Applying Ventricular Wall Shape and Motion Features from CMRI for Aiding Diagnosis of Cardiomyopathies

Stephani S. H. Costa[1], **Vagner Mendonça Gonçalves**[1,2], **Fátima L. S. Nunes**[1]

[1]Laboratory of Computer Applications for Health Care,
Escola de Artes, Ciências e Humanidades, Universidade de São Paulo
Rua Arlindo Bettio, 1000, São Paulo – SP, Brazil, 03828-000

[2]Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Campus São Paulo
Rua Pedro Vicente, 625, São Paulo – SP, Brazil, 01109-010

{stephani.henrique, fatima.nunes}@usp.br, vagner.goncalves@ifsp.edu.br

Abstract. *Cardiomyopathies are diseases usually characterized by dilation or hypertrophy of the heart muscle. Left Ventricle (LV) is the heart chamber most affected in most cases. Cardiac Cine Magnetic Resonance Imaging (CMRI) is a powerful tool applied for diagnosis of cardiomyopathies. Although some studies define descriptors based on CMRI images, usually they are related to clinical metrics. In this paper, we explored shape and motion features from the LV ventricular wall to define descriptors based on a priori knowledge about heart anomalies to build Supervised Machine Learning-based classification models capable of discriminating cases of dilated cardiomyopathy, hypertrophic cardiomyopathy, or those ones without anomalies associated with these diseases. The best classification model built and evaluated achieved* F_1*-score* $= 0.85 \pm 0.05$*, accuracy* $= 0.85 \pm 0.04$*, and AUC* $= 0.94 \pm 0.02$*. Our results are promising, indicating the potential of the approach for applications in computer-aided diagnosis systems.*

Keywords. *Cardiac Cine Magnetic Resonance Imaging, CMRI, Shape Features, Motion Features, Supervised Machine Learning, Left Ventricle, Ventricular Wall, Computer-Aided Diagnosis, Dilated Cardiomyopathy, Hypertrofic Cardiomyopathy.*

1. Introduction

The human heart works as an efficient, durable and reliable pump, being responsible for the blood that circulates throughout the body, providing tissue oxygenation, nutrition and waste removal [Kumar et al. 2020]. Among the heart chambers, the Left Ventricle (LV) is the chamber responsible for pumping blood out of the heart [Kumar et al. 2020; Whiteman et al. 2021]. The ventricular wall is formed by the heart muscle, the myocardium, covered by an inner membrane, the endocardium, and an outer membrane, the pericardium. The myocardium contracts during systole and relaxes during diastole to pump blood [Kumar et al. 2020].

Cardiomyopathies correspond to a heterogeneous group of diseases of the myocardium associated with electrical and/or mechanical dysfunctions that are usually, but not invariably, characterized by ventricular dilation or hypertrophy [Maron et al. 2006]. Cardiomyopathies and myocarditis resulted together in 410 thousand deaths worldwide in 2021 [Martin et al. 2024]. Among the different types of cardiomyopathies, the most common correspond to Dilated Cardiomyopathy (DC) and Hypertrophic Cardiomyopathy (HC), conditions that can lead to consequences such as heart failure, heart transplantation, or sudden death [Braunwald 2017; Izquierdo et al. 2021; Kumar et al. 2020; Maron et al. 2006].

DC is characterized by the enlargement of the cavities of the heart chambers, mainly the LV cavity, as well as systolic dysfunction. It is the most common cardiomyopathy in the world, being the main motivation for heart transplants [Braunwald 2017; Maron et al. 2006]. HC, in turn, can be characterized by hereditary factors related to abnormal thickening of the myocardium, without dilation of the LV, which leads to reduction of the ventricular cavity. It is a common cause of sudden death in adolescents and young adults, especially athletes [Elliott et al. 2014; Maron et al. 2006].

CMRI is an important tool to support the diagnosis of cardiomyopathies, providing a spatio-temporal analysis of the cardiac structure [Maron et al. 2006; Menchón-Lara et al. 2019]. In a CMRI exam, sequences of images of the patient's heart structure over a period of time are generated and analyzed by specialists. Slices are two-dimensional images of different sections of the heart at the same time point in the cardiac cycle. Frames, in turn, are two-dimensional images of the same section of the heart throughout the entire cardiac cycle [Menchón-Lara et al. 2019]. Visual fatigue caused by the large amount of images that must be continuously inspected is an aspect that has challenged the medical community, leading to less accuracy in the diagnoses [Waite et al. 2017]. In this context, Computer-Aided Diagnosis (CAD) approaches have been developed with the purpose of supporting specialists in decision-making regarding diagnosis [Ammar et al. 2021; Antonopoulos et al. 2021; Bhatia et al. 2022; Izquierdo et al. 2021; Liu et al. 2023; Moreno et al. 2019; Peña et al. 2021; Xiao et al. 2020; You et al. 2021; Zhang et al. 2023].

In this research, our main objective is to build and evaluate Supervised Machine Learning (SML)-based classification models for discriminating cases of DC, HC, and those ones without characteristic anomalies of these diseases (NA, from No Anomalies) through the application of shape and motion features from the LV ventricular wall segmented in CMRI images. We have applied feature descriptors presented by Delmondes [2022], as well as new feature descriptors implemented during this research. Thus, the main contribution of this article is a comparative study of the application of different SML algorithms in conjunction with features extracted from CMRI images, based on *a priori* knowledge about heart anomalies. Our results are promising and demonstrate the effectiveness of the approach for applications in CAD systems for cardiomyopathies.

In addition to this introductory section, this paper is organized as follows: in Section 2, we present related works; in Section 3, we present the materials and methods; in Section 4, we present and discuss the results; and, finally, in Section 5, we present our conclusions.

2. Related Work

The works summarized in this section were selected through an exploratory analysis of the scientific literature in the ACM Digital Library and IEEE Xplore Digital Library, using keywords related to cardiac magnetic resonance, such as MRI, CMR, cardiology, left ventricle, computer-aided diagnosis, and feature descriptors.

Moreno et al. [2019], Ammar et al. [2021], Peña et al. [2021], and Bhatia et al. [2022] proposed shape, motion, or spatio-temporal feature descriptors for CMRI images applied in conjunction with SML algorithms to build classification models capable of discriminate among cases of various heart diseases, including DC and HC. These researchers applied SML algorithms such as k-Nearest Neighbours (kNN), Random Forest (RFO), and Support

Vector Machine (SVM), obtaining models with good mean performance values (greater than 0.75) considering traditional metrics such as accuracy, precision and F_1-score.

Neisius et al. [2019], Antonopoulos et al. [2021], Izquierdo et al. [2021], Liu et al. [2023], and Zhang et al. [2023] applied radiomic features from CMRI T1 maps also in conjunction with SML algorithms for building classification models that, among other heart diseases, were also able to discriminate among DC, HC, and NA cases. SML algorithms such as AdaBoost, Decision Tree, Ensemble Learning, Gradient Boost (GBO), kNN, Logistic Regression (LRE), Multilayer Perceptron, Naive Bayes, RFO, and SVM were applied in these studies. Mean performance values greater 0.75 for metrics such as accuracy, precision, and Area Under the ROC Curve (AUC) were also reported by the researchers.

Deep Learning applications using CMRI images as input data for classifying cases of heart disease, including DC and HC, has also been explored in recent scientific research, as we can see in Snaauw et al. [2019], Xiao et al. [2020], and You et al. [2021]. These researchers applied algorithms such as Convolutional Neural Networks, Recurrent Neural Networks, and their variations, obtaining mean performance values greater than 0.90 for accuracy and AUC.

In general, results presented in related works have shown us that CMRI image features have great potential to enable effective CAD approaches for cardiomyopathies. Although several approaches have already been proposed, the development of new feature descriptors and the comparative study of different classification models are relevant contributions to the area and increase the list of approaches and techniques available that can be applied and reused in different contexts.

3. Materials and Methods

In this section, we describe the materials and methods applied in this research. In Section 3.1, we present essential information about the CMRI exams used. In Section 3.2, we present the features that made up the dataset. In Section 3.3, we present the classification model building process. Finally, in Section 3.4, we present the algorithms, strategies, and frameworks applied.

3.1. Materials

We used in this research a total of 384 CMRI exams from the archive of the Heart Institute of the *Hospital das Clínicas* of the *Faculdade de Medicina da Universidade de São Paulo* (InCor/HCFMUSP). To use these images, we obtained approval and authorization from the Committee for Ethics in Research on Human Beings of the School of Arts, Sciences and Humanities of the University of São Paulo (Certificate of Presentation for Ethical Consideration No. 49049021.1.0000.5390), as well as from the Committee for Ethics in Research of the HCFMUSP (Certificate of Presentation for Ethical Consideration No. 49049021.1.3001.0068).

Each exam has a diagnosis associated with it given by a specialist from InCor/HCFMUSP. HC, DC, and NA correspond to the diagnosis associated to 171, 112, and 101 cases, respectively. We used the slices and frames from the short axis view of each exam throughout the entire cardiac cycle. The region of interest of each image has dimensions of 256x256, covering the LV ventricular wall.

To compose the dataset, in addition to image features (Section 3.2), we also used the demographic data of patients' age and sex (Table 1) available in the CMRI exam metada.

Bergamasco et al. [2022] presented more details about the CMRI exams used, such as imaging specifications and the distribution of cases by age and sex.

Table 1. Demographic features.

ID	Feature	Description
$F1$	Age	Patient's age in years (min.=2, max.=87, avg.=46.9, std. dev.=17.1, med.=46).
$F2$	Sex	Patient's sex according exam metadata (#female=265, #male=119).

3.2. Dataset Features

We represented each CMRI exam through a vector of 31 features, including age and sex (Table 1), accompanied by its target class (the diagnosis associated with the exam). Eighteen of these features were computed using the descriptors presented by Delmondes [2022] (Table 2), which extract shape and motion features from the LV ventricular wall segmented in slices and frames of the exam.

Table 2. Features computed using the descriptors presented by Delmondes [2022].

ID	Features	Description
$F3$ to $F8$	Mean endocardium motion	Mean endocardium motion and respective standard deviation along the frames during the complete cardiac cycle ($F3$ and $F4$), systole ($F5$ and $F6$) and diastole ($F7$ and $F8$) considering all slices of the exam.
$F9$ to $F17$	Endocardium motion	Endocardium motion along the frames of the basal, medial and apical slices, considering the complete cardiac cycle ($F9$, $F10$, and $F11$), systole ($F12$, $F13$, and $F14$), and diastole ($F15$, $F16$, and $F17$).
$F18$ to $F20$	Mean ventricular wall area	Mean ventricular wall area considering all exam slices at the time points of end-diastole ($F18$), mid-systole ($F19$), and end-systole ($F20$).

Source: adapted from Delmondes [2022, p. 53].

We also developed 11 new feature descriptors with the objective of complementing and extending the discriminative ability of the descriptors presented by Delmondes [2022], as presented in Table 3.

Table 3. Features computed using the new descriptors developed in this research.

ID	Features	Description
$F21$ to $F23$	LV cavity volume	LV cavity volume at the time points of end-diastole ($F21$), mid-systole ($F22$) and end-systole ($F23$).
$F24$ to $F25$	Variation in LV cavity volume	Difference between the largest and the smallest LV cavity volume observed in systole ($F24$) and diastole ($F25$).
$F26$	Mean endocardium area	Mean endocardium surface area considering all exam frames.
$F27$ to $F30$	Mean endocardium minimum and maximum diameters	Mean and respective standard deviation of the endocardium minimum diameters ($F27$ and $F28$) and maximum diameters ($F29$ and $F30$) computed throughout the entire cardiac cycle.
$F31$	Variation in ventricular wall thickness	Difference between the maximum thicknesses of the ventricular wall computed in systole and diastole.

3.3. Classification Model Building Process

We applied the general classification model building process presented in Figure 1. This process is composed of five subprocesses, namely: A) segmentation; B) feature extraction; C) dataset preprocessing and splitting; D) SML algorithm hyperparameter tuning; and, E) classification model building. We conducted subprocesses D and E through an adaptation of the nested (or double) cross-validation [Stone 1974].

Segmentation (Subprocess A). Bergamasco et al. [2022] performed the segmentation of the LV ventricular wall applying a semi-automatic method to identify the contours

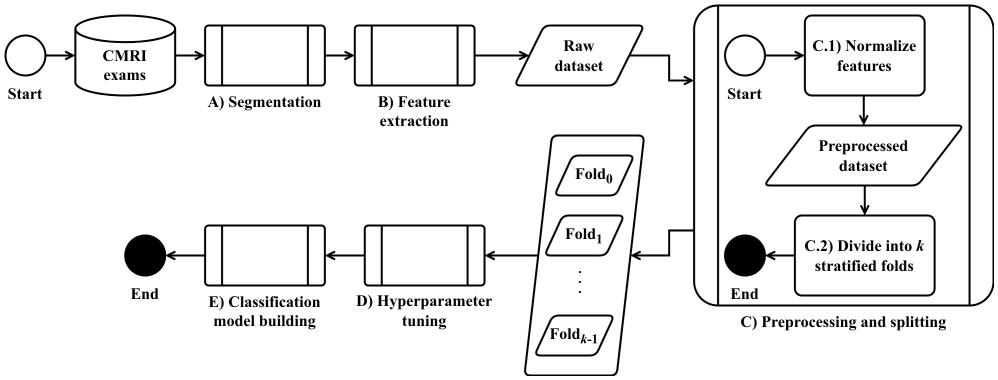
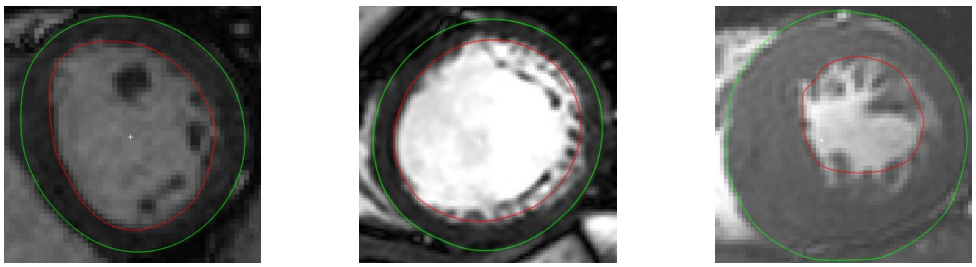


Figure 1. Classification model building process.

of the endocardium and epicardium (outermost layer of the pericardium), using the Medviso Segment software [Heiberg et al. 2010]. In Figure 2, we present examples of regions of interest from CMRI slices highlighting a NA case (Figure 2a), a DC case (Figure 2b), and a HC case (Figure 2c). The contours of the ventricular wall are shown in the images (epicardium in green and endocardium in red).



(a) NA (b) DC (c) HC

Figure 2. Region of interest from a CMRI slice for each diagnosis.

Feature extraction (Subprocess B). From the segmented images, we extract features applying the descriptors presented in Section 3.2.

Dataset preprocessing and splitting (Subprocess C). We normalized each dataset feature (Step C.1) adjusting the values in the range $[0.0, 1.0]$ using the Min-Max strategy [Pedregosa et al. 2011]. For the nominal attribute sex, we replaced the nominal values "female" and "male" with the numerical values 0.0 and 1.0, respectively. Next, we split the preprocessed dataset into k stratified folds (Step C.2). We applied $k = 10$ for all process executions.

SML algorithm hyperparameter tuning (Subprocess D). For each SML algorithm hyperparameter to be tuned, we defined a set of values to be tested. Then, we computed all possible combinations among the predefined hyperparameters values. With each combination of values, we performed a k-fold cross validation run with fold reservation.

In each cross validation round, we reserved a fold that was not used in either the training or testing of the classifier. From the remaining $k - 1$ folds, we used the union of $k - 2$ folds to train the SML algorithm in the round and the last remaining fold to test the classifier resulting from training. In each of the k rounds, a distinct fold was reserved. Still in each cross validation round, based on the test of the classifier built, we computed the

weighted macro-averaged F_1-score, considering the One-versus-the-Rest (OvR) approach for analyzing confusion matrices [Manning et al. 2008; Bishop 2006]. At the end of executing the k rounds, we computed the simple arithmetic mean of the k values of weighted macro-averaged F_1-score.

After executing all k-fold cross validations, one for each combination of hyperparameter values, we selected the set of hyperparameter values that maximized the mean of the k values of weighted macro-averaged F_1-score.

Classification model building (Subprocess E). We built the final classification model to be evaluated through a new k-fold cross validation, this time without fold reservation, applying the combination of SML algorithm hyperparameter values selected in Subprocess D.

In each round of this new cross validation, we trained the SML algorithm with the set of samples resulting from the union between the training and testing folds used in the corresponding round of Subprocess D cross validations. For testing the classifier built in the round, we used the reserved fold in the corresponding round of Subprocess D cross-validations.

From the predictions provided by the classifier built in each round, we computed the weighted macro-averaged precision, recall, and F_1-score values, as well as the overall accuracy and the macro-averaged Receiver Operating Characteristic (ROC) curve, with its respective AUC value [Manning et al. 2008; Bishop 2006]. After the execution of the k rounds, we computed the simple arithmetic mean of the k values of each performance metric, as well as the mean macro-averaged ROC curve.

3.4. Algorithms, Strategies, and Frameworks

In this research, we used the Python programming language, version 3.11.0, as well as the open source libraries Scikit-Learn [Pedregosa et al. 2011], version 1.3.0, and Imbalanced-Learn [Lemaître et al. 2017], version 0.11.0. Through different executions of the process presented in Section 3.3, we built, evaluated and compared 120 different classification models, exploring all possible combinations among the algorithms and strategies presented in Table 4.

Table 4. Algorithms and strategies applied in this research.

SML algorithm	Feature subset	Strategy of dimension reduction	Strategy of class balancing
GBO GPR LRE RFO SVM	A = { Image features } B = { Age, Image features } C = { Sex, Image features } D = { Age, Sex, Image features }	None LDA	None Undersamplig Upsampling

We applied five SML algorithms, namely: Gaussian Process (GPR) [Rasmussen and Nickisch 2010], Gradient Boosting (GBO) [Zhang et al. 2023], Logistic Regression (LRE) [Izquierdo et al. 2021; Zhang et al. 2023], Random Forest (RFO) [Antonopoulos et al. 2021; Izquierdo et al. 2021; Liu et al. 2023; Zhang et al. 2023], and Support Vector Machine (SVM) [Izquierdo et al. 2021; Liu et al. 2023; Zhang et al. 2023]. The feature subsets tested consisted of combinations among the union of the features presented in Tables 2 and 3 (image features) and patients' demographic features (Table 1). To better reference the

feature subsets, we named the four combinations of features explored in this research using the letters A, B, C and D, as presented in Table 4.

Since the dataset is unbalanced, we also performed variations of the process by applying class balancing strategies only to the training folds in the cross-validation rounds. To this end, we applied the random undersampling and the SMOTE [Chawla et al. 2002] approaches. We also applied different dimensionality reduction strategies to the training and testing folds in each cross-validation round. In this paper, we only address the dimensionality reduction strategy that generated the best results, Linear Discriminat Analysis (LDA) [Hastie et al. 2009].

In order to analyze the statistical significance of the differences observed among the performances of the different classification models built and evaluated, we applied the pairwise Mann-Whitney U Test [MacFarland and Yates 2016], testing the veracity of the null hypothesis H_0: the arithmetic mean among the k performance values achieved in the respective k cross-validation rounds applied to build the classification model is equal for the two classification models compared to each other. We applied a two-sided alternative hypothesis and p-value < 0.05. For each classification model, we considered as a sample the set of ten mean performance values computed, respectively, in the ten cross-validation rounds referring to the classification model building subprocess (Subprocess E). In this paper, we highlight the results of the pairwise Mann-Whitney U test applied to the performance values computed through the weighted macro-averaged F_1-score.

4. Results and Discussion

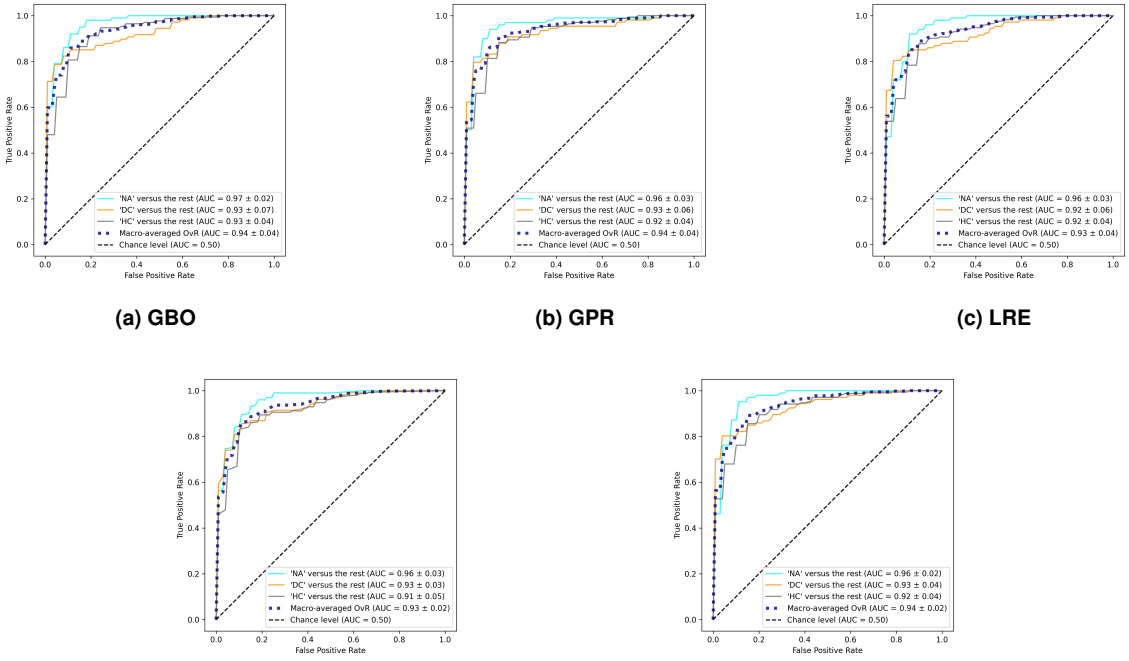
In Table 5, we present the performance values achieved by the best classification model per SML algorithm and per feature subset applied. In general, most models achieved good mean classification performance, with mean values of F_1-score, precision, recall, and accuracy greater than or equal to 0.80, as well as AUC greater than or equal to 0.90. The metrics used are traditional in classification studies and were chosen based on the literature [Moreno et al. 2019; Liu et al. 2023; Izquierdo et al. 2021].

The individual or joint application of the demographic features of age and sex combined with the image features (feature subsets B, C or D) resulted in classification models with higher mean performance values when compared to the models built with the application of only image features (feature subset A). The best classification models resulted from the joint application of age, sex and image features (feature subset D). The best model overall was the one built on the SVM algorithm that achieved mean performance values F_1-score $= 0.85 \pm 0.05$, accuracy $= 0.85 \pm 0.04$, and AUC $= 0.94 \pm 0.02$. Once a specific feature subset is fixed, we can observe that the difference among the mean performance values achieved by the classification models is low, comparing the different SML algorithms. This difference is at most 0.30 for the models represented in Table 5.

In Fig. 3, we present the mean ROC curves per class and the mean macro-averaged ROC curve computed for the best classification models highlighted in gray shading in Table 5. Considering the mean macro-averaged ROC curves of the best classification models, the performance variability when comparing the different applied SML algorithms was low. The ROC analysis by class also allowed us to observe that, although most classification models performed better in discriminating NA cases in relation to the other classes, the discriminative ability for all classes was good for all models (AUC greater than 0.90).

Table 5. Mean performance values and respective standard deviations achieved by the best classification model per SML algorithm and per feature subset. In gray shading we highlight the line referring to the best model per SML algorithm. In bold, we highlight the values referring to the best classification model overall.

SML algorithm	Features	F_1-score	Precision	Recall	Accuracy	AUC
GBO	A	0.76 ± 0.10	0.78 ± 0.11	0.76 ± 0.10	0.76 ± 0.10	0.89 ± 0.06
GBO	B	0.83 ± 0.06	0.84 ± 0.06	0.83 ± 0.06	0.83 ± 0.06	0.92 ± 0.04
GBO	C	0.79 ± 0.06	0.81 ± 0.06	0.79 ± 0.06	0.79 ± 0.06	0.91 ± 0.03
GBO	D	0.83 ± 0.06	0.85 ± 0.06	0.84 ± 0.06	0.84 ± 0.06	0.94 ± 0.04
GPR	A	0.79 ± 0.03	0.81 ± 0.04	0.79 ± 0.03	0.79 ± 0.03	0.91 ± 0.03
GPR	B	0.81 ± 0.05	0.83 ± 0.06	0.82 ± 0.05	0.82 ± 0.05	0.92 ± 0.03
GPR	C	0.80 ± 0.05	0.82 ± 0.05	0.80 ± 0.05	0.80 ± 0.05	0.91 ± 0.04
GPR	D	0.84 ± 0.07	0.86 ± 0.06	0.84 ± 0.07	0.84 ± 0.07	0.94 ± 0.04
LRE	A	0.79 ± 0.07	0.81 ± 0.06	0.79 ± 0.07	0.79 ± 0.07	0.90 ± 0.04
LRE	B	0.82 ± 0.06	0.83 ± 0.06	0.82 ± 0.06	0.82 ± 0.06	0.92 ± 0.04
LRE	C	0.81 ± 0.04	0.84 ± 0.04	0.81 ± 0.04	0.81 ± 0.04	0.91 ± 0.04
LRE	D	0.84 ± 0.07	0.86 ± 0.06	0.84 ± 0.07	0.84 ± 0.07	0.93 ± 0.04
RFO	A	0.77 ± 0.08	0.79 ± 0.07	0.77 ± 0.08	0.77 ± 0.08	0.90 ± 0.06
RFO	B	0.82 ± 0.06	0.83 ± 0.06	0.82 ± 0.05	0.82 ± 0.05	0.92 ± 0.04
RFO	C	0.78 ± 0.09	0.80 ± 0.09	0.79 ± 0.09	0.79 ± 0.09	0.90 ± 0.05
RFO	D	0.83 ± 0.05	0.84 ± 0.05	0.83 ± 0.05	0.83 ± 0.05	0.93 ± 0.02
SVM	A	0.78 ± 0.05	0.80 ± 0.04	0.78 ± 0.05	0.78 ± 0.05	0.91 ± 0.04
SVM	B	0.81 ± 0.08	0.83 ± 0.07	0.81 ± 0.08	0.81 ± 0.08	0.93 ± 0.04
SVM	C	0.81 ± 0.05	0.83 ± 0.05	0.81 ± 0.05	0.81 ± 0.05	0.91 ± 0.05
SVM	**D**	$\mathbf{0.85 \pm 0.05}$	$\mathbf{0.87 \pm 0.05}$	$\mathbf{0.85 \pm 0.04}$	$\mathbf{0.85 \pm 0.04}$	$\mathbf{0.94 \pm 0.02}$



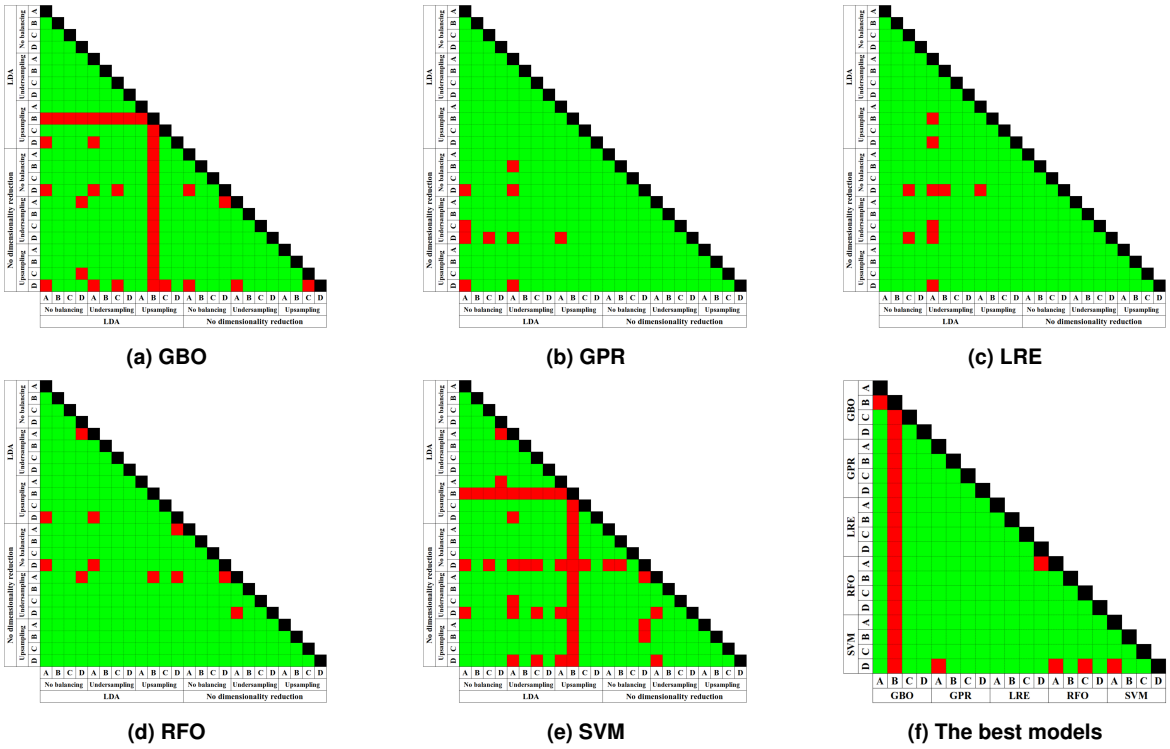
(a) GBO **(b) GPR** **(c) LRE**

(d) RFO **(e) SVM**

Figure 3. Mean ROC curves per class and mean macro-averaged ROC curve computed for the best classification model by SML algorithm.

In figures from 4a to 4e, we illustrate the results of the pairwise Mann-Whitney U tests that we applied to all classification models built and evaluated per SML algorithm. Green and red colors indicate statistically non-significant and significant differences, re-

spectively. These models are the results of different combinations among feature subsets, dimensionality reduction strategies, and class balancing strategies, presented in Table 4. In Figure 4f, we illustrate the results of the application of the pairwise Mann-Whitney U test to the 20 classification models represented in Table 5.



(a) GBO (b) GPR (c) LRE

(d) RFO (e) SVM (f) The best models

Figure 4. Results of the pairwise U test of the null hypothesis H_0 on all classification models built and tested by SML algorithm (figures from 4a to 4e), and on the 20 models presented in Table 5 (Figure 4f).

The U tests we carried out showed us that the difference among the mean performances of most classification models, obtained through different combinations between SML algorithm, feature subset, dimension reduction strategies, and class balancing strategies, was not significant. This fact highlights the robustness of the presented results in terms of performance of the classification models.

In previous work, Delmondes [2022] presented a Content-Based Image Retrieval (CBIR) approach to retrieve DC, HC or NA cases similar to a case given as a reference. The researcher applied the same images that we used in this work and the features presented in the Tables 2 and 1. One of the limitations that Delmondes [2022] observed regarding the CBIR approach was the need to stratify cases by patients' age and sex to achieve good retrieval performance.

In Table 6 we present the difference between the performance values observed in the CBIR approach of Delmondes [2022] and in our classification approach when the application of only image features is compared with the application of image features in conjunction with demographic features.

Even though this is a qualitative and indirect comparison, since the approaches are different, we observed that our classification approach enabled models with higher performance. Furthermore, our classification approach is less dependent on the demographic fea-

Table 6. Best performance values achieved in the CBIR approach of Delmondes [2022] and in our classification approach for each feature subset applied.

Task	Performance metric	Features	DC	HC	NA	Mean
CBIR [Delmondes 2022]	PxR* AUC	Table 2	0.48	0.50	0.33	0.44
		Tables 1 and 2	0.83	0.69	0.46	0.66
Classification (our approach)	ROC AUC	Table 2†	0.83	0.80	0.95	0.86
		Tables 1 and 2†	0.83	0.81	0.95	0.86
		Tables 1, 2, and 3	0.93	0.92	0.96	0.94

* Precision versus Recall curve.
† We do not present or discuss these results in detail in this paper.

tures of age and sex, since most classification models did not present statistically significant performance differences when compared to each other, as we evidenced through the application of the pairwise Mann-Whitney U test. Our best mean classification performance value, considering only the features presented in tables 2 and 1 applied together (AUC $= 0.86$), was 30.3% better than the best mean retrieval performance observed by Delmondes [2022] with the application of the same features (PxR AUC $= 0.66$). The difference between the mean performance values, comparing the application of only the image features from Table 2 with their application together with the demographic features, was 50.0% in the CBIR approach (PxR AUC increased from 0.44 to 0.66), and 0.0% in our classification approach (the mean AUC value maintained at 0.86). Thus, the classification approach that we present in this paper also contributes as a better option for application in CAD systems for cardiomyopathies than CBIR approaches, taking CMRI images as input.

Furthermore, it is important to acknowledge some limitations of our study. Firstly, our research is constrained by a limited number of cases, which hinders the use of more advanced techniques such as Deep Learning (DL), for example. Additionally, we face the challenge of class imbalance, which we attempted to mitigate by applying a balancing technique with upsampling. Another point to be highlighted is that our study is single-center, and studies with more diversified and multicentric databases are necessary for a better assessment of the models' generalization capability.

5. Conclusion

We evaluated and compared 120 SML classification models for discriminating DC, HC, and NA cases based on CMRI image features. Results showed us that, even though the demographic features of age and sex increasing performance values when comparing models that used them with models that did not use them, there was no statistically significant difference among the performances achieved by the most classification models built and compared to each other.

The best classification model built and evaluated, based on SVM and the application of all available features, achieved F_1-score $= 0.85 \pm 0.05$, accuracy $= 0.85 \pm 0.04$, and AUC $= 0.94 \pm 0.02$. Our results are promising, indicating the potential of the approach for practical applications. An important next step would be the development of clinically viable interfaces for effective implementation. As future work, we intend to combine the approach presented in this paper with LV segmentation and content-based retrieval approaches for cardiomyopathy cases, also under development in other researches in our laboratory, to implement a CAD system of cardiomyopathies applicable in medical routine.

Acknowledgements

This work was supported in part by the Brazilian National Council for Scientific and Technological Development (CNPq) [grant numbers 307710/2022-0 and 138378/2023-1] and São Paulo Research Foundation (FAPESP) – National Institute of Science and Technology – Medicine Assisted by Scientific Computing (INCT-MACC) [grant 2014/50889-7].

References

Ammar, A., Bouattane, O., and Youssfi, M. (2021). Automatic cardiac cine MRI segmentation and heart disease classification. *Comput. Med. Imaging Graph.*, 88.

Antonopoulos, A. S., Boutsikou, M., Simantiris, S., Angelopoulos, A., and et al. (2021). Machine learning of native T1 mapping radiomics for classification of hypertrophic cardiomyopathy phenotypes. *Sci. Rep.*, 11(1).

Bergamasco, L. C. C., Lima, K. R. P. S., Rochitte, C. E., and Nunes, F. L. S. (2022). A bipartite graph approach to retrieve similar 3D models with different resolution and types of cardiomyopathies. *Expert Syst. Appl.*, 193.

Bhatia, D., Kanakatte, A., and Ghose, A. (2022). Cardiac anomaly detection from cine MRI images using physiological features and random forest classifier. In *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* IEEE.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, New York, 1 edition.

Braunwald, E. (2017). Cardiomyopathies: an overview. *Circ. Res.*, 121(7).

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-Sampling Technique. *J. Artif. Intell. Res.*, 16(1).

Delmondes, P. H. M. (2022). Sistemas de auxílio ao diagnóstico de cardiomiopatias: uma abordagem baseada em descritores multi-slice e multi-frame. Dissertação, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo. [In Portuguese].

Elliott, P. M., Anastasakis, A., Borger, M. A., Borggrefe, M., and et al. (2014). 2014 ESC guidelines on diagnosis and management of hypertrophic cardiomyopathy: the task force for the diagnosis and management of hypertrophic cardiomyopathy of the European Society of Cardiology (esc). *Eur. Heart J.*, 35(39).

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer, 2 edition.

Heiberg, E., Sjögren, J., Ugander, M., Carlsson, M., Engblom, H., and Arheden, H. (2010). Design and validation of Segment - freely available software for cardiovascular image analysis. *BMC Med. Imaging*, 10(1).

Izquierdo, C., Casas, G., Martin-Isla, C., Campello, V. M., and et al. (2021). Radiomics-based classification of left ventricular non-compaction, hypertrophic cardiomyopathy, and dilated cardiomyopathy in cardiovascular magnetic resonance. *Front. Cardiovasc. Med.*, 8.

Kumar, V., Abbas, A. K., and Aster, J. C. (2020). *Robbins & Cotran, Pathologic Basis of Disease*. Elsevier, 10 edition.

Lemaître, G., Nogueira, F., and Aridas, C. K. (2017). Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Mach. Learn. Res.*, 18(17).

Liu, Q., Lu, Q., Chai, Y., Tao, Z., Wu, Q., Jiang, M., and Pu, J. (2023). Papillary-muscle-derived radiomic features for hypertrophic cardiomyopathy versus hypertensive heart disease classification. *Diagnostics*, 13(9).

MacFarland, T. W. and Yates, J. M. (2016). Mann–Whitney U test. In MacFarland, T. W. and Yates, J. M., editors, *Introduction to Nonparametric Statistics for the Biological Sciences Using R*, pages 103–132. Springer, Cham, 1 edition.

Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, Cambridge.

Maron, B. J., Towbin, J. A., Thiene, G., Antzelevitch, C., and et al. (2006). Contemporary definitions and classification of the cardiomyopathies: an American Heart Association scientific statement from the Council on Clinical Cardiology, Heart Failure and Transplantation Committee; Quality of Care and Outcomes Research and Functional Genomics and Translational Biology interdisciplinary working groups; and Council on Epidemiology and Prevention. *Circulation*, 113(14).

Martin, S. S., Aday, A. W., Almarzooq, Z. I., Anderson, C. A. M., and et al. (2024). 2024 heart disease and stroke statistics: a report of US and global data from the American Heart Association. *Circulation*, 149(8).

Menchón-Lara, R. M., Wattenberg, F. S., Higuera, P. C., Fernández, M. M., and López, C. A. (2019). Reconstruction techniques for cardiac cine MRI. *Insights Imaging*, 10.

Moreno, A., Rodriguez, J., and Martínez, F. (2019). Regional multiscale motion representation for cardiac disease prediction. In *Symp. Image Signal Process. Artif. Vis.* IEEE.

Neisius, U., El-Rewaidy, H., Nakamori, S., Rodriguez, J., Manning, W. J., and Nezafat, R. (2019). Radiomic analysis of myocardial native T1 imaging discriminates between hypertensive heart disease and hypertrophic cardiomyopathy. *JACC: Cardiovasc. Imaging*, 12(10).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., and et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.*, 12(85).

Peña, H., Gómez, S., Romo-Bucheli, D., and Martinez, F. (2021). Cardiac disease representation conditioned by spatio-temporal priors in cine-MRI sequences using generative embedding vectors. In *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* IEEE.

Rasmussen, C. E. and Nickisch, H. (2010). Gaussian processes for machine learning (GPML) toolbox. *J. Mach. Learn. Res.*, 11.

Snaauw, G., Gong, D., Maicas, G., van den Hengel, A., Niessen, W. J., Verjans, J., and Carneiro, G. (2019). End-to-end diagnosis and segmentation learning from cardiac magnetic resonance imaging. In *IEEE Int. Symp. Biomed. Imaging*. IEEE.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc.*, 36(2):111–133.

Waite, S., Kolla, S., Jeudy, J., Legasto, A., Macknik, S. L., Martinez-Conde, S., Krupinski, E. A., and Reede, D. L. (2017). Tired in the reading room: the influence of fatigue in Radiology. *J. Am. Coll. Radiol.*, 14(2).

Whiteman, S., Alimi, Y., Carrasco, M., Gielecki, J., Zurada, A., and Loukas, M. (2021). Anatomy of the cardiac chambers: a review of the left ventricle. *Transl. Res. Anat.*, 23.

Xiao, J., Liu, X., Tao, Q., and Chen, J. (2020). Learning motion based auxiliary task for cardiomyopathy recognition with cardiac magnetic resonance images. In *Int. Conf. Comput. Sci. Appl. Eng.*

You, Y., Viktorovich, L. A., Qiu, J., Nikolaevich, K. A., and Vladimirovich, B. Y. (2021). Cardiac magnetic resonance image diagnosis of hypertrophic obstructive cardiomyopathy based on a double-branch neural network. *Comput. Methods Programs Biomed.*, 200.

Zhang, X., Cui, C., Zhao, S., Xie, L., and Tian, Y. (2023). Cardiac magnetic resonance radiomics for disease classification. *Eur. Radiol.*, 33(4).