

# Deep Learning for ECG-Based Diagnosis and Computational Modeling of Ventricular Dynamics: A Two-Part Investigation

I.Expósito Castro, T.Pazos Puig, and X.Miret Esqué

**Abstract— Objective:** This study presents a two-part methodology for automated ECG classification and electrophysiological modeling of arrhythmic risk. **Methods:** In the first part, deep learning models, including One-Dimensional Residual Network (ResNet1D) and Long Short-Term Memory (LSTM), were trained on raw ECG signals from the PTB-XL dataset to classify ECGs into three diagnostic super classes: normal sinus rhythm (NORM), myocardial infarction (MI), and ST-segment changes (STTC). Dimensionality reduction was performed using Principal Component Analysis (PCA) on patient metadata (age, sex, height, and weight), enabling visualization of the distribution of diagnostic superclasses and biomarkers. In the second part, 3D anatomical modeling was used to segment MRI data and build mesh models of the left ventricle. Electrical propagation across the myocardium was simulated to study the relationship between individual ventricular geometry and arrhythmias. **Results:** Both models achieved strong performance on the test set, with ResNet1D outperforming LSTM in terms of accuracy and recall for NORM. However, LSTM demonstrated more stability during training, showing less fluctuation in test performance over epochs. Dimensionality reduction with PCA revealed that metadata features, particularly age and weight, correlated with diagnostic labels. In the simulation, specific stimulation protocols reliably induced reentrant activity, demonstrating the influence of tissue properties and anatomical heterogeneity on arrhythmia formation. **Conclusion:** The integration of deep learning for ECG classification and 3D modeling for arrhythmia risk assessment provides a powerful and comprehensive tool for cardiovascular disease diagnosis and understanding. **Significance:** This methodology advances computational cardiology by combining data-driven machine learning techniques with biophysical modeling to improve arrhythmia prediction and enhance personalized healthcare.

**Index Terms—** 3D anatomical modeling, Arrhythmia risk, Deep learning, Dimensionality reduction, Electrocardiography (ECG), Long Short-Term Memory (LSTM), Multi-label classification, Myocardial infarction (MI), Principal Component Analysis (PCA), PTB-XL, Reentrant circuits, Residual Network (ResNet1D), Signal to Noise Ratio (SNR), ST-segment changes (STTC), Ventricular simulation.

## I. INTRODUCTION

ELECTROCARDIOGRAPHY (ECG) is one of the most widely used and cost-effective tools in clinical cardiology, enabling noninvasive evaluation of cardiac electrical activity. With the growing global burden of cardiovascular diseases, automated systems for ECG interpretation have become critical in supporting clinicians with the timely diagnosis of

arrhythmias and other cardiac abnormalities. Machine learning (ML) and Deep Learning techniques have significantly advanced biomedical signal processing by enabling models to learn directly from raw ECG data, bypassing the need for handcrafted feature extraction. Several studies [1], [2] have demonstrated the potential of such models to classify complex ECG patterns with accuracy comparable to that of expert cardiologists.

This lab project focuses on the development of classification models to categorize ECG signals into three clinically significant superclasses: normal sinus rhythm (NORM), myocardial infarction (MI), and ST changes (STTC). All three conditions are highly relevant in both emergency and clinical settings, where timely detection and monitoring can lead to improved patient outcomes. Although there has been significant progress in ECG classification, many existing systems still rely on classical approaches or handcrafted features, which often fail to fully capture the complex spatiotemporal patterns inherent in ECG signals.

Our work differs by using raw ECG signals as input to train deep learning models, specifically, One-Dimensional Residual Network (ResNet1D) and Long Short-Term Memory (LSTM) architectures, providing an end-to-end solution that eliminates the need for manual feature engineering.

The ResNet1D model is a specialized version of the original Residual Network (ResNet), which has proven highly effective in image classification tasks. A key innovation of this architecture is the use of residual connections, which help prevent vanishing gradients and enable deeper networks to train more efficiently. In our context, the 1D variant is tailored to sequential data like ECG signals, enabling the model to learn discriminative temporal patterns such as QRS complexes, P waves, and T waves, key components for clinical ECG interpretation.

LSTM (Long Short-Term Memory) networks, a type of Recurrent Neural Network (RNN), are designed to capture long-range dependencies in time-series data. In ECG classification, LSTMs are advantageous because they can retain and utilize historical information over longer time windows, helping to identify subtle transitions or recurring abnormalities

across multiple cardiac cycles. These capabilities make LSTMs particularly suited to tasks where temporal dynamics play a central role, such as distinguishing between NORM and pathological rhythms.

In addition to model development, we performed dimensionality reduction using Principal Component Analysis (PCA) on selected clinical metadata, such as age, sex, height, and weight. This analysis provided insights into the distribution of patient characteristics across diagnostic classes and helped visualize potential separability between conditions in a low-dimensional space. The resulting projections also supported exploratory data analysis and informed our understanding of the relationships between physiological traits and cardiac pathologies.

Our project expands the scope of ECG analysis by integrating 3D anatomical modeling and electrophysiological simulation. Patient-specific left ventricular models derived from MRI data are segmented and meshed to simulate electrical propagation across the myocardium. This modeling approach allows us to investigate how anatomical variability affects arrhythmic risk under controlled stimulation protocols [3].

This two-part methodology—combining data-driven ECG classification with simulation-based electrophysiological modeling—offers a more comprehensive and personalized framework for studying cardiac disorders. While many existing studies focus on either signal classification or biophysical simulation, our work bridges the two, contributing to the emerging field of computational cardiology. By uniting deep learning and mechanistic modeling, we move toward a more integrative and individualized approach to arrhythmia detection and cardiovascular risk assessment.

## II. METHODS

### A. Model Building for Classification of ECG Superclass Diagnoses

In this task, we built and trained two deep learning models—ResNet1D and LSTM—to classify ECG signals into three diagnostic superclasses: NORM, MI, and STTC. The choice of architectures was motivated by benchmark results reported in [4], where ResNet1D and LSTM models achieved AUCs above 0.92 when distinguishing among the five superclass categories [NORM, MI, STTC, Conduction Disturbance (CD), and Right Ventricular Hypertrophy (HYP)] on the PTB-XL dataset. Our implementations were adapted from the official GitHub repository provided by the authors [5]. Dimensionality reduction with PCA was also performed using patient metadata (age, sex, weight, and height) as input features. The resulting latent space was used to visualize the distribution of both the diagnostic super classes (NORM, MI, STTC) and to explore trends in underlying patient biomarkers.

The PTB-XL dataset is a large, publicly available clinical ECG dataset comprising 21,799 12-lead ECG records from

18,869 patients, each 10 seconds long. It includes raw waveform data (sampled at both 100 Hz and 500 Hz), annotated SCP-ECG diagnostic statements, and extensive metadata such as patient demographics (age, sex, height, weight), ECG recording information (device, site, date), signal quality indicators (noise types, baseline drift), and clinical labels reviewed by cardiologists. Each record is uniquely identified by an `ecg_id` and associated with a `patient_id`. The dataset is organized into a metadata CSV file (`ptbx1_database.csv`), a diagnostic label reference (`scp_statements.csv`), and two directories (`records100/` and `records500/`) containing down-sampled and high-resolution ECG signal files, respectively.

In our deep learning pipeline, both the LSTM and ResNet1D models were trained directly on the raw down-sampled ECG signals, without applying explicit denoising or handcrafted feature extraction. To ensure generalization, we followed the official 10-fold stratified split provided by the PTB-XL dataset authors. This predefined split ensures that all ECG records from the same patient are assigned to the same fold, which prevents data leakage. Specifically, we used folds 1–9 for training and fold 10 for testing. This setup preserves label distribution across folds and enables realistic evaluation of model performance on unseen patients. The only preprocessing applied was z-score normalization per lead, standardizing each ECG signal to zero mean and unit variance. To assess the quality of the raw data, we computed the Signal-to-Noise Ratio (SNR) per lead across the training dataset. SNR was estimated by comparing each raw ECG signal with a filtered version obtained through high-pass (0.5 Hz) and low-pass (45 Hz) filtering, linear detrending, and mirrored edge padding to mitigate boundary artifacts.

Two distinct labeling strategies were adopted to meet the specific needs of model training and data visualization. For model training, a multi-label classification approach was used, as each ECG record can be annotated with multiple diagnostic superclasses (e.g., both MI and STTC). Allowed the models to learn richer and potentially overlapping diagnostic patterns, improving their robustness and flexibility. However, for dimensionality reduction and biomarker visualization, a single-label strategy was employed. In this case, each ECG was assigned a dominant diagnostic label—selected from NORM, MI, or STTC—based on the highest confidence score in the SCP-ECG annotations. This simplification was necessary for clarity in scatter plot visualizations, where each point must correspond to a single class. By separating labeling strategies for modeling and visualization, we preserved both interpretability and diagnostic richness.

To prepare the labels for training, we processed the SCP-ECG diagnostic statements (`scp_codes`) provided with each ECG record. These are dictionaries mapping ECG labels to confidence scores. Using the `scp_statements.csv` reference file, we filtered the SCP codes to retain only those associated with diagnostic classes, and mapped each code to its corresponding superclass (NORM, MI, STTC). Since each ECG can have multiple valid labels, we adopted a multi-label classification

setup. The MultiLabelBinarizer from scikit-learn was used to one-hot encode the presence of the three target superclasses. Only records containing at least one of the target labels were retained for model training and testing. This labeling strategy allowed both LSTM and ResNet1D models to learn from complex or ambiguous ECG patterns while preserving the multi-label nature of the dataset.

We implemented a lightweight ResNet1D architecture in PyTorch for multi-label classification of ECGs into NORM, MI, and STTC. The model architecture consisted of an initial convolutional layer followed by four residual blocks with increasing filter widths, batch normalization, ReLU activations, and global average pooling. A dropout layer ( $p = 0.5$ ) was added before the fully connected output layer to mitigate overfitting. ECGs were input as 12-lead, 10-second signals sampled at 100 Hz and normalized per lead. The model was trained using the Adam optimizer (learning rate =  $1e-3$ ) and the binary cross-entropy loss function (BCEWithLogitsLoss). Performance was tracked using training/testing loss and accuracy curves, classification reports with per-class precision, recall, F1-score, and Hamming Loss. Early stopping was applied with a patience of five epochs to prevent overfitting. While multiple hyperparameter configurations were explored, some overfitting persisted and was best mitigated through early stopping.

We also implemented a unidirectional LSTM model in PyTorch for the same classification task. The architecture consisted of two stacked LSTM layers with 128 hidden units each, followed by adaptive average pooling and a fully connected output layer with three output units (one per target class). Dropout ( $p=0.3$ ) was applied within the LSTM layers to mitigate overfitting. The model accepted normalized 12-lead ECGs and was trained using BCEWithLogitsLoss and the Adam optimizer (learning rate =  $1e-3$ ). Early stopping was employed based on test loss, again with a patience of five epochs. As with the ResNet1D model, training dynamics were visualized, and final performance was evaluated using standard multi-label metrics.

Finally, we explored dimensionality reduction techniques on the PTB-XL dataset using patient metadata features. Initially, we considered extracting ECG features using the NeuroKit2 library, aiming to merge these with patient metadata to form a fixed-length feature vector for each sample. However, the process of feature extraction using NeuroKit2 on all 12 leads proved computationally infeasible, taking over 8 hours to process the full dataset and often yielding unreliable results with insufficient peaks detected. To address this, we attempted a targeted approach by selecting the most informative leads through PCA on a random sample of 2,000 ECGs. Despite this, similar issues persisted, including repeated warnings of incomplete signal decomposition.

Due to these constraints, we based our final dimensionality reduction on metadata only—specifically, age, sex, height, and weight. While height and weight had substantial missingness (68% and 57%, respectively), our missingness analysis revealed

no significant bias linked to diagnostic labels, sex, or age. Therefore, we imputed the missing values using the global median. Features such as infarction stage, pacemaker, and extra beats were excluded from the analysis due to high missingness or lack of standardized annotations. Dimensionality reduction was performed using both Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP). While UMAP yielded compact clusters, it offered little insight into how individual features contributed to the projections, making interpretation difficult. In contrast, PCA, although linear, provided interpretable loadings and explained variance ratios, allowing us to directly link principal components to biomarkers. Therefore, PCA was selected for the final visualization and discussion. Scatter plots of the PCA latent space colored by age, sex, height, weight, and dominant diagnostic label revealed potential trends between patient profiles and ECG pathologies, supporting the utility of metadata in exploratory ECG analysis.

### *B. Heart Segmentation and Simulation*

The segmentation process began by importing the MRI data in .vtk format into 3D Slicer as a volume, ensuring the file path was correctly specified. The left ventricle was identified across the coronal, transverse, and sagittal planes. Segmentation was performed using the Paint and Grow from Seeds tools within the Segment Editor module. The initial segmentation was refined using the Smoothing and Scissors tools to produce a closed, anatomically accurate, and smooth 3D geometry.

The resulting left ventricular geometry was exported as an .STL file. Post-processing was performed in ParaView, applying a sequence of filters—Extract Surface, Tessellate, Extract Surface again, Decimate, and Triangulate—to reduce the model complexity to approximately 20,000 cells and 10,000 points. Additional smoothing was applied. While further mesh refinement could have been performed using tools like Blender or MeshLab, the ParaView pipeline was deemed sufficient for the scope of this course.

For the simulation component, we used Processing, a lightweight Java-based environment particularly well-suited for visualizing cardiac electrophysiology. Its accessible syntax and built-in graphical capabilities enabled dynamic rendering of electrical wave propagation, activation patterns, and reentrant circuits across the ventricular mesh. This environment allowed for interactive, real-time visualization of electrophysiological dynamics under different pacing protocols.

Using simulation-based approaches for arrhythmia research provides several advantages. It allows for precise control over parameters such as stimulation frequency, number of beats, conduction velocity restitution (CVR), and action potential duration restitution (APDR)—factors known to influence the emergence of reentrant activity. Since many life-threatening arrhythmias, including ventricular tachycardia and fibrillation, are driven by reentrant mechanisms, simulation offers a

powerful and ethical alternative to in vivo or in vitro experimentation. Hundreds of physiological and pathological conditions can be tested computationally, enabling the identification of tissue zones—such as border regions, epicardial, or endocardial segments—that are more prone to sustaining reentrant loops.

Overall, this modeling framework not only enhances the understanding of arrhythmic dynamics but also serves as a foundation for developing predictive tools and therapeutic strategies based on individual patient anatomy and electrophysiological properties.

### III. RESULTS

#### A. Model Building for Classification of ECG Superclass Diagnoses

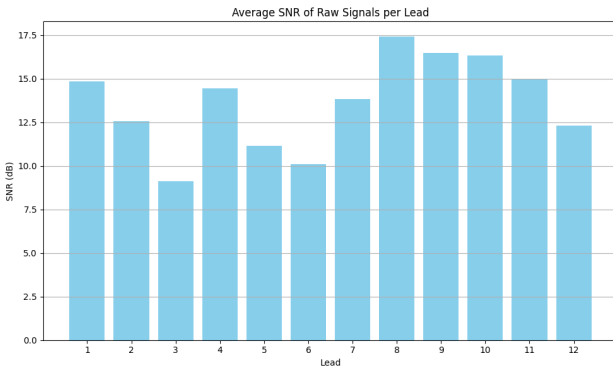


Fig. 1. Average signal-to-noise ratio (SNR) per lead computed from the raw ECG signals in the training set.

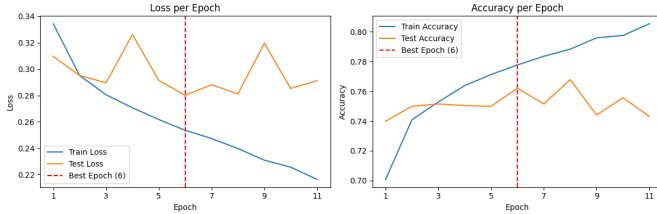


Fig. 2. Training and test loss and accuracy curves for ResNet1D. The red dashed line indicates the best epoch selected by early stopping.

TABLE I  
OVERALL PERFORMANCE METRICS (RESNET1D)

Metric	Value
Test Loss	0.2801
Test Subset Accuracy	0.7620
Hamming Loss	0.1215

TABLE II  
CLASSIFICATION REPORT BY CLASS (RESNET1D)

Class	Precision	Recall	F1-score	Support
NORM	0.8637	0.9346	0.8978	963
MI	0.8107	0.7164	0.7606	550
STTC	0.8118	0.7121	0.7587	521
Macro Avg	0.8287	0.7877	0.8057	2034
Weighted Avg	0.8289	0.8186	0.8251	2034

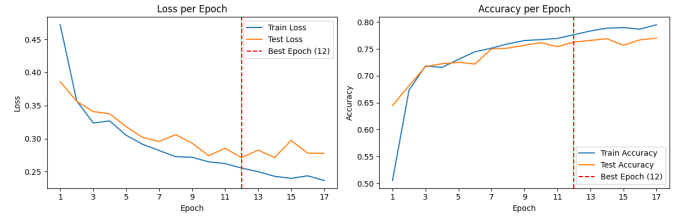


Fig. 3. Training and test loss and accuracy curves for LSTM. The red dashed line indicates the best epoch selected by early stopping.

TABLE III  
OVERALL PERFORMANCE METRICS (LSTM)

Metric	Value
Test Loss	0.2713
Test Subset Accuracy	0.7626
Hamming Loss	0.1162

TABLE IV  
CLASSIFICATION REPORT BY CLASS (LSTM)

Class	Precision	Recall	F1-score	Support
NORM	0.8815	0.9034	0.8923	963
MI	0.7947	0.7673	0.7808	550
STTC	0.8239	0.7543	0.7876	521
Macro Avg	0.8334	0.8083	0.8202	2034
Weighted Avg	0.8433	0.8284	0.8353	2034

TABLE V  
CLASS DISTRIBUTION IN THE TRAINING SET

Class	Training Samples
NORM	8551
MI	4919
STTC	4714

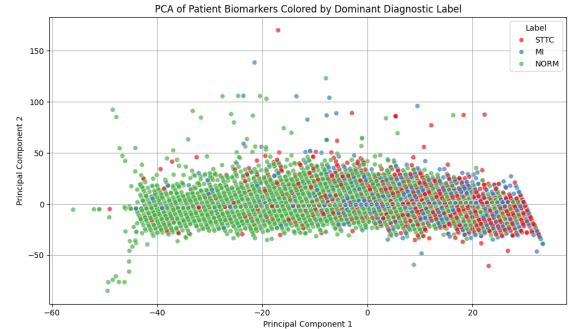


Fig. 4. PCA projection of patient metadata (age, sex, height, and weight), colored by dominant diagnostic label.

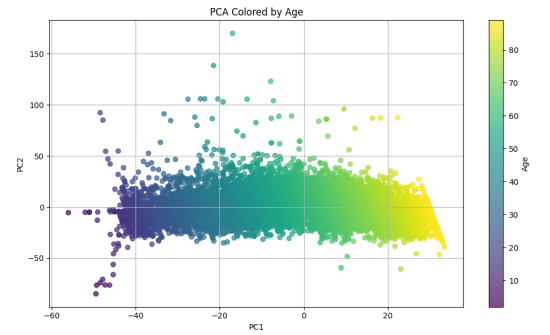


Fig. 5. PCA projection of patient biomarkers colored by Age.

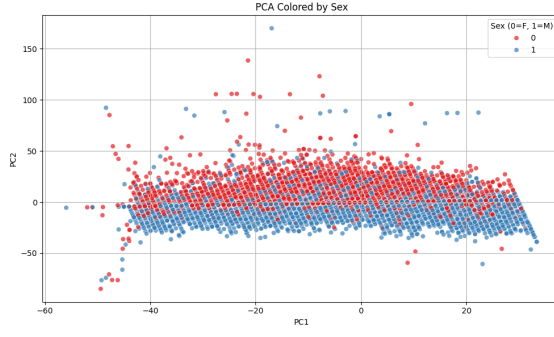


Fig. 6. PCA projection of patient biomarkers colored by Sex.

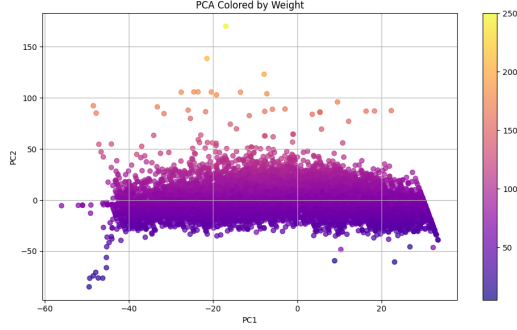


Fig. 7. PCA projection of patient biomarkers colored by Weight.

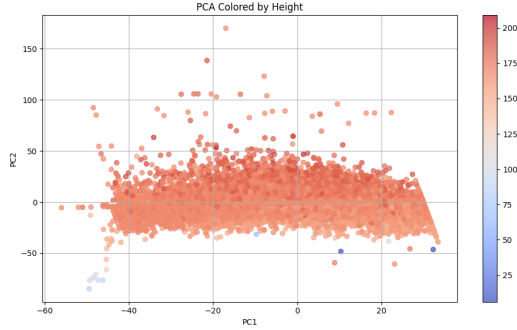


Fig. 8. PCA projection of patient biomarkers colored by Height.

TABLE VI PCA LOADINGS AND EXPLAINED VARIANCE		
Feature	PC1 Loading	PC2 Loading
Age	0.9960	0.0888
Sex	0.0020	-0.0141
Height	-0.0348	0.3036
Weight	-0.0821	0.9485

Component	Explained Variance Ratio
PC1	0.6627
PC2	0.2708
Total	0.9336

## B. Heart Segmentation and Simulation

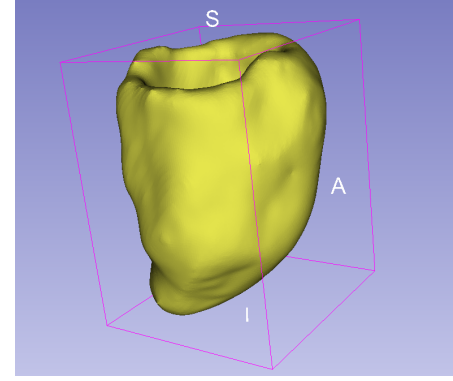


Fig. 9. Geometry of the ventricle segmentation.

At 1800.0 ms, an external stimulus triggered activation, with the next stimulus scheduled at 2400.0 ms. The maximum action potential duration (APD MAX) recorded was 397.3428 ms at a midmyocardial (M) node located in a border zone (BZ) region (node ID 203737), which had been activated at 1325.0122 ms. Subsequent activations followed stimuli at 2400.0 ms, 3000.0 ms, 3600.0 ms, and 3885.999 ms, with similar characteristics until a notable drop in APD to 367.30017 ms and a new ID 212537 at time 4170.0 ms. At 4456.0 ms, another drop to 365.77936 ms and a new ID 64182 was recorded. After that, multiple spontaneous activations occurred without new external stimuli, indicating possible reentry activity, starting from 4776.1416 ms and recurring consistently at intervals until at least 28705.959 ms. These reactivations repeatedly involved the initial nodes and suggested a self-sustained reentrant circuit, which could be pathological and indicative of arrhythmogenic behavior in the simulated tissue.

The simulation results showed that several parameter combinations were capable of inducing reentrant activity, which is a hallmark of cardiac arrhythmias. From the total of 612 simulations run, at least five simulations demonstrated either sustained or multi-cycle reentries, indicating that under certain conditions, the cardiac tissue model was highly prone to developing reentrant circuits. For instance, simulation 123 produced sustained reentry when stimulated with an S1 frequency of 600 ms and a faster S2 frequency of 285 ms, following a protocol of 6 S1 stimuli and 3 S2 stimuli. This reentry occurred with a stimulation node (id\_extra) at 96899, and was recorded in segment 12.0, associated with AHA region 7.0 on the endocardial layer. The same parameters, when applied with a different stimulation node (90551), led to 6 reentrant cycles in simulation 135, this time in AHA segment 11.0.

The model also displayed similar reentrant behavior in simulation 150, again with the same stimulation frequencies and memory settings, but using node 32481, which led to sustained reentry in AHA region 16.0. In simulation 189, the model produced 7 reentrant cycles using node 94301, highlighting this region's particular sensitivity to reentry.

Interestingly, this case occurred on the epicardial layer, suggesting that reentry could emerge differently depending on the tissue depth.

Finally, simulation 495, which reused node 94301 but increased the CVR factor to 1.25, also produced sustained reentry, reinforcing the idea that this node and the surrounding segment are particularly susceptible to arrhythmogenic behavior. The change in the CVR factor likely altered the conduction restitution properties, making the tissue more favorable to sustaining reentrant loops.

Overall, the results confirmed that specific combinations of stimulation frequency, stimulus location, and tissue memory parameters were critical in generating reentry. In particular, using an S1 of 600 ms and an S2 of 285 ms, coupled with moderate memory values and precise stimulation sites, consistently resulted in either sustained or prolonged reentrant circuits. These findings suggested that both anatomical and electrophysiological heterogeneity played crucial roles in arrhythmia formation within the simulated heart model.

TABLE VII  
SIMULATION RESULTS AND REENTRY INDUCTION ANALYSIS

Reentry Type	Simulation Node (id_extra)	AH A Segment	Layer	S1 (ms)	S2 (ms)	nStims S1	nStims S2	CVR Factor	Reentry Note
Sustained Reentry	96899	7.0	Endo	600	285	6	3	1.0	Classic sustained reentry initiation
6 Reentry Cycles	90551	11.0	Endo	600	285	6	3	1.0	Transient reentry, self-terminating
Systained Reentry	32481	16.0	Endo	600	285	6	3	1.0	Robust reentry in different region
7 Reentry Cycles	94301	12.0	Epi	600	285	6	3	1.0	Epi reentry—sensitive region
Sustained Reentry	94301	12.0	Epi	600	285	6	3	1.25	Reentry persists even with increased CVR

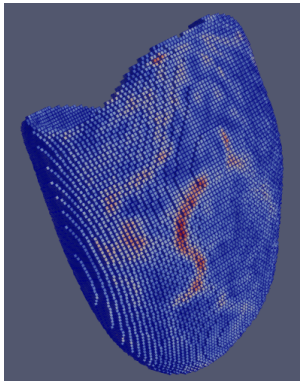


Fig. 10. Scalars Paraview visualization.

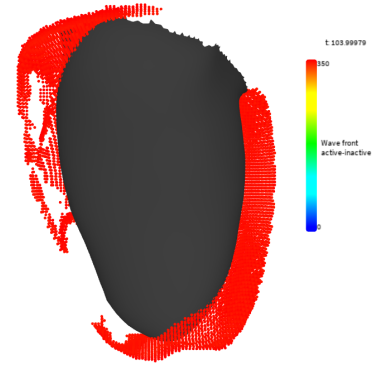


Fig. 11. Wave Front active-reactive screenshot of Processing simulation.

#### IV. DISCUSSION

##### A. Model Building for Classification of ECG Superclass Diagnoses

The decision to train on raw ECG data, without extensive filtering or handcrafted feature extraction, was supported by our SNR analysis. As shown in Fig. 1, average SNR values ranged from 9 dB to 17 dB across leads, indicating that the raw ECG signals had sufficient quality for training. Given these results, we determined that the raw data provided a strong foundation for the deep learning models, enabling them to effectively learn meaningful patterns in the ECG signals.

Both the ResNet1D and LSTM models achieved strong performance results in the multi-label classification of ECGs into NORM, MI, and STTC superclasses. As shown in Tables I and III, both models attained comparable subset accuracy (0.760 for Resnet1D and 0.7626 for LSTM), with Hamming Loss slightly lower in the LSTM (0.1162 vs. 0.1215), indicating fewer label errors per sample.

When examining per-class performance (Tables II and IV), ResNet1D achieved slightly higher recall for the NORM class (0.93 vs 0.90), while LSTM outperformed ResNet1D on the STTC class, with a higher F1-score (0.79 vs. 0.76). Both models showed lower recall on MI and STTC compared to NORM, reflecting the difficulty of detecting less prevalent and more heterogeneous pathological patterns. For instance, in the ResNet1D model, recall was 0.9346 for NORM but dropped to 0.7164 for MI and 0.7121 for STTC. Similarly, in the LSTM model, recall was 0.9034 for NORM versus 0.7673 for MI and 0.7543 for STTC. These gaps highlight the models' stronger sensitivity to normal patterns and comparatively reduced ability to detect certain cardiac abnormalities.

Macro-averaged F1-scores were 0.8057 for ResNet1D and 0.8202 for LSTM, while weighted F1-scores were also marginally higher in the LSTM (0.8353 vs. 0.8251), suggesting that the LSTM model achieved slightly better overall balance across class predictions. These results support that both architectures are effective, with the LSTM showing a small edge in handling class imbalance.

Figures 2 and 3 show that the ResNet1D model reached its



optimal performance faster, with early stopping triggering at epoch 6, compared to epoch 12 for the LSTM. Although early stopping was effective for both models in identifying an epoch before severe performance degradation on the test set occurred, both models show signs of overfitting, as evidenced by the divergence between training and test metrics. Focusing on overfitting characteristics, the ResNet1D model demonstrated an earlier onset, with divergence between training and test metrics becoming apparent around epoch 4. Following this point, the model's generalization performance showed considerable volatility, particularly visible as fluctuations in the test loss and accuracy curves. Conversely, the LSTM model maintained generalization for more epochs, with divergence typically starting around epoch 8-12. The subsequent decline in the LSTM's test performance followed a more gradual trend with lower metric volatility compared to ResNet1D.

In addition to these findings, the results suggest complementary strengths between the two models: while ResNet1D achieved higher recall for the NORM class (0.9346), the LSTM performed better on MI and STTC in terms of F1-score (0.7808 and 0.7876, respectively), suggesting that an ensemble approach could potentially enhance overall performance by leveraging their respective advantages. Notably, the MI class exhibited the lowest precision in both models (0.8107 for ResNet1D and 0.7947 for LSTM), indicating a tendency toward false positives, possibly due to overlapping features with other pathological classes. Given the strict nature of subset accuracy, where a prediction is considered correct only if *all* assigned labels exactly match the ground truth, achieving around 76% for both models indicates strong performance in a challenging multi-label setting. Complementarily, the Hamming Loss values obtained (0.1215 for ResNet1D and 0.1162 for LSTM) reflect the average fraction of misclassified labels per sample, with lower values indicating better performance. These low Hamming Loss values suggest that even when full label matches were not achieved (as penalized by subset accuracy), the number of incorrect label assignments remained relatively low, supporting the models' reliability in multi-label classification.

From a clinical perspective, high recall on NORM ECGs is reassuring, but the relatively lower recall for MI (0.7614 in ResNet1D and 0.7673 in LSTM) underscores the importance of further improving sensitivity to potentially life-threatening conditions. The class distribution in the training set (Table V) shows that the NORM class has a significantly higher number of samples compared to MI and STTC. This imbalance likely contributed to the better performance observed for NORM across both models (ResNet1D and LSTM), and potential solutions could include class augmentation techniques to mitigate this issue. These techniques could involve oversampling underrepresented classes to increase the number of samples in the minority classes or using weighted loss functions, which penalize misclassifications of underrepresented classes more heavily during training. These strategies could help improve model performance on less

frequent classes and enhance overall classification accuracy, particularly for MI and STTC.

To analyze the relationship between patient biomarkers and diagnostic outcomes, we performed PCA on four metadata features: age, sex, height, and weight. Table VI shows that the 2D latent space (PC1 vs PC2) captures 93.36% of the total variance in this metadata subset, validating the use of the first two principal components for visualization. This high explained variance supports the interpretability of the scatter plots and confirms the robustness of the analysis.

As illustrated in Fig. 4, when the PCA latent space is colored according to the dominant diagnostic label, a clear structure emerges. The NORM class shows a tendency to cluster predominantly towards the left side of the plot (lower PC1 values), exhibiting a distribution distinct from the pathological classes, MI and STTC. Conversely, while MI and STTC cases largely overlap, they are more concentrated towards the right side of the plot (higher PC1 values). This visual segregation, particularly between healthy controls and patients with MI or STTC, highlights that metadata features carry discriminatory information regarding the presence or absence of these cardiopathies.

To understand which specific metadata features drive this separation, we examined the individual PCA loadings (Table VI) and the scatter plots colored by each biomarker (Figures 5-8). Fig. 5 clearly shows a strong gradient of age aligning almost perfectly with the PC1 axis, transitioning from younger individuals on the left to older individuals on the right. This visual observation is strongly supported by the very high loading of 0.9960 for age on PC1 (Table VI). This direct relationship implies that the observed separation between NORM and pathological classes along PC1 is largely driven by age, reflecting the well-documented fact that conditions like MI and STTC are significantly more prevalent in older populations.

Fig. 7 confirms that weight has the highest loading on PC2 (0.9485). However, while weight contributes significantly to the overall variance, its effect on diagnostic class separation is less pronounced compared to age.

Fig. 6 (sex) shows clustering along the PC2 axis, with females (sex=0) towards higher PC2 values and males (sex=1) towards lower values. Yet, the clustering of sex does not strongly align with the diagnostic label distribution in Fig. 4, suggesting that sex alone is not a strong predictor of these specific cardiopathies.

Fig. 8 (height) shows an even less distinct pattern, reinforcing its lower relevance in distinguishing between the diagnostic labels in this metadata space.

The clear separation between NORM and pathological states in the PCA space suggests that traditional machine learning models using age and weight could achieve reasonable

performance for binary classification (NORM vs. Pathology). However, the overlap between MI and STTC indicates that these features alone are insufficient for distinguishing between these two conditions. This emphasizes the need to incorporate richer data sources, such as the raw ECG signals, to improve classification accuracy for more granular tasks like differentiating between cardiovascular diseases.

### B. Heart Segmentation and Simulation

The simulation results reveal clear patterns in the emergence of reentrant activity within the cardiac model, demonstrating that specific combinations of stimulation timing, tissue location, and electrophysiological properties can reliably induce arrhythmic behavior. Across five notable simulations, reentry was consistently triggered using an S1-S2 stimulation protocol with S1 at 600 ms and S2 at 285 ms, along with a modest variation in conduction velocity restitution (CVR). Sustained and multi-cycle reentries appeared in both endocardial and epicardial layers, notably involving nodes 96899, 90551, 32481, and 94301.

Segment 12.0, particularly node 94301, emerged as a hotspot for reentry, especially when CVR was increased to 1.25, indicating a region of heightened susceptibility. The presence of reentry in both layers and across diverse anatomical segments underscores the importance of spatial heterogeneity and tissue memory in arrhythmogenesis.

These results support the use of parameter sweeps with computational cardiac models as an effective tool to identify vulnerable regions and conditions conducive to arrhythmias. This approach contributes meaningfully to risk stratification and therapeutic planning in cardiac electrophysiology.

## V. CONCLUSION

This study presents a comprehensive approach to the classification of arrhythmias using deep learning models, alongside a 3D simulation of electrophysiological behavior in the heart. We developed and evaluated two deep learning architectures, ResNet1D and LSTM, for classifying ECG signals into three clinically relevant superclasses: normal sinus rhythm (NORM), myocardial infarction (MI), and ST-segment changes (STTC). Both models demonstrated strong performance, with the LSTM model showing slightly better overall balance and the ResNet1D model achieving higher recall for NORM classification. These results underscore the potential of deep learning models for automated ECG classification and suggest that both architectures offer robust solutions for arrhythmia detection.

In addition to the classification models, we incorporated a 3D segmentation and simulation workflow to explore the underlying mechanisms of arrhythmias. By segmenting MRI data to construct patient-specific 3D models of the left ventricle, we simulated electrical wave propagation to identify conditions conducive to reentry—a hallmark of many arrhythmias. The simulations revealed that specific combinations of stimulation protocols and tissue properties

could reliably induce reentrant activity, highlighting the importance of spatial and electrophysiological heterogeneity in arrhythmogenesis.

The integration of raw ECG signal analysis with advanced 3D modeling offers a unique and comprehensive tool for both diagnosis and understanding of arrhythmias. This two-pronged methodology not only provides insights into the detection of cardiac abnormalities but also enhances our understanding of the electrophysiological processes underlying these conditions. Ultimately, this work contributes to the growing field of computational cardiology, where machine learning and biophysical simulations combine to improve both the prediction and treatment of cardiac diseases.

## REFERENCES

- [1] A. W. Hannun, P. Rajpurkar, M. Haghighpanahi, G. T. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," *Nature Medicine*, vol. 25, no. 1, pp. 65–69, Jan. 2019.
- [2] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Deep convolutional neural network model for classification of heartbeats using ECG signals," *Comput. Biol. Med.*, vol. 89, pp. 389–396, Dec. 2017.
- [3] A. Prakosa et al., "Personalized virtual-heart technology for guiding the ablation of infarct-related ventricular tachycardia," *Nat. Biomed. Eng.*, vol. 2, no. 10, pp. 732–740, Oct. 2018.
- [4] P. M. Rojas, N. A. Götz, J. G. Arias-Castro, and P. T. Arribas, "Benchmarking ECG classification models with the PTB-XL dataset," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 4, pp. 984–992, Apr. 2020.
- [5] H. Helme, "ECG classification with PTB-XL Benchmarking," GitHub repository, [Online]. Available: [https://github.com/helme/ecg\\_ptbxl\\_benchmarking/](https://github.com/helme/ecg_ptbxl_benchmarking/)