Cardiac Pathology Prediction

IMA205 - Challenge 2025

Apprentissage pour l'image et la reconnaissance d'objets

User: ***gochanson***
CABRERA Joel - angel.cabreradechia@telecom-paris.fr

Contents

1 Context

The goal of this challenge is to classify subjects based on their cardiac magnetic resonance imaging (CMRI) scans into one of five diagnostic categories:

1. Healthy controls
2. Myocardial infarction
3. Dilated cardiomyopathy
4. Hypertrophic cardiomyopathy
5. Abnormal right ventricle

These categories reflect a variety of structural and functional heart conditions that can be differentiated through careful analysis of cardiac morphology and function as captured in CMRI data.

To tackle this classification task, we first aim to extract a set of meaningful features from the medical images. These features are designed to reflect relevant physiological properties such as ventricular volumes, myocardial thickness, and ejection fractions. Once extracted, these features will serve as input to various machine learning models, including Random Forests, XGBoost, Support Vector Machines, and Logistic Regression, to predict the corresponding class label.

This pipeline emulates a clinically-inspired approach, where diagnostic decisions rely on interpretable and quantifiable markers derived from imaging, while leveraging the predictive power of modern machine learning classifiers.

2 Segmentation Methodology

In this section, we detail the method implemented for segmenting the left ventricle (LV), which is critical for extracting robust features.

Unlike more complex methodologies such as iterative thresholding for endocardial contour extraction or active contour models (ACM or snakes) for epicardial contour extraction, described in [1], our method simplifies significantly by directly leveraging provided myocardium segmentation masks.

2.1 Method Overview

As illustrated in Figure 1, the LV (blue region) is surrounded by the myocardium (green region). Given that myocardium segmentation is available, we will only use that information to easily extract the desired area.

2.2 Segmentation Steps

The simplified segmentation process involves the following sequential steps:

1. **Myocardium Mask Utilization:** Directly use the provided myocardium segmentation mask.

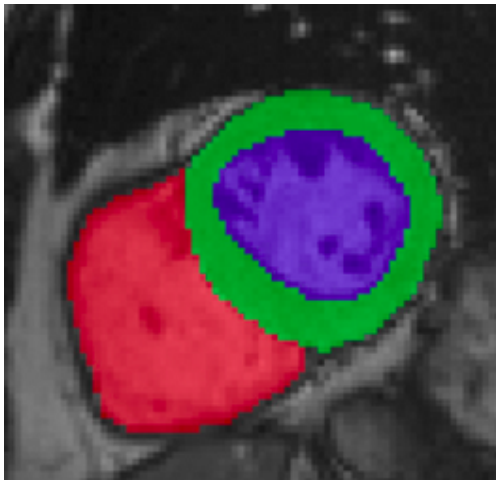

Figure 1: Example of the LV (blue) surrounded by the myocardium (green).

2. **Convex Hull Computation:** Compute a convex hull around the myocardium to define a smooth enclosed region.
3. **Myocardium Exclusion:** Subtract the myocardium region from its convex hull, effectively isolating the LV cavity.
4. **Connected Component Filtering:** Retain only the largest connected component to ensure anatomical continuity.
5. **Hole Filling and Final Convex Hull:** Apply morphological operations to fill internal holes, followed by a final convex hull to yield a clean and anatomically plausible LV mask.

The steps of the proposed segmentation method are illustrated in Figure 2, showing the evolution from the original image to the final LV mask.

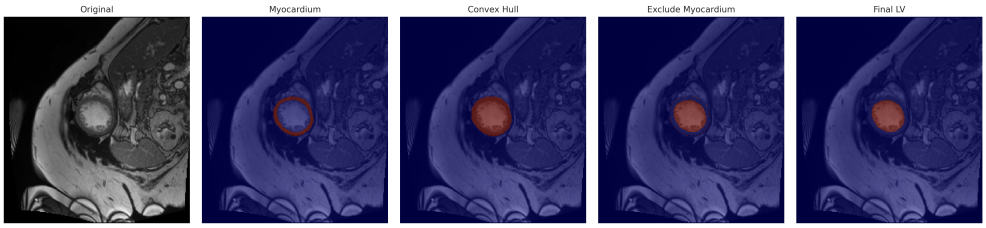

Figure 2: Visual representation of the simplified left ventricle segmentation steps. From left to right: (1) original image, (2) myocardium mask, (3) convex hull computation, (4) myocardium exclusion, and (5) final LV mask after morphological refinement.

It is pertinent to mention that this is easily computable due to the availability of the segmentation of the myocardium. In other contexts, we should have applied more complex methodologies such as iterative thresholding, region-growing based on intensity, among others. Our approach provides a simpler, computationally efficient, and robust alternative.

2.3 Evaluation of LV Segmentation Accuracy

To evaluate the performance of our proposed left ventricle (LV) segmentation method, we computed the Dice similarity coefficient (DSC) on the training set, where full ground truth annotations are available. The Dice score is defined as:

$$DSC = \frac{2|P \cap G|}{|P| + |G|} \tag{1}$$

where P is the predicted LV mask and G is the ground truth mask (label 3 in our dataset).

Importantly, our method assumes that a segmentation of the myocardium (label 2) is available, which is consistent with the information provided at test time. Thus, the evaluation measures how well the LV cavity can be reconstructed from the myocardium mask alone, without relying on image intensities or learned models.

After running our method across all slices of the training set, and excluding slices where the ground truth LV cavity is not present, we obtained a mean Dice score of **0.9999**. This near-perfect score confirms that, under the assumption of a known myocardium mask, our approach produces highly accurate and geometrically consistent LV segmentations.

3 Classification

The selected features for classification encompass both functional and anatomical cardiac descriptors. These include end-diastolic and end-systolic volumes for the left and right ventricular cavities (`LVC_ED_Volume`, `LVC_ES_Volume`, `RVC_ED_Volume`, `RVC_ES_Volume`), as well as the left and right ejection fractions (`Ejection_Fraction_Left`, `Ejection_Fraction_Right`). These variables are core indicators of ventricular performance and are widely used in clinical practice to assess contractility and diagnose heart failure [2, 3, 4].

To compute the **ejection fraction (EF)** of the left and right ventricles, we used the standard clinical formula:

$$EF = \frac{V_{ED} - V_{ES}}{V_{ED}} \times 100 \tag{2}$$

where:

- V_{ED} is the end-diastolic volume (maximum ventricular volume),
- V_{ES} is the end-systolic volume (minimum ventricular volume).

These volumes were estimated from the segmentation masks by counting the number of voxels in the cavity at the ED and ES phases, and multiplying by the physical voxel volume:

$$V = VoxelCount \times VoxelSize_x \times VoxelSize_y \times SliceThickness \tag{3}$$

In addition, we included myocardial thickness at both ED and ES phases (`Myocardium_Thickness_ED`, `Myocardium_Thickness_ES`) as structural features. Patient-specific factors such as height and weight were also included.

3.1 Classifier Performance without Data Augmentation

In this experiment, we compare four classifiers: Random Forest, XGBoost, SVM and Logistic Regression. These classifiers were splitted on an 80 %/20 % train/validation of the raw feature set. Each model was tuned via 5-fold grid search over its hyperparameters. Table 1 summarizes the selected parameters and the validation accuracies.

Table 1: Validation accuracy and best hyperparameters for each classifier, without data augmentation.

Model	**Best Hyperparameters**	**Val. Accuracy**
Random Forest	`n_estimators = 100`, `max_depth = None`, `min_samples_split = 2`	90.0 %
XGBoost	`n_estimators = 200`, `max_depth = 2`, `learning_rate = 0.1`	95.0 %
SVM (linear)	`C = 0.1`	100.0 %
Logistic Regression (lbfgs)	`C = 10`	100.0 %

These results show that, even without any data augmentation, the feature space is highly separable: Random Forest achieves 90.0 % accuracy, XGBoost 95.0 %, and both SVM and Logistic Regression reach perfect scores on the validation split. We can see the confusion matrices in the Figure 3.

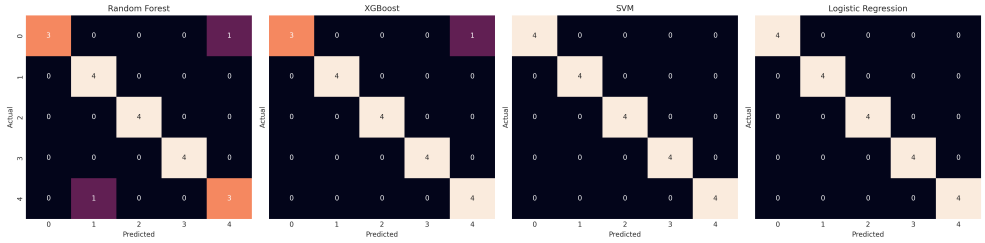

Figure 3: Confusion matrices for the four tuned classifiers (Random Forest, XGBoost, SVM and Logistic Regression) on the held-out validation set.

While the perfect validation accuracy suggests excellent results, it may also indicate potential overfitting, which can be tackled using data augmentation techniques.

3.2 Classifier Performance with Data Augmentation

To assess the impact of data augmentation on classifier robustness, we applied the following transformations to each training subject (five augmentations per original sample). The transformations were based on the ones used in [4].

1. **RandomAffine:** scaling in the range $[0.6, 1.4]$, rotation up to $\pm 5°$, translation up to 5mm in the LR and AP axes.
2. **RandomFlip:** horizontal (LR) or vertical (AP) flip with probability 0.5.

Features were then extracted from both original and augmented subjects using the same LV segmentation pipeline. We retrained and tuned the four classifiers under the identical grid-search setup, using an 80/20 train/validation split. Table 2 summarizes the best hyperparameters and validation accuracies with data augmentation.

Table 2: Validation accuracy and best hyperparameters for each classifier, with data augmentation.

Model	**Best Hyperparameters**	**Val. Accuracy**
Random Forest	`n_estimators=200`, `max_depth=None`, `min_samples_split=2`	93.33 %
XGBoost	`n_estimators=200`, `max_depth=4`, `learning_rate=0.1`	97.50 %
SVM (RBF)	`C=10, kernel='rbf'`	95.00 %
Logistic Regression (lbfgs)	`C=10`	90.83 %

These results indicate that data augmentation improves generalization for XGBoost and Random Forest, while SVM and Logistic Regression remain competitive. The highest validation accuracy (97.5 %) is achieved by XGBoost, suggesting that it benefits most from the diversified training set. Figure 4 shows the confusion matrices for each classifier.

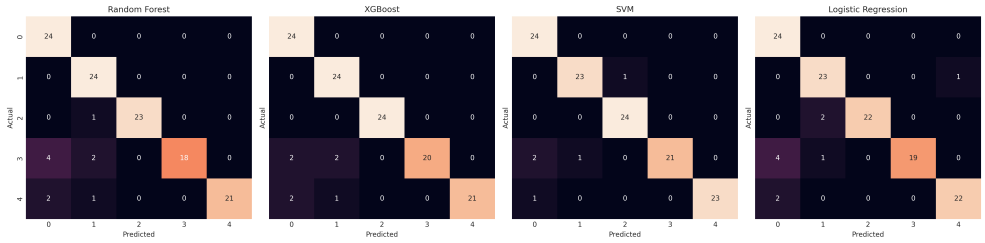

Figure 4: Confusion matrices for the four classifiers (Random Forest, XGBoost, SVM, and Logistic Regression) trained on the augmented dataset.

3.3 Features Importance

To gain insight into which features contributed most to the classification task, we examined the feature importances obtained from Random Forest and XGBoost.

As shown in Figure 5, both models strongly emphasize the relevance of the left ventricular ejection fraction (`Ejection_Fraction_Left`), followed by the right ventricular ejection fraction and end-systolic volumes of both ventricles. These features are physiologically meaningful, as they capture the contractile performance and volumetric status of the heart during different phases of the cardiac cycle. In addition, myocardial thickness measurements at both ED and ES phases were also among the most important features, highlighting their value in assessing structural heart function. The central role of left ventricular ejection fraction (LVEF) in assessing and

managing heart failure has been well established in clinical literature. In particular, reduced LVEF ($\leq 40\%$) is a defining criterion for heart failure with reduced ejection fraction (HFrEF), a condition characterized by progressive ventricular dilation and adverse cardiac remodeling [5].

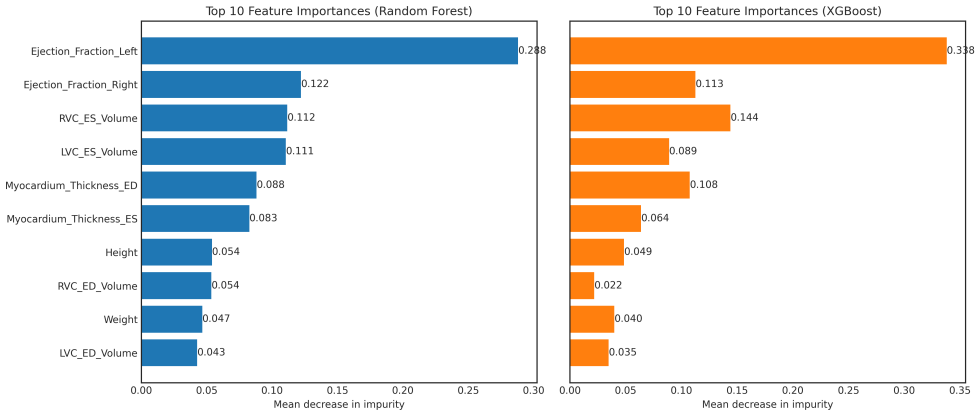

Figure 5: Top 10 most important features according to Random Forest (left) and XGBoost (right), based on mean decrease in impurity. Both models highlight `Ejection_Fraction_Left`, `Ejection_Fraction_Right`, and ventricular volumes as the most discriminative features. The importance values reflect how much each feature contributes to reducing classification uncertainty across all trees.

We can see that there is a consistency between the two models in prioritizing cardiac dynamic features (particularly `Ejection_Fraction_Left`). This supports the robustness of the extracted descriptors and their ability in this classification task.

3.4 Classifier Decision Boundaries in PCA Space

To better understand how each classifier separates the classes in feature space, we projected the data onto the first two principal components and visualized the decision boundaries of the four main models used in our study (Figure 6). The first two components capture a substantial portion of the variance in the data, specifically 68.0% for PC1 and 28.4% for PC2, making this 2D projection informative for class separation analysis.

The decision surfaces show distinct patterns: Random Forest and XGBoost produce non-linear, axis-aligned partitions; SVM defines smooth non-linear boundaries; and Logistic Regression results in linear separation. These visualizations qualitatively support the numerical performance metrics presented earlier, highlighting the consistency between the learned decision boundaries and the underlying class structure.

4 Possible Improvements

While the proposed pipeline demonstrates strong classification performance using a relatively small set of interpretable features, several avenues exist to further enhance the model.

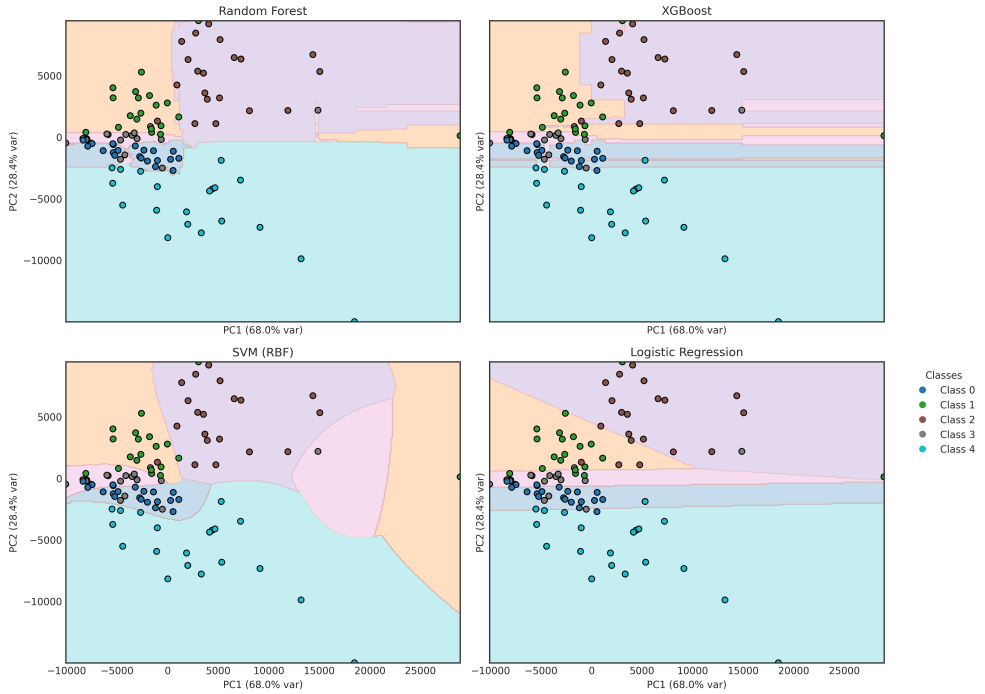

Figure 6: Decision boundaries of the four main classifiers projected onto the PCA-reduced 2D feature space. The axes indicate the percentage of variance explained by each principal component (PC1 and PC2).

First, the current feature set includes volumetric and functional descriptors such as ejection fractions, cavity volumes, and myocardial thickness at two key timepoints (ED and ES). However, prior work [3] has demonstrated that additional domain-specific features can further improve model robustness and interpretability. These include instant anatomical features (such as circularity or minimum and maximum thickness), and dynamic features derived from the full time series.

Incorporating such features, particularly dynamic descriptors beyond ED and ES, could allow models to capture subtle temporal patterns indicative of early or complex cardiac dysfunction.

5 Conclusion

In this work, we developed a segmentation-based pipeline for classifying cardiac pathologies from CMRI images. By combining geometric and functional features extracted from the segmented left and right ventricles and myocardium, we trained several supervised classifiers and achieved high validation performance. Among the tested models, XGBoost and SVM showed the most consistent results, both with and without data augmentation.

Beyond accuracy, we emphasized feature interpretability and anatomical relevance. Features such as ejection fraction, ventricular volumes, and myocardial thickness emerged as key predictors, aligning with established clinical knowledge.

References

[1] Hae-Yeoun Lee et al. "Automatic Left Ventricle Segmentation Using Iterative Thresholding and an Active Contour Model With Adaptation on Short-Axis Cardiac MRI". In: *IEEE Transactions on Biomedical Engineering* 57.4 (2010), pp. 905–913. DOI: 10.1109/TBME.2009.2014545.

[2] Jelmer M. Wolterink et al. "Automatic Segmentation and Disease Classification Using Cardiac Cine MR Images". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Ed. by Mihaela Pop et al. Cham: Springer International Publishing, 2018, pp. 101–110. ISBN: 978-3-319-75541-0.

[3] Fabian Isensee et al. "Automatic Cardiac Disease Assessment on cine-MRI via Time-Series Segmentation and Domain Specific Features". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Ed. by Mihaela Pop et al. Cham: Springer International Publishing, 2018, pp. 120–129. ISBN: 978-3-319-75541-0.

[4] Mahendra Khened, Varghese Alex, and Ganapathy Krishnamurthi. "Densely Connected Fully Convolutional Network for Short-Axis Cardiac Cine MR Image Segmentation and Heart Diagnosis Using Random Forest". In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Ed. by Mihaela Pop et al. Cham: Springer International Publishing, 2018, pp. 140–151. ISBN: 978-3-319-75541-0.

[5] Sean P. Murphy, Nasrien E. Ibrahim, and Jr Januzzi James L. "Heart Failure With Reduced Ejection Fraction: A Review". In: *JAMA* 324.5 (Aug. 2020), pp. 488–504. ISSN: 0098-7484. DOI: 10.1001/jama.2020.10262. eprint: https://jamanetwork.com/journals/jama/articlepdf/2768982/jama_murphy_2020_rv_200007_1605896147.19162.pdf. URL: https://doi.org/10.1001/jama.2020.10262.