IMA205 Challenge 2025

Cardiac Pathology Prediction from CMRI

Pedro de Almeida Marim

May 2025

Abstract

This report presents a complete pipeline for automatic cardiac pathology classification based on cardiac MRI data. From anatomical segmentations, clinically relevant features were extracted, including ejection fractions, chamber volumes, BSA-normalized metrics, and myocardial wall thickness statistics. Multiple classification algorithms were explored—such as Logistic Regression, Support Vector Machines, Random Forests, and XGBoost—with a strong focus on model generalization and interpretability. Feature selection and cross-validation strategies were used to refine the models and reduce overfitting. The final submission, based on a Random Forest classifier, achieved an accuracy of 88% on the challenge test set. The Kaggle nickname used for submission was **Rust-eze**.

1 Introduction

Cardiovascular diseases remain one of the leading causes of mortality worldwide. Early and accurate diagnosis of cardiac pathologies is essential to prevent serious complications such as heart failure or sudden cardiac arrest. Among the various imaging techniques available, cardiac magnetic resonance imaging (CMRI) provides high-resolution, non-invasive visualizations of cardiac anatomy and function, making it an invaluable tool for clinical assessment.

However, interpreting CMRI data manually is time-consuming and requires specialized expertise. This has motivated the development of computer-aided diagnosis (CAD) systems to assist clinicians by automatically extracting relevant features and predicting cardiac conditions. The IMA205 Challenge 2025 focuses on building such a system to classify subjects into five diagnostic categories: healthy controls, myocardial infarction, dilated cardiomyopathy, hypertrophic cardiomyopathy, and abnormal right ventricle.

In this project, I address the challenge by designing a complete processing pipeline that includes preprocessing CMRI data, extracting anatomical and functional features from segmentations, and training machine learning models for classification. The goal is to build an interpretable and generalizable model capable of predicting the correct pathology from unseen CMRI data using only the information provided in the challenge.

2 Data Description

The dataset provided in the IMA205 Challenge 2025 consists of cardiac magnetic resonance imaging (CMRI) scans from 150 subjects. Each subject belongs to one of five diagnostic categories:

- Class 0: Healthy controls
- Class 1: Myocardial infarction
- Class 2: Dilated cardiomyopathy
- Class 3: Hypertrophic cardiomyopathy
- Class 4: Abnormal right ventricle

The dataset is split into a training-validation set (100 subjects) and a test set (50 subjects). Only the training-validation set includes ground truth labels for model development. Each subject is represented by two 3D CMRI volumes acquired at two key time points in the cardiac cycle:

- End-Diastole (ED) — when the heart is fully relaxed
- End-Systole (ES) — when the heart is fully contracted

For training subjects, each volume is accompanied by a corresponding segmentation mask in the same shape, labeling three anatomical regions:

- Label 1: Right ventricle cavity (RVC)
- Label 2: Myocardium (LVM)
- Label 3: Left ventricle cavity (LVC)

In addition to image data, metadata including subject height and weight is also available for all patients, allowing for the computation of body surface area (BSA) and normalization of anatomical features. In the test set, only partial segmentation is provided — specifically, the left ventricle cavity label (3) is missing — motivating the need for left ventricle segmentation as an intermediate step.

2.1 Data Format and Dimensions

All CMRI volumes and segmentation masks are provided in the NIfTI format (with the `.nii` extension), a standard used in medical imaging that supports compressed 3D data with associated metadata. Each file contains a volumetric image with three spatial dimensions: height, width, and number of slices along the axial (z) direction.

A notable characteristic of the dataset is the variation in spatial resolution and volume shape across different subjects. Each subject has distinct image dimensions depending on their anatomy and acquisition protocol. For example:

- Case 113: ED shape $(216, 256, 10)$, ES shape $(216, 256, 10)$
- Case 114: ED shape $(232, 288, 15)$, ES shape $(232, 288, 15)$
- Case 115: ED shape $(232, 256, 9)$, ES shape $(232, 256, 9)$

This variability must be taken into account during feature extraction and model design. Additionally, voxel size information (in millimeters) is embedded in the NIfTI headers and can vary across subjects. This information is critical for computing real-world physical quantities such as volume (in milliliters) and thickness (in millimeters) from the segmentation masks.

3 Preprocessing: Segmentation

A complete segmentation is provided for the training-validation set, with three labels identifying the right ventricle cavity (RVC), myocardium (LVM), and left ventricle cavity (LVC). These segmentation masks were used to extract anatomical and functional features for model training.

In contrast, the segmentation masks in the test set are incomplete: the label corresponding to the LVC (label 3) is entirely missing. Therefore, as a preprocessing step, I implemented a method to infer the LVC region in test images using the available myocardium mask.

The LVC is typically enclosed within the myocardium and occupies the interior of the ring-shaped structure formed by it. I leveraged this anatomical property by applying a morphological hole-filling operation on the myocardium mask in each axial slice. This process recovers the closed region inside the myocardium, which I then assign to the LVC. The method was applied slice-by-slice and produced consistent and anatomically plausible segmentations.

This inferred LVC mask was then used in the same way as the original label in the training set to extract volumetric and functional features for classification. This step ensures that both training and test data are treated uniformly in the downstream pipeline.

3.1 Experimental LVC Segmentation Using Iterative Thresholding

After inspecting the myocardium segmentation in the test set, it became evident that the left ventricle cavity (LVC) could be accurately and consistently recovered by simply filling the hole enclosed by the myocardium in each axial slice. This observation led to the adoption of the hole-filling approach as the main preprocessing method due to its simplicity and perfect coherence across all test subjects.

However, the apparent simplicity of this solution motivated an attempt to explore a more general and challenging method for LVC segmentation — one that does not rely on the availability of the myocardium mask or any other label. For this purpose, I implemented a method based on the paper *Automatic Left Ventricle Segmentation Using Iterative Thresholding and an Active Contour Model With Adaptation on Short-Axis Cardiac MRI*. This technique segments the LVC directly from the MRI intensities by iterative thresholding.

As expected, this approach lacked the consistency of the hole-filling strategy that takes full advantage of the given partial segmentation. As such, it was ultimately not included in the final pipeline. Nevertheless, the exploration is documented here to highlight an alternative anatomical segmentation strategy and to demonstrate initiative toward building more general-purpose methods.

The method proposed in the referenced paper segments the left ventricle cavity (LVC) in short-axis cardiac MRI through a six-step process that progressively refines the segmentation starting from raw image intensities. The first step involves identifying an initial seed point within the LVC by applying the Hough Transform to difference images between diastolic and systolic phases. Once the seed is established, the second step performs a region-growing operation to extract a region assumed to contain mostly blood, from which the mean and standard deviation of the LVC intensity are estimated. The third step compensates for coil sensitivity artifacts by fitting a planar bias field to the intensities of the detected blood pool and correcting the image accordingly. In the fourth step, the signal intensity of the myocardium is estimated by observing the behavior of a threshold-based region growing as it begins to expand into surrounding tissues, identifying the threshold just before this "effusion". With these intensity statistics in hand, the fifth step performs a final segmentation of the LVC using an absolute threshold-based region-growing algorithm, which produces a clean and compact mask of the cavity. Finally, in the sixth step, the seed point is propagated across all slices by minimizing an energy function that combines spatial proximity and intensity similarity, allowing a full volumetric segmentation of the LVC without relying on prior anatomical labels.

3.1.1 Step 1: Initial Seed Detection

The first step of the method consists of estimating a seed point located inside the left ventricle cavity. To do so, a circular Hough Transform is applied to the difference between the end-diastolic (ED) and end-systolic (ES) phases of a mid-ventricular slice. The rationale is that the LVC exhibits significant motion between these two phases, making it one of the most prominent dynamic structures in the image. By subtracting the ES slice from the ED slice, static anatomical structures are suppressed while the moving blood pool of the LVC is enhanced.

After preprocessing the difference image via edge detection, the Hough Transform is applied over a predefined range of radii. The circle with the strongest accumulator score is selected, and its center is used as the seed point. This approach performed well in most cases, particularly when the LVC presented regular geometry and good contrast. However, some challenges were encountered, especially in cases where the right ventricle (RV) appeared more circular and better defined than the LVC. In those cases, the Hough Transform sometimes mistakenly selected the RV instead.

To mitigate this, a post-filtering step was introduced to select the best candidate among the top three circles detected. The final seed was chosen as the one whose center had the lowest vertical coordinate (i.e., closest to the inferior half of the image), under the anatomical assumption that the LVC is usually located lower than the RV in axial views. This adjustment significantly improved the robustness of the seed detection step across major part of test subjects.

This step is crucial and also one of the most challenging, as it sets the foundation for the entire segmentation pipeline. Many test subjects exhibit poor contrast or non-standard anatomy, which complicates the detection process significantly. To address this, various preprocessing strategies were tested, including histogram matching and histogram equalization. Although they improved contrast in problematic cases, they also degraded performance in well-behaved images, introducing inconsistencies across the dataset. So, ultimately, we proceded with the previously described approach because it provided the best balance between robustness and simplicity, even though it still leads to seed misplacements in a few cases.

This step remains one of the key limitations of the current pipeline and represents an important area for future improvement, as any error at this stage directly compromises the subsequent segmentation process.

3.1.2 Step 2: Blood Pool Region Growing

After detecting the initial seed point, the next step consists of extracting a soft segmentation of the left ventricle blood pool by applying a region-growing algorithm. The goal is to isolate a region composed primarily of blood pixels, from which the intensity distribution (mean and standard deviation) of the LVC can be estimated for subsequent steps.

Starting from the seed, the region is grown by iteratively adding neighboring pixels whose intensities are within a fixed threshold from the region's current mean intensity. The implementation maintains a dynamic list of pixel values in the region to continuously update the local mean as new pixels are included. This makes the growth process sensitive to local intensity trends and more adaptive than using a static reference.

A major challenge in this step is that region growing can easily "leak" into surrounding tissues when contrast is weak, especially in slices with anatomical ambiguities or poor signal-to-noise ratio. To manage this, the algorithm monitors the total region size. If this value exceed predefined limit, the region is flagged as *leaked*. The algorithm reacts by reducing the threshold parameter and restarting the process. This loop continues until the region grows stably without leaking. This strategy proved to be both flexible and effective in adapting the region-growing behavior to the wide variability present across the dataset.

The final mask is then used to compute the mean and standard deviation of the blood signal intensity, which are essential parameters for the next steps.

3.1.3 Step 3: Coil Sensitivity Compensation

Magnetic resonance images often exhibit spatial intensity inhomogeneities caused by non-uniform coil sensitivity. These biases can affect the accuracy of intensity-based segmentation, especially in structures like the blood pool that are expected to have relatively uniform signal intensity. To correct this, the method models the intensity variation across the blood region as a linear bias field.

Specifically, a planar surface is fitted to the intensity values of the segmented blood region obtained in the previous step. This is done via least-squares regression, estimating the best-fitting plane that minimizes the residuals across the selected pixels. Once the bias plane is computed, it is subtracted from the original image to produce a corrected slice where the blood region exhibits more homogeneous intensity.

Although this correction improves local uniformity, it can also introduce negative intensity values in regions outside the blood pool. To handle this, a normalization step is applied after bias subtraction to ensure all values are non-negative. Two strategies were tested: simply shifting the entire image to have zero as its minimum, and normalizing the intensities to the $[0, 1]$ range. The former was ultimately preferred, as it preserved the intensity scale and avoided distortions during threshold-based processing in subsequent steps.

3.1.4 Step 4: Myocardium Signal Estimation

Once the blood pool has been segmented and corrected for bias, the next step is to estimate the typical signal intensity of the myocardium. This is a crucial parameter for defining anatomical boundaries and initializing models in later stages. The approach adopted in the paper relies on a dynamic thresholding process that simulates the behavior of region-growing under increasingly permissive intensity conditions.

The idea is to perform multiple region-growing operations starting from the LVC seed, using progressively larger values of the threshold parameter, which controls the allowed intensity deviation from the region mean. Initially, only blood pixels are included. As the threshold increases, the region begins to include adjacent myocardial tissue, and eventually leaks into other surrounding structures such as fat, muscle, or the right ventricle. This behavior is monitored by tracking the volume of the grown region at each iteration.

To detect the moment of "effusion" into non-myocardial tissue, the algorithm looks for a sudden discontinuity in the volume growth curve. Specifically, if the increase in region size between two successive thresholds is significantly larger than the previous increment, the region is flagged as *leaked*, and the process stops. The last valid mask before the leak is assumed to contain only blood and possibly a thin myocardial border.

This mask is then dilated using a morphological disk to define a surrounding annular region, and the difference between the dilated mask and the original is used as a proxy for the myocardium. The mean and standard deviation of this ring region are taken as estimates for the myocardial signal.

This technique proved effective in providing a robust and adaptive estimation of the myocardium intensity. However, the method is sensitive to noisy contours and sharp anatomical transitions, and in some cases may under- or over-estimate the myocardium depending on how cleanly the region grows before leaking. Still, it offers a reasonable approximation with no prior anatomical labels.

3.1.5 Step 5: Final Absolute Threshold Region Growing

With the blood and myocardium signal statistics estimated in the previous steps, the final segmentation of the left ventricle cavity is performed using an absolute threshold-based region-growing algorithm. Unlike earlier stages, which relied on adaptive thresholds and feedback mechanisms, this final step applies a hard intensity threshold derived from the myocardial and blood signal distribution.

The segmentation starts at the previously detected seed and includes all connected pixels whose intensity is above a fixed threshold. In practice, the threshold used is `myoc_mean + lv_std`, which aims to isolate the full-blood region while excluding myocardial and other surrounding tissues. This strategy leverages the known brightness gap between blood and myocardium to produce a clean binary segmentation.

The algorithm iteratively grows the region by checking all 8-connected neighbors around the seed, adding new pixels if their intensity meets the threshold criterion. At the end, post-processing is applied, including hole-filling and connected component analysis to retain only the largest coherent region.

This stage delivers the final LV cavity mask for a given slice. While simple, this approach was found to be quite effective when the intensity statistics were accurately estimated. However, as with previous steps, its reliability still depends heavily on the correctness of the seed point and the quality of the bias-corrected image. When well initialized, this final region-growing pass yields compact and anatomically plausible LVC segmentations.

3.1.6 Step 6: Seed Propagation Across Slices

Once the left ventricle cavity (LVC) is segmented in a mid-ventricular slice, the method proceeds by propagating the segmentation to adjacent slices in both directions. This is done recursively by reusing the final LV mask from each processed slice to determine the seed for the next one. The propagation continues until the beginning or end of the volume is reached.

The propagation mechanism relies on an energy minimization criterion defined in the original paper. For a given slice, the center of gravity of the previous mask is computed and used as the center of an 11×11 search window. Inside this window, a pixel-wise energy function is evaluated for each candidate position:

$$E(p) = \left(\sqrt{\frac{2\sigma_{\text{prev}}}{w-1}} \cdot \|p - p_{\text{CoG}}\| \right)^2 + (I(p) - \mu_{\text{prev}})^2$$

Here, p is a candidate pixel, p_{CoG} is the center of gravity of the previous mask, μ_{prev} and σ_{prev} are the mean and standard deviation of the previous slice's blood region, and $I(p)$ is the intensity of the candidate pixel in the current slice. The pixel with the lowest energy is selected as the new seed for the next slice.

This recursive seed propagation is done in both directions starting from the central slice. At each step, the newly found seed is used to run the final absolute threshold region-growing algorithm, producing a new LV mask. This mask then feeds into the next propagation step.

This strategy proved effective for smoothly varying sequences of slices and offered a coherent way to extend segmentation throughout the 3D volume without relying on ground truth. However, because it depends on local consistency, the method is sensitive to poor initial seed placement or rapidly changing anatomy across slices. Despite these limitations, it allowed for robust volumetric segmentation in a majority of cases.

3.1.7 Evaluation and Results

The performance of the proposed segmentation method was evaluated using the Dice similarity coefficient between the predicted masks and the ground truth annotations available in the training set. The metric was computed slice-by-slice and then averaged across the volume for each case. This evaluation was performed over a subset of training subjects for which ground truth masks were available.

The segmentation outcomes were found to fall into three main categories:

- **Fully successful segmentations:** In several cases (e.g., case 65), the segmentation was accurate across all slices. The predicted masks aligned well with the ground truth, and the small differences observed were mostly due to the absence of artificial smoothing in the predicted contours.
- **Partial failures in apical slices:** In cases like 62 and 64, the segmentation was consistent in basal and mid-ventricular slices but degraded toward the apex. The method often failed to detect the disappearance of the LVC, continuing to grow into irrelevant tissue. This limitation stems from the lack of an explicit stopping condition based on anatomical plausibility.
- **Incorrect segmentations due to seed misplacement or leakage:** Some cases, such as 49 and 68, were affected by incorrect initial seed placement or uncontrolled leakage during the final region-growing step. In case 49, the seed was placed in the myocardium, preventing proper segmentation. In case 68, although the seed was correctly placed, the absolute threshold allowed the region to leak into adjacent structures due to lack of growth control at this stage.

The average Dice coefficients across slices varied significantly between cases. Fully successful cases achieved average Dice scores above 0.95, while partially successful cases typically ranged from 0.78 to 0.85. Failure cases dropped well below 0.1. Results of representative cases are summarized in Table 1.

Table 1: Representative Segmentation Results

Case	**Outcome**	**Mean Dice**	**Remarks**
65	Fully successful	0.957	Accurate in all slices
62	Apical degradation	0.841	Final slice oversegmented
64	Apical degradation	0.780	Apex included non-LV tissue
63	Mixed performance	0.825	Minor leakage, undergrowth in some slices
68	Leakage failure	0.027	Final mask exploded due to threshold
49	Seed misplacement	0.000	Seed located in myocardium

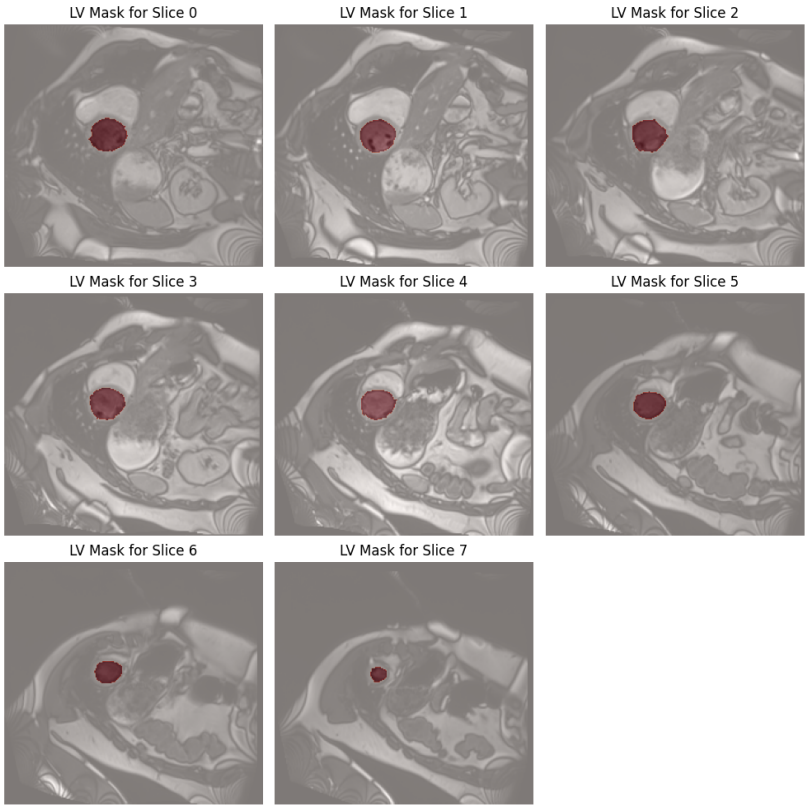

Figure 1: Example of successful segmentation in case 65. The predicted LVC masks are shown overlaid on the original MRI slices. The segmentation aligns closely with the ground truth, especially in basal and mid slices.

3.1.8 Discussion and Limitations

The segmentation method explored in this section demonstrates that it is possible to obtain anatomically coherent LVC masks using only MRI intensities, without relying on ground truth labels. The method follows a fully unsupervised approach based on iterative thresholding, regional statistics, and recursive propagation, and was able to achieve accurate results in a significant portion of the evaluated cases.

However, several limitations were encountered throughout the pipeline. The most critical issue is the strong dependence on the initial seed point. When the seed is mislocated — for example, inside the right ventricle or the myocardium — the entire segmentation fails, as all subsequent steps build upon that initial position. While anatomical heuristics helped improve robustness, the absence of a learned or adaptive seed detection strategy remains a key limitation.

Another challenge lies in the handling of apical slices. The current pipeline assumes that a region resembling the LVC is always present in the next slice, and continues propagating segmentation accordingly. In reality, the LVC becomes vanishingly small near the apex, and the algorithm lacks a stopping criterion based on anatomical disappearance. As a result, some segmentations continue beyond the actual end of the cavity, leading to false positives.

Additionally, the final absolute-threshold region-growing step does not include protection mechanisms against leakage. Once a threshold is selected, the region may grow uncontrollably if the seed is near an intensity transition. This behavior was observed in several failure cases and suggests that even simple validation mechanisms — such as comparing region size with prior slices — could improve robustness.

Despite these issues, the method proved valuable in demonstrating the viability of rule-based, interpretable segmentation for cardiac structures. It also served as a foundation for understanding the intensity distributions and geometry of the LVC, and can potentially be integrated into more complex or supervised segmentation frameworks in future work.

4 Feature Extraction

The goal of feature extraction is to quantify anatomical and functional characteristics of the heart from segmented cardiac MRI volumes.

Once the segmentation masks were obtained for both end-diastole (ED) and end-systole (ES) phases, I extracted anatomical and functional features directly from these masks. The feature design was guided by prior work in the literature, ensuring that the selected attributes were both clinically meaningful and useful for classification.

4.1 Features Based on Prior Work

I first implemented the set of 10 features proposed in the paper *Densely Connected Fully Convolutional Network for Short-Axis Cardiac Cine MR Image Segmentation and Heart Diagnosis Using Random Forest*. These features are standard clinical descriptors and were directly computed from the segmentations and available metadata:

- Ejection fraction (EF) of the left and right ventricles
- Volume of the left ventricle at ED and ES
- Volume of the right ventricle at ED and ES
- Volume of the myocardium at ED (used as a proxy for myocardial mass)
- Patient height and weight

All volumes were obtained the sum of the number of voxels in all slices of the respective structure's mask, and multiplied by the voxel volume (in mm^3), as specified in the NIfTI header, to obtain anatomically meaningful measurements, since each voxel represents a different real-world size depending on the acquisition parameters. Ejection fractions were computed using the standard formula based on ED and ES volumes.

Ejection fraction (EF), a widely used index of cardiac function, is among the most critical descriptors. The EF quantifies how much blood a ventricle ejects relative to its filled volume, and is defined as:

$$\text{EF} = \frac{V_{\text{ED}} - V_{\text{ES}}}{V_{\text{ED}}}$$

where V_{ED} is the volume at end-diastole and V_{ES} is the volume at end-systole. A normal left ventricular EF typically ranges from 50% to 70%. Lower values are indicative of systolic dysfunction (e.g., dilated cardiomyopathy), while certain pathologies like hypertrophic cardiomyopathy may exhibit preserved or even elevated EF values due to reduced cavity size.

4.2 Additional Features for Diagnostic Enhancement

To complement the initial features, I added several additional descriptors inspired by the paper *Automatic Cardiac Disease Assessment on Cine-MRI via Time-Series Segmentation and Domain Specific Features*. These features are intended to enhance the model's ability to detect ventricular imbalance and morphological abnormalities:

- Volume ratios:
  - LV/RV at ES
  - LV/Myocardium
  - RV/Myocardium
- BSA-normalized volumes:
  - Volume of LV at ED and ES divided by BSA
  - Volume of RV at ED and ES divided by BSA
  - Volume of myocardium at ED and ES divided by BSA
- Myocardial thickness (mean, standard deviation, minimum, maximum)
- Apical RVC/LVC ratio (area ratio at the most apical slice containing myocardium)

These features provide additional insight into the relative sizes and proportions of cardiac structures, which are often altered in pathological cases. Normalization by BSA allows for better comparison across subjects of different body sizes.

4.3 Myocardial Thickness: Two Estimation Approaches

Estimating the thickness of the myocardium was a good morphological descriptor in many cases. I implemented two distinct strategies:

- **Distance transform method:** for each myocardium voxel, I calculated the distance to the nearest left ventricle cavity pixel in the same slice using a Euclidean distance transform. This provides an approximate thickness measurement from the inner wall outward. Although simple and efficient, this method only accounts for one side of the myocardial wall and tends to underestimate thickness, especially in asymmetric or dilated cases.
- **Contour-based method:** I extracted the inner and outer contours of the myocardium in each slice and measured the shortest distances between them. This provides a more accurate anatomical thickness estimate but is computationally more expensive and sensitive to contour quality. The values were aggregated across slices using mean, standard deviation, minimum, and maximum statistics.

In the final pipeline, I prioritized the contour-based method due to its anatomical coherence.

5 Classification Models

The extracted features were used to train supervised classification models in order to automatically predict the cardiac pathology for each subject. Given the nature of the dataset and the constraints of the challenge, several methodological considerations influenced the modeling choices.

First, the dataset is relatively small, consisting of only 100 labeled subjects, which imposes a high risk of overfitting. With such limited data, even a few misclassified samples can strongly affect performance estimates. To mitigate this, I employed cross-validation strategies and focused on interpretable, low-complexity models that are known to perform well on small datasets.

Importantly, the dataset is perfectly balanced, with 20 subjects for each of the five diagnostic categories. While this simplifies training—eliminating the need for class reweighting or sampling correction—it also increases the importance of preserving class proportions during any train-test split. Therefore, all my experiments used stratified sampling to ensure that class balance was maintained in training and validation subsets. This was particularly important when holding out a portion of the small training data to simulate a even smaller internal test set.

5.1 Training Algorithms

In this section, I present the different classification algorithms tested throughout the project, alongside the motivations for their use and the results obtained. All models were trained using the extracted features, with preprocessing strategies adapted to each algorithm. A portion of the training data (20%) was held out as an internal test set, and all experiments were evaluated both using cross-validation and this internal set. The test set provided by the challenge was not used during training, as it contains no labels and is reserved for final submission.

5.1.1 Logistic Regression

I began my experiments with Logistic Regression, which is often the first choice when working with small datasets. Its simplicity, low model complexity, and interpretability make it a reliable baseline. When the data is close to linearly separable, Logistic Regression can perform remarkably well without the risk of overfitting. To ensure compatibility with the model, I applied min-max scaling to the features and used grid search to tune the regularization strength. Both a 5-fold and a stratified 5-fold cross-validation setup were tested, yielding scores of 0.9375 and 0.90, respectively. The final test accuracy was 95%.

While this result is promising, it should be interpreted with caution, as the small size of the test set could lead to optimistic performance estimates. Nonetheless, the strong performance of this simple linear model suggests that the problem is not highly non-linear, and that very complex algorithms may not be necessary.

5.1.2 k-Nearest Neighbors (k-NN)

Next, I evaluated the k-Nearest Neighbors algorithm. As a non-parametric, instance-based learner, k-NN is useful for identifying whether similar patients (in terms of features) tend to belong to the same class. The model was trained using min-max scaled features and tuned using cross-validation. The best performance was achieved with $k = 3$ neighbors, Euclidean distance, and distance-based weighting. The model obtained a test accuracy of 85% and a cross-validation score of 0.9125.

5.1.3 Support Vector Machine (SVM)

To explore the separation margin between classes more explicitly, I tested a Support Vector Machine classifier. SVM is a strong choice for small datasets, as it focuses on maximizing the decision boundary margin and can operate well in high-dimensional feature spaces. Features were standardized before training, and grid search was used to tune the regularization and kernel parameters. The best model used a linear kernel with $C = 0.1$ and achieved both a cross-validation score and test accuracy of 95%.

Interestingly, the best-performing configuration used a linear kernel, further supporting the idea that the classes are closer to being linearly separable in the extracted feature space. This reinforces the earlier observations from the Logistic Regression results and highlights the strong discriminative power of the chosen features.

5.1.4 XGBoost

As an alternative to Random Forests, I trained an XGBoost classifier. XGBoost is a gradient-boosted tree ensemble known for its strong performance on structured data. It builds trees sequentially to correct the errors of previous ones, and is often more expressive than traditional bagged models. The classifier achieved a cross-validation score of 0.9125 and a test accuracy of 90%, making it a strong non-linear contender.

5.1.5 Random Forest

I also trained a Random Forest classifier, which was ultimately selected as the final model for generating predictions on the challenge test set. Random Forest is a widely used ensemble method that combines the predictions of multiple decision trees trained on different random subsets of the data. This approach is particularly well-suited to small- to medium-sized tabular datasets, where it can model complex relationships without overfitting, especially when the number of features is moderate.

Random Forests are also highly robust to noise, resistant to feature scaling issues, and provide valuable insights through built-in feature importance metrics. These strengths align well with the goals of this challenge, which involves a modest number of samples, structured anatomical features, and the need for model interpretability. Hyperparameter tuning was performed via grid search, and the final model achieved a test accuracy of 95% and a cross-validation score of 0.9125. In addition to strong quantitative performance, Random Forest was recommended in one of the reference papers, further supporting its relevance for this task. These factors together made it the preferred choice for the final submission.

5.2 Feature Importance and Feature Selection

To better understand which features contributed most to the classification task, I analyzed feature importances using the Random Forest model. The built-in feature importance scores reflect how often each feature is used in informative splits across the ensemble of trees. While the rankings varied slightly across executions due to the model's inherent randomness, certain features consistently appeared near the top of the list. These included the ejection fraction of the left ventricle, multiple volume ratios (e.g., LV/RV, LV/myocardium), and statistics related to myocardial thickness. Their repeated presence among the most important features reinforces their clinical relevance and strong predictive value.

In contrast, some features were consistently ranked near the bottom. These included patient height, weight, and the apical RVC/LVC ratio. The latter, in particular, was often zero across a large portion of the dataset, making it uninformative. Additionally, I observed that having both raw volumes and BSA-normalized volumes introduced redundancy. Keeping both could confuse the model or artificially inflate the importance of those measurements. As a result, I removed low-importance and potentially redundant features to simplify the input space and reduce noise. Specifically, I removed the raw volume features (e.g., volume of LV, RV, and myocardium at ED and ES), patient height and weight, and the apical RVC/LVC ratio.

After pruning the least informative and redundant features, I retrained the best classification models on the reduced feature set. This allowed me to evaluate whether the simplified representation retained enough discriminative power. Logistic Regression and SVC achieved a 90% accuracy on the internal test set, while Random Forest achieved 95% (respective cross validation scores: 0.912, 0.949, 0.911). Although the same or slightly lower than the earlier results obtained using all features, these models are likely to generalize better due to reduced risk of overfitting. Moreover, the retained features are clinically interpretable and more robust, which further supports their use in the final pipeline.

6 Final Classification and Submission

After completing the model selection and evaluation process, I trained a final Random Forest classifier using the entire labeled dataset, including the 20% portion that had previously been held out for internal testing. This decision was made to maximize the amount of training data available.

The final Random Forest model was used to generate predictions for the 50 test subjects provided by the challenge. The submission achieved an accuracy of 88% on the public leaderboard, confirming the model's strong generalization capabilities and the effectiveness of the selected feature set.

7 Conclusion

In this project, I developed a complete pipeline for automatic cardiac pathology classification using cardiac MRI data. Starting from segmentation-based feature extraction, I implemented both clinically motivated and literature-inspired features, including ejection fractions, volume ratios, BSA-normalized volumes, and myocardial thickness statistics. Multiple classification algorithms were tested and compared, with careful attention to validation strategies, preprocessing, and feature selection.

Logistic Regression and Support Vector Machines demonstrated strong performance, indicating that the data is linearly separable in the chosen feature space. Ultimately, a Random Forest classifier was selected for final submission due to its strong accuracy, robustness, and interpretability. The final model achieved 88% accuracy on the challenge test set, confirming its generalization capability. This project highlights the importance of thoughtful feature design and model selection, especially when working with limited annotated data in the medical domain.

For future work, a natural next step would be to replace the segmentation-based pipeline with a fully automatic model that can segment all relevant cardiac structures—left ventricle, right ventricle, and myocardium—directly from raw CMRI volumes. While this report explored partial approaches to that problem, a robust, fully supervised segmentation model would reduce reliance on provided labels and significantly extend the pipeline's applicability to real clinical scenarios.

References

[1] Xue, W., et al. *Densely Connected Fully Convolutional Network for Short-Axis Cardiac Cine MR Image Segmentation and Heart Diagnosis Using Random Forest.* 2018.

[2] Zheng, Y., et al. *Automatic Left Ventricle Segmentation Using Iterative Thresholding and an Active Contour Model With Adaptation on Short-Axis Cardiac MRI.* Computers in Biology and Medicine, 2008.

[3] Isensee, F., et al. *Automatic Cardiac Disease Assessment on Cine-MRI via Time-Series Segmentation and Domain Specific Features.* In MICCAI, 2017.

[4] Le Folgoc, L., and Gori, P. *IMA205 Challenge 2025: Cardiac Pathology Prediction*, https://www.kaggle.com/competitions/ima-205-challenge-2025.