Development of Deep Learning Models for Predicting Cardiac Related Outcomes

Jonathan Le

Abstract

Cardiovascular disease remains the leading cause of death globally, with heart failure affecting approximately 6.2 million Americans and carrying a five-year mortality rate of nearly 50% despite modern therapies. This paper addresses three specific challenging clinical decision-making problems: selecting patients for cardiac resynchronization therapy (CRT), identifying those who would benefit from implantable cardioverter-defibrillators (ICDs) for primary prevention of sudden cardiac death, and distinguishing genetic cardiomyopathy from cardiac sarcoidosis and myocarditis. We develop deep learning approaches that move beyond simplistic clinical heuristics toward precision medicine in cardiac care, leveraging rich information contained in cardiac electrical and imaging data through sophisticated computational techniques to improve patient selection for these important but costly and invasive therapies, ultimately enhancing outcomes while reducing unnecessary procedures.

1 Introduction

Cardiovascular disease remains the leading cause of death globally, with heart failure affecting approximately 6.2 million Americans and carrying a staggering five-year mortality rate of nearly 50% despite modern therapies (Heart Failure Society of America (2023)). Within this broad spectrum of cardiac dysfunction, there are three specific challenging clinical decision-making problems: selecting patients for cardiac resynchronization therapy (CRT), identifying those who would benefit from implantable cardioverter-defibrillators (ICDs) for primary prevention of sudden cardiac death (SCD), and distinguishing genetic cardiomyopathy from cardiac sarcoidosis and myocarditis.

1.1 Predicting CRT Response

Heart failure with reduced ejection fraction (HFrEF) affects ventricular synchrony, leading to inefficient cardiac contraction. CRT has revolutionized treatment by resynchronizing ventricular contraction through biventricular pacing, improving symptoms, quality of life, and survival in selected patients. However, despite guideline-recommended selection criteria based primarily on QRS duration and morphology, 30-40% of patients show minimal benefit from this invasive and costly intervention. The current selection paradigm relies heavily on QRS duration (typically $\geq$150 ms) and left bundle branch block (LBBB) morphology as primary indicators for CRT candidacy. This approach has significant limitations, as these electrocardiographic parameters inadequately capture the complex electrical and mechanical dyssynchrony patterns that may predict response. Current guidelines do not account for the heterogeneity of cardiac dysfunction among patients with similar ECG findings, leading to unnecessary procedures in likely non-responders while potentially missing patients who might benefit despite not meeting conventional criteria. The disconnect between electrical markers on traditional ECG analysis and actual mechanical dyssynchrony highlights the need for more sophisticated approaches to patient selection. This gap represents a critical unmet need in heart failure management and motivates previous work using advanced ECG analysis through functional principal component decomposition (FPCD). For CRT patient selection, functional principal component decomposition (FPCD) of ECG waveforms could capture subtle electrical patterns associated with mechanical dyssynchrony and CRT response, however remains limited in that it relies on the creation of manually created features. Because of these limitations, we propose an approach that uses ECG analysis using convolutional neural networks (CNNs) to interpret ECG enabling accurate prediction of CRT response.

1.2 Assessing Arrhythmia Risk

Similarly problematic is the current approach to primary prevention of SCD in patients with ischemic cardiomyopathy (ICM). Ischemic cardiomyopathy (ICM) is a weakening of the heart muscle

due to poor blood supply that results from coronary artery disease (CAD) - a condition where the blood vessels supplying the heart muscle become narrowed or blocked due to plaque buildup. When coronary arteries become blocked, they cause heart attacks (myocardial infarctions) that damage the heart muscle, creating areas of scar tissue that can serve as substrates for dangerous arrhythmias. CAD is the most frequent cause of sudden cardiac arrest, accounting for nearly 70% of cases. Currently, implantation of ICDs are the gold standard in preventing SCD. Guidelines recommend prophylactic ICD implantation based primarily on a left ventricular ejection fraction (LVEF) threshold of $\leq$35%, a criterion derived from landmark trials conducted decades ago. This dichotomous LVEF-based approach suffers from significant limitations. First, LVEF has suboptimal predictive value for arrhythmic events, with studies showing that the majority of sudden cardiac deaths occur in patients with LVEF >35%. Additionally, advances in medical therapy since the original ICD trials have decreased SCA rates, altering the risk-benefit calculus. There is also increased recognition of the competing risk of non-SCD mortality, particularly in patients with severely reduced LVEF. This results in many ICD implantations in patients who never experience life-threatening arrhythmias while failing to protect many patients at genuine risk.

Interestingly, the presence, extent, and characteristics of myocardial scar tissue, especially its heterogeneity and distribution, are more closely associated with arrhythmic risk than LVEF alone. However, these factors are not integrated into current clinical decision-making. The significant limitations of the LVEF-based approach highlight the need for more sophisticated risk stratification methods. We aimed to develop more accurate predictive models that go beyond the limitations of LVEF alone. In order to do this, we propose the development of Convolutional Neural Network (CNN) models in order to evaluate clinical outcomes from the myocardial scar tissue.

1.3 Genetic Cardiomyopathies

The last problem involves the accurate differentiation between genetic cardiomyopathy, cardiac sarcoidosis, and non-genetic myocarditis. This task is crucial because misdiagnosis leads to significant harm through inappropriate treatments (unnecessary immunosuppression, risky biopsies) or missed opportunities for life-saving interventions (ICD placement, exercise restrictions, family genetic testing). Standard diagnostic tools lack specificity, while indiscriminate genetic testing introduces risks of uncertain findings. We propose an approach that uses ECG analysis using convolutional neural networks (CNNs) to establish reliable phenotypic signatures, enabling accurate distinction between these conditions and appropriate clinical management without the drawbacks of broad genetic testing.

All of these approaches share a common goal: moving beyond simplistic clinical heuristics toward precision medicine in cardiac care. By leveraging the rich information contained in cardiac electrical and imaging data through sophisticated computational techniques, we sought to improve patient selection for these important but costly and invasive therapies, ultimately enhancing outcomes while reducing unnecessary procedures. Improved patient selection for CRT could prevent unnecessary procedures in likely non-responders, reducing healthcare costs and avoiding procedural complications. Similarly, more accurate risk stratification for ICDs could protect high-risk patients currently missed by LVEF criteria while sparing low-risk patients from unnecessary device implantation. The potential healthcare impact is substantial. With approximately 200,000 CRT devices and 150,000 ICDs implanted annually worldwide (Mela et al. (2013)), even modest improvements in patient selection algorithms could affect tens of thousands of clinical decisions annually. This paper examines these two related but distinct approaches to improving cardiac care through advanced computational analysis of cardiac data (ECGs and MRIs), demonstrating how machine learning techniques can address long-standing challenges in cardiovascular medicine and move us closer to truly personalized cardiac care.

2 Methods

2.1 Predicting CRT Response

We developed a deep learning approach to predict cardiac resynchronization therapy (CRT) response using standard 12-lead electrocardiogram (ECG) data. Our dataset comprised pre-procedure and post-procedure ECGs from patients who underwent CRT implantation, allowing us to identify electrical patterns associated with successful therapy outcomes.

The dataset consisted of 12-lead ECG recordings from patients who underwent CRT implantation. For each patient, we collected both pre-CRT and post-CRT ECG recordings, along with clinical outcome data indicating whether the patient responded to therapy (binary classification). A critical aspect of our preprocessing was the extraction of the QRS complex from each lead's waveform, which was cropped to a standardized 250 sample points. This crop-

ping focused our analysis specifically on ventricular depolarization patterns most relevant to CRT response. Each ECG recording was stored as a multi-dimensional array with dimensions (12, 250, 1), where 12 represents the number of leads (I, II, III, aVR, aVL, aVF, V1-V6), 250 represents the cropped QRS complex points, and 1 represents the voltage dimension.

Our preprocessing pipeline included importing ECG signals from the clinical database in Excel format, converting the signals to standardized tensors with consistent dimensions, normalizing the ECG voltage values to improve model training stability, and creating a binary response label for each patient based on established clinical criteria for CRT response.

The composite model architecture (ECGCompositeNet) simultaneously processed both pre-CRT and post-CRT ECGs to identify patterns associated with successful therapy. The core ECG processing network incorporated initial per-lead feature extraction using three convolutional blocks, each with batch normalization, ReLU activation, and max pooling. Intermediate feature dimensions progressed through 32, 64, and 128 channels. Global feature combination was achieved through fully connected layers with dropout regularization (rates of 0.8 and 0.5) to prevent overfitting given the limited sample size. The model concluded with a binary classification output using sigmoid activation.

We implemented a cross-validation strategy to robustly evaluate model performance. The dataset was split into 5 folds using stratified sampling to maintain class balance. For each fold, we trained the model on the training set and evaluated on the validation set using binary cross-entropy loss and the Adam optimizer with a learning rate of 0.0001. Early stopping was implemented with a patience of 10 epochs to prevent overfitting. Models were trained for a maximum of 100 epochs with a batch size of 32.

For model interpretability, we created a simpler and more clinically intuitive method that plotted characteristic waveform features from the top three and bottom three probability cases. By comparing these extreme examples, we could identify pattern differences between the patients most likely and least likely to respond to CRT according to our model. These visualizations provided clinically relevant insights into the electrophysiological patterns associated with CRT response and potential mechanisms underlying therapy effectiveness.

Performance was evaluated using area under the receiver operating characteristic curve (AUC) and accuracy metrics. The final reported performance represents the average across all 5 cross-validation folds.

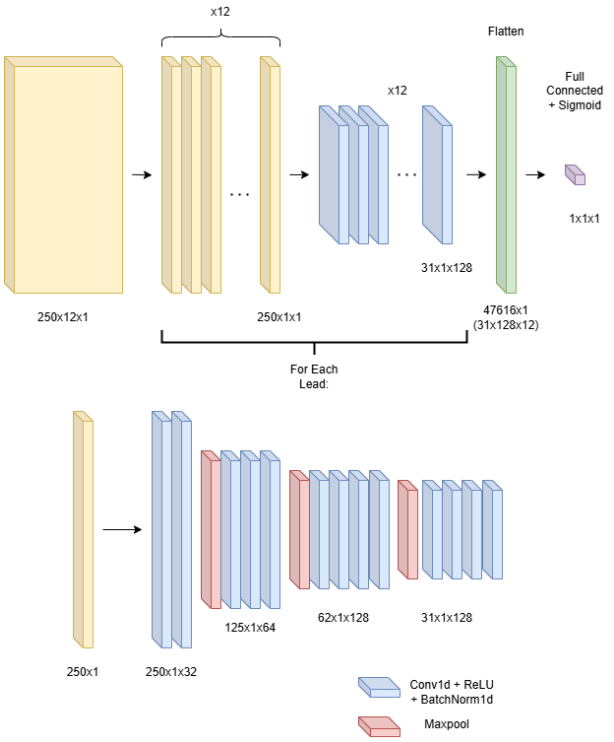

Figure 1: CNN Model used for processing ECG Leads

2.2 Ventricular Arrhythmia

We trained a CNN model on a dataset of 235 patients with each patient containing a set of CMR short axis slices of the left ventricle with Late Gadolinium Enhancement (LGE). Images were annotated by recording the contours of the epicardium, the endocardium, and the LV insertion point, all of which are coordinates, which serves as metadata for specifying the regions of interest for the model. For each short axis slice, the pixel values were normalized by considering the maximum and minimum values within the myocardium (region between epicardium and endocardium contours). The short axis slices had their orientation corrected by rotating the image such that the LV insertion point will always be oriented downwards with the center of the LV as the rotation point.

For labels, we defined a binary composite outcome where a positive label is assigned if any of the three outcomes Sudden Cardiac Death (SCD), Cardiac Arrest or if the patient had a sustained episode of Ventricular Tachycardia.

For annotating the scar, we used a pretrained Swin-Unet model trained from Professor Zhang's lab that

does binary segmentation of the myocardium region to label scar maps. A sample result of the label is shown in Fig. 14.

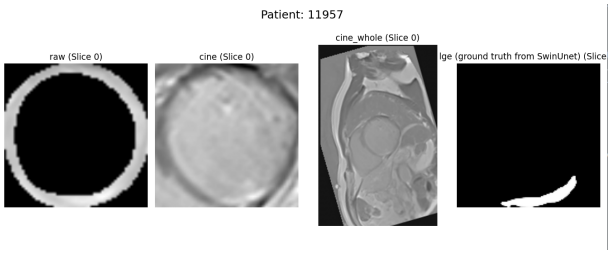

Figure 2: From Left to Right: Masked + Normalized LGE Image, Cropped SA Image, Full SA Image (orientation corrected), LGE Scar Annotation

Therefore, our dataset consists of cardiac MRI data with dimensions (235, 2, 128, 128, 1). This 5-dimensional tensor represents 235 patients. The second dimension contains two distinct elements: (1) a binary mask of the myocardium region derived from metadata delineating the endocardial and epicardial boundaries of short-axis (SA) slices, and (2) the corresponding scar tissue segmentation. Each element is represented as a 128×128 single-channel image capturing the regions of interest.

For the model, we trained on a CNN model that evaluates each patient by processing each slice with convolutions, before processing all the patient slices with a standard MLP network.

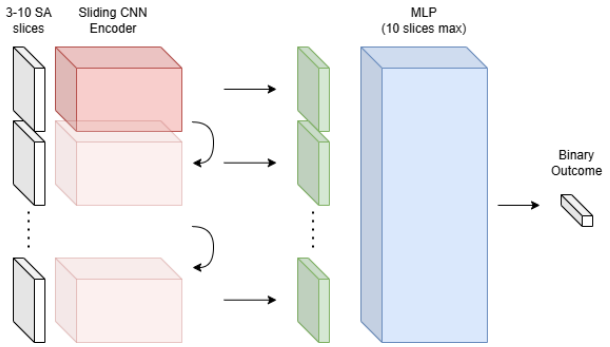

Figure 3: Model Architecture for Scar Evaluation

The model architecture follows a sliding CNN approach that processes each cardiac MRI slice individually before aggregating the results. The steps to process a patient are shown below:

1. The model accepts input data consisting of 3-10 short-axis (SA) slices per patient. Each slice contains two channels: one representing the myocardium mask and one representing the LGE scar segmentation, each with dimensions of 128×128 pixels.

2. For each slice, the Sliding CNN Encoder processes the 2-channel input through a series of convolutional layers. This encoder extracts relevant features from each slice independently, capturing spatial patterns associated with myocardial scarring.

3. After processing each slice individually, the model produces a fixed-length feature vector for each slice. These feature vectors are then fed into an MLP (Multi-Layer Perceptron) network that can handle up to 10 slices maximum.

4. The MLP network aggregates the information across all slices, learning to weight the importance of different slices and their features in making the final prediction.

5. Finally, the model outputs a binary prediction indicating whether the patient is at risk for ventricular arrhythmia based on the scar tissue characteristics observed across all slices.

This sliding CNN approach effectively handles the variable number of slices per patient while maintaining the spatial relationships within each slice that are critical for accurate prediction of cardiac outcomes.

We evaluated by utilizing 5-fold crossfold validation on AUC as well as accuracy.

2.3 Genetic Cardiomyopathy

We trained on 500 patients distinguishing between genetic cardiomyopathy and cardiac sarcoidosis. For this classification task, we employed the same composite model architecture as the previous ECG network described in Section 2.1. This approach allowed us to leverage the same deep learning framework to identify distinct ECG patterns associated with each condition, potentially enabling non-invasive differentiation between these clinically similar but etiologically distinct cardiomyopathies.

We developed separate binary classification models to distinguish between three clinically similar but pathologically distinct cardiac conditions: genetic cardiomyopathy (GEN), cardiac sarcoidosis (SARC), and myocarditis (MYO). Our dataset comprised 500 patients with 12-lead ECG recordings. Similar to our CRT response prediction methodology, we extracted and cropped the QRS complex from each lead, but additionally included RR interval data to capture

timing-related features that might distinguish these conditions.

3 Results

3.1 CRT Response

Table 1: AUC Results for ECG-based CNN Models for CRT Response Prediction

Model Input	AUC
Pre-CRT ECG	0.84
Post-CRT ECG	0.76
Combined Pre-CRT + Post-CRT ECG	0.87

Below are the plots of the ECG waveforms showing the top 3 and the bottom three waveforms. These plots allow one to intepreet the model's behavior in classifying what features or aspects the model attends to in the waveform to give the waveform a high probability or a low probability. Especially in the lead 4 morphology, correlations in what the model classifies as high probability or low probability can be observed through the

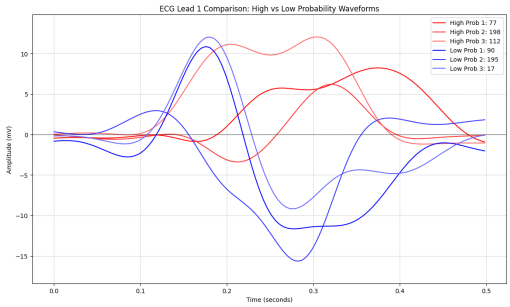

Figure 4: ECG Lead 1 Morphologies for top 3 classified probabilities + bottom 3 probabilities

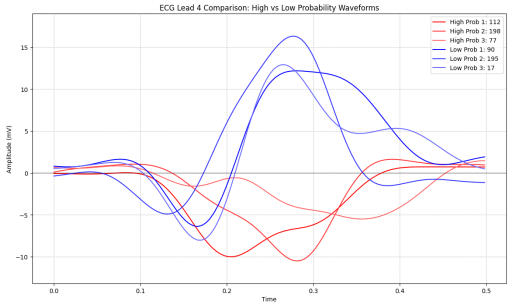

Figure 5: ECG Lead 4 Morphologies for top 3 classified probabilities + bottom 3 probabilities

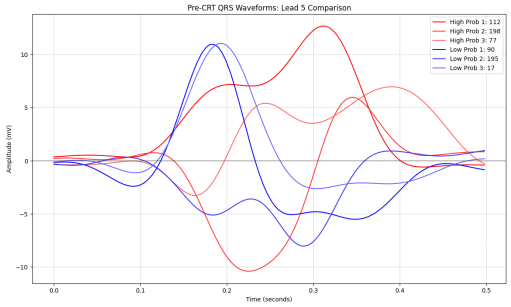

Figure 6: ECG Lead 5 Morphologies for top 3 classified probabilities + bottom 3 probabilities

3.2 Ventricular Arrhythmia

Shown below are the reported AUCs for for 5-Fold Cross fold Validation

Table 2: 5-Fold Cross-Validation AUC Results for Ventricular Arrhythmia Prediction

Fold	AUC
Fold 1	0.411
Fold 2	0.631
Fold 3	0.567
Fold 4	0.456
Fold 5	0.598
Mean AUC	0.533
Standard Deviation	0.090

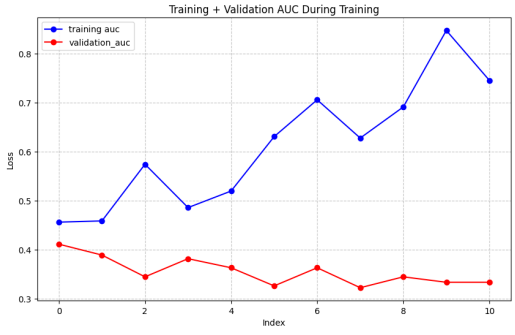

Figure 7: Fold 1 Training AUC Validation AUC Training Convergence Graph

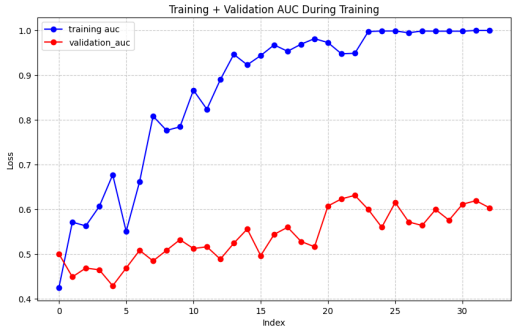

Figure 8: Fold 2 Training AUC Validation AUC Training Convergence Graph

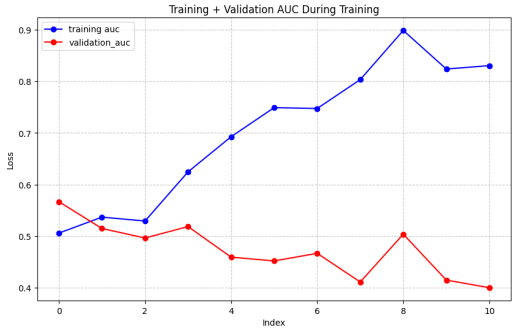

Figure 9: Fold 3 Training AUC Validation AUC Training Convergence Graph

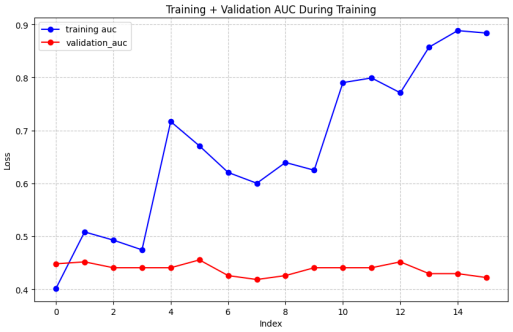

Figure 10: Fold 4 Training AUC Validation AUC Training Convergence Graph

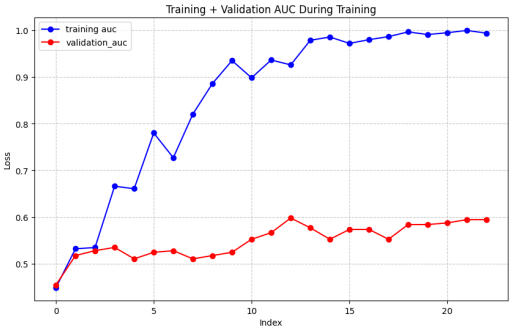

Figure 11: Fold 5 Training AUC Validation AUC Training Convergence Graph

For the model, the reported AUC score for average validation was 0.53. Additionally during training, the training AUC was consistently higher than the validation AUC across all folds.

3.3 Genetic Cardiomyopathy

Shown below is the AUC values trained from CV Validation when discriminating between the two classes.

Table 3: Test Set AUC Values

	QRS	RR (extended)
GEN vs. MYO	0.89	0.82
GEN vs. SARC	0.64	0.50

Shown below are the waveforms for the GEN vs MYO binary classification task:

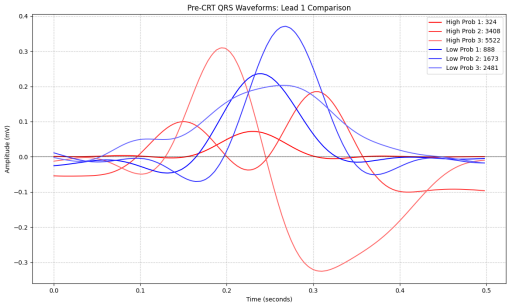

Figure 12: ECG Lead 1 Morphologies for top 3 classified probabilities + bottom 3 probabilities

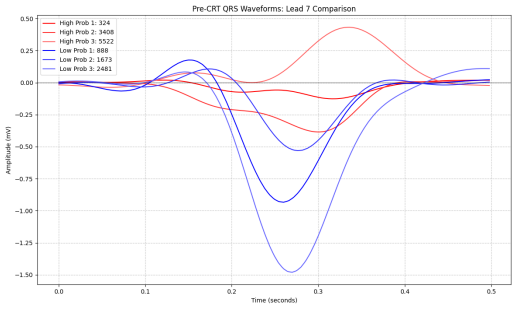

Figure 13: ECG Lead 7 Morphologies for top 3 classified probabilities + bottom 3 probabilities

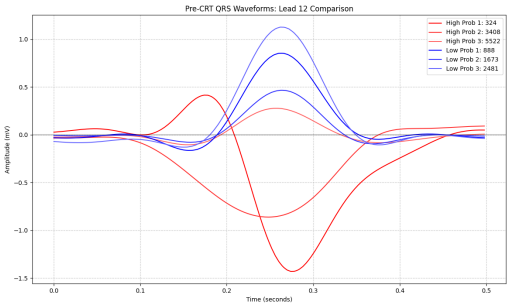

Figure 14: ECG Lead 12 Morphologies for top 3 classified probabilities + bottom 3 probabilities

4 Discussion

Our study demonstrates varying degrees of success in applying deep learning approaches to three distinct cardiac prediction tasks, each with important clinical implications.

4.1 CRT Response Prediction

The CRT response prediction model achieved promising performance, with the combined pre/post-CRT model yielding the highest AUC of 0.87, followed by the pre-CRT model (AUC 0.84) and post-CRT model (AUC 0.76). These results suggest that while pre-CRT ECG data alone contains substantial predictive information, the integration of post-implantation electrical patterns further enhances prediction accuracy. This finding has important clinical implications, as it indicates that baseline ECG characteristics before device implantation can provide meaningful insight into likely response, potentially allowing for better patient selection prior to this invasive procedure.

The morphological analysis of ECG waveforms in Figures 5-7 reveals distinctive patterns between likely responders and non-responders. These visual differences align with the model's probability assignments, suggesting the CNN has successfully identified clinically relevant electrical patterns. Such interpretable features provide a partial window into the "black box" of the neural network and could potentially inform clinical decision-making by highlighting specific ECG characteristics that correlate with therapeutic success.

The ability to predict CRT response with this level of accuracy represents a significant improvement over current clinical criteria based primarily on QRS duration and morphology. By capturing more subtle and complex electrical patterns through deep learn-

ing, our approach highlights the potential for use of deep learning methods to extract relevant features in ECG for diagnosis.

4.2 Ventricular Arrhythmia Prediction

The ventricular arrhythmia prediction model demonstrated modest performance with a mean AUC of 0.533 (SD 0.090) across five cross-validation folds. The substantial variability between folds (ranging from 0.411 to 0.631) suggests model instability that requires careful interpretation. This inconsistency likely stems from several factors that warrant discussion.

First, the training convergence graphs (Figures 9-13) consistently show higher training AUC compared to validation AUC across all folds, a classic indicator of overfitting. This pattern reveals that the model effectively learns patterns within the training data but struggles to generalize these insights to unseen examples. The relatively small dataset size (235 patients) likely contributes to this limitation, particularly given the complexity of the model architecture and the high-dimensional nature of cardiac MRI data.

Second, the variability across folds suggests that the dataset may contain heterogeneous patterns or subgroups of patients with distinct risk profiles. This heterogeneity is a known challenge in arrhythmia prediction, as the pathophysiological processes leading to arrhythmic events can vary substantially between patients, even with similar scar patterns. The model's performance fluctuations across different subsets of the data highlight the need for larger, more diverse datasets that can better represent the full spectrum of cardiac pathology.

Despite the modest overall performance, our approach represents an important step toward image-based arrhythmia risk stratification that moves beyond the limitations of LVEF-based criteria. The integration of myocardial scar characteristics through deep learning offers a foundation for future refinement, potentially leading to more accurate risk assessment tools with larger datasets and model optimization.

4.3 Genetic Cardiomyopathy Classification

The genetic cardiomyopathy classification models showed distinct performance patterns across different diagnostic comparisons. The GEN vs. MYO classification achieved strong discriminative capability with an AUC of 0.89 using QRS complex data and 0.82 using extended RR interval data. This suggests robust electrocardiographic differences between genetic cardiomyopathies and myocarditis that can be captured through deep learning approaches.

In contrast, the GEN vs. SARC classification showed more modest performance (AUC 0.64 with QRS data and 0.50 with RR data), indicating greater similarity in ECG patterns between these conditions. This aligns with clinical experience, as cardiac sarcoidosis can often mimic genetic cardiomyopathies in its electrical manifestations, making differentiation challenging even for experienced clinicians.

The morphological analysis of ECG leads (Figures 14-16) shows consistent in waveform characteristics between conditions. Lead 12 (V1) shows particularly distinctive patterns, with genetic cardiomyopathy cases

These visual patterns correspond well with the model's classification probabilities, suggesting the CNN has identified clinically relevant distinguishing features.

The superior performance of QRS-based models compared to RR interval-based models across both classification tasks highlights the primary importance of ventricular depolarization patterns in distinguishing these conditions. This finding provides valuable insight for further refinement of diagnostic algorithms and potential clinical application.

4.4 Limitations and Future Directions

Several limitations warrant consideration. First, our datasets remain relatively small for deep learning applications, potentially limiting generalizability. The CRT dataset (196 patients) and VA dataset (235 patients), while substantial for clinical studies, are modest by machine learning standards. This limitation is particularly evident in the VA prediction task, where model overfitting was apparent despite regularization efforts.

Second, the binary outcome definitions, while clinically relevant, may oversimplify the complex spectrum of treatment responses and disease manifestations. Future work might benefit from more nuanced outcome measures or multi-class approaches that better capture the heterogeneity of cardiac conditions and treatment responses.

Third, while our interpretability approaches provide some insight into model decision-making, deep neural networks remain partially opaque. Further development of visualization techniques and feature attribution methods could enhance clinical trust and adoption of these predictive tools.

Future work should focus on external validation with independent cohorts to assess generalizability, integration of multiple data modalities (combining ECG with imaging and clinical variables), and prospective evaluation of these models in clinical decision-making scenarios. Additionally, exploration of transfer learning approaches might help overcome the limited dataset sizes by leveraging knowledge from larger, related datasets.

4.5 Conclusion

This study demonstrates that deep learning approaches can provide valuable predictive insights across multiple challenging cardiac decision problems. Our models for CRT response prediction, ventricular arrhythmia risk stratification, and cardiomyopathy classification each showed varying degrees of success, with the CRT and cardiomyopathy classification models achieving particularly promising performance. These approaches offer potential paths toward more precise, personalized cardiac care that moves beyond the limitations of current clinical heuristics, potentially improving patient outcomes while reducing unnecessary procedures and their associated costs and risks.

References

Heart Failure Society of America (2023). Heart failure epidemiology and outcomes statistics: A report of the heart failure society of america. *Journal of Cardiac Failure*.

Mela, T., Singh, J., and Rush, A. (2013). Device-related infections. *Journal of Cardiac Implantable Electronic Devices*.