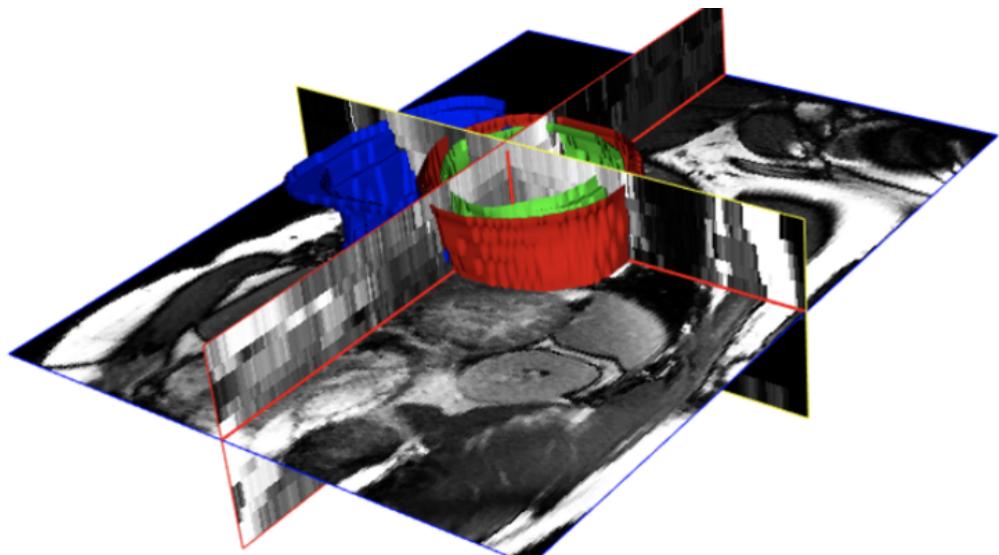


**Cardiovascular disease diagnosis using cardiac
magnetic resonance radiomics and machine
learning**



Submitted by:

Joan Arxer I Vargas

Sofía González Estrada

Roger Puig I Arxer

Ravneet-Rahul Sandhu Singh

Frances Scarlett Thomas

Machine Learning
Group 5

May 4, 2025

1. Introduction

Advances in medical imaging and analytical techniques have long supported the diagnosis and management of Cardiovascular Diseases (CVDs). The increasing availability of high-resolution imaging and computational tools presents an opportunity to enhance clinical decision-making through data-driven approaches.

1.1 Motivation

CVDs remain a major global health burden, causing nearly one-third of annual deaths worldwide [1]. Early and reliable diagnosis is key to managing progression and improving outcomes. Cardiac Magnetic Resonance (CMR) is the gold standard for assessing heart morphology and function due to its high resolution and non-invasive nature [2]. However, traditional diagnostics often rely on limited parameters, potentially overlooking patterns critical for certain conditions [3].

1.2 State of the Art

Radiomics has recently emerged as a promising method to extract data from medical images, providing a more detailed representation of tissue characteristics. By quantifying traits like shape, texture, and intensity, it enables the creation of datasets for automated disease classification. Combined with Machine Learning (ML) models, these features offer new opportunities for predictive tools beyond traditional image interpretation [4].

1.3 Objectives

For this challenge, we have been provided with the Automated Cardiac Diagnosis Challenge (ACDC) dataset [5], derived from cine-MRI scans acquired at the University Hospital of Dijon (France). It offers standardized imaging data capturing real-time heart movement across the cardiac cycle. Radiomic features were extracted from these segmented heart structures using the PyRadiomics library [6]. Each patient is assigned to one of five diagnostic categories, making this a multi-class classification problem. Therefore, the objective is to compare different supervised machine learning algorithms and identify the best-performing model.

1.4 Goals

To address this objective, we define a set of specific goals that guide the project workflow:

- ✓ Perform an exploratory analysis to understand the data.
- ✓ Compare Simple Split and Stratified K-Fold techniques.
- ✓ Apply feature selection and dimensionality reduction.
- ✓ Implement ML models for supervised classification.
- ✓ Evaluate and compare model performance using robust metrics.
- ✓ Assess the clinical relevance of the best-performing model.

2. Methodology

The workflow of our project can be divided into five different stages: Exploratory Data Analysis (EDA), Data Partitioning, Dimensionality Reduction, Models, and Metrics. Figure 1, displays a generalized view on these stages.

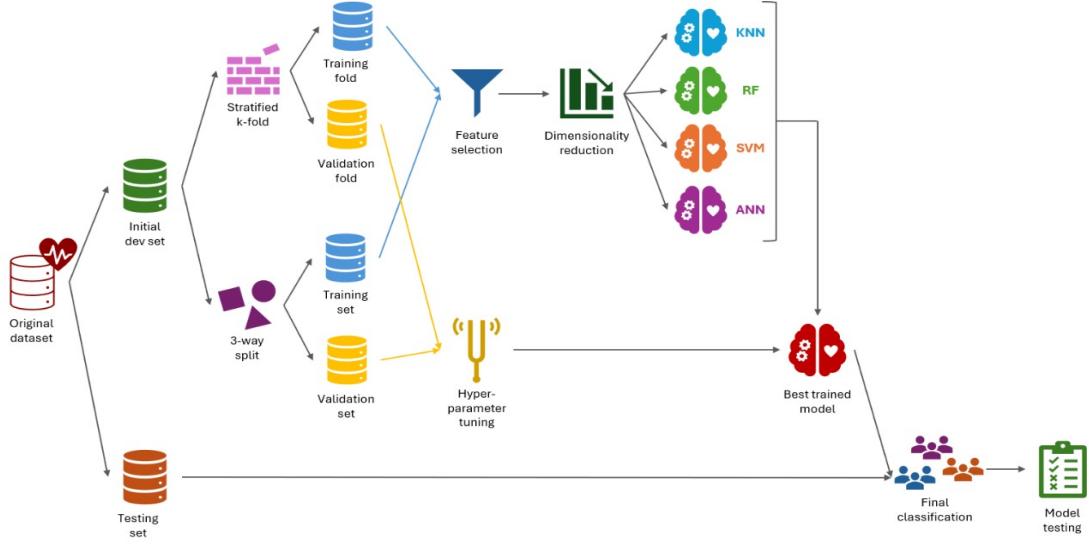


Figure 1: Methodology schema.

2.1 Exploratory Data Analysis (EDA)

Dataset Description

The ACDC dataset is well-curated, with no missing values and 20 patients per class across five categories: Dilated Cardiomyopathy (DCM), Hypertrophic Cardiomyopathy (HCM), Myocardial Infarction (MINF), Normal (NOR), and Right Ventricular abnormality (RV). Only weight and height were provided for demographics, so we computed the Body Mass Index (BMI) to analyze their distribution across classes, defined as:

$$\text{BMI} = \frac{\text{Weight (kg)}}{\text{Height (m)}^2}. \quad (1)$$

The BMI distributions across classes are shown in Figure A.1. We observe that most patients across all classes fall within the normal and overweight BMI ranges. HCM patients tend to have higher BMI values, with several falling into the obese categories, while RV patients generally show lower BMI values compared to other groups.

Outliers Detection

Next, we checked for the presence of outliers using the Interquartile Range (IQR) method. A value x is defined as an outlier if it satisfies the condition:

$$x < Q_1 - 1.5 \times \text{IQR} \quad \text{or} \quad x > Q_3 + 1.5 \times \text{IQR}, \quad (2)$$

where the IQR is given by $Q_3 - Q_1$. In Figure A.2, we show the number of outliers detected per class for the top ten features with the highest total outlier counts. Given that each class has 20 observations per feature, an average of approximately one outlier per class suggests that outliers are sparse and evenly distributed. Therefore, we conclude that no outlier removal is necessary.

Normalization

For normalization, we applied the `StandardScaler` method. Prior to this, we used the Shapiro-Wilk test to verify that the features approximately followed a Gaussian distribution. Figure A.3 presents Q-Q plots and corresponding p -values, showing that most features reasonably follow a normal distribution. `StandardScaler` transforms each feature x according to:

$$z = \frac{x - \mu}{\sigma}, \quad (3)$$

where μ is the mean and σ is the standard deviation of the feature. This ensures that each feature has a mean of zero and a standard deviation of one.

Correlation Analysis

Finally, we computed the Pearson correlation coefficient between each radiomic feature and the class labels. Since the target variable is categorical with five classes, we applied a One-vs-Rest (OvR) strategy, where each class was binarized and analyzed independently against the rest. For a given class, the correlation was computed as:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}}, \quad (4)$$

where x represents the feature values and y the binary labels (one for the class of interest, zero otherwise). Figure A.4 shows the top five most correlated features for each class. Shape and texture descriptors are the most prominent among the highly correlated features. However, what is evident is the high level of multicollinearity among the radiomic features, as shown in Figure A.5.

2.2 Data Partitioning

Simple Split

In the Simple Split strategy, the dataset was initially divided into 80% for training and validation, and 20% for testing. The 80% training-validation portion was further split into 75% for training and 25% for validation, resulting in final proportions of 60% training, 20% validation, and 20% testing (Figure A.6). This 80/20 split is a widely accepted convention in machine learning, particularly for small to medium-sized datasets, as it provides a practical balance between allocating sufficient data for model training and retaining enough data to evaluate performance on previously unseen cases [7].

Stratified K-Fold

For the Stratified K-Fold strategy, we again used the initial 80% of the data for training and validation, resulting in 80 samples, corresponding to 16 samples per class ($80/5 = 16$). We then applied cross-validation while maintaining class balance. Although $k = 5$ is often considered a gold standard for small medical imaging datasets to balance bias and variance [8], we selected $k = 4$ to better match the dataset structure. This ensured that each class contributed 12 samples for training and 4 samples for validation across the folds, as shown in Figure A.7.

2.3 Dimensionality Reduction

Modified Least Absolute Shrinkage and Selection Operator (LASSOmodf)

To address the strong multicollinearity and redundancy typical of radiomic features, we adopted a multi-stage feature selection strategy [9]. First, variance filtering was applied to remove features with zero variance. Then, an ANOVA F-test was performed to retain features with statistically significant differences across classes, selecting those with $p < 0.05$. Subsequently, a sparse model was obtained by solving the LASSO optimization problem:

$$J(\mathbf{h}) = \sum_{i=1}^m \left(y^{(i)} - \sum_{j=0}^n h_j x_j^{(i)} \right)^2 + \lambda \sum_{j=0}^n |h_j|, \quad (5)$$

where $x^{(i)}$ and $y^{(i)}$ represent the features and label of the i -th sample, \mathbf{h} are the regression coefficients, and λ controls sparsity. The LASSO coefficient trajectories across different values of λ are illustrated in Figure A.8.

Both the regularization parameter λ and the coefficient threshold `coe_thr` were optimized through a similar grid search procedure: for each candidate value, the Mean Squared Error (MSE) on a validation subset was computed and the value minimizing MSE was selected, as shown in Figures A.9 and A.10. Specifically, the validation MSE for a given parameter θ was computed as:

$$\text{MSE}(\theta) = \frac{1}{|D_{\text{val}}|} \sum_{i \in D_{\text{val}}} (y^{(i)} - \hat{y}^{(i)}(\theta))^2, \quad (6)$$

where D_{val} is the validation subset, θ corresponds either to λ or to `coe_thr`, and $\hat{y}^{(i)}(\theta)$ denotes the prediction under the given parameter. For coefficient thresholding, after fitting the LASSO model with the optimal λ_{opt} , the threshold `coe_thr` was varied between the minimum and maximum absolute coefficients:

$$h_{\max} = \max_j |h_j|, \quad h_{\min} = \min_j |h_j| \quad (7)$$

retaining features satisfying $|h_j| > \text{coe_thr}_{\text{opt}}$ for the final feature subset.

Linear Discriminant Analysis (LDA)

Following feature selection, the retained features were further reduced in dimensionality using Linear Discriminant Analysis (LDA). LDA is a supervised technique that projects the data onto a lower-dimensional space by maximizing the separation between classes while minimizing the variance within each class. Since our dataset contains five classes, LDA can reduce the feature space to at most $5 - 1 = 4$ dimensions. Alternative methods like Principal Component Analysis (PCA) were not considered optimal, as our goal was not to interpret individual features but to maximize class separability. Additionally, collinearity had already been addressed during feature selection, making PCA's variance-based transformation unnecessary.

2.4 Models

k-Nearest Neighbors (KNN)

The first supervised model applied was the k-Nearest Neighbors (KNN) classifier. KNN predicts the class of a new sample based on the majority class among its k closest neighbors in the feature space. It is a non-parametric, instance-based learning algorithm that relies on distance comparisons rather than a formal training phase. While not expected to be the most accurate model, its simplicity makes it a useful baseline for comparison. In [10], KNN was selected for its interpretability and effectiveness on small radiomic datasets, particularly when feature space was already reduced like our case.

Random Forest (RF)

The second supervised model applied was the Random Forest (RF) classifier. RF is an ensemble learning method that constructs multiple decision trees using random subsets of the data and features, combining their predictions through majority voting. It is well-suited for modeling non-linear interactions and is robust to feature correlations, making it effective in handling complex relationships within radiomic data. Li et al. [11] demonstrated similar benefits when applying RF to predict treatment outcomes from radiomic features in cervical cancer, citing the model's stability and good generalization on structured imaging data.

Support Vector Machine (SVM)

The third supervised model applied was the Support Vector Machine (SVM) classifier. SVM is a supervised learning algorithm that aims to find the optimal hyperplane that separates classes by maximizing the margin between the closest data points of different classes, known as support vectors. In our pipeline, SVM was a suitable choice because the dimensionality reduction ensured a compact feature space, allowing SVM to effectively maximize class separation. This is supported by the findings of Xu et al. [12], who employed SVM for lymph node metastasis prediction using radiomics.

Artificial Neural Network (ANN)

The last supervised model applied was the Artificial Neural Network (ANN). ANNs consist of interconnected layers of neurons, where each neuron computes a weighted sum of its inputs followed by a non-linear activation function. However, typically they require careful tuning and regularization to prevent overfitting, especially when working with limited data. In a related study, Mulet de los Reyes et al. [13] demonstrated the ability of ANNs to perform robust tumor segmentation from radiomic inputs by leveraging feature learning. Their findings suggest that when paired with appropriate preprocessing and dimensionality reduction, ANNs can effectively learn from structured imaging-derived datasets.

2.5 Metrics

Model Selection

During hyperparameter tuning, models were evaluated using a custom score designed to balance generalization and overfitting:

$$\text{Score} = \text{Validation Acc.} - p \times |\text{Training Acc.} - \text{Validation Acc.}|, \quad (8)$$

where $p = 0.5$ controls the penalty for overfitting. The best hyperparameters were selected by maximizing this score, encouraging models with high validation performance and low overfitting.

Model Evaluation

After training, the following evaluation metrics were computed to evaluate the model's performance on the test set (Table 1):

Table 1: Evaluation metrics used for model assessment.

Metric	Definition	Formula
Accuracy (Macro)	Measures the overall proportion of correctly classified samples, averaged equally across all classes.	$\frac{1}{C} \sum_{c=1}^C \frac{\text{TP}_c + \text{TN}_c}{\text{TP}_c + \text{FP}_c + \text{FN}_c + \text{TN}_c}$
Precision (Macro)	Measures the ability of the model to correctly identify positive instances across all classes.	$\frac{1}{C} \sum_{c=1}^C \frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c}$
Recall (Macro)	Measures the ability of the model to capture all relevant instances for each class.	$\frac{1}{C} \sum_{c=1}^C \frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c}$
F1-Score (Macro)	Harmonic mean between precision and recall, providing a balance between the two.	$\frac{1}{C} \sum_{c=1}^C 2 \times \frac{\text{Precision}_c \times \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c}$

In addition to the overall metrics, the confusion matrix, Receiver Operating Characteristic (ROC-AUC) curves, and class-wise performance metrics were also evaluated.

3. Experimental Design

The experimental design of the project was structured into two pipelines: the Simple Split pipeline and the K-Fold pipeline (Figure 2).

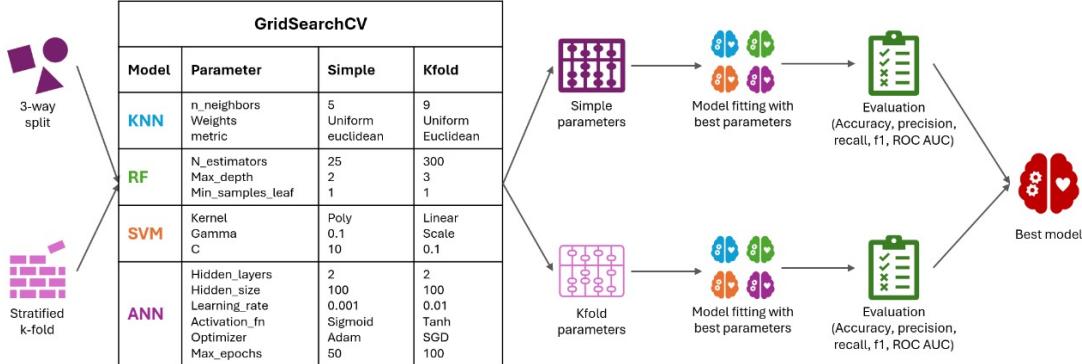


Figure 2: Experimental design schema.

3.1 PyRadiomics

Radiomic features were computed for three cardiac structures: Left Ventricle (LV), Myocardium (MYO), and Right Ventricle (RV), at two different time points of the cardiac cycle: End-Systole (ES) and End-Diastole (ED). In Table 2, we present the various feature classes generated by PyRadiomics, along with brief descriptions of each.

Table 2: PyRadiomics feature classes and their descriptions.

Feature Class	Description
First-order	Represents statistical properties of voxel intensities, such as mean, variance, and skewness.
Shape	Describes the geometric properties of an ROI, including surface area, volume, and sphericity.
GLCM	Quantifies texture by analyzing the spatial relationships between pixel intensities.
GLRLM	Measures the distribution of pixel intensities, focusing on the length of consecutive pixels with the same intensity.
GLSZM	Quantifies the size of homogeneous regions by counting connected voxels with similar intensities.
NGTDM	Captures local texture variations by measuring differences between a voxel intensity and its neighbors.
GLDM	Describes the dependency between gray levels in an image, focusing on their spatial arrangement from the central voxel.

3.2 Feature Selection and Transformation

The final features selected by the LASSOmodf method were (10): Shape Mesh Volume LV ED, GLCM Idmn LV ED, GLSZM Large Area High Gray Level Emphasis LV ED, Shape Elongation LV ES, Shape Mesh Volume RV ES, First Order Kurtosis RV ES, First Order Skewness RV ES, GLDM Gray Level Non Uniformity RV ES, Shape Minor Axis Length MYO ES, and NGTDM Coarseness MYO ES. These features were then used as input for LDA transformation.

3.3 Model Hyperparameter Tuning

Hyperparameter optimization was performed using GridSearchCV. In the Annexes (Figures A.11 - A.14), we illustrate the evolution of selected hyperparameters across the Simple Split and Stratified K-Fold pipelines. The Simple Split strategy often results in large discrepancies between training and validation performance, suggesting overfitting due to reliance on a single training set. In contrast, the Stratified K-Fold approach yields more consistent and generally lower scores, indicating improved generalization and greater stability.

For KNN, the hyperparameters were:

- `n_neighbors`: Number of neighbors used for classification; higher values yield smoother decision boundaries but may reduce model sensitivity.
- `weights`: Strategy to assign weights to neighbors—either equally (`uniform`) or inversely proportional to distance (`distance`).
- `metric`: Distance measure used to compute similarity between data points.

For RF, the hyperparameters were:

- `n_estimators`: Total number of decision trees in the ensemble; more trees generally improve performance but increase computation time.
- `max_depth`: Maximum allowed depth of each tree; deeper trees can capture complex patterns but risk overfitting.
- `min_samples_leaf`: Minimum number of samples required to form a leaf node; higher values constrain model complexity and reduce overfitting.

For SVM, the hyperparameters were:

- `kernel`: Function that defines the feature transformation used to separate classes (e.g., linear, RBF, polynomial).
- `gamma`: Defines how far the influence of a single training example reaches; high values create more flexible, complex boundaries.
- `C`: Regularization parameter that balances margin maximization and classification error; lower values favor wider margins.

For ANN, the hyperparameters were:

- `hidden_layers`: Number of hidden layers in the network; deeper networks can model more complex relationships.
- `hidden_size`: Number of neurons per hidden layer; larger sizes can capture more nuanced features.
- `learning_rate`: Step size for updating weights during training; controls speed and stability of convergence.
- `activation_fn`: Non-linear function applied to neurons (e.g., ReLU, sigmoid); enables learning of complex patterns.
- `optimizer`: Optimization algorithm used for training (e.g., Adam, SGD); affects convergence rate and stability.
- `max_epochs`: Maximum number of complete passes through the training data; sets the upper limit for training duration.

4. Results

This section presents the main results of the study. First, we compare the performance of the evaluated models. Then, we analyze the classification performance of the best model through confusion matrices, ROC-AUC curves, and class-wise metrics. Figures for the rest of the models can be found attached in Annexes.

4.1 Model Comparison

Table 3 presents a detailed comparison of four classification models—KNN, RF, SVM, and ANN—under three validation strategies: Baseline, Simple Split, and Stratified K-Fold.

Under the Baseline setting, where no validation is applied, overall performance is low. KNN and SVM both show an accuracy of 0.35 with wide 95% confidence intervals (0.14, 0.56), while ANN performs worst with 0.20 (0.02, 0.38), indicating instability. RF performs noticeably better, reaching 0.75 (0.56, 0.94), clearly higher than the others.

In the Simple Split strategy, all models improve. KNN and ANN both reach 0.70 (0.50, 0.90), and SVM increases to 0.60 (0.39, 0.81). RF remains the top performer with 0.80 (0.62, 0.98). This suggests that even a basic train-test split brings performance gains over the baseline.

With Stratified K-Fold, performance increases further and the ranking among models shifts. SVM now achieves the highest accuracy at 0.85 with a relatively narrow interval (0.69, 1.00), showing more consistent performance. KNN and Random Forest both reach 0.80 with the same interval (0.62, 0.98), while ANN remains at 0.70 (0.50, 0.90). The intervals across models are generally narrower here, which may point to more reliable estimates under this strategy.

Table 3: Performance metrics under different validation strategies (Baseline, Simple Split, and Stratified K-Fold). KNN = k-Nearest Neighbors, RF = Random Forest, SVM = Support Vector Machine, ANN = Artificial Neural Network.

Split	Metric	KNN	RF	SVM	ANN
Baseline	Accuracy (95% CI)	0.35 (0.14, 0.56)	0.75 (0.56, 0.94)	0.35 (0.14, 0.56)	0.20 (0.02, 0.38)
	Precision (macro)	0.45	0.74	0.57	0.04
	Recall (macro)	0.35	0.75	0.35	0.20
	F1 Score (macro)	0.38	0.73	0.35	0.07
	ROC-AUC (macro)	0.72	0.84	0.50	0.50
Simple	Accuracy (95% CI)	0.70 (0.50, 0.90)	0.80 (0.62, 0.98)	0.60 (0.39, 0.81)	0.70 (0.50, 0.90)
	Precision (macro)	0.78	0.82	0.73	0.73
	Recall (macro)	0.70	0.80	0.60	0.70
	F1 Score (macro)	0.71	0.80	0.60	0.69
	ROC-AUC (macro)	0.88	0.95	0.87	0.91
K-Fold	Accuracy (95% CI)	0.80 (0.62, 0.98)	0.80 (0.62, 0.98)	0.85 (0.69, 1.00)	0.70 (0.50, 0.90)
	Precision (macro)	0.83	0.83	0.89	0.80
	Recall (macro)	0.80	0.80	0.85	0.70
	F1 Score (macro)	0.80	0.80	0.85	0.67
	ROC-AUC (macro)	0.95	0.95	0.98	0.95

4.2 Confusion Matrix and ROC-AUC Curve

Selecting our best-performing model—SVM evaluated using the stratified K-Fold strategy—we observe robust and well-balanced performance across all five cardiac conditions, as shown in Figure 3.

The confusion matrix shows perfect classification for all cases of MINF and RV. On the other hand, DCM and HCM are misclassified as MINF, which may indicate that the model tends to associate overlapping structural features from these cardiomyopathies with those of MINF. This is in line with prior observations by Bellotti et al. [14], who noted that DCM and HCM can exhibit morphologic characteristics that may overlap not only with each other but also with features observed in MINF, particularly in early or atypical stages. One NOR case is also misclassified as RV, but overall the predictions remain highly accurate. Interestingly, the strong performance of the RV class stands out, especially considering that the right ventricle is often described as more difficult to assess and characterize due to its complex geometry, thin walls, and highly variable contraction patterns [15].

The ROC-AUC curves reinforce these results, with all classes achieving values above 0.94. Notably, RV reaches 1.00, while DCM, NOR, and HCM each exceed 0.97.

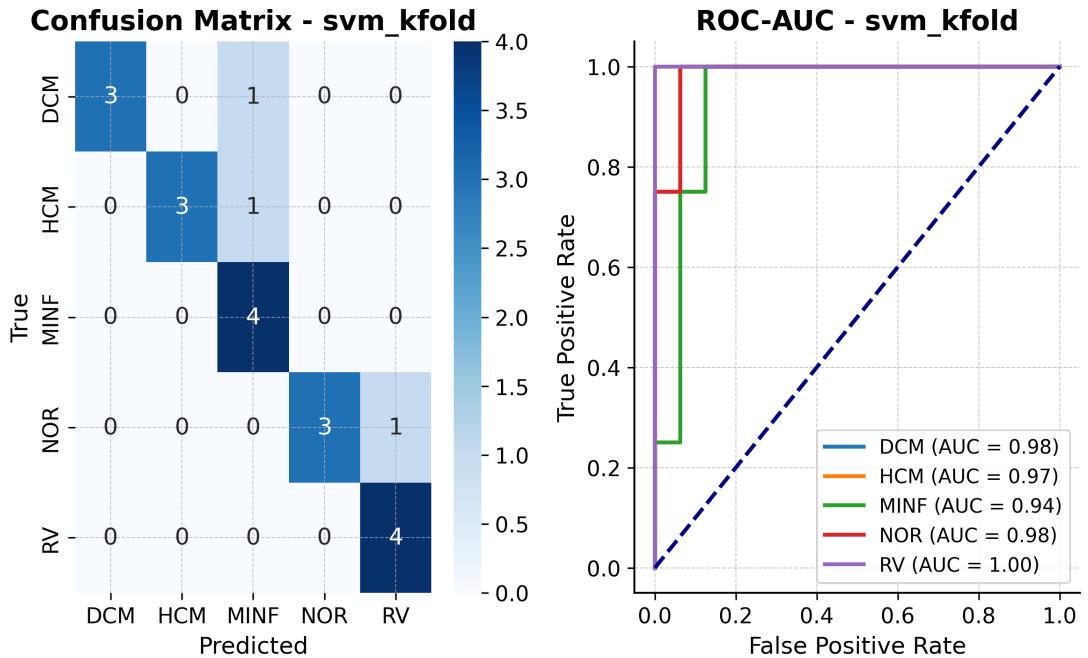


Figure 3: Confusion matrix and ROC curves for the Support Vector Machine (SVM) model trained with Stratified K-Fold cross-validation. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

4.3 Class-wise Metrics

To ensure that no single class disproportionately influenced the overall performance metrics, we analyzed class-specific precision, recall, and F1-scores, as presented in Figure 4.

DCM and HCM show high precision but lower recall, suggesting that the model is selective when assigning these labels and may miss some true cases. Both classes were previously misclassified as MINF, indicating potential overlap in their structural features. Therefore, even though MINF achieves perfect recall, it shows lower precision, reflecting a tendency to attract false positives. NOR shows balanced performance, with only a slight drop in recall due to one misclassified case. RV maintains strong and consistent values across all metrics, indicating that it was both reliably identified and not confused with other classes. This final analysis further supports the patterns observed in the confusion matrix and ROC-AUC curves.

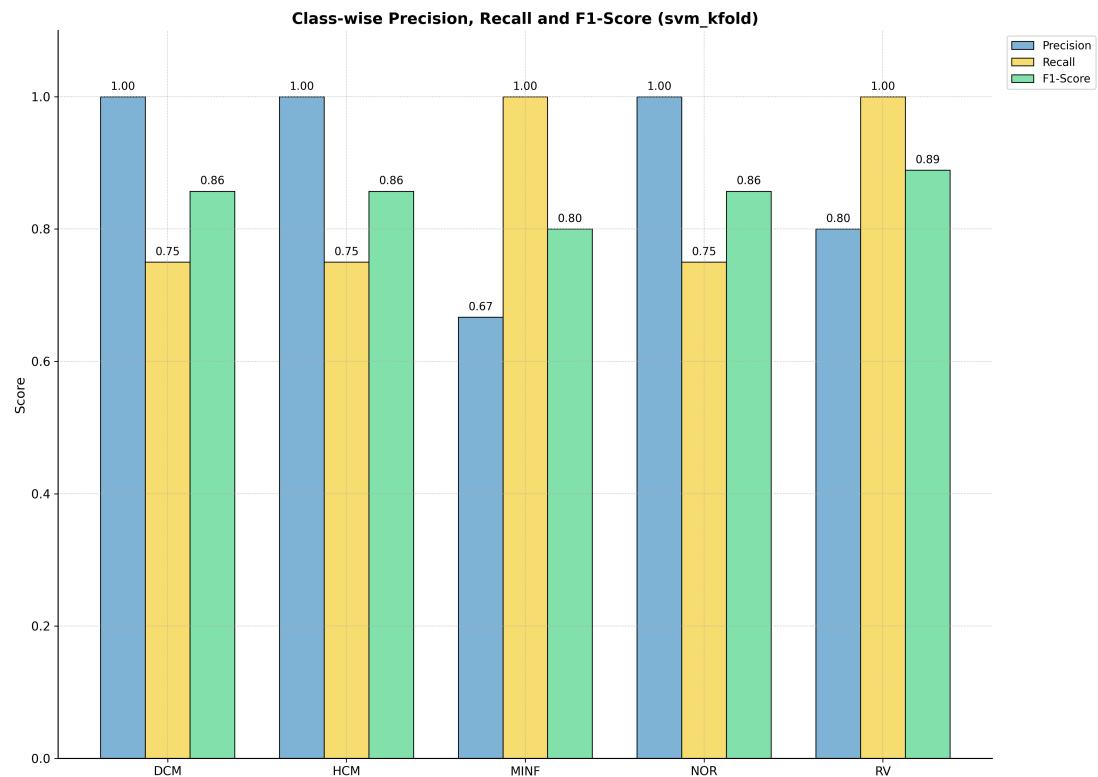


Figure 4: Class-wise Precision, Recall, and F1-Score for the Support Vector Machine (SVM) model trained with Stratified K-Fold cross-validation. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

5. Discussion and Conclusions

In this first challenge, we worked with the Automated Cardiac Diagnosis Challenge (ACDC) dataset with the objective of classifying patients into five cardiac conditions: Dilated Cardiomyopathy (DCM), Hypertrophic Cardiomyopathy (HCM), Myocardial Infarction (MINF), Normal (NOR), and Right Ventricular abnormality (RV). Our team developed and compared different pipelines using State of the Art (SoA) methodological approaches to evaluate their performance and identify the best-performing model.

The ACDC dataset, being well-curated and balanced across classes, simplified preprocessing. Our initial exploratory analysis confirmed the absence of major issues such as class imbalance or extreme outliers. Normalization was verified and applied appropriately. We also noted a considerable degree of multicollinearity, which is expected given that many features extracted via the PyRadiomics library describe closely related structural and texture-based characteristics.

To address these challenges, we implemented two primary classification pipelines in addition to a baseline model. Due to the high dimensionality and correlated nature of the features, we applied an enhanced version of the Least Absolute Shrinkage and Selection Operator (LASSO) method that not only performs feature selection but also eliminates variables that contribute weakly to model performance. Following this step, the selected features were projected into a lower-dimensional space using Linear Discriminant Analysis (LDA). While this transformation improves class separability, it introduces a limitation in interpretability, as the original radiomic features are replaced by linear combinations. Given that the main goal of this project was classification rather than clinical interpretability, this trade-off was considered acceptable. However, in future work—particularly in clinical contexts—methods focusing solely on feature selection may be more appropriate to preserve interpretability.

We evaluated four classifiers—K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN)—optimized using GridSearchCV. Hyperparameter grids were manually defined but restricted to reasonable ranges (e.g., excluding very low values for k in KNN to mitigate overfitting). The tuning process was computationally demanding, especially for ANN, which required more time to converge due to the model's depth and the need to learn complex feature interactions. This helps explain the ANN's low accuracy under the baseline setting, where the model had a shallow architecture and very limited training data.

Regarding the performance of the selected data splitting strategies and the feature reduction pipeline based on LASSOmodf combined with LDA, the progression from baseline to simple split and then to stratified K-Fold clearly highlights the importance of choosing the proper methods and parameters before entering into modelling. The baseline approach, which lacked structured validation, resulted in reduced reliability and increased variability. The simple split

improved stability by introducing a fixed validation set, while stratified K-Fold further enhanced generalization by maintaining class balance across folds. The main advantage of using stratified K-Fold over baseline and simple split is its ability to generate multiple training and validation partitions, allowing for more reliable performance estimates and better use of limited data.

Regarding model comparisons, KNN was not expected to perform best due to its sensitivity to noise and overfitting, but it still produced reasonably stable results, achieving up to 0.80 accuracy under K-Fold validation. RF outperformed the other models in the baseline and simple split settings with accuracies of 0.75 and 0.80, respectively, and ranked second in the K-Fold setting. ANN, while more computationally intensive to train, showed improved results with more data and deeper structure, reaching 0.70 accuracy under K-Fold. Ultimately, SVM emerged as the top-performing model, achieving the highest accuracy of 0.85 under stratified K-Fold and consistently high performance across all metrics. This may be attributed to SVM's capacity to handle high-dimensional, correlated input spaces effectively, particularly when paired with LDA, which generates linearly separable components. The clear separation in the transformed feature space likely enabled the SVM to construct optimal decision boundaries, making it particularly suitable for this classification task.

In summary, our approach combining advanced feature selection, dimensionality reduction, and model evaluation proved effective. While interpretability was partially compromised by the use of LDA, the classification results were robust and align with expected trends, supporting the pipeline's validity.

References

- [1] World Health Organization. Cardiovascular diseases (cvds), 2021. Accessed: 2025-04-03. URL: https://www.who.int/cardiovascular_diseases.
- [2] Z. Raisi-Estabragh, C. Izquierdo, V. M. Campello, C. Martin-Isla, A. Jaggi, N. C. Harvey, K. Lekadir, and S. E. Petersen. Cardiac magnetic resonance radiomics: Basic principles and clinical perspectives. *European Heart Journal - Cardiovascular Imaging*, 21(4):349–356, 2020. doi:10.1093/ejhci/jeaa028.
- [3] Philippe Lambin, Rianne T. H. Leijenaar, Thomas M. Deist, Jef Peerlings, Edwin E. C. de Jong, Justin van Timmeren, Silvia Sanduleanu, Romain T. H. M. Larue, Alexander J. G. Even, Aniek Jochems, Yvonne van Wijk, Henry Woodruff, Jesse van Soest, Tessa Lustberg, Erik Roelofs, Wouter van Elmpt, Andre Dekker, Felix M. Mottaghay, Joachim E. Wildberger, and Sean Walsh. Radiomics: the bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology*, 14(12):749–762, December 2017. Epub 2017 Oct 4. doi:10.1038/nrclinonc.2017.141.
- [4] R. J. Gillies, P. E. Kinahan, and H. Hricak. Radiomics: Images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2016. doi:10.1148/radiol.2015151169.
- [5] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, and et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Transactions on Medical Imaging*, 37(11):2514–2525, 2018. doi:10.1109/TMI.2018.2837502.
- [6] J. F. S. et al. R. Van Griethuysen. Pyradiomics: A python-based radiomics library, 2024. URL: <https://pyradiomics.readthedocs.io/en/latest/>.
- [7] Will Koehrsen. Beyond 80/20: A practical guide to train/test splits in machine learning, 2018. Accessed: 2025-04-29. URL: <https://medium.com/data-science/beyond-80-20-a-practical-guide-to-train-test-splits-in-machine-learning-5fc62ebe276f>.
- [8] Andrius Vabalas, Emma Gowen, Ellen Poliakoff, and Alexander J. Casson. Machine learning algorithm validation with a limited sample size. *PLOS ONE*, 14(11):e0224365, 2019. URL: <https://app.dimensions.ai/details/publication/pub.1122390471>, doi:10.1371/journal.pone.0224365.
- [9] Ke Wang, Ying An, Jiancun Zhou, Yuehong Long, and Xianlai Chen. A novel multi-level feature selection method for radiomics. *Alexandria Engineering Journal*, 66:993–999, 2023. URL: <https://www.sciencedirect.com/science/article/pii/S1110016822007268>, doi:10.1016/j.aej.2022.10.069.

- [10] Anil Pandey, Jagrati Chaudhary, Sumit Garg, Aditi Khurana, Param Sharma, Rachna Seth, Chetan Patel, and Rakesh Kumar. A machine learning model for differentiating lytic lesion of Ich and tuberculosis using haralick texture features and k-nearest neighbour (knn) as a classifier. *Journal of Nuclear Medicine*, 64(supplement 1):P592, 2023. URL: https://jnm.snmjournals.org/content/64/supplement_1/P592.
- [11] Xiaoyan Li, Xiaoyan Wang, Yanfeng Wang, Yuhong Zhang, Xiaoyan Liu, Yanfang Zhang, and Yufeng Wang. Optimisation and evaluation of the random forest model in the prediction of treatment effect of neoadjuvant chemotherapy–radiation therapy on advanced cervical cancer. *Archives of Gynecology and Obstetrics*, 301:1095–1104, 2020. doi:10.1007/s00404-020-05908-5.
- [12] Lihong Xu, Peng Yang, Wen Liang, Wen Liu, Wen Wang, Chun Luo, Xiaoying Zhang, Xiaoyan Deng, Lei Zhang, and Jian Wang. A radiomics approach based on support vector machine using mr images for preoperative lymph node status evaluation in intrahepatic cholangiocarcinoma. *European Radiology*, 29:6607–6617, 2019. doi:10.1007/s00330-019-06237-0.
- [13] Alexander Mulet de los Reyes, Victoria Hyde Lord, María Elena Buen, Daniel Gandia, Luis Gómez Déniz, Maikel Noriega Alemán, and Cecilia Suárez. Combined use of radiomics and artificial neural networks for the three-dimensional automatic segmentation of glioblastoma multiforme. *Expert Systems*, 40(3):e12986, 2023. doi:10.1111/exsy.12986.
- [14] Giorgio Bellotti et al. Comparison of hypertrophic and dilated cardiomyopathy: echocardiographic and morphologic findings. *International Journal of Cardiology*, 85(1):83–91, 2002.
- [15] Bjoern Sievers, Paulus Kirchhof, and Jan Bogaert. Right ventricular morphology and function: from conventional methods to new imaging techniques. *Heart*, 100(4):262–270, 2014.

Annexes



Figure A.1: Distributions of Body Mass Index (BMI, kg/m²) across classes. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

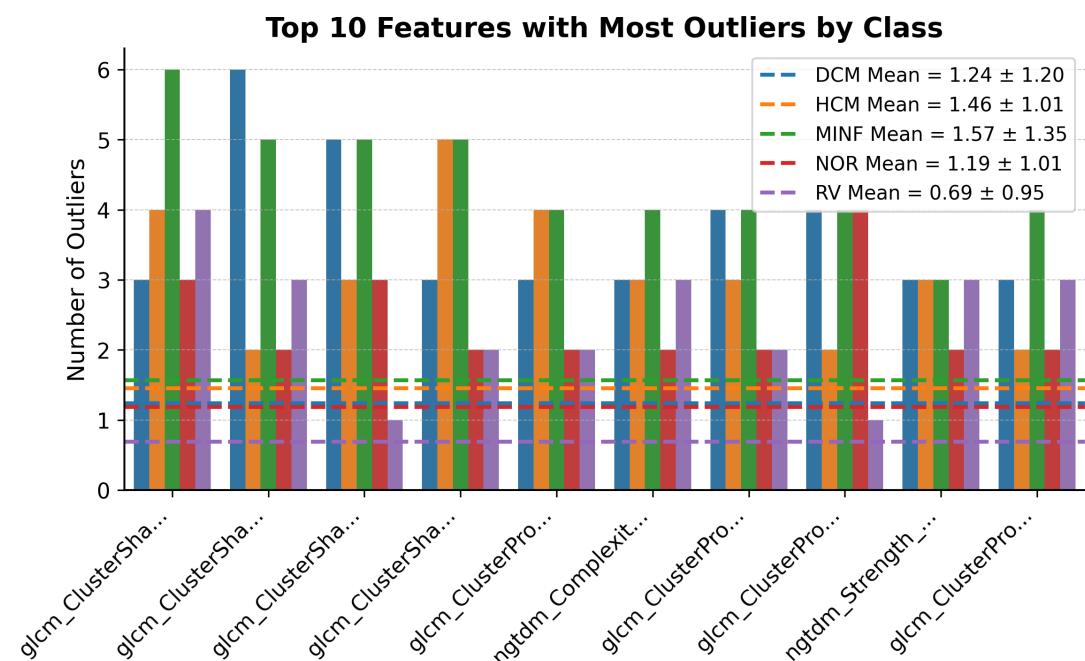


Figure A.2: Top 10 features with the highest total number of outliers. Bars are grouped by class and represent the number of outliers per feature-class pair. The red dashed line marks the overall mean \pm standard deviation.

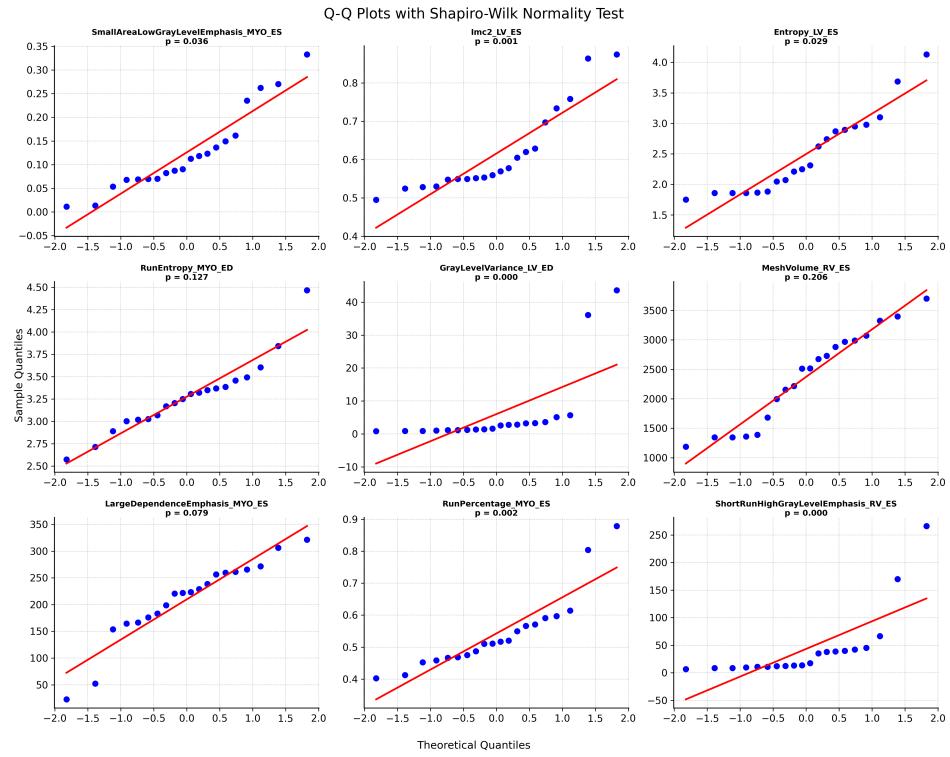


Figure A.3: Q-Q plots for randomly selected features, along with Shapiro-Wilk test p -values. Statistically significant is $p < 0.05$.

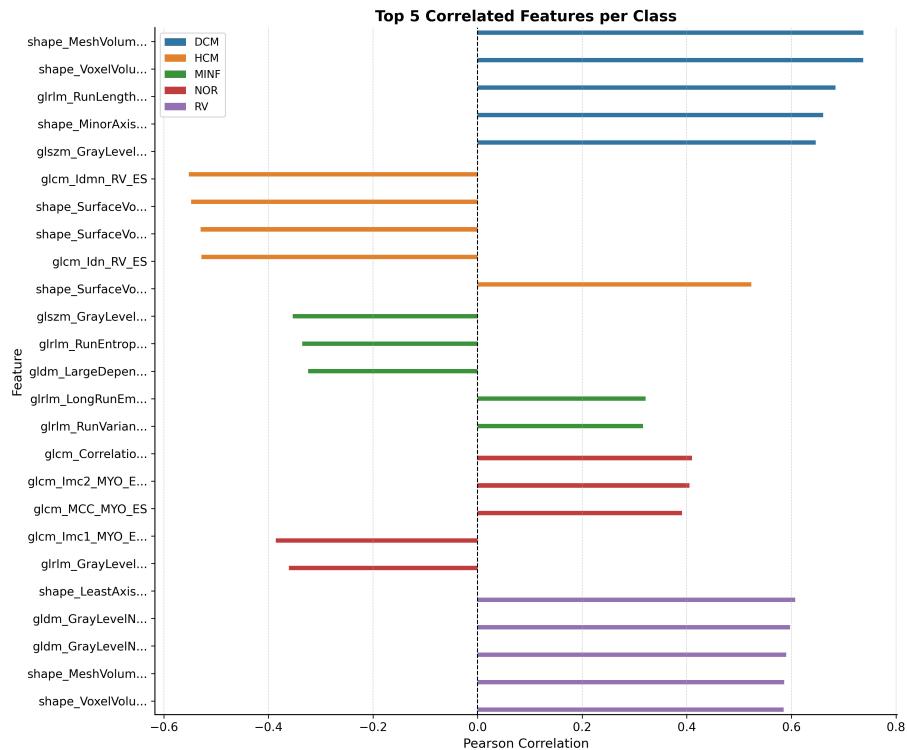


Figure A.4: Top 5 radiomic features most correlated with each diagnostic class based on Pearson correlation. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

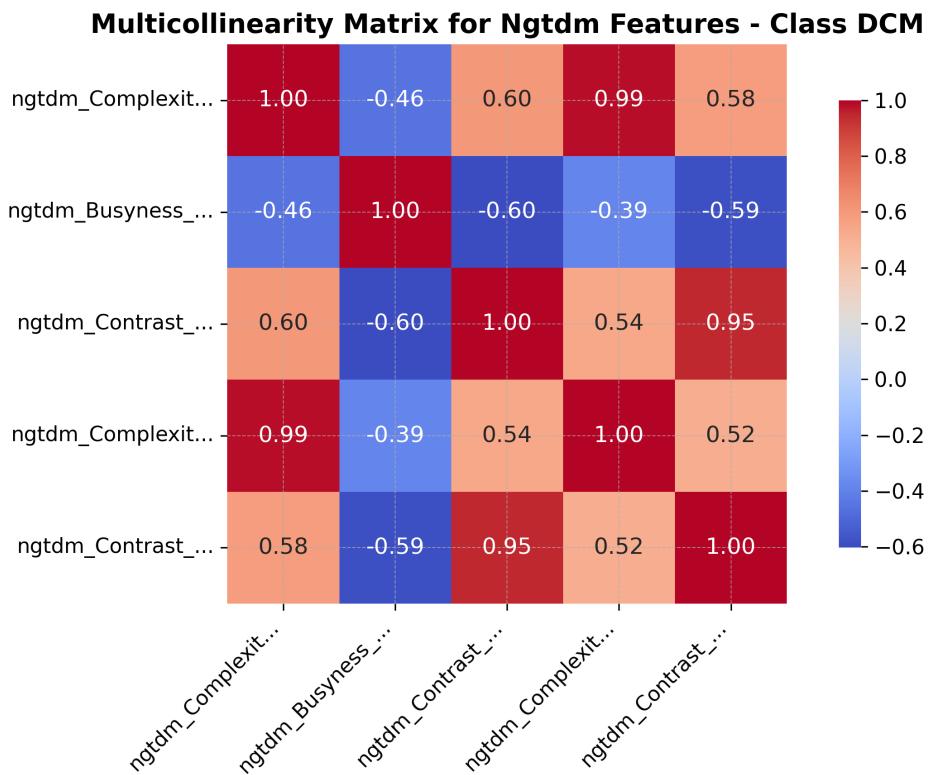


Figure A.5: Multicollinearity matrix for five randomly selected Ngtdm features in the Dilated Cardiomyopathy (DCM) class.



Figure A.6: Visualization of the Simple Split strategy. Each color represents a different class, and hatched bars indicate validation samples. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

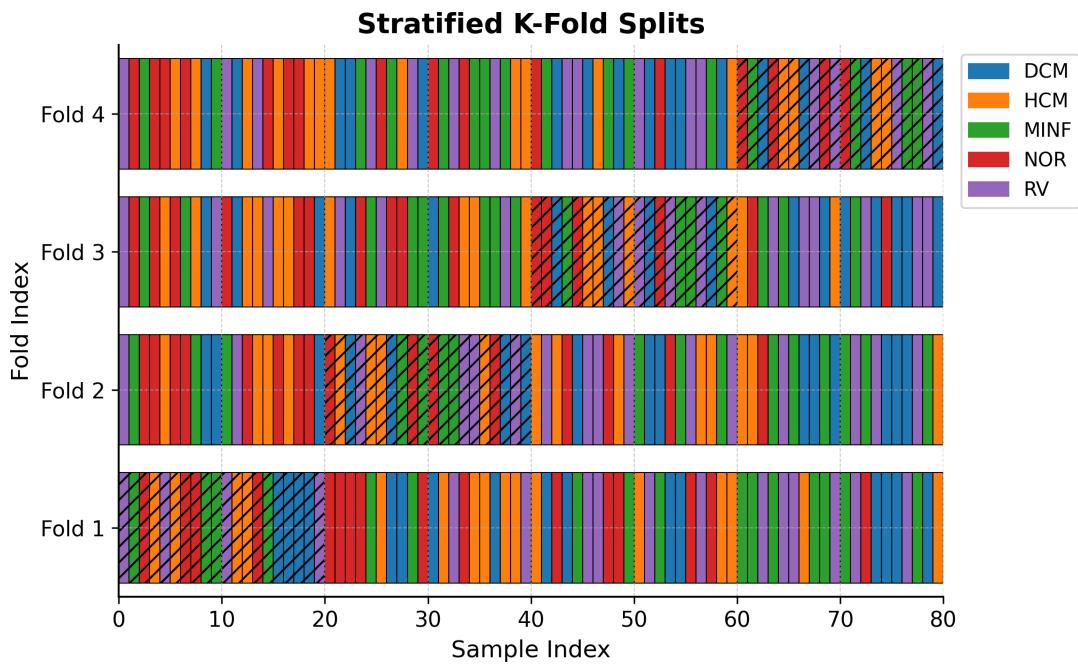


Figure A.7: Visualization of the Stratified K-Fold cross-validation splits. Each color represents a different class, and hatched bars indicate validation samples. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

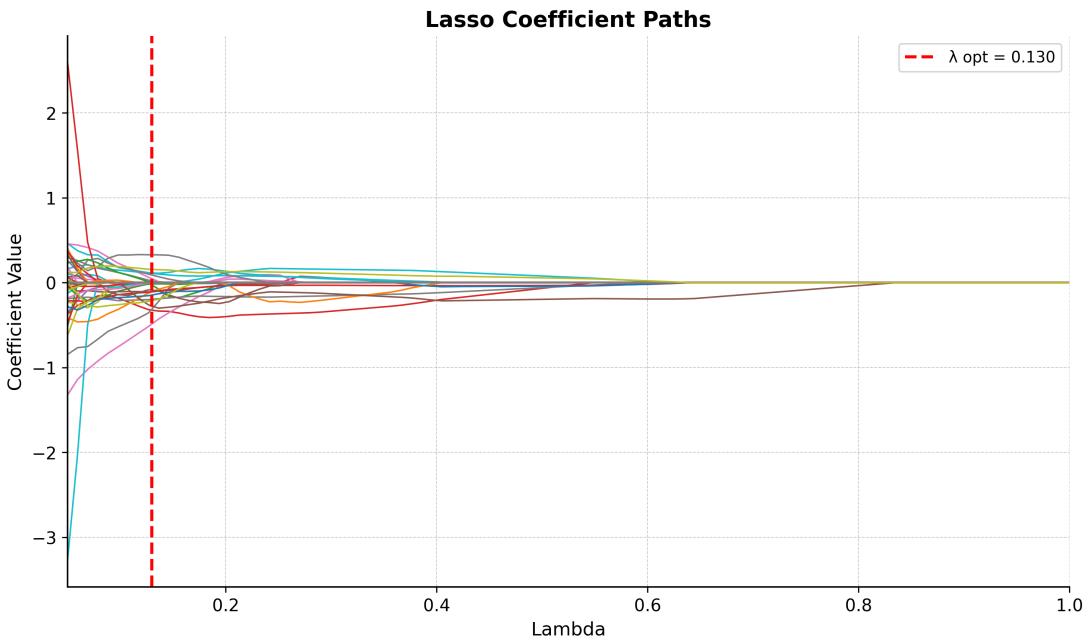


Figure A.8: LASSO coefficient paths across varying λ values. The red dashed line indicates the selected optimal λ_{opt} minimizing validation error.

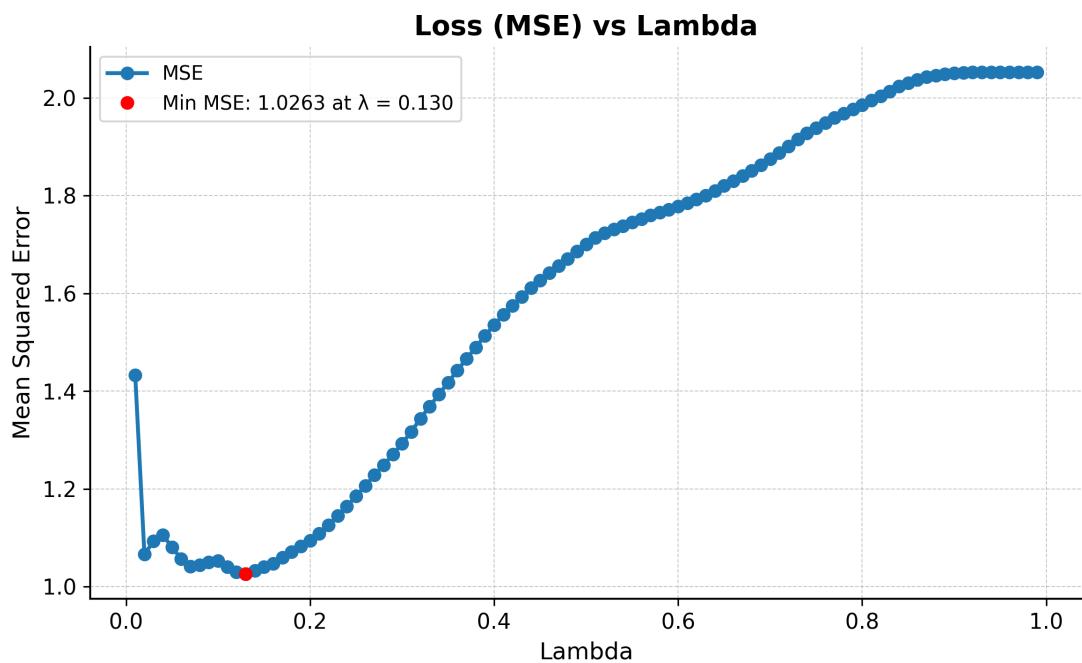


Figure A.9: Mean Squared Error (MSE) as a function of λ . The minimum MSE is marked at the optimal regularization parameter λ_{opt} .

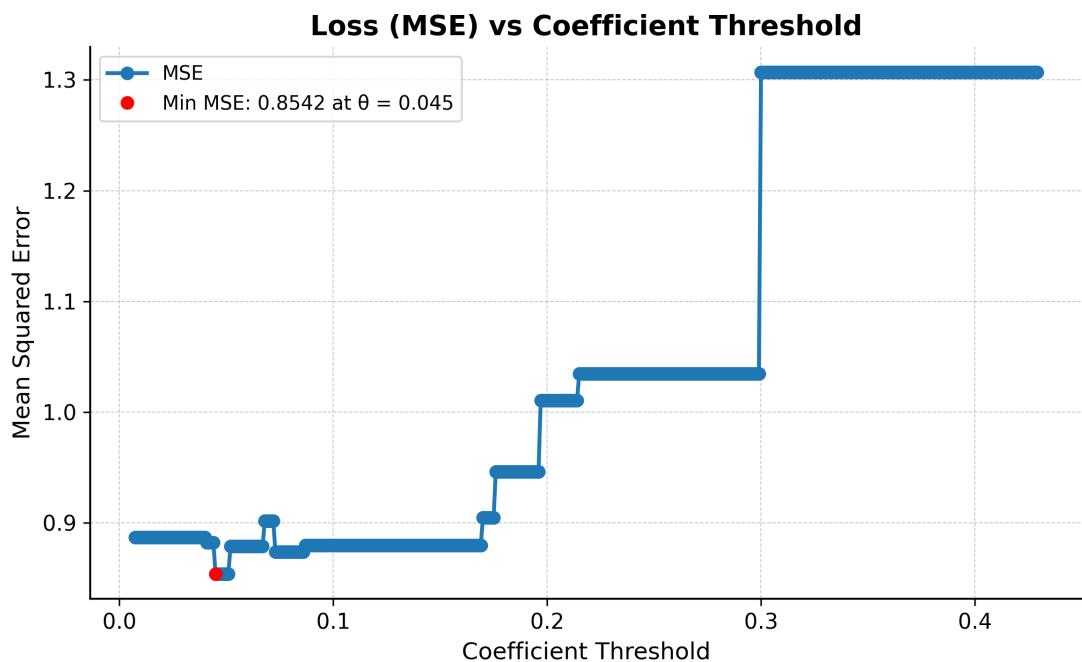


Figure A.10: Mean Squared Error (MSE) as a function of the coefficient threshold coe_thr after LASSO fitting. The optimal threshold $\text{coe_thr}_{\text{opt}}$ corresponds to the minimum MSE.

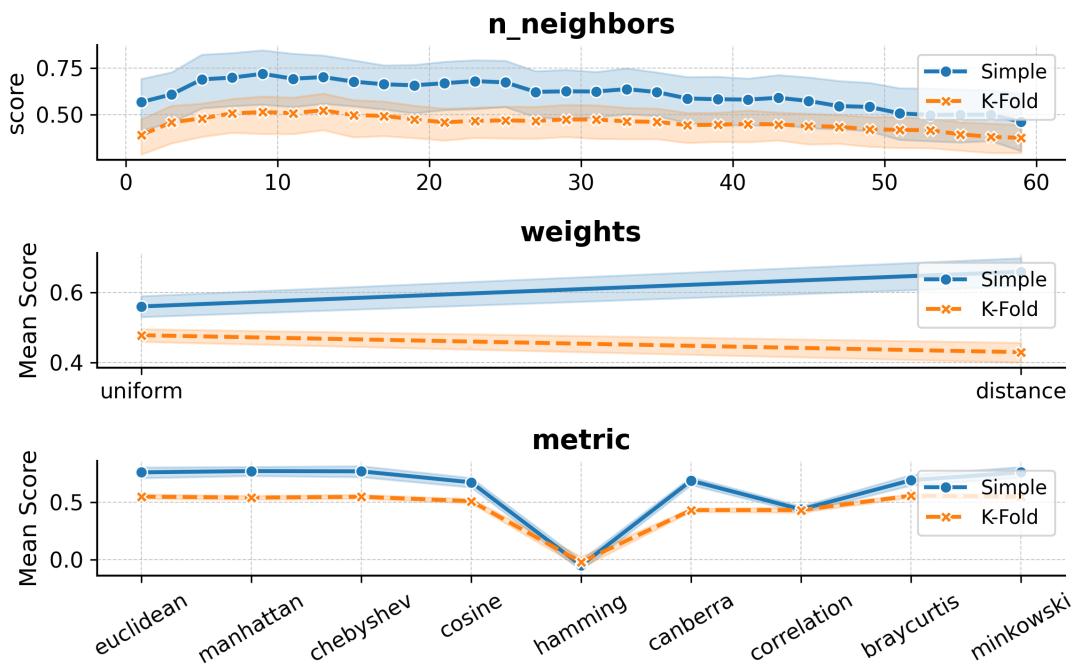


Figure A.11: Hyperparameter evolution for the k-Nearest Neighbors (KNN) model across different values.

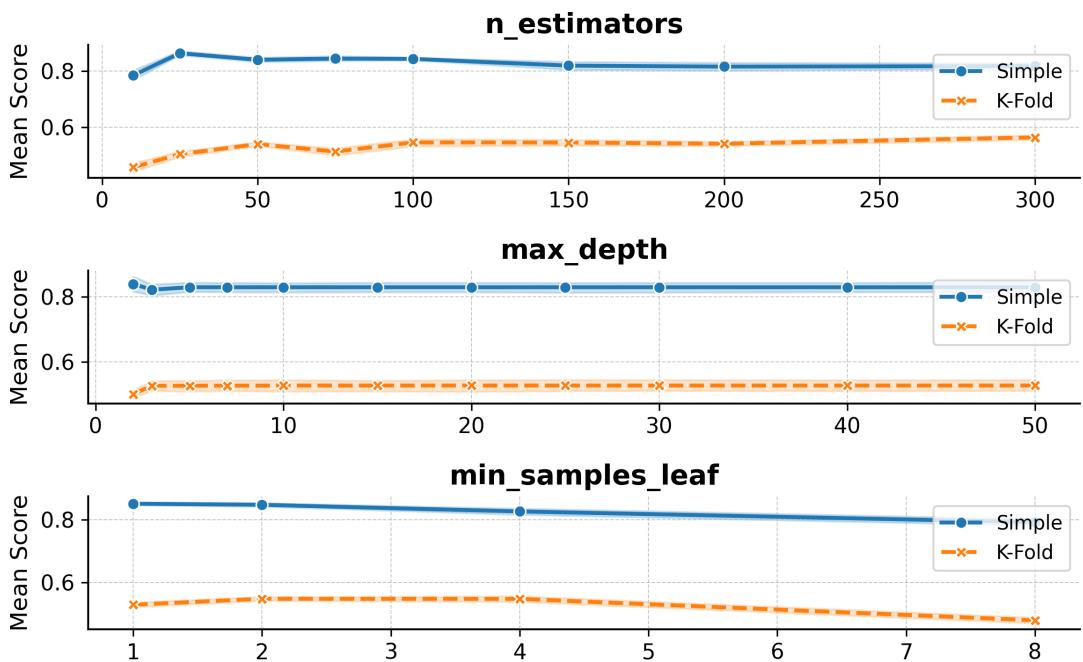


Figure A.12: Hyperparameter evolution for the Random Forest (RF) model across different values.

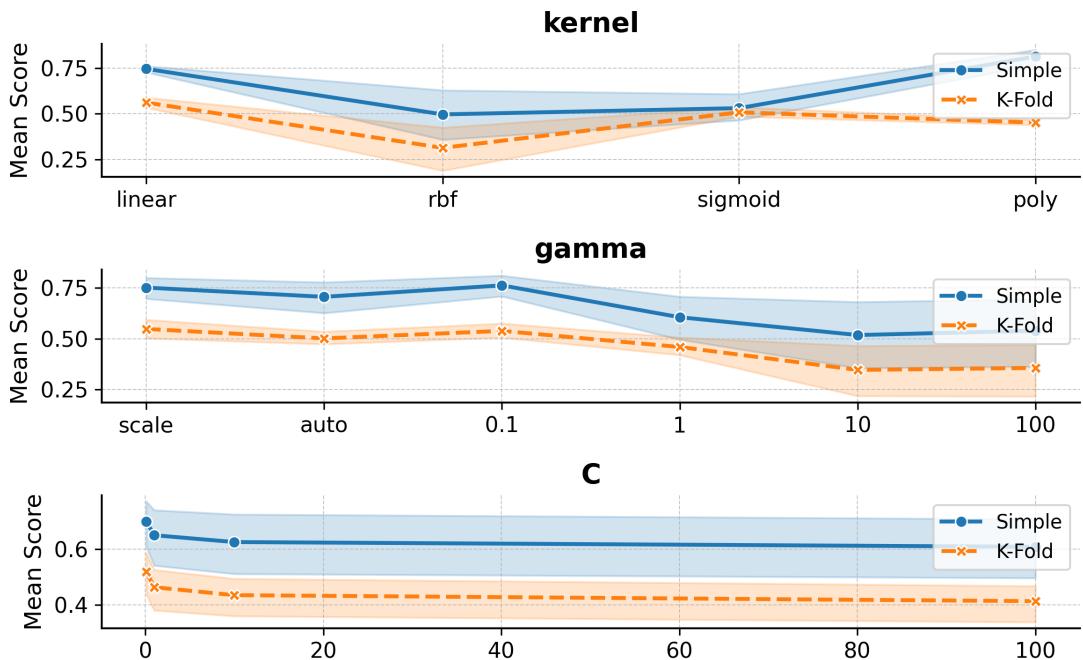


Figure A.13: Hyperparameter evolution for the Support Vector Machine (SVM) model across different values.

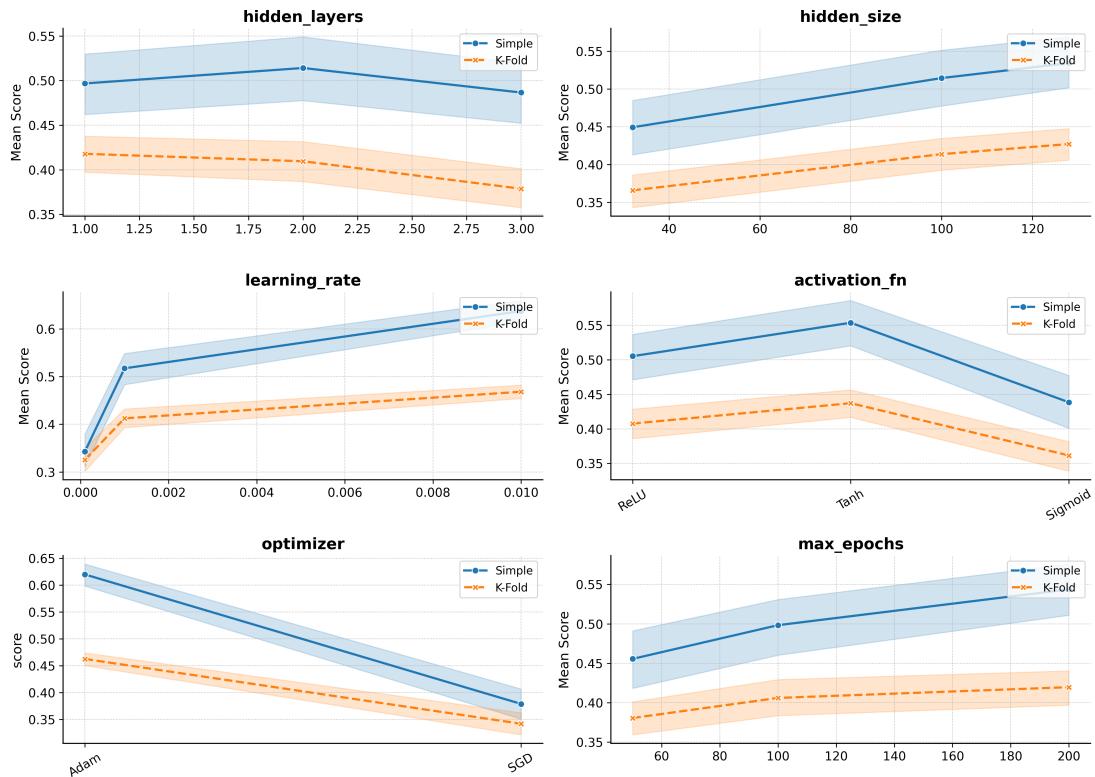


Figure A.14: Hyperparameter evolution for the Artificial Neural Network (ANN) model across different values.

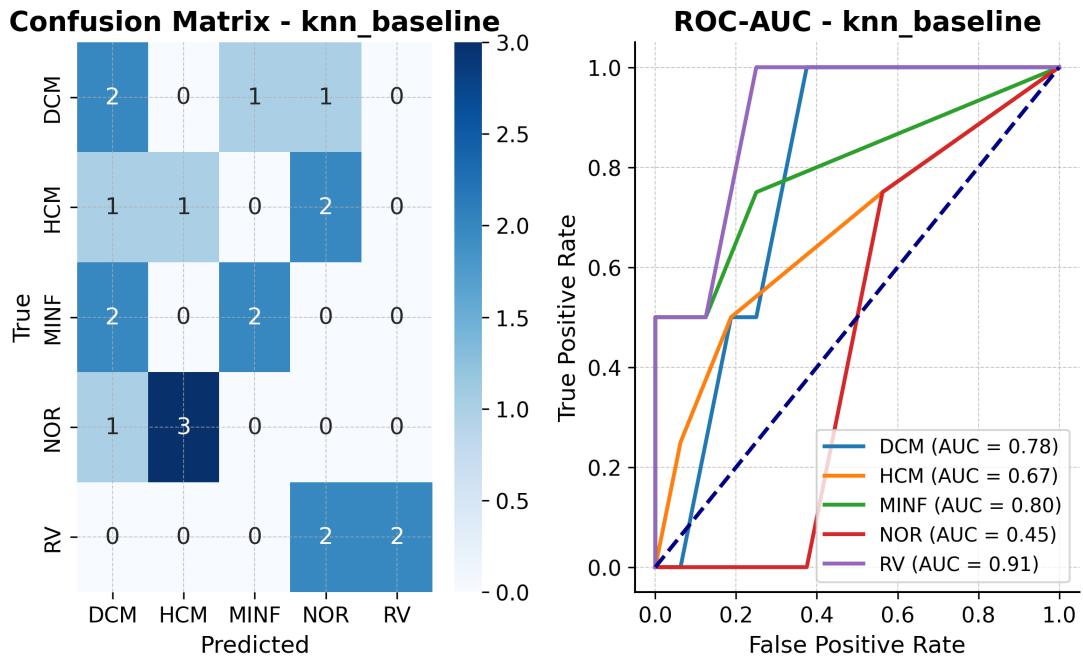


Figure A.15: Confusion matrix and ROC curves for the k-Nearest Neighbors (KNN) model trained using no strategy. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

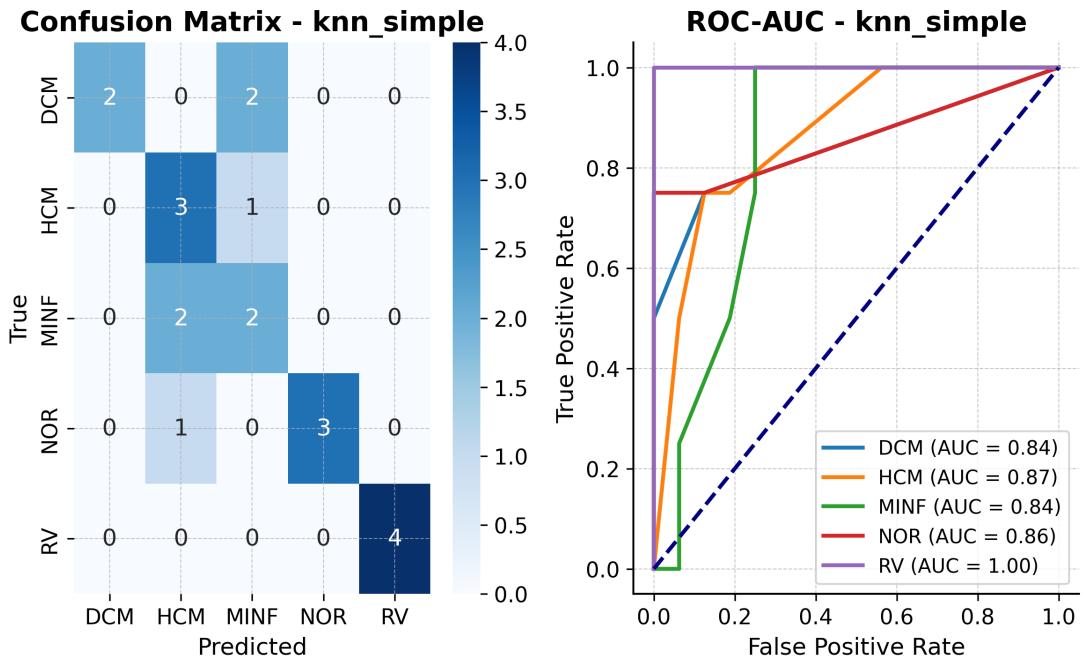


Figure A.16: Confusion matrix and ROC curves for the k-Nearest Neighbors (KNN) model trained using the Simple Split strategy. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

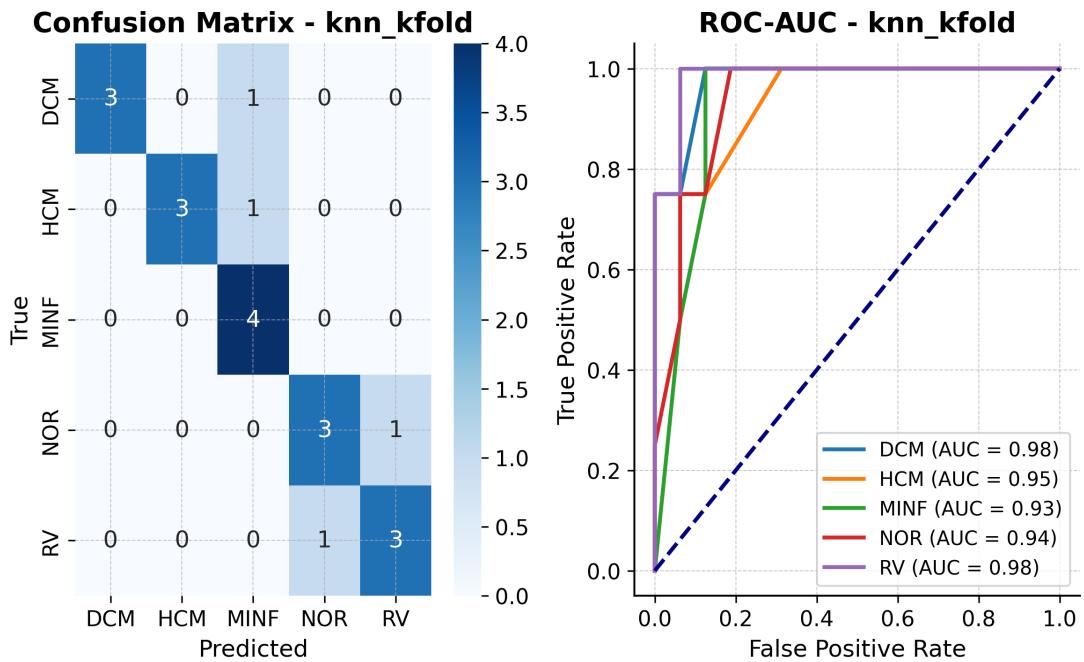


Figure A.17: Confusion matrix and ROC curves for the k-Nearest Neighbors (KNN) model trained with Stratified K-Fold cross-validation. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

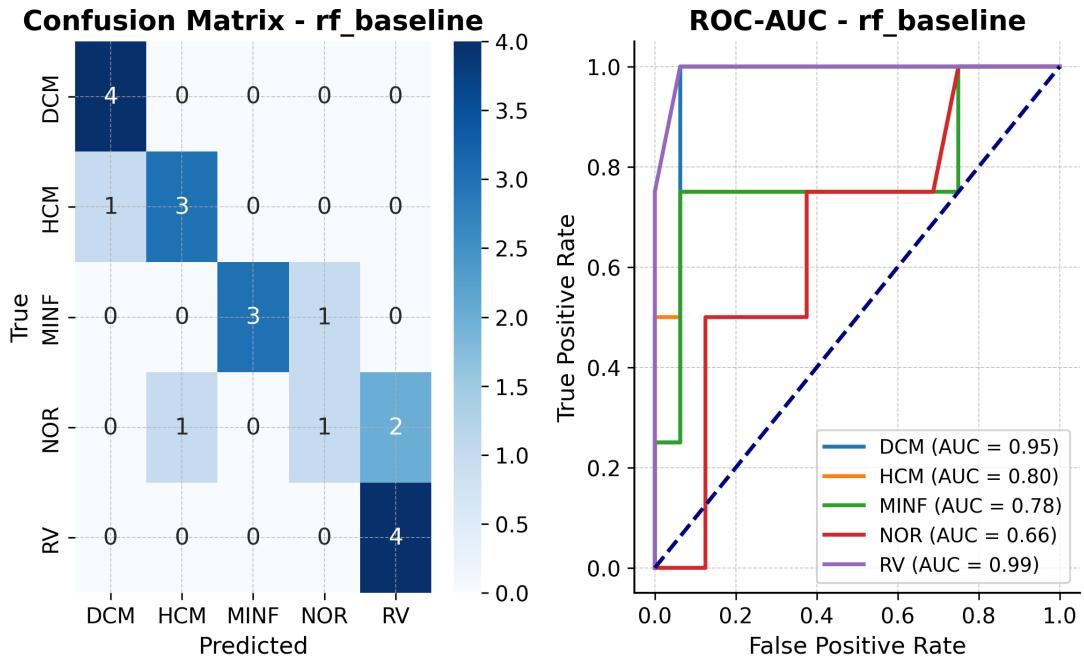


Figure A.18: Confusion matrix and ROC curves for the Random Forest (RF) model trained using no strategy. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

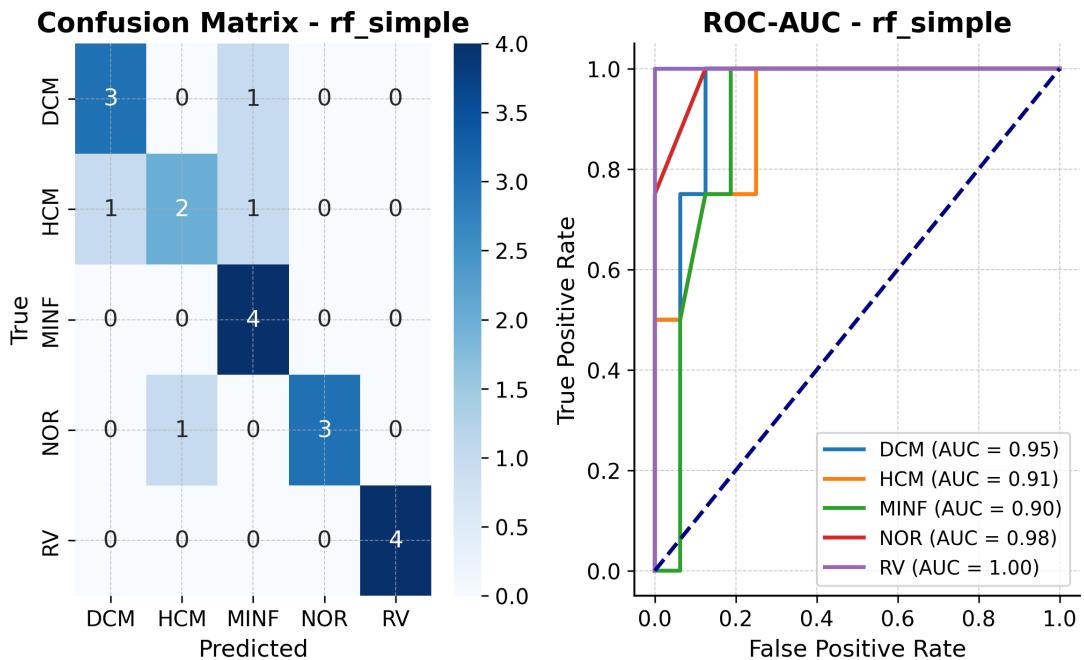


Figure A.19: Confusion matrix and ROC curves for the Random Forest (RF) model trained using the Simple Split strategy. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

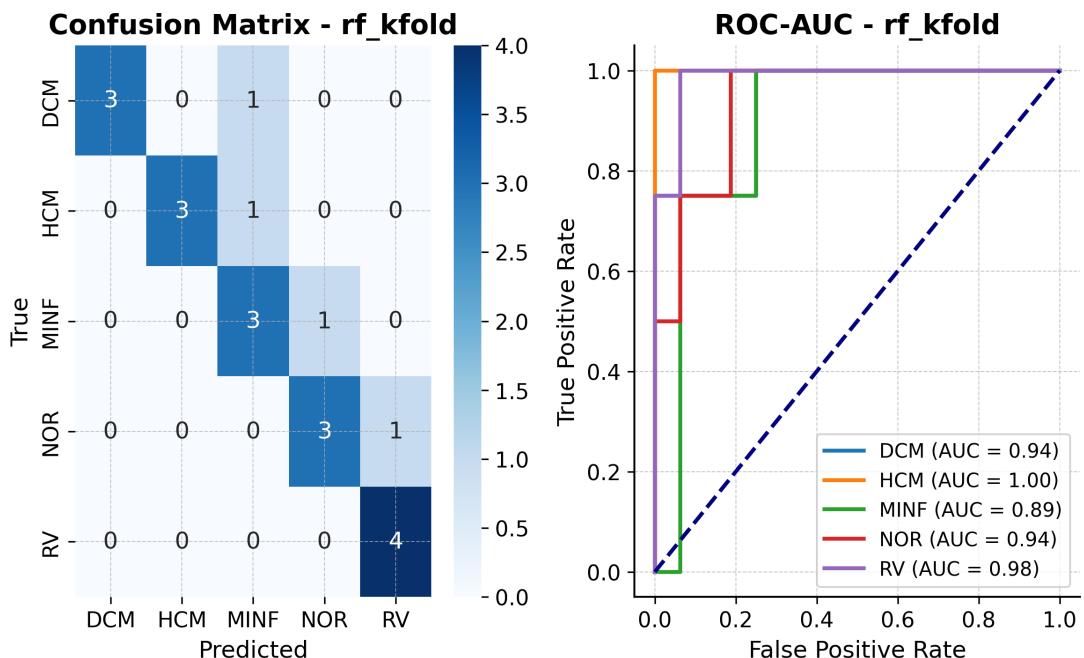


Figure A.20: Confusion matrix and ROC curves for the Random Forest (RF) model trained with Stratified K-Fold cross-validation. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

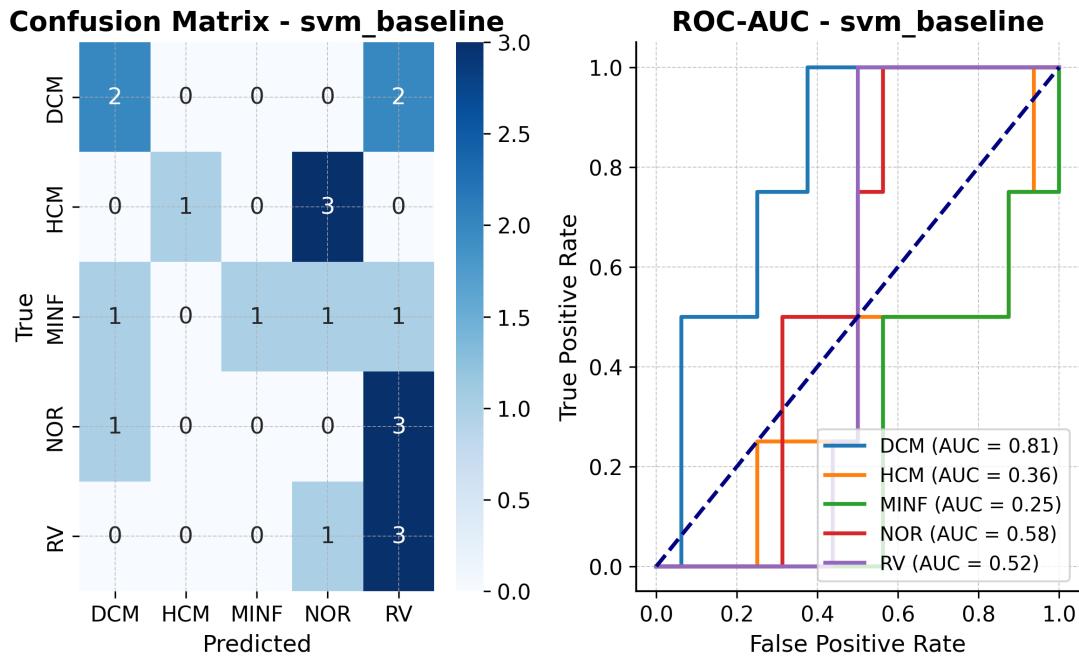


Figure A.21: Confusion matrix and ROC curves for the Support Vector Machine (SVM) model trained using no strategy. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

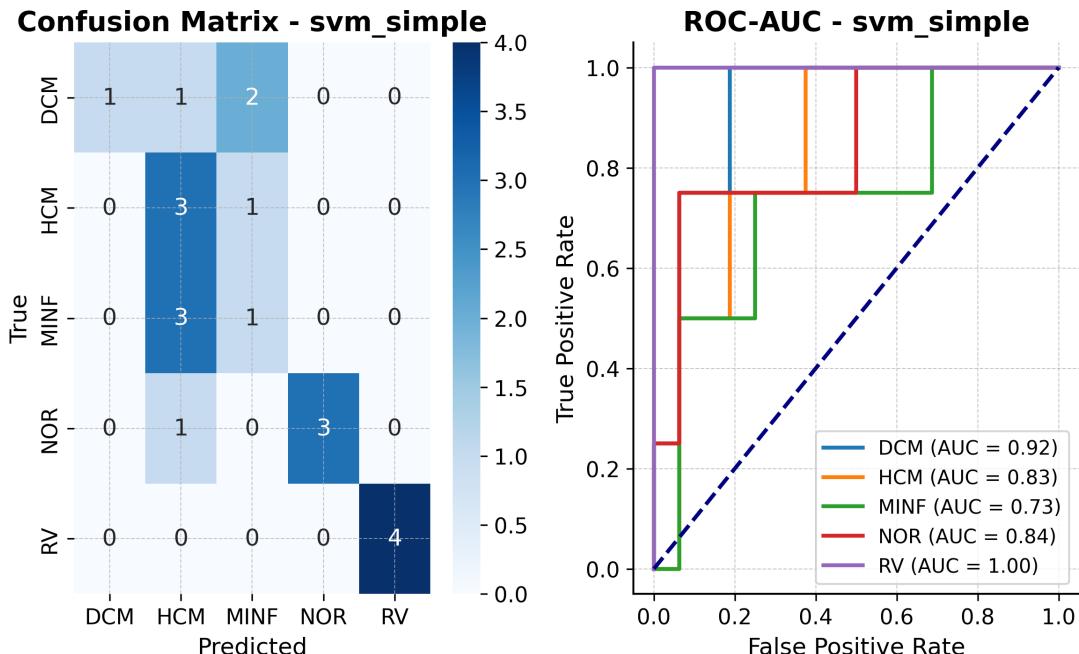


Figure A.22: Confusion matrix and ROC curves for the Support Vector Machine (SVM) model trained using the Simple Split strategy. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

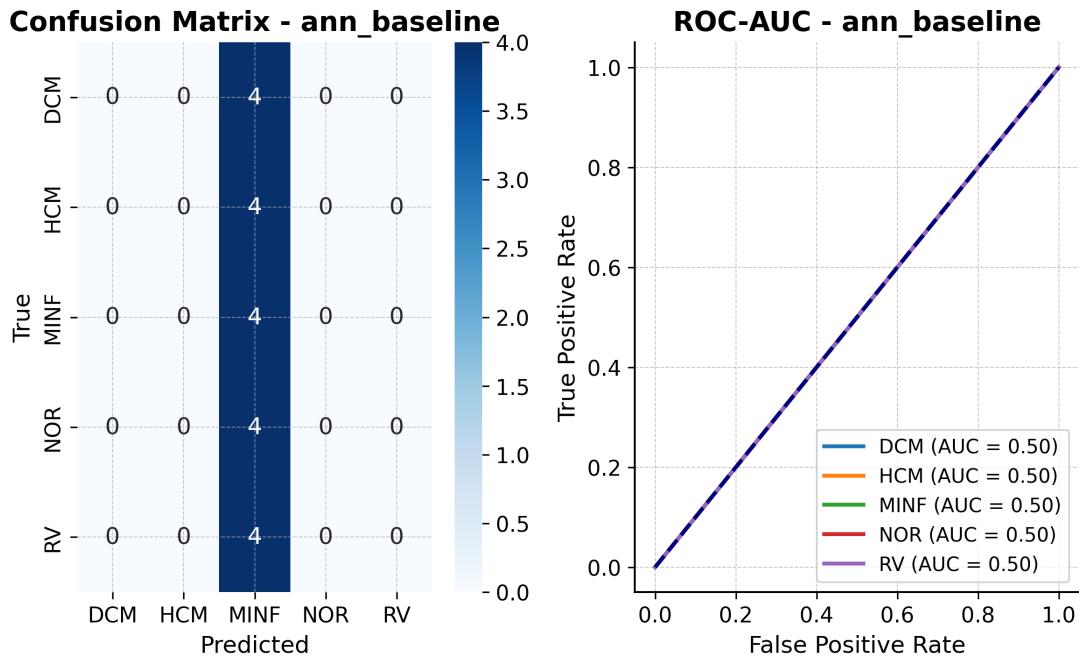


Figure A.23: Confusion matrix and ROC curves for the Artificial Neural Network (ANN) model trained using no strategy. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

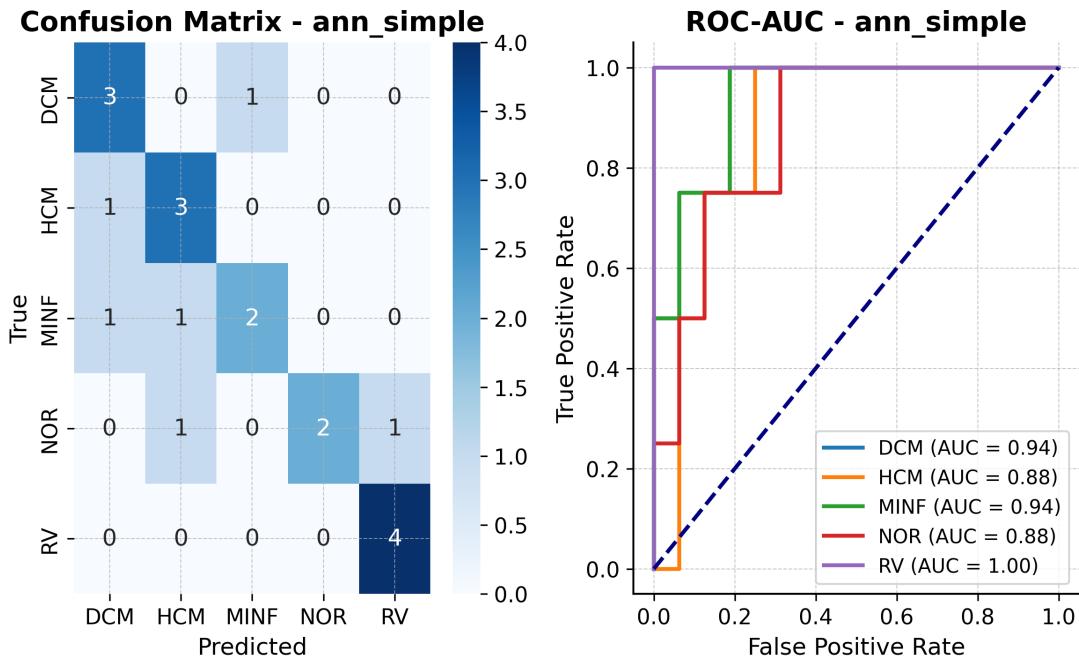


Figure A.24: Confusion matrix and ROC curves for the Artificial Neural Network (ANN) model trained using the Simple Split strategy. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

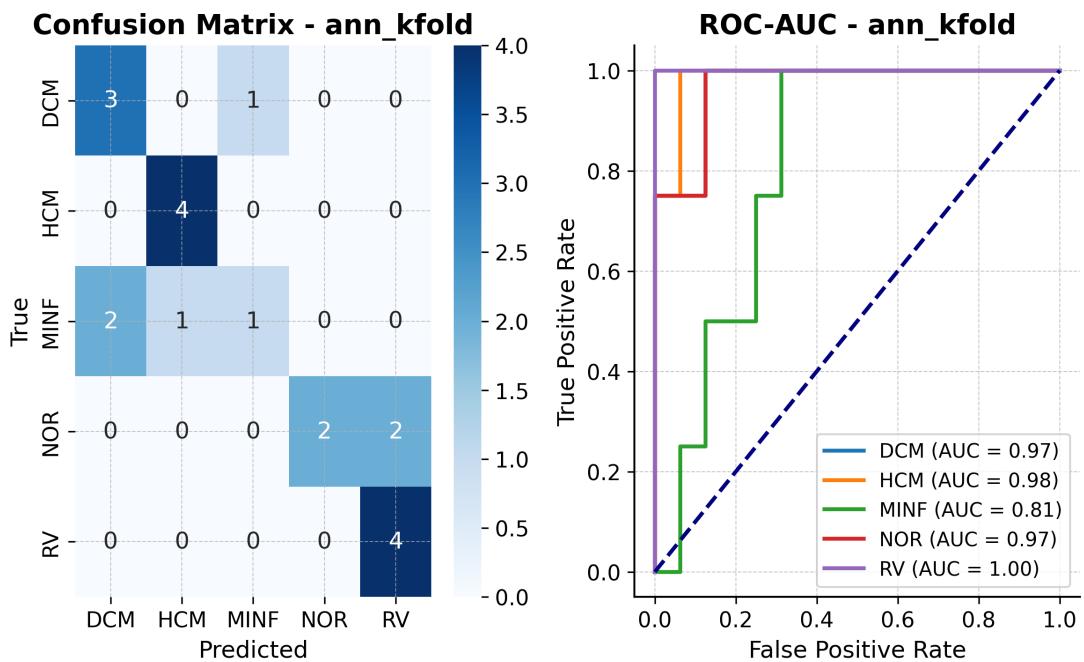


Figure A.25: Confusion matrix and ROC curves for the Artificial Neural Network (ANN) model trained with Stratified K-Fold cross-validation. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

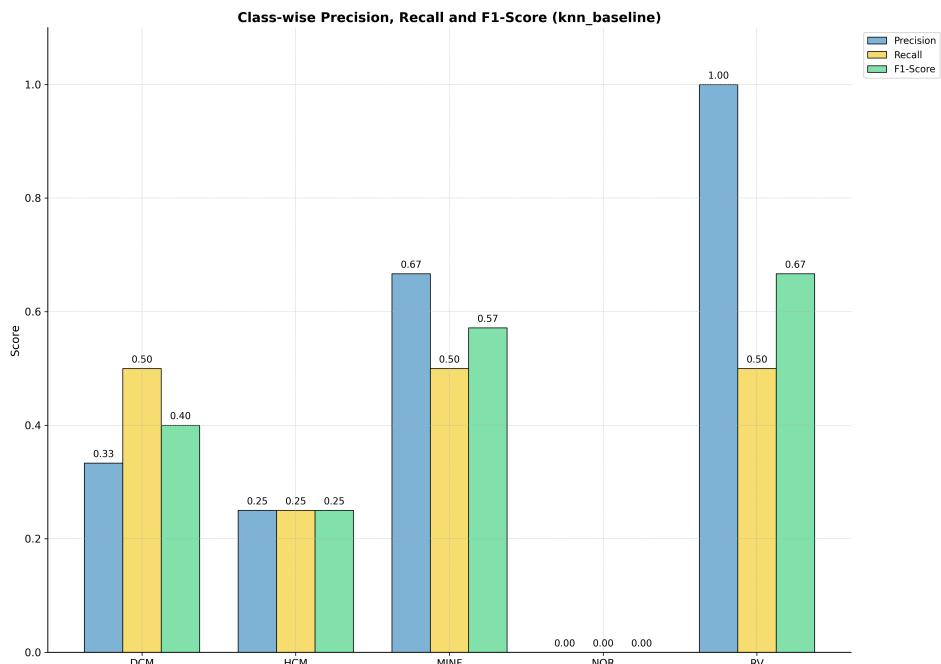


Figure A.26: Class-wise Precision, Recall, and F1-Score for the k-Nearest Neighbors (KNN) model trained using no strategy. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

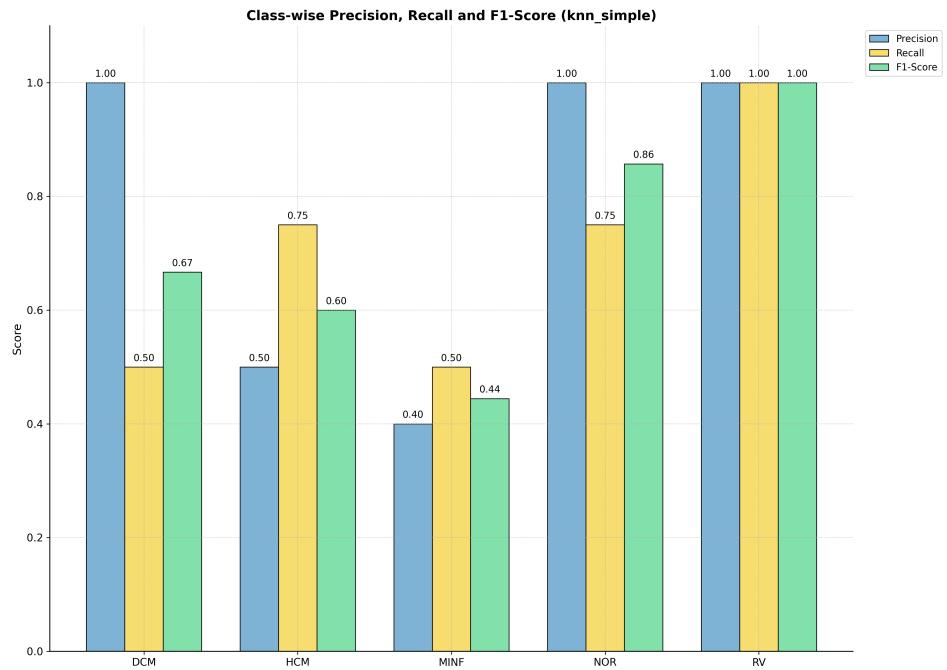


Figure A.27: Class-wise Precision, Recall, and F1-Score for the k-Nearest Neighbors (KNN) model trained using the Simple Split strategy. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

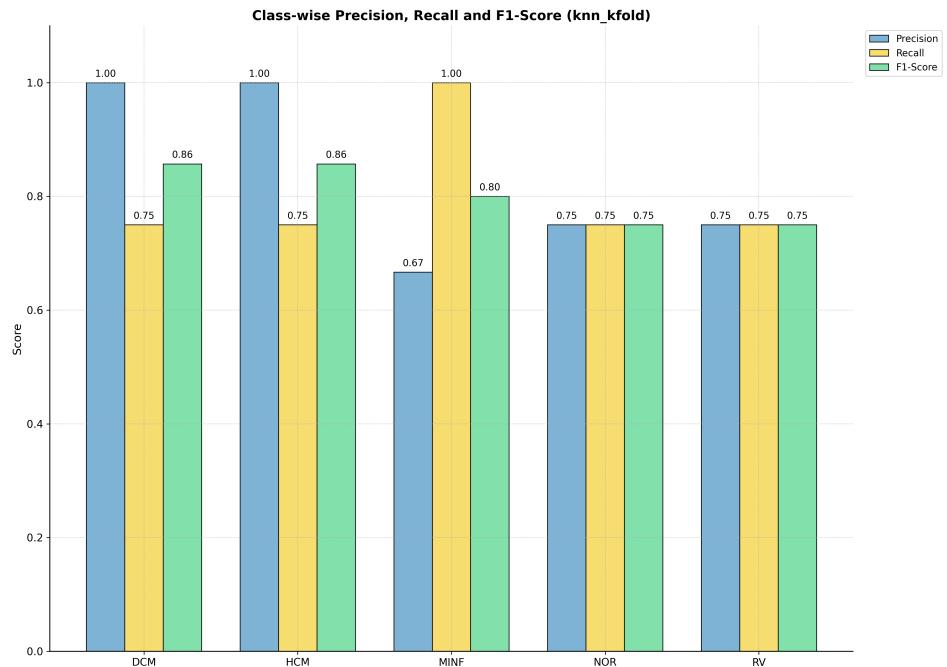


Figure A.28: Class-wise Precision, Recall, and F1-Score for the k-Nearest Neighbors (KNN) model trained with Stratified K-Fold cross-validation. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

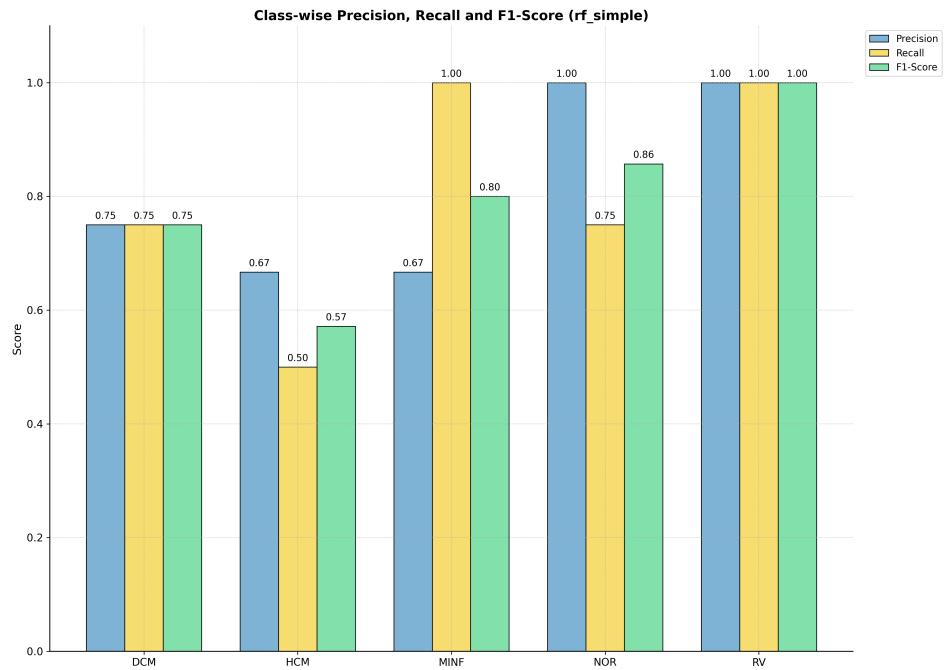


Figure A.29: Class-wise Precision, Recall, and F1-Score for the Random Forest (RF) model trained using the Simple Split strategy. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

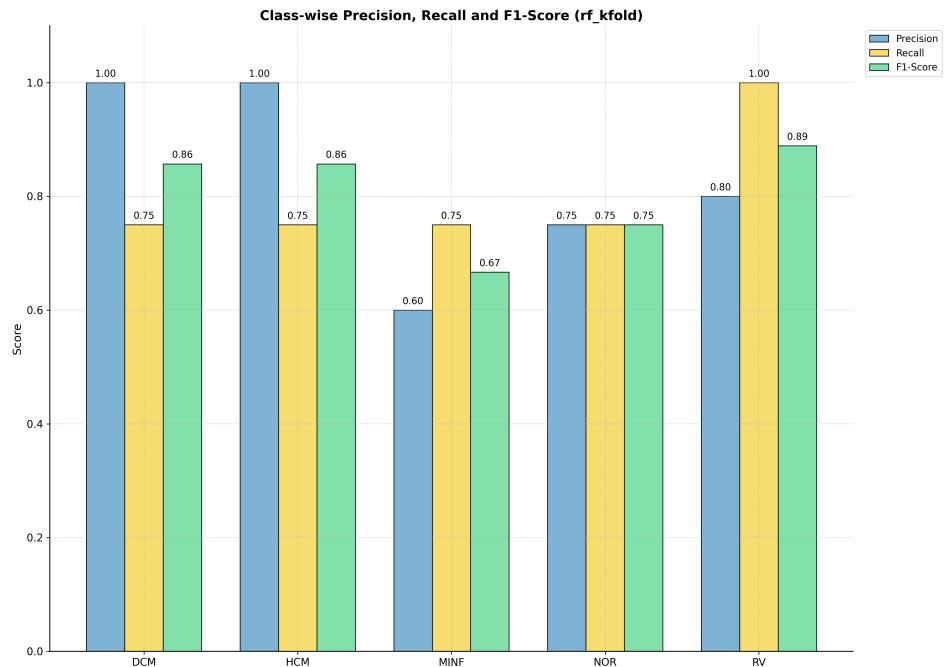


Figure A.30: Class-wise Precision, Recall, and F1-Score for the Random Forest (RF) model trained with Stratified K-Fold cross-validation. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

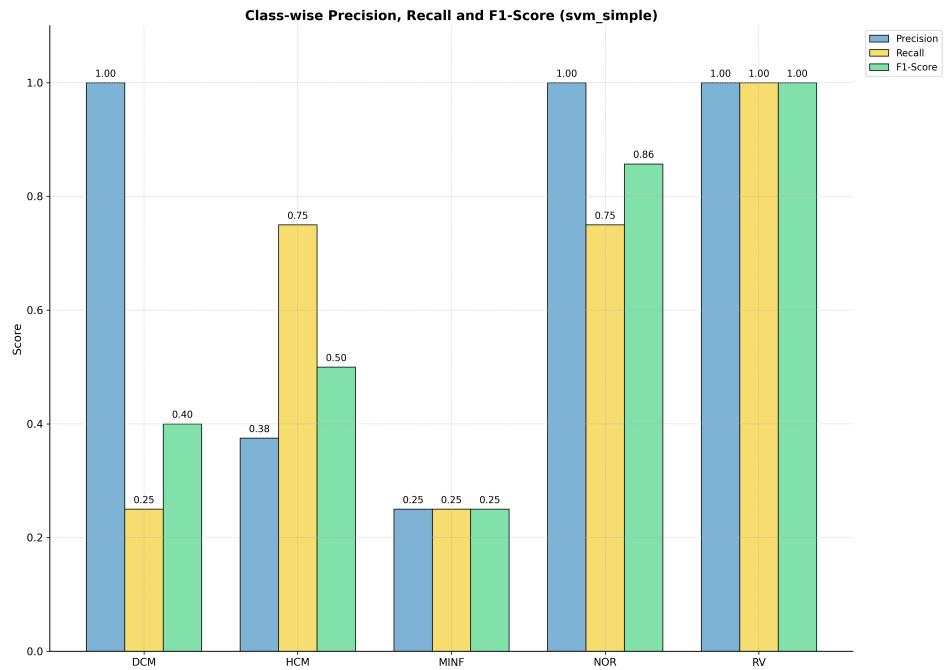


Figure A.31: Class-wise Precision, Recall, and F1-Score for the Support Vector Machine (SVM) model trained using the Simple Split strategy. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

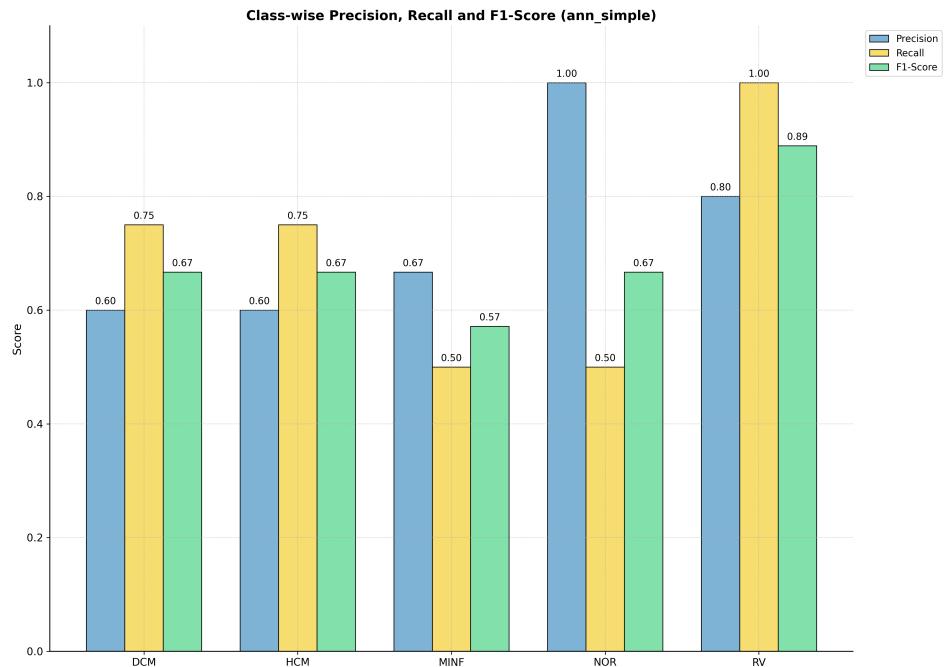


Figure A.32: Class-wise Precision, Recall, and F1-Score for the Artificial Neural Network (ANN) model trained using the Simple Split strategy. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.

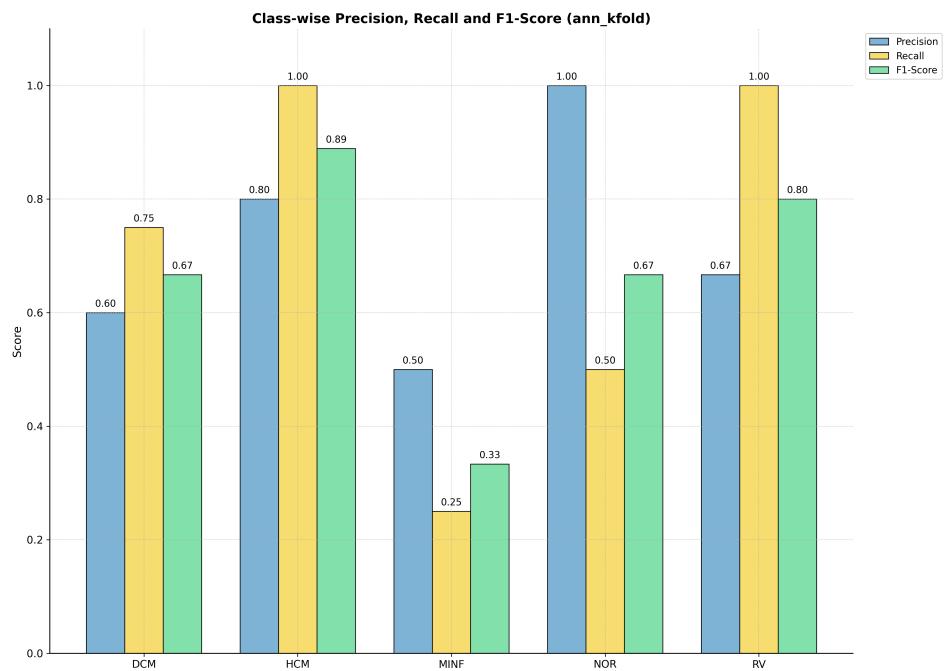


Figure A.33: Class-wise Precision, Recall, and F1-Score for the Artificial Neural Network (ANN) model trained with Stratified K-Fold cross-validation. DCM = Dilated Cardiomyopathy, HCM = Hypertrophic Cardiomyopathy, MINF = Myocardial Infarction, NOR = Normal, RV = Right Ventricular abnormality.