quality of the MRI scanner, position of the body during the scan etc.) may be the signal detected thus wiping out any interesting clinical data that may be hidden within the images. It is possible that with labelled data the model would be much better at detecting clinical-related issues, whereas for now the auto-encoder loss trained on reconstruction only (without a supervised component) is much more focused on addressing graphical issues than we want it to. Another possible expansion of the training whilst staying in the unsupervised realm is to instantiate substantial regularization which would be focused on the spatial loss and not only an MSE-based reconstruction loss. Yet another option is using a pre-trained medically-oriented model as the basis for fine-tuning with the mechanism we implemented already.

Study Limitations

This study has several important limitations. First, this is a retrospective based study on a relatively healthy population that does not necessarily represent LVH among high-

risk patients. Second, cause-specific mortality was not investigated, and the primary outcome was a composite of several, yet important, cardiovascular events. Third, important clinical and laboratory predictors of poor survival such as clinical signs of right heart failure, NYHA functional class, and BNP were not available. Lastly, quantitative indices of LV hypertrophy, such as LV interventricular septum width were not readily available for most patients and were not included in the analysis. Finally, while independent associations have been demonstrated, causality could not be established due to study design.

Conclusions

Our analysis has yielded the identification of four distinct subgroups among patients with LVH through the application of unsupervised clustering techniques. These subgroups exhibit significance in relation to the primary outcome of all-cause mortality and significant cardiovascular events. Unfortunately, clustering of CMR images and the subsequent comparison of clusters, aim to capture the interplay between the morphological characteristics of LVH and associated comorbidities, did not yielded meaningful results.

REFERENCES

1. Lazzeroni, Davide, et al. “From Left Ventricular Hypertrophy to Dysfunction and Failure.” *Circulation Journal*, vol. 80, no. 3, 2016, pp. 555–64, https://doi.org/10.1253/circj.CJ-16-0062.
2. Haider, Agha W., et al. “Increased Left Ventricular Mass and Hypertrophy Are Associated with Increased Risk for Sudden Death.” *Journal of the American College of Cardiology*, vol. 32, no. 5, Nov. 1998, pp. 1454–59, https://doi.org/10.1016/S0735-1097(98)00407-0.
3. Ghali, Jalal K., et al. “Influence of Left Ventricular Geometric Patterns on Prognosis in Patients With or Without Coronary Artery Disease.” *Journal of the American College of Cardiology*, vol. 31, no. 7, June 1998, pp. 1635–40, https://doi.org/10.1016/S0735-1097(98)00131-4.
4. Bluemke, David A., et al. “The Relationship of Left Ventricular Mass and Geometry to Incident Cardiovascular Events.” *Journal of the American College of Cardiology*, vol. 52, no. 25, Dec. 2008, pp. 2148–55, https://doi.org/10.1016/j.jacc.2008.09.014.
5. Milani, Richard V., et al. “Left Ventricular Geometry and Survival in Patients With Normal Left Ventricular Ejection Fraction.” *American Journal of Cardiology*, vol. 97, no. 7, Apr. 2006, pp. 959–63, https://doi.org/10.1016/j.amjcard.2005.10.030.
6. Levy, Daniel, et al. “Prognostic Implications of Echocardiographically Determined Left Ventricular Mass in the Framingham Heart Study.” *New England Journal of Medicine*, vol. 322, no. 22, May 1990, pp. 1561–66, https://doi.org/10.1056/NEJM199005313222203.

7. Gardin, Julius M., and Michael S. Lauer. “Left Ventricular HypertrophyThe Next Treatable, Silent Killer?” *JAMA*, vol. 292, no. 19, Nov. 2004, pp. 2396–98, https://doi.org/10.1001/jama.292.19.2396.

8. Myerson, Saul G., et al. “Assessment of Left Ventricular Mass by Cardiovascular Magnetic Resonance.” *Hypertension*, vol. 39, no. 3, Mar. 2002, pp. 750–55, https://doi.org/10.1161/hy0302.104674.

9. Khouri, Michel G., et al. “A 4-Tiered Classification of Left Ventricular Hypertrophy Based on Left Ventricular Geometry.” *Circulation: Cardiovascular Imaging*, vol. 3, no. 2, Mar. 2010, pp. 164–71, https://doi.org/10.1161/CIRCIMAGING.109.883652.

10. Ganau A, Devereux RB, Roman MJ, de Simone G, Pickering TG, Saba PS, Vargiu P, Simongini I, Laragh JH. Patterns of left ventricular hypertrophy and geometric remodeling in essential hypertension. J Am Coll Cardiol. 1992; 19: 1550–1558.

11. Krumholz HM, Larson M, Levy D. Prognosis of left ventricular geometric patterns in the Framingham Heart Study. J Am Coll Cardiol. 1995; 25: 879–884.

12. Bang, Casper N., et al. “Four-Group Classification of Left Ventricular Hypertrophy Based on Ventricular Concentricity and Dilatation Identifies a Low-Risk Subset of Eccentric Hypertrophy in Hypertensive Patients.” *Circulation: Cardiovascular Imaging*, vol. 7, no. 3, May 2014, pp. 422–29, https://doi.org/10.1161/CIRCIMAGING.113.001275.

13. Cikes, Maja, et al. “Machine Learning-Based Phenogrouping in Heart Failure to Identify Responders to Cardiac Resynchronization Therapy.” *European Journal of Heart Failure*, vol. 21, no. 1, 2019, pp. 74–85, https://doi.org/10.1002/ejhf.1333.

14. Perry, Andrew, et al. “Unsupervised Cluster Analysis of Patients with Recovered Left Ventricular Ejection Fraction Identifies Unique Clinical Phenotypes.” *PLOS ONE*, vol. 16, no. 3, Mar. 2021, p. e0248317, https://doi.org/10.1371/journal.pone.0248317.

15. Segar, Matthew W., et al. “Phenomapping of Patients with Heart Failure with Preserved Ejection Fraction Using Machine Learning-Based Unsupervised Cluster Analysis.” *European Journal of Heart Failure*, vol. 22, no. 1, 2020, pp. 148–58, https://doi.org/10.1002/ejhf.1621.

16. Petersen, Steffen E., et al. “UK Biobank’s Cardiovascular Magnetic Resonance Protocol.” *Journal of Cardiovascular Magnetic Resonance*, vol. 18, no. 1, Feb. 2016, p. 8, https://doi.org/10.1186/s12968-016-0227-4.

FIGURES and TABLES

Table 1 – Clinical characteristics of LVH patients

	Increased LVM (N=4,255)
Age	64.2 ± 7.4
Female	(57.5)
BMI, kg/m^2	28.2 ± 4.96
BSA, m^2	1.9 ± 0.218
Diabetes Mellitus	(8.3)
Hypertension	(47.3)
Dyslipidemia	(30.3)

	Increased LVM (N=4,255)
Prior MI	(3.9)
CHF	(1.4)
PVD	(26)
Chronic Lung Disease	(2.8)
CKD	(0.9)
Aortic Stenosis	(0.8)
Mitral Regurgitation	(0.6)
Atrial Fibrillation	(3)
LVEF, %	55.4 ± 7.3

	Increased LVM (N=4,255)
LVEDV, mL	148 [125-175]
LVESV, mL	64 [52-80]
LVM index, g/m^2	70 [60-78]
AST, U/L	24.4 [21-29]
ALT, U/L	20.5 [15.5-27.7]
Urea, mmol/L	5.28 [4.53-6.12]
Creatinine, µmol/L	69 [60.6-78.7]
HbA1c, mmol/L	34.9 [32.5-37.4]
HDL, mmol/L	1.40 [1.17-1.69]

	Increased LVM (N=4,255)
LDL, mmol/L	3.52 [2.99-4.11]
Urate, µmol/L	302 [251-360]

Values are mean ± SD, (%), or median [Interquartile range].

ALT – alanine transaminase, AST – aspartate transaminase, BMI – body mass index, BSA – body surface area, CHF – congestive heart failure, CKD – chronic kidney disease, HbA1c – glycated haemoglobin, HDL – high density lipoprotein, LDL – low density lipoprotein, LVEDV – left ventricular end diastolic volume, LVEF – left ventricular ejection fraction, LVESV – left ventricular end systolic volume, LVM – left ventricle mass, MI – myocardial infraction, PVD – peripheral vascular disease

Table 2 – Measures for quality of clustering

# of clusters	Silhouette Coefficient	Calinski Harabasz	Davies Bouldin
C = 3	0.235	**1983.25**	1.37
C = 4	**0.244**	1893.32	**1.25**

Table 3 – Clinical characteristics across clusters

	Cluster 1 (N=1,578)	**Cluster 2** (N=1,296)	**Cluster 3** (N=824)	**Cluster 4** (N=557)
Age	64.2 ± 7.29	65.0 ± 7.2	63.8 ± 7.53	63.3 ± 7.84
Female	(93.1)	(62.7)	(10.7)	(14)
BMI, kg/m^2	26.3 ± 4.33	29.6 ± 5.25	30.1 ± 4.8	27.7 ± 4.06
BSA, m^2	1.77 ± 0.168	1.91 ± 0.194	2.08 ± 0.185	2.03 ± 0.176
Diabetes Mellitus	(6.1)	(10.5)	(10.3)	(6.5)
Hypertension	(39.4)	(52.6)	(54.5)	(46.5)
Dyslipidemia	(21.6)	(33.3)	(41)	(32.5)

	Cluster 1 (N=1,578)	Cluster 2 (N=1,296)	Cluster 3 (N=824)	Cluster 4 (N=557)
Prior MI	(1.5)	(3.5)	(6.3)	(8.3)
CHF	(0.5)	(1.4)	(1.9)	(3.4)
PVD	(23.7)	(27.9)	(28.9)	(23.7)
Chronic Lung Disease	(1.8)	(4.1)	(2.9)	(2.5)
CKD	(0.3)	(0.9)	(1.8)	(1.3)
Aortic Stenosis	(0.4)	(1.2)	(0.7)	(0.9)
Mitral Regurgitation	(0.4)	(0.7)	(0.7)	(1.1)
Atrial Fibrillation	(2.3)	(2.1)	(4.0)	(5.9)

	Cluster 1 (N=1,578)	Cluster 2 (N=1,296)	Cluster 3 (N=824)	Cluster 4 (N=557)
LVEF, %	56.6 ± 6.18	56.7 ± 6.32	54 ± 7.38	50.7 ± 9.58
LVEDV, mL	133 [116-150]	137 [120-158]	169 [151-192]	209 [193-231]
LVESV, mL	56 [48-66]	59 [50-70.3]	77 [65.8-91]	99 [86-117]
LVM index, g/m^2	60.1 [57-65]	67.4 [59-75.9]	77.5 [74-83.2]	78.8 [74.2-85.7]
AST, U/L	22.8 [19.8-26.2]	24.3 [21.1-29.2]	27.5 [23.8-33.1]	25.5 [22.2-29.3]

	Cluster 1 (N=1,578)	Cluster 2 (N=1,296)	Cluster 3 (N=824)	Cluster 4 (N=557)
ALT, U/L	16.9 [13.6-21.9]	22 [16.4-28.9]	26.8 [20.1-36]	21.1 [16.6-27.8]
Urea, mmol/L	4.94 [4.2-5.74]	5.37 [4.68-6.25]	5.7 [4.9-6.54]	5.36 [4.66-6.14]
Creatinine, µmol/L	61.8 [55.9-68]	69.2 [61.3-78.2]	81.2 [73.5-89.1]	76.5 [68.8-82.5]
HbA1c, mmol/L	34.6 [32.2-37.1]	35.3 [32.8-38]	35.3 [32.9-37.9]	34.3 [32.3-36.8]
HDL, mmol/L	1.6 [1.34-1.86]	1.35 [1.14-1.59]	1.23 [1.04-1.44]	1.33 [1.13-1.61]
LDL, mmol/L	3.48 [2.99-4.03]	3.57 [3.01-4.2]	3.62 [3.04-4.16]	3.39 [2.87-3.93]

	Cluster 1 (N=1,578)	Cluster 2 (N=1,296)	Cluster 3 (N=824)	Cluster 4 (N=557)
Urate, µmol/L	236 [208-260]	323 [303-348]	412 [388-452]	310 [276-337]

Values are mean ± SD, (%), or median [Interquartile range].

All variables had $p < 0.05$

ALT – alanine transaminase, AST – aspartate transaminase, BMI – body mass index, BSA – body surface area, CHF – congestive heart failure, CKD – chronic kidney disease, HbA1c – glycated haemoglobin, HDL – high density lipoprotein, LDL – low density lipoprotein, LVEDV – left ventricular end diastolic volume, LVEF – left ventricular ejection fraction, LVESV – left ventricular end systolic volume, MI – myocardial infraction, PVD – peripheral vascular disease

Table 4a – Univariate Cox analysis for composite endpoint

	HR	95% CI	P value
Unsupervised clusters* (n events [%]):			
Cluster 1 (n=80 [5%])	reference		
Cluster 2 (n=99 [7.6%])	1.6	1.2-2.16	**<0.001**
Cluster 3 (n=97 [11.7%])	2.04	1.49-2.78	**<0.001**
Cluster 4 (n=88 [15.8%])	2.64	1.92-3.62	**<0.001**

* Patients in cluster 1 are serving as baseline

Table 4b – Multivariate Cox analysis for composite endpoint

	HR	95% CI	P value
Unsupervised clusters* (n events [%]):			
Cluster 1 (n=80 [5%])	reference		
Cluster 2 (n=99 [7.6%])	1.38	1.02-1.88	**0.039**
Cluster 3 (n=97 [11.7%])	1.49	1.06-2.1	**0.022**
Cluster 4 (n=88 [15.8%])	1.94	1.36-2.78	**<0.001**
LVM indexed, per 10 g/m^2	1.33	1.23-1.44	**<0.001**

* Patients in cluster 1 are serving as baseline; LVM – left ventricle mass

Table 5 – Univariate Cox analysis for each endpoint

Unsupervised clusters* (n events [%]):	**HR**	**95% CI**	**P value**
All-cause mortality:			
Cluster 1 (n=22 [1.4%])	reference		
Cluster 2 (n=27 [2%])	1.50	0.85-2.63	0.2
Cluster 3 (n=30 [3.6%])	2.64	1.53-4.59	**<0.001**
Cluster 4 (n=21 [3.7%])	2.70	1.48-4.91	**<0.001**
HF admissions:			
Cluster 1 (n=22 [1.4%])	reference		
Cluster 2 (n=24 [1.9%])	1.32	0.74-2.36	0.3
Cluster 3 (n=32 [3.9%])	2.84	1.65-4.89	**<0.001**
Cluster 4 (n=40 [7.2%])	5.27	3.13-8.86	**<0.001**
Ventricular arrhythmia:			
Cluster 1 (n=3 [0.2%])	reference		
Cluster 2 (n=8 [0.6%])	3.23	0.86-12.2	0.083
Cluster 3 (n=6 [0.7%])	3.85	0.96-15.4	0.057
Cluster 4 (n=12 [2.1%])	11.3	3.20-40.2	**<0.001**
Atrial fibrillation			
Cluster 1 (n=46 [3%])	reference		
Cluster 2 (n=62 [4.8%])	1.65	1.12-2.41	**0.01**
Cluster 3 (n=51 [6.2%])	2.17	1.46-3.23	**<0.001**
Cluster 4 (n=49 [8.8%])	3.10	2.07-4.63	**<0.001**

Figure 1a – SVD and explained variance

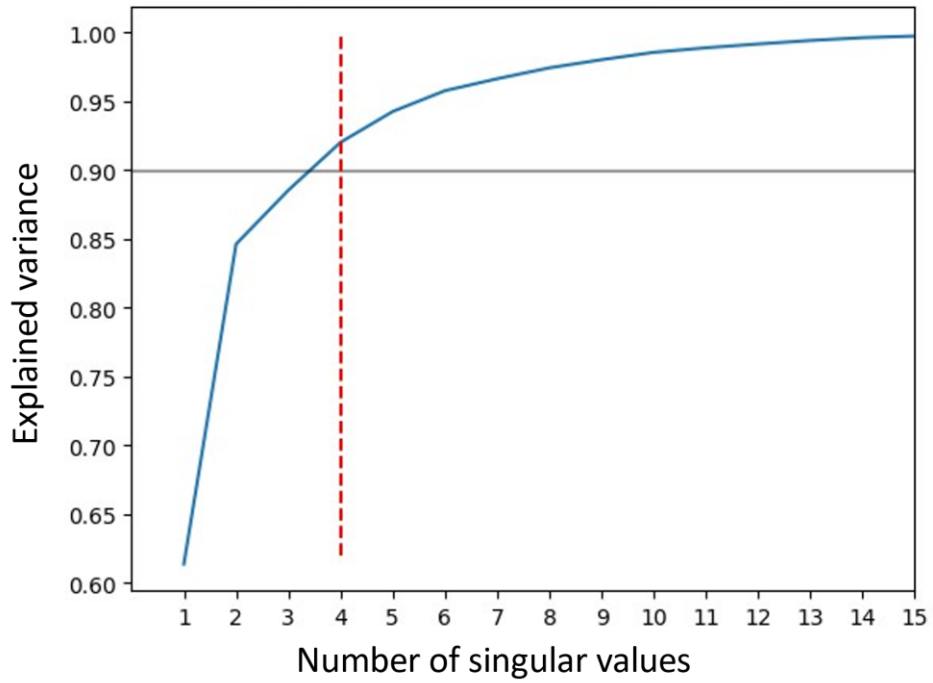


Figure 1b – Elbow plot

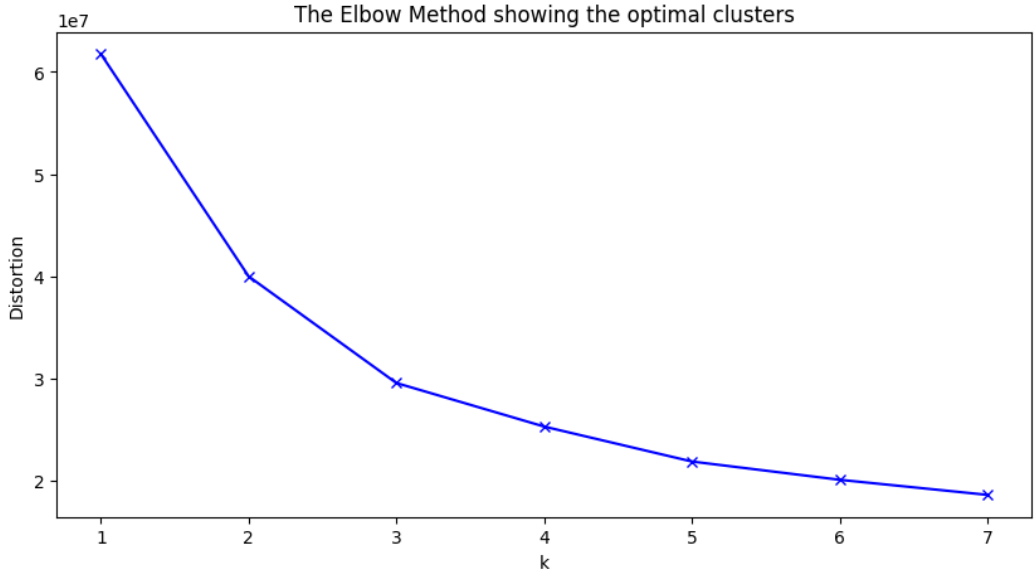


Figure 2 – Dendrogram

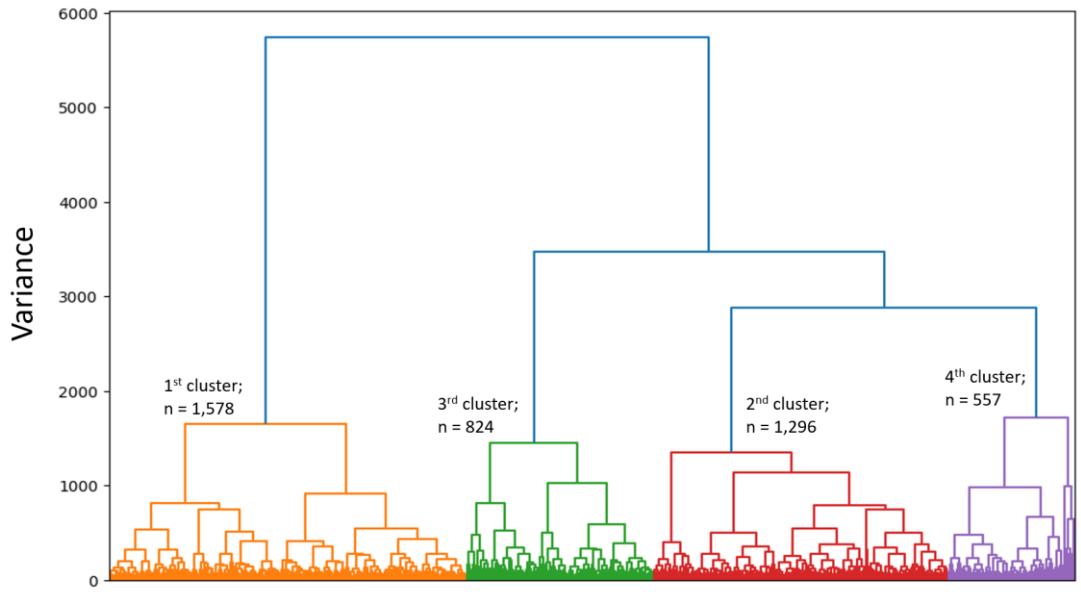


Figure 3 – Heatmap of variables across clusters

Figure 4 – Kaplan-Meier for composite endpoint

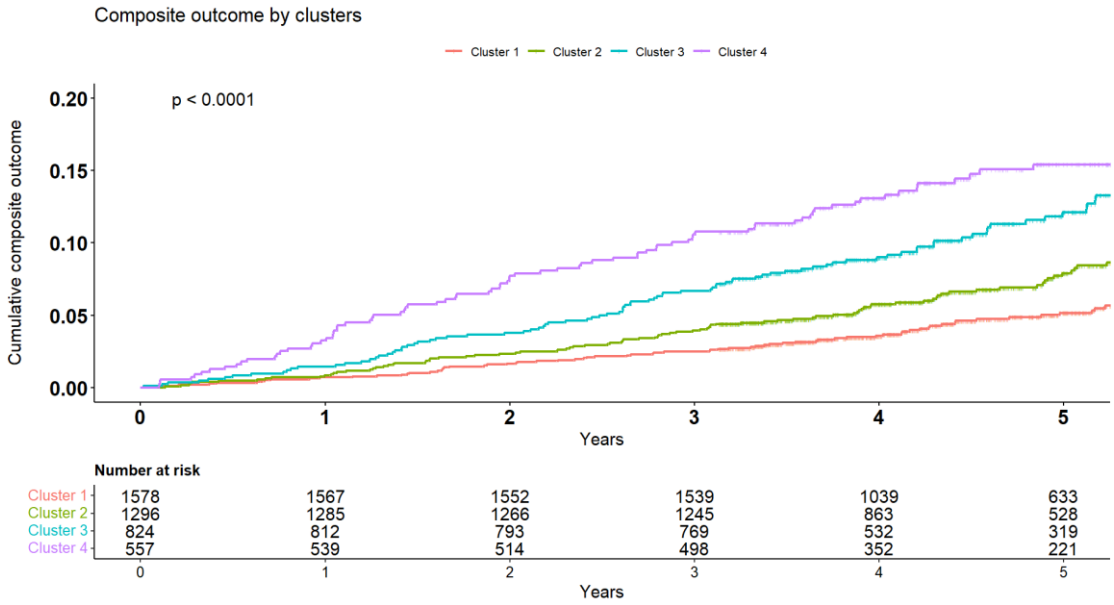


Figure 5 – Kaplan-Meier for each outcome

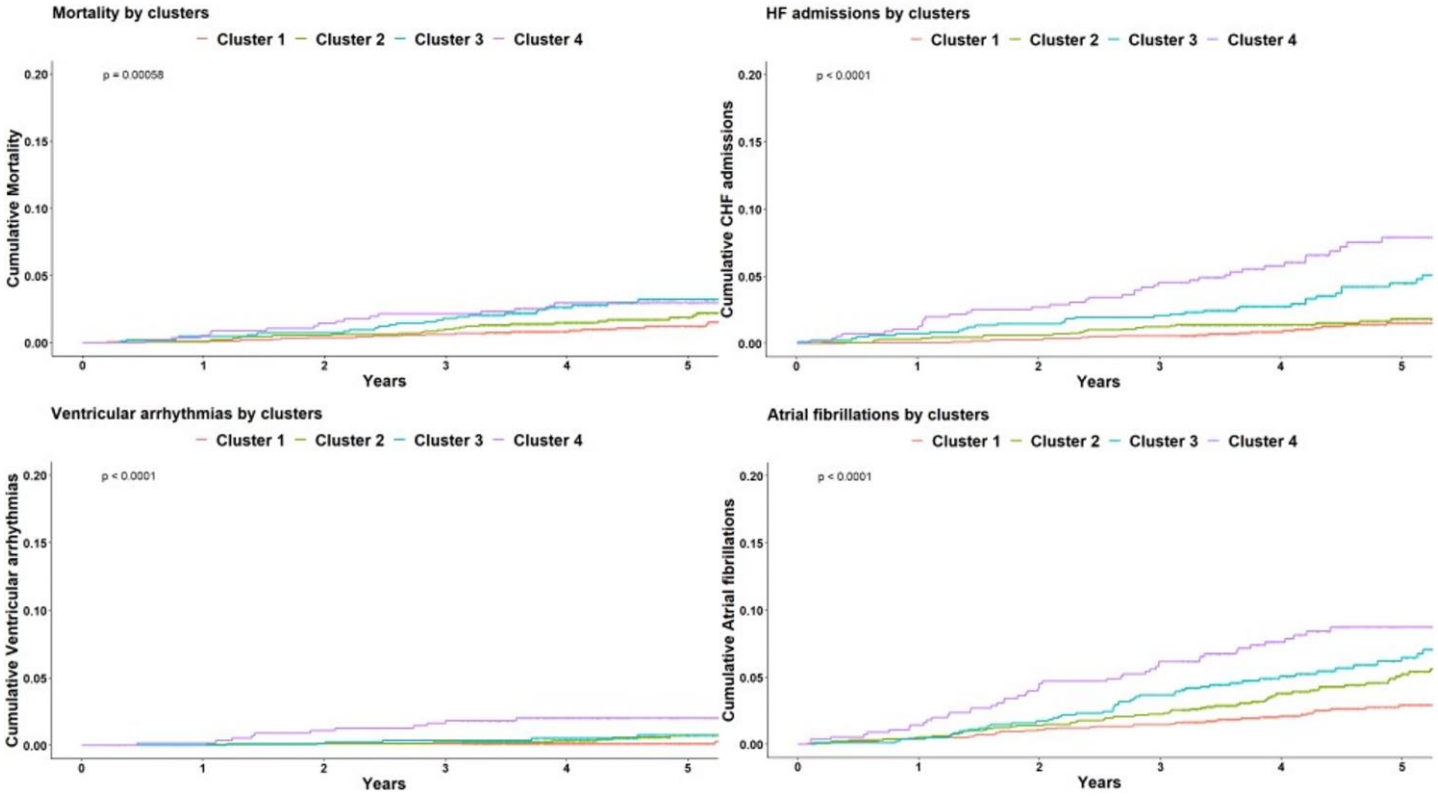


Figure 6a – AE reconstruction loss for training and test sets

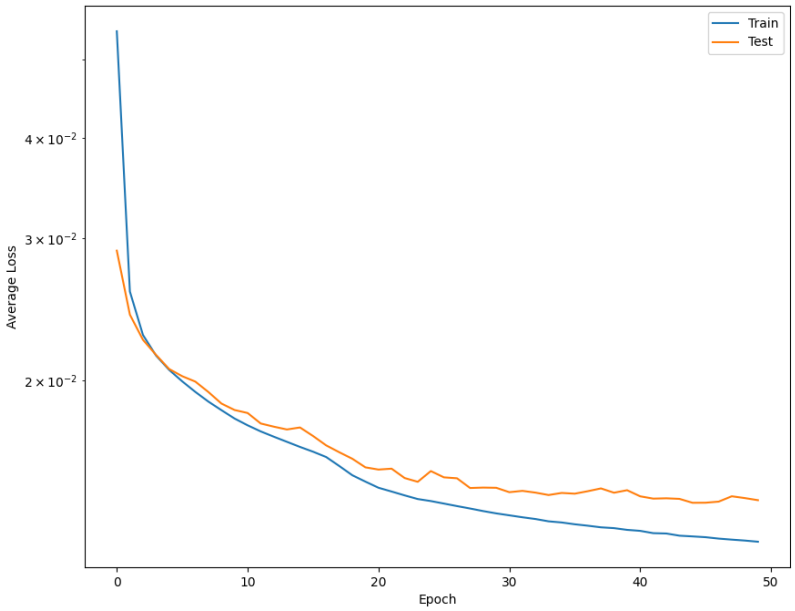
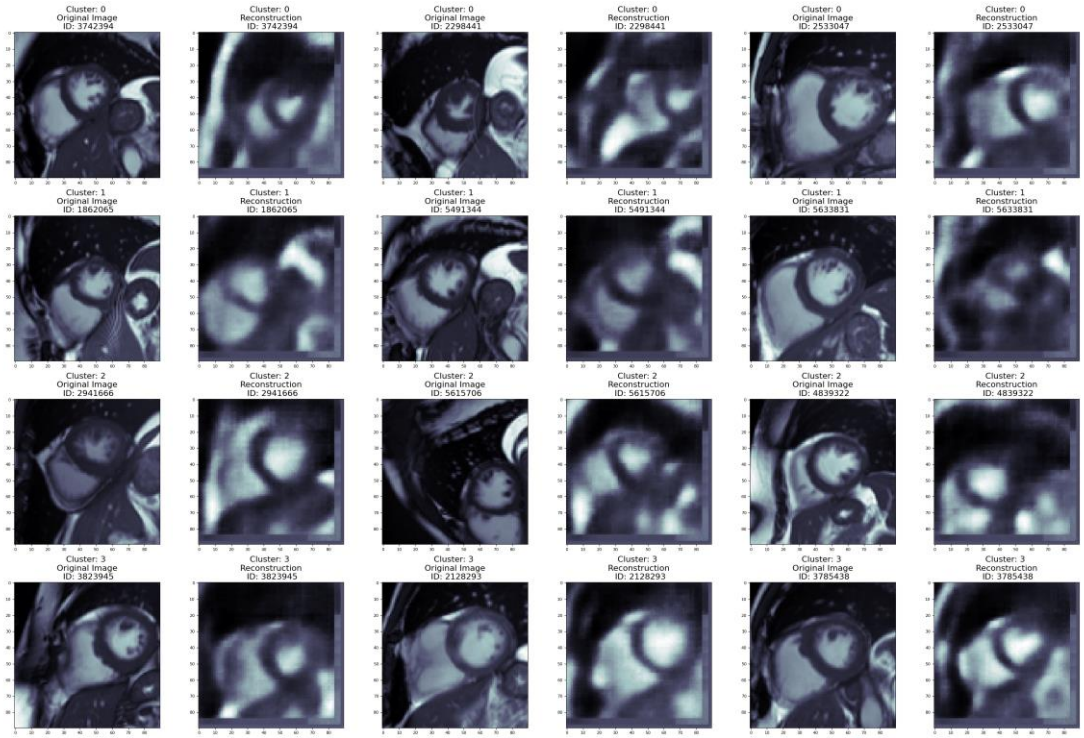
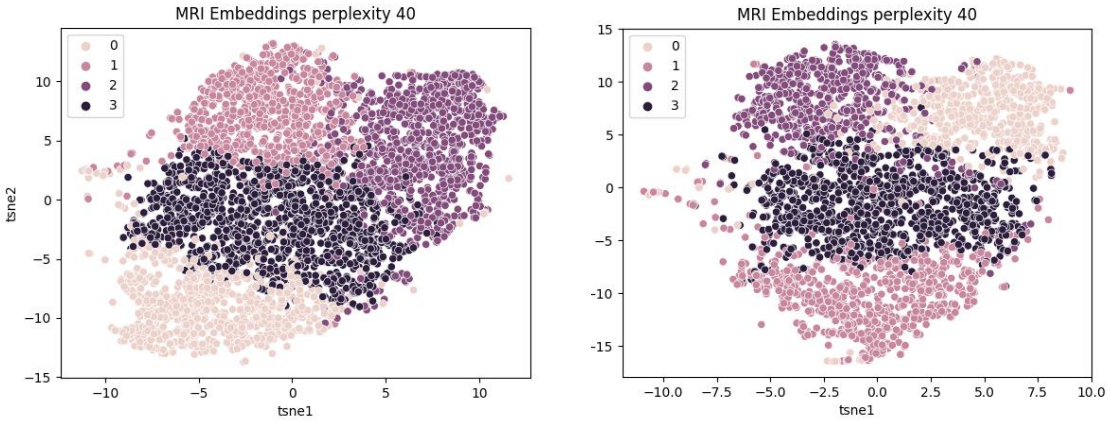


Figure 6b – Original and reconstructed images across all clusters



Example of reconstruction through Auto-Encoder, each row issues a different clusters (#clusters=4) and presented are original and reconstructed image interchangeably

Figure 7 – t-SNE on the clusters of embeddings



Left: T-SNE on "mean" method (20 epochs, series #5); Right: T-SNE on "series" method (20 epochs, series#15

Figure 8 – clusters comparison