A Hybrid Reinforcement and Deep Learning Framework for ECG and PCG Based Cardiovascular Signal Classification

Islam D. S. Aabdalla[1], Dr.D.Vasumathi[2]
[1]Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad, India
engislam2021@gmail.com
[2] Professor of Computer Science and Engineering, Jawaharlal Nehru Technological University
Hyderabad, India
Vasukumar_devara@jntuh.ac.in

Abstract
This study presents an innovative approach that combines Deep Reinforcement Learning with machine learning techniques to classify cardiovascular disease conditions using ECG and PCG signals from the EPNOGRAM dataset. The core aim is to boost classification accuracy by leveraging Deep Reinforcement Learning(DRL) for enhanced feature extraction, which is then fed into conventional models like CNN, RNN, LSTM, BiLSTM, and hybrid architectures such as CNN-RNN and LSTM &BiLSTM. Experimental results reveal notable performance gains when Deep Reinforcement Learning(DRL) is integrated. Specifically, the CNN-RNN hybrid achieved 90% accuracy, 89% precision, 88% recall, 88% F1-score, and a 92% AUC-ROC. The LSTM-BiLSTM hybrid outperformed with 92% accuracy, 91% precision, 90% recall, and 90% F1-score, and a 94% AUC. These findings highlight the effectiveness of merging Deep Reinforcement Learning(DRL) with supervised models, particularly in addressing the complexity of cardiopulmonary signal classification, and demonstrate the superior performance of hybrid models over standard classifiers in both accuracy and interpretability.

Keywords
Signal Processing, Machine Learning, Reinforcement Learning, Supervised Learning, Deep Q-Network (DQN)

1. Introduction
Cardiovascular diseases remain the leading cause of death worldwide, responsible for approximately 17.9 million fatalities each year[1]. Early detection is essential for improving patient outcomes and enabling timely medical intervention. Electrocardiogram and phonocardiogram signals are widely utilized non-invasive diagnostic methods that offer complementary insights into the heart's electrical and mechanical activities[2]. However, these signals are often contaminated with noise, exhibit significant variability, and can be influenced by patient-specific factors, making classification a challenging task. Traditional supervised learning models, while effective in controlled environments, struggle to generalize in dynamic, noisy, and real-time settings[3]. Recent advancements in reinforcement learning, particularly deep Q-networks, have shown promise in addressing such challenges. Reinforcement Learning(RL) is especially well-suited for applications that require optimized decision-making, as it continuously improves performance through feedback from the environment involving sequential data and uncertain conditions[4]. By integrating RL with SL, hybrid models can capitalize on the strengths of both paradigms, achieving enhanced classification accuracy while dynamically adapting to the evolving nature of the signals[5]. On the other hand, the controlling agent governs the workflow by deciding whether to continue using the backbone network or activate the discriminator, using reinforcement learning techniques[6] within a POMDP framework[7]. Each spatial-temporal hidden state of the Electrocardiogram signal snippets serves as a state. Actions are chosen through a learned policy to maximize long-term rewards based on the discriminator's performance. The state is represented by the hidden vector output from the policy and backbone network is modelled via a neural network that maps states to actions. The agent's actions determine whether to continue processing or interrupt and classify. Rewards are assigned based on classifier correctness, with the discriminator ultimately generating final label predictions through a fully connected network and softmax normalization[5].

This paper develops a hybrid framework that combines RL with SL to classify ECG and PCG signals and improve early CVD detection. The proposed model integrates a deep Q-learning agent with supervised classifiers to jointly optimize prediction accuracy and decision timing. A domain-specific reward mechanism is designed to guide the RL agent toward clinically relevant behavior. Our extensive experiments on framework cardiovascular datasets demonstrate that the hybrid framework consistently outperforms both standalone SL and RL models across multiple performance metrics, including accuracy, F1-score, and inference latency. The Key Contributions:

- Development of a hybrid Reinforcement Learning-Supervised Learning (RL-SL) architecture specifically designed for CVD classification based on ECG and PCG signals.
- Introduction of a domain-aware reward function that aligns with clinical diagnostic objectives to enhance RL learning.
- Empirical results confirm that the proposed hybrid model outperforms conventional supervised and reinforcement learning baselines on publicly available cardiovascular datasets.

The paper is organized as follows: Section 2 covers related research, Section 3 details the proposed framework and methodology, Section 4 presents experimental findings and analysis, and the final section concludes the study and outlines future research directions.

2. Related Work

Early prediction and analysis of heart diseases are critical for timely diagnosis and intervention. Recent advancements explore deep reinforcement learning for efficient and accurate early classification of varied-length heart signals.

Yu Huang et al[5]. This study presents the Snippet Policy Network (SPN), a deep reinforcement learning framework for the early classification of varied-length ECGs. SPN integrates a snippet generator, a controlling agent, a backbone network, and a discriminator to optimize both accuracy and earliness. It achieves over 80% accuracy, outperforming existing methods by at least 7% across multiple metrics. This is the first known work to target early cardiovascular disease classification using varied-length ECG data, offering a generalizable solution for early classification in time series tasks.

Ismail et al[8]. Introduced RL-ECG Net, a reinforcement learning-based framework designed for multi-class arrhythmia detection using raw ECG signals. This framework tackles issues such as class imbalance and high computational demands by utilizing a Markov Decision Process to automatically optimize both training procedures and architectural hyperparameters of deep learning models, including MLP, CNN, LSTM, and GRU. RL-ECGNet achieved classification accuracies between 88.45% and 96.41% across nine arrhythmia classes, including underrepresented categories, while significantly reducing training time and memory usage. This study demonstrates the potential of reinforcement learning to enhance both the performance and efficiency of ECG-based arrhythmia classification in a scalable and resource-efficient manner.

Pandey et al[9]. Proposed a DL framework for multiclass heart disease classification using a modified mixed attention mechanism integrated with a BLSTM network (M2AM-Deep BiLSTM). Addressing limitations in prior methods, such as misclassification, noise, and inefficient feature extraction, their model leverages advanced pre-processing with bandpass filtering and wavelet transformation, followed by R-point detection and HRV feature extraction. The hybrid attention mechanism combines zero-parameter and triplet attention to enhance feature relevance and convergence achieving a high accuracy of 93.82%.

Babu, et al[10]. Focused on a hybrid approach that integrates advanced feature selection with classification techniques to enhance the prediction of heart disease using ECG signals. Their method focuses on analysing morphological variations in ECG waveforms, especially the QRS intervals, through a multi-stage pipeline involving noise reduction, thresholding, and signal interval analysis. To optimize feature selection, a modified chicken swarm optimization algorithm was introduced, which effectively identified the most relevant attributes for classification. The prediction was performed using a multi-module neural network system, which was capable of detecting abnormalities associated with cardiac conditions. Tested on ECG data from the UCI repository, the approach demonstrated strong predictive performance, though challenges related to parameter optimization were noted. This work highlights the efficacy of combining intelligent feature selection with neural network models for accurate and automated diagnosis of cardiovascular diseases.

Alboghbaish and Eleyan[11]. Proposed a deep learning framework that integrates CNN and LSTM networks for the classification of ECG signals, aiming to improve early detection of cardiovascular diseases such as arrhythmia and heart failure. The model was evaluated using ECG datasets from MIT-BIH and BIDMC under two experimental scenarios involving five-class and three-class classifications. Before model input, the FFT transformation was applied to pre-process the ECG signals. The proposed LSTM architecture demonstrated superior performance over other existing machine learning and deep learning models, achieving classification accuracies of 97.6% and 99.2% in the respective scenarios. Unlike traditional approaches that rely on interval-based features, this method uses raw heartbeat signals, making it more robust to signal variability and noise.

Christian et al[12]. Proposed a model-driven framework for ECG signal analysis using reinforcement learning, aiming to enhance the understanding of cardiac rhythm and its modulation through physiologically meaningful modelling. Their approach involves decomposing ECG signals into overlapping lognormal components and utilizing deep reinforcement learning to estimate the underlying parameters. Applied to ECG data from infants aged 1 to 24 months, the framework revealed age-dependent variations in modelling parameters, with over 750,000 PQRST complexes analysed, most meeting signal quality requirements. Several parameters demonstrated statistically significant correlations with age, highlighting the method's sensitivity to subtle physiological changes.

Temesgen et al[13]. Conducted a comprehensive study of self-supervised representation learning for clinical 12-lead signals data to address the persistent issue of label scarcity in medical datasets. Also demonstrated the feasibility and effectiveness of extracting discriminative features from ECG signals without relying on extensive labeled data. Their experiments showed that CPC-based models achieved linear evaluation performance nearly matching supervised models within 0.5 and offered up to 1% improvement when fine-tuned. Furthermore, self-supervised learning significantly enhanced label efficiency, reaching comparable performance with only 50 to 60% of the training data and improved robustness to physiological noise. This study highlights the potential of self-supervised learning as a powerful tool for developing scalable, data-efficient, and robust ECG-based diagnostic systems, especially for rare or underrepresented cardiac conditions.

Chen et al[14]. Addressed the challenges of ECG signal classification, particularly the difficulty of obtaining large amounts of labelled data and the prevalence of class imbalance and noisy labels. They proposed a robust active learning framework that clusters ECG data in a low-dimensional embedded space and selects the most informative instances from local clusters based on local average minimal distances. This method enhances label diversity and improves classification performance while minimizing labelling effort. Additionally, a novel noisy label reduction strategy is integrated to strengthen model robustness further. Their approach demonstrated strong results on the MIT-BIH arrhythmia database, highlighting its effectiveness in realistic ECG classification scenarios.

Sowmya and Deepa Jose[15] investigated the use of a CNN-LSTM deep learning architecture for classifying arrhythmia from ECG signals, leveraging both convolutional layers for feature extraction and LSTM layers for capturing temporal dependencies in the data. Given the widespread use of ECG as a non-invasive diagnostic tool for detecting heart abnormalities, the study focuses on the early identification of arrhythmias, a significant cause of cardiovascular complications and mortality, particularly among middle-aged and older individuals. Drawing on existing datasets such as MIT-BIH, the authors examined current methods for ECG signal classification and found that many achieved accuracies above 94%. Their proposed CNN-LSTM model demonstrated improved classification performance over standalone CNN architectures. However, they noted limitations, such as the need for a larger dataset to ensure generalizability and the difficulty of classifying multiple disease symptoms simultaneously. They suggested that future enhancements involve hybrid CNN-LSTM models integrated with edge computing to facilitate real-time anomaly detection and communication with healthcare services, especially valuable during pandemics or in remote healthcare scenarios.

Luo et al[16]. Introduced a hybrid convolutional and recurrent neural network architecture designed to improve multi-class arrhythmia classification from ECG signals. Addressing the significant challenge of class imbalance commonly found in ECG datasets, particularly the MIT-BIH atrial fibrillation database, the authors employed effective data pre-processing strategies to enhance model robustness. Trained on a large-scale nine-class ECG dataset and validated through 10-fold cross-validation, HCRNet achieved an impressive average accuracy of 99.01%. Their results demonstrate the model's effectiveness in accurately detecting various types of arrhythmias and underscore its potential for clinical application in rapid and reliable cardiac diagnosis.

Sepahvand and A Mohammadi[17] proposed a novel method to simplify arrhythmia classification by reducing the reliance on 12-lead ECG signals through a teacher-student knowledge distillation framework. Their approach addresses the limitations of traditional deep learning models, which, although accurate, are computationally demanding and not ideal for deployment on resource-constrained devices such as those used in the Internet of Medical Things. In this framework, a teacher model trained on multi-lead ECG data transfers its knowledge to a more compact student model that operates using only single-lead ECG inputs. The distillation process involves guiding the student model to mimic the teacher's output and internal feature representations. Despite being over 262 times smaller, the student model achieved nearly equivalent classification performance, with only a 0.81% drop in

accuracy on the Chapman ECG dataset. This technique demonstrates the potential for deploying efficient yet accurate arrhythmia detection systems in real-world, low-resource clinical, and remote monitoring settings.

Nainwal et al[18]. Introduced a deep learning framework integrating Improved Monarch Butterfly Optimization (IMBO) with Convolutional CNN for arrhythmia classification from ECG signals. Their method leverages morphological and wavelet-based features extracted from long-duration ECG recordings, addressing the challenge of detecting infrequent and irregular heart rhythms. IMBO is applied for feature selection to reduce dimensionality and enhance classifier performance, after which the refined features are processed by a CNN for final classification. Evaluated on a standard ECG dataset, the proposed IMBO-CNN model achieved a high accuracy of 99.49%, outperforming traditional CNN models and existing metaheuristic approaches.

Ballinger et al[19]. Introduced Deep Heart, a semi-supervised, multi-task LSTM model trained on over 57,000 person-weeks of wearable heart rate data to predict major cardiovascular and metabolic conditions, including diabetes, hypertension, high cholesterol, and sleep apnea. By comparing two semi-supervised techniques, sequence learning and heuristic pertaining, the study demonstrated superior performance over traditional hand-engineered biomarkers, achieving high AUC scores of 84.51. Previous works in biomedical signal classification relied heavily on CNNs, LSTMs, and other classifiers. Reinforcement learning in clinical signal processing remains underexplored. Works using DQN for medical image navigation or treatment planning have emerged, but integrating DRL with signal-based ML classification has not been extensively validated.

3. Methodology

Figure 1 outlines the proposed methodology, which begins with loading and preprocessing the dataset to prepare it for analysis. Various deep learning models are then applied, including CNN, RNN, LSTM, BiLSTM, and hybrid architectures such as LSTM-BiLSTM and CNN-RNN. Reinforcement learning is integrated using Deep Q-Network(DQN) algorithms, which optimize the learning process through a reward-based feedback mechanism. These deep learning and reinforcement learning components are combined into a unified predictive framework. Final predictions are generated using this integrated model and evaluated using key performance metrics, including reward, precision, recall, F1-score, and accuracy.

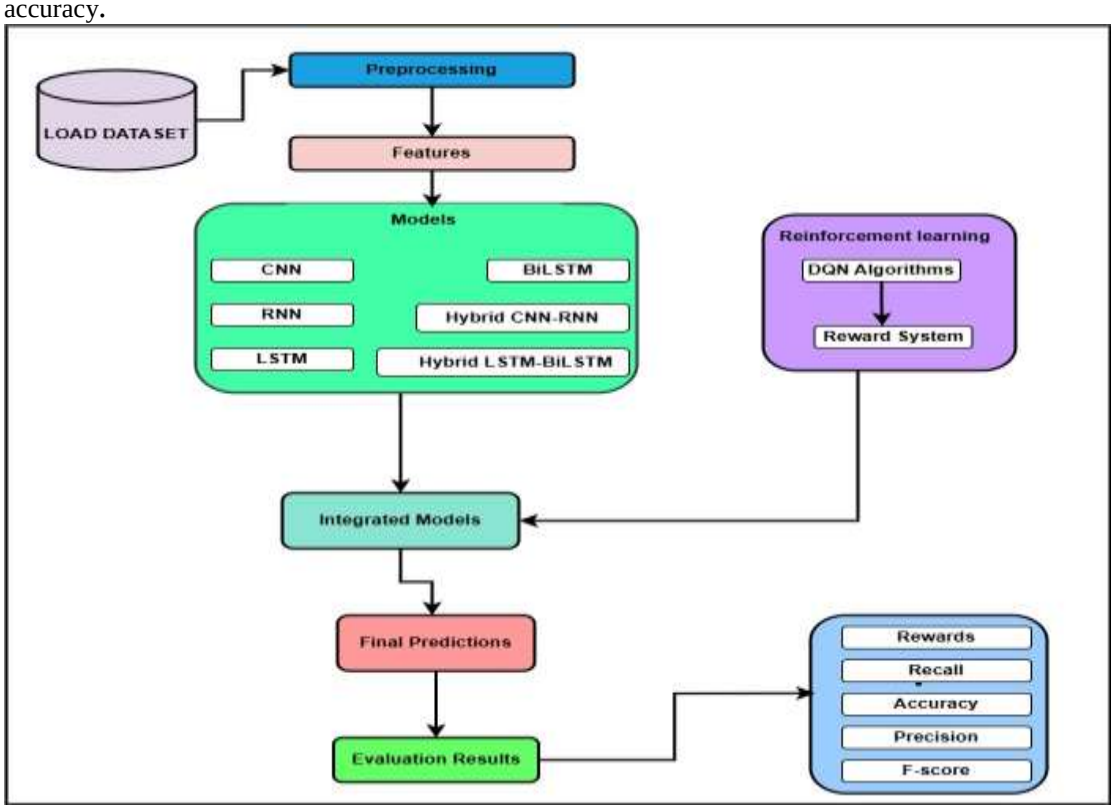

Figure 1: A hybrid framework that integrates supervised machine learning models with reinforcement learning

3.1 Description of the Dataset and Pre-processing

The EPNOGRAM dataset contains synchronized ECG and PCG signals sampled every millisecond, totaling 14,400,000 data points per type in Figure 2. Each entry includes a normalized ECG value -0.62 to 0.82, a PCG value -0.84 to 0.99, a timestamp 0 the 1800s, and a binary label normal and abnormal. As shown in Figure 3, the dataset is imbalanced, with 12,982,535 normal and 1,417,465 abnormal samples. Named for its integration of electrical (ECG) and acoustic (PCG) signals, EPNOGRAM offers a dual perspective that improves cardiovascular abnormality detection over single-modality approaches.

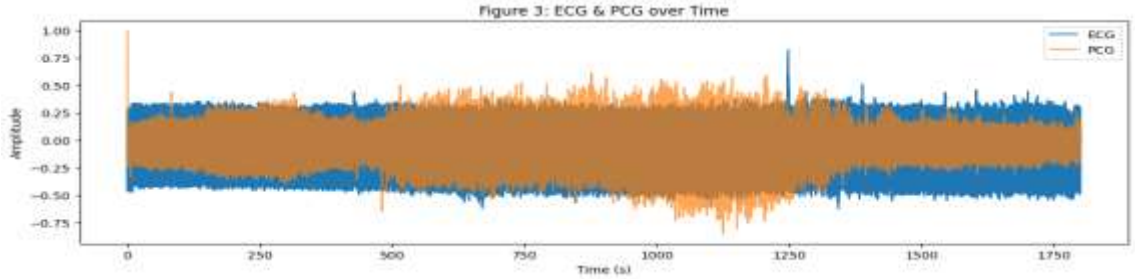

Figure 2: Original ECG and PCG signal samples from the dataset

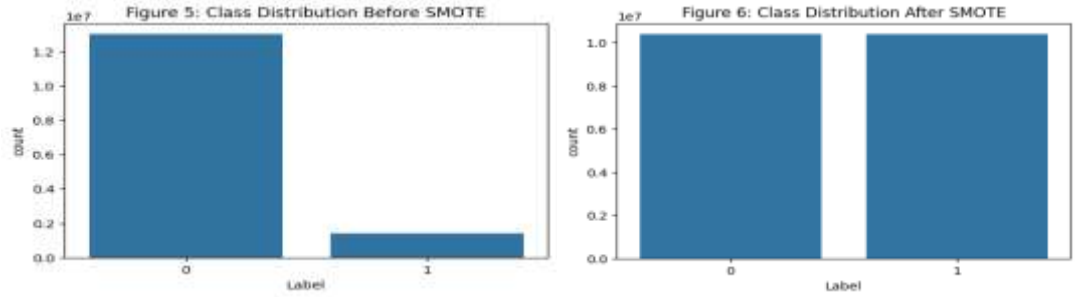

Figure 3: Dataset distribution balanced vs. Imbalanced

The dataset was preprocessed by applying signal augmentation, noise, scaling, and time-shifting. Signals were segmented into 5-second windows, bandpass filtered 0.5–40 Hz for ECG, 20–200 Hz for PCG, resampled at 1,000 Hz, and denoised to improve classification quality. The dataset was cleaned, normalized, and split into testing and training sets for supervised and reinforcement learning. DQN training was analyzed in three phases: early, middle, and late, with key learning milestones identified.

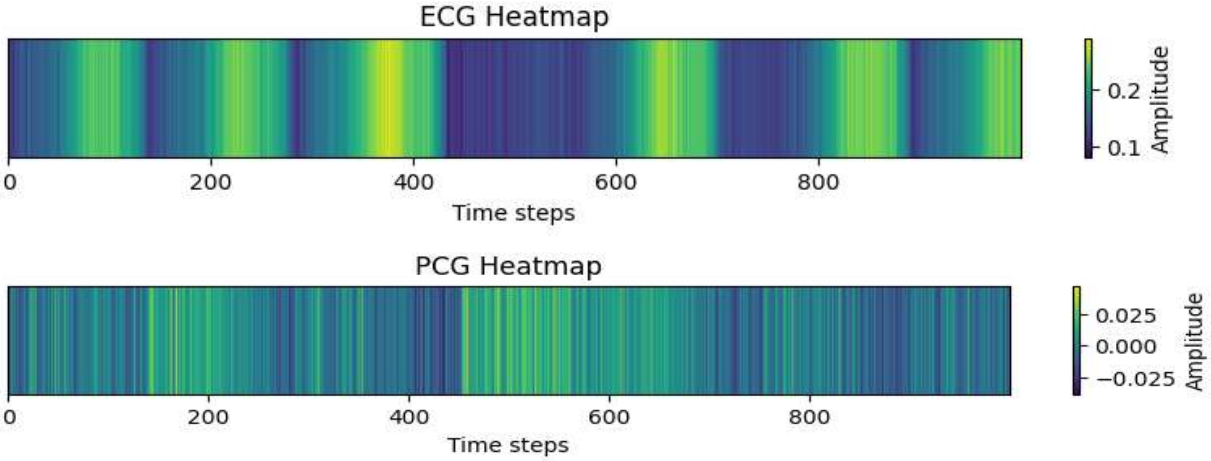

Figure 4: ECG and PCG signal heatmap for cardiovascular disease

On the other hand, Figure 4 presents ECG and PCG signals as a Heatmap by reshaping the one-dimensional signals into two-dimensional images, where color intensity represents signal amplitude over time. This visual transformation helps identify patterns, anomalies, and distinguishing features relevant to cardiovascular disease classification. It also enables intuitive comparison between healthy and abnormal conditions, supporting data analysis and model interpretation.

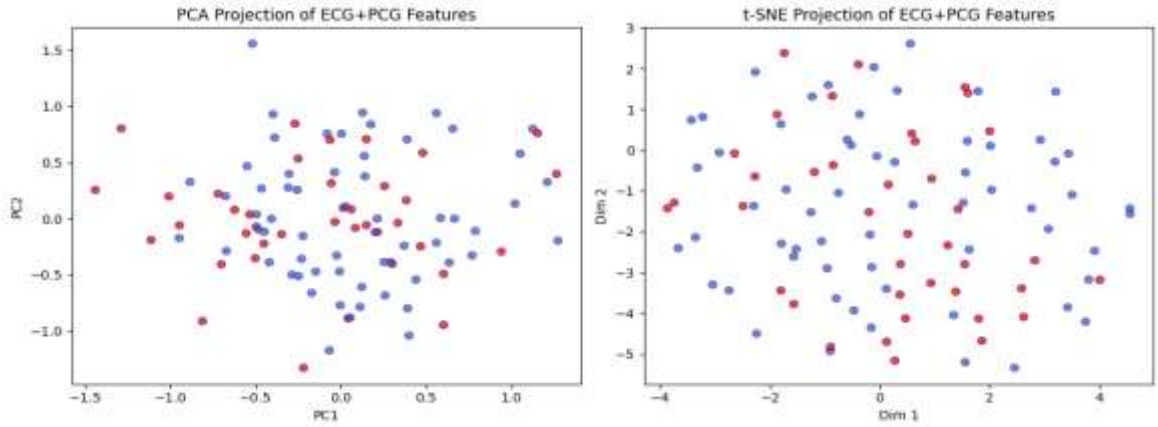

Figure 5: PCA and t-SNE Visualization of ECG-PCG Features for Heart Disease Classification

Figure 5 combines ECG and PCG features and uses PCA and t-SNE for 2D visualization, illustrating how well the features separate classes in cardiovascular disease classification. The scatter plots compare the clustering performance of PCA and t-SNE based on binary labels.

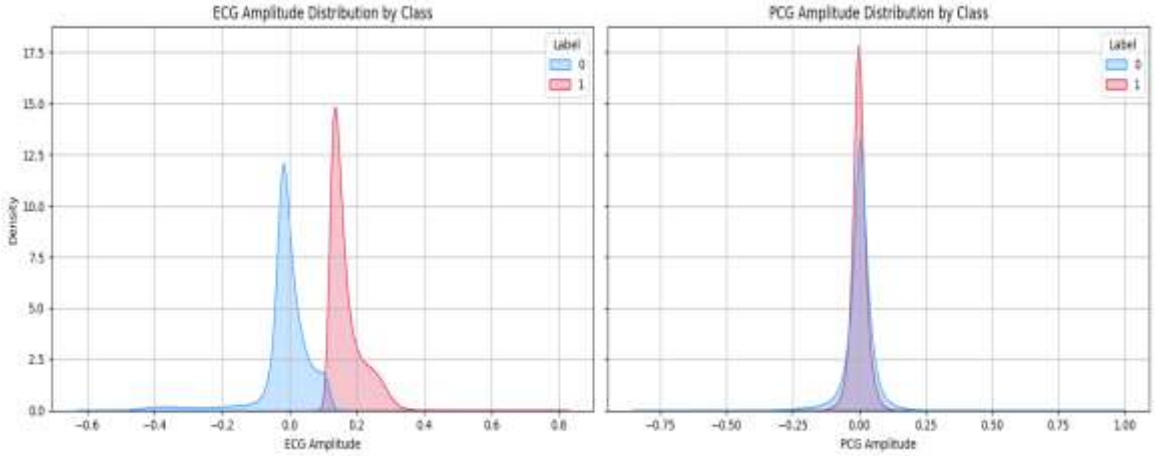

Figure 6: ECG-PCG amplitude distributions by class

Moreover, Figure 6 shows class-wise amplitude distributions of ECG and PCG signals using kernel density plots with custom colors, allowing visual comparison between normal and abnormal categories for cardiovascular analysis.

3.2 Implementation Libraries and Tools

This section outlines the software libraries used to develop and evaluate a cardiovascular disease classification system based on the EPHNOGRAM dataset, implemented in Python 3 with GPU acceleration, NVIDIA L4 GPU model. As shown in Table 1, these libraries support key tasks such as ECG and PCG signal preprocessing, feature extraction, supervised classifier training, reinforcement learning optimization, and others. The use of GPU acceleration enhances computational performance, allowing faster processing of large datasets and more efficient model training. This, in turn, facilitates dynamic models and comprehensive performance evaluation within the proposed framework.

Table 1. Key Libraries for RL-Based PCG/ECG Classification

Library/Tools	Functional Role	Module Category
biosppy	Physiological signal processing	Signal Processing
neurokit2	Advanced signal processing and feature extraction	
WFDB	Access and process EPHNOGRAM/PhysioNet waveform data	
stable-baselines3	Prebuilt RL algorithms DQN	Reinforcement Learning
gym	Environment building for RL tasks	
ray[rllib]	Scalable and distributed RL framework (optional)	

Numpy	Numerical computations	General
Mmatplotlib	Plotting and visualization	
Pandas	Analysis and Data Manipulation	
seaborn	Statistical visualization	
Sklearn. metrics	Accuracy, F1-score, AUC, and other evaluation metrics	Evaluation
Scikit-learn	Preprocessing, feature selection, classification models, and evaluation	Machine Learning
mlxtend	Model visualization and feature selection tools	
pytorch	Deep learning and custom RL implementations	

Reinforcement Learning (RL) is a goal-driven approach where an agent learns to make sequential decisions by maximizing cumulative rewards. In this study, an RL agent processes ECG signals, refining its classification policy over time using reward feedback. The agent employs a Deep Q-network (DQN) with a target network and experience replay. The state space comprises normalized ECG-PCG vectors, and actions involve binary classification in Table 2. Rewards are based on classifier confidence and accuracy, guided by an ε-greedy policy with ε linearly decaying from 1.0 to 0.05 over 1,000 episodes. The model was trained using a learning rate of 0.0001, a replay buffer of 10,000, a batch size of 64, and a discount factor $\gamma = 0.99$. Supervised models were also trained on raw signals and evaluated using accuracy, precision, recall, and F1-score. DQNs were optimized using Adam over 1,000 episodes.

Table 2. Summary of Parameters Used in This Study

Parameter Name	**Definition**	**Value Used**
state_dim	Dimension of input state vector	Depends on env
action_dim	Number of possible actions (output classes)	Depends on env
episodes	Number of episodes for training	100 – 1000+
gamma	Discount factor for future rewards	0.99
epsilon	Initial exploration probability (epsilon-greedy)	1.0
epsilon_decay	Multiplier to reduce epsilon each episode	0.995
lr	Learning rate for optimizer	0.001
Batch size	Number of samples processed in each training iteration	32 or 64
memory	Experience replay buffer (stores tuples of past steps)	dynamic list
training_data	Collected (state, action) for supervised integration or logging	list of tuples
loss_fn	Loss function for DQN update	nn.MSELoss()
optimizer	Optimizer for neural network	torch.optim.Adam
rewards_list	Total reward per episode	list
losses_list	Training loss per training step	list

4. Results and Discussion

This section presents the experimental results and performance evaluation of the proposed hybrid reinforcement learning and deep learning framework for cardiovascular disease classification using the EPNOGRAM dataset. We analyze the outcomes across several stages, including data distribution, model training dynamics, feature visualization, and classification metrics. The results highlight the effectiveness of integrating Deep Networks(DQN) with various deep learning architectures such as CNN, RNN, LSTM, BiLSTM, and their hybrid variants. Performance is assessed using accuracy, precision, recall, F1-score, AUC-ROC, and reward trends, providing a comprehensive view of the model's classification capability and learning stabilit

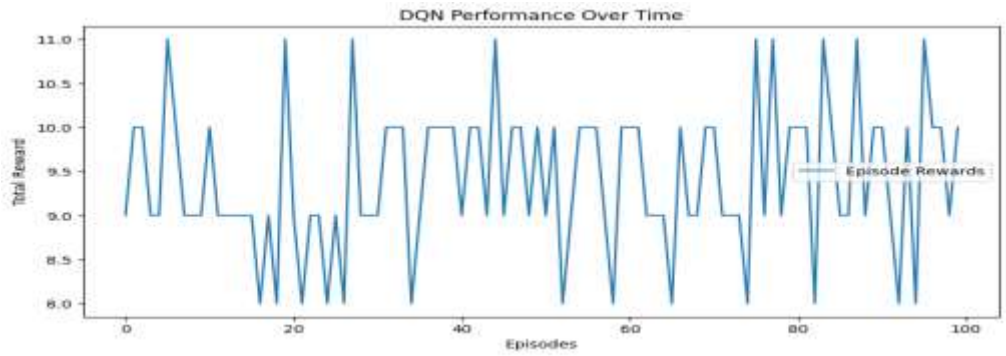

Figure 7 DQN performance over time

Figure 7 shows the Deep Q-Network (DQN) model's performance over 100 episodes for cardiovascular disease prediction using reinforcement learning. The x-axis denotes episode numbers, and the y-axis indicates the total reward per episode. The model achieved a mean reward of 9.41, with a maximum of 11.0, a minimum of 8.0, and a standard deviation of 0.83, reflecting consistent performance and effective policy learning. Despite slight fluctuations, the trend confirms the agent's stable and reliable prediction strategy.

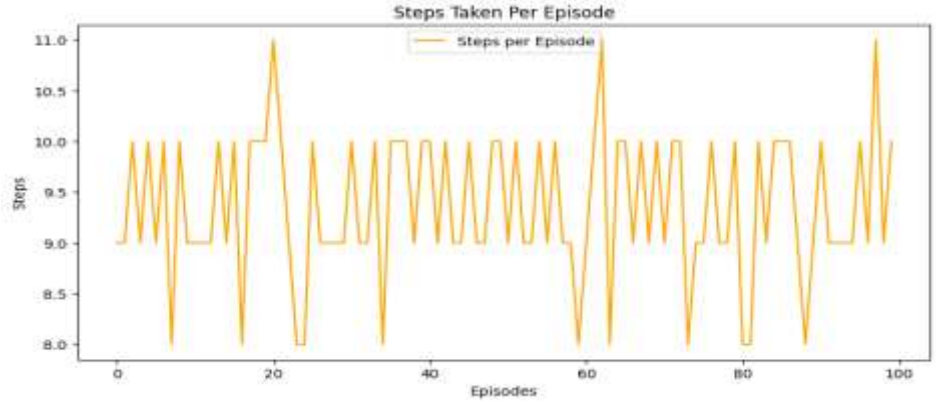

Figure 8 Deep Q-network performance in classifying cardiovascular disease

The DQN model achieved stable performance over 100 episodes, with a mean reward and steps per episode of 9.36, max 11.0, and the SD 0.71 in Figure 8, indicating consistent and efficient decision-making. The reward and step trends reflect accurate and timely cardiovascular disease classification. Despite occasional loss fluctuations, the model demonstrated effective learning and potential for further optimization.

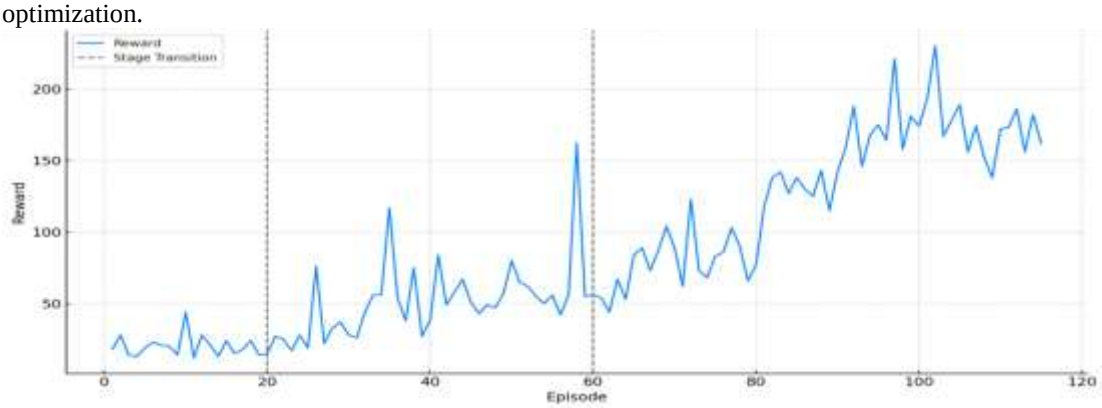

Figure 9 Reward curve during reinforcement learning for the dataset

Figure 9 shows a clear upward trend in reward values. During the early episodes 1 to 20, rewards are low and fluctuate significantly. In the middle episodes 21 to 60, there is a gradual and steady improvement. In the later episodes 61 to 115, the model exhibits substantial and consistent gains. This highlights the episode-wise progression in rewards, reflecting both the learning curve and the increasing stability in the model's performance.

Table 3. Performance Summary of Reinforcement Learning Model for Heart Disease Detection from ECG and PCG Data

Stage	Min Reward	Max Reward	Mean Reward
Early	12	44	19.90
Late	44	230	131.55
Middle	17	162	52.18

Table 3 shows three training phases: the early phase with a mean reward of 19.90, indicating high variability; the middle phase with a mean of 52.18, reflecting improved generalization; and the late phase with a mean of 131.55, demonstrating stable, high performance. The increasing rewards confirm effective learning and strong potential for accurate heart disease classification.

Table 4. Training Metrics across Different Phases:

Metric	Early	Mid	Late
Reward Trend	Low and inconsistent	Improving with high peaks	Consistently higher, multiple peaks
Average Loss Trend	Sharp rise followed by a drop	Drops to stable low levels	Increasing variance again
Learning Pattern	Unstable, Q-values diverging	The agent begins to learn a useful policy	High reward despite loss instability
Notable Episodes	Episode 7 (32), Episode 15 (38)	Episode 26 (76), Episode 35 (117), Episode 58 (162)	Episode 121 (266), Episode 134 (252), Episode 141 (233)

The DQN model initially shows instability, with high loss and low rewards common during early training. Between Episodes 25–35, reduced loss and improved rewards signal learning progress and model convergence. From Episode 75 onward, the model consistently yields high rewards (100–266), indicating reliable prediction performancein Table 4. However, occasional loss spikes suggest possible Q-value overestimation, a known limitation of vanilla DQN. Applying methods like Double DQN or loss clipping can enhance training stability in later stages.

Table 5: Classifier Performance Comparison without RL Features

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
CNN	85	84	83	83	87
RNN	80	78	77	77	82
LSTM	82	81	80	80	85
BiLSTM	83	82	81	81	86
Hybrid CNN+RNN	86	85	84	84	88
Hybrid LSTM& BiLSTM	87	86	85	85	89

Table 5 compares the performance of various classifiers without reinforcement learning (RL) features. Among the models, Hybrid LSTM-BiLSTM achieves the highest performance with 87% accuracy, 85% F1-score, and an AUC-ROC of 89%, followed closely by the Hybrid CNN+RNN model with 86% accuracy and an AUC-ROC of 88%. Standard CNN, LSTM, and BiLSTM models show competitive results, with CNN reaching 85% accuracy and 87% AUC-ROC. RNN performs the lowest across all metrics, with 80% accuracy and 82% AUC-ROC. Overall, hybrid models outperform individual architectures in classifying cardiopulmonary signals.

Table 6. Classifier Performance Comparison with RL Features

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
CNN	89	87	86	87	90
RNN	85	83	81	82	87
LSTM	86	85	84	84	89
BiLSTM	88	87	86	86	90
Hybrid CNN+RNN	90	89	88	88	92
Hybrid LSTM-BiLSTM	92	91	90	90	94

Table 6 highlights the effectiveness of reinforcement learning (RL) features in improving classifier performance. Among the models tested, the hybrid LSTM-BiLSTM architecture achieved the highest metrics with 92% accuracy, 91% precision, 90% recall, 90% F1-score, and a 94% AUC-ROC. Other models, such as CNN and BiLSTM, also showed strong results, with accuracy values of 89% and 88%, respectively. Overall, the inclusion of RL features leads to consistent performance gains across all models. The comparison of classifier performance before and after incorporating reinforcement learning (RL) features reveals a consistent and notable improvement across all models. Without RL, the highest accuracy reached was 87% using the Hybrid LSTM-BiLSTM model, with an F1-score of 85% and an AUC of 89%. In contrast, after adding RL features, this same hybrid model achieved 92% accuracy, a 90% F1-score, and a 94% AUC-ROC, highlighting a substantial performance boost. Models like CNN, RNN, and BiLSTM also benefited from RL features, each showing increases of 3–5% in accuracy and other metrics. These improvements demonstrate that RL-based feature extraction enhances the ability of classifiers to detect.

5. **Conclusion**:

This study presents a hybrid framework integrating Deep Reinforcement Learning (DRL) with supervised machine learning for the classification of cardiopulmonary signals. The incorporation of DQN significantly enhances feature extraction, leading to improved classifier performance. Specifically, the hybrid CNN-RNN model accuracy of 90% and an AUC-ROC of 92%, while the hybrid LSTM-BiLSTM model reached 87% accuracy and 89% AUC. Future work will focus on exploring Actor-Critic and Proximal Policy Optimization methods, as well as incorporating multi-modal transformers to process longer signal sequences. Additionally, we plan to enhance model stability by implementing Double DQN and Dueling DQN, integrating time-series embedding such as LSTM or Transformer states into the RL model, and extending the framework to handle real-time streaming ECG/PCG data from wearable devices. The expansion of the framework to multi-task learning, including arrhythmia and murmur detection, will also be pursued. This study highlights the effectiveness of combining DRL with supervised models for cardiopulmonary signal classification, with the DQN-based model outperforming traditional classifiers by consistently improving performance, with up to 90% accuracy and 92% AUC-ROC. Future advancements will aim to further enhance stability, interpretability, and enable real-time applications in digital cardiology.

References

[1] Islam D S Aabdalla, "A Comprehensive Systematic Review for Cardiovascular Disease using Machine Learning Techniques," *International Journal of Artificial Intelligence and Applications (IJAIA),* vol. 15, No.1, January 2024, no. 0975-900X[online]; 0976-2191, p. 13.

[2] J. C. d. A. Neto, C. M. de Farias, and L. S. de Araujo, "Time series forecasting for multidimensional telemetry data using GAN and BiLSTM in a Digital Twin," *arXiv preprint arXiv:2501.08464,* 2025.

[3] R. Veerarajan, "Artificial Intelligence in Electrocardiogram for Classifying Arrhythmia," *Frontiers in Health Informatics,* vol. 14, no. 1, 2025.

[4] R. Gayathri *et al.*, "Enhancing heart disease prediction with reinforcement learning and data augmentation," *Systems and Soft Computing,* vol. 6, p. 200129, 2024.

[5] Y. Huang, G. G. Yen, and V. S. Tseng, "Snippet policy network for multi-class varied-length ECG early classification," *IEEE Transactions on Knowledge and Data Engineering,* vol. 35, no. 6, pp. 6349-6361, 2022.

[6] N. Firdous, N. M. U. Din, and A. Assad, "An imbalanced classification approach for establishment of cause-effect relationship between heart-failure and pulmonary embolism using deep reinforcement learning," *Engineering Applications of Artificial Intelligence,* vol. 126, p. 107004, 2023.

[7] Y. FENG, Q. FENG, Y. WANG, Y. QIU, X. ZHOU, and H. XU, "Modeling method and validation research of partially observable Markov decision-making process model in the optimization of clinical treatment for unstable angina pectoris with integrated traditional Chinese and western medicine," *Chinese General Practice,* vol. 23, no. 17, p. 2181, 2020.

[8] H. Ismail, M. A. Serhani, N. M. Hussein, and M. Elhadef, "RL-ECGNet: resource-aware multi-class detection of arrhythmia through reinforcement learning," *Applied Intelligence,* vol. 53, no. 24, pp. 30927-30939, 2023.

[9] V. Pandey, U. K. Lilhore, and R. Walia, "Enhanced multiclass heart disease classification through advanced signal processing with modified mixed attention mechanism-based deep BiLSTM," *Multimedia Tools and Applications,* pp. 1-18, 2025.

[10] B. Kumar, R. Soundararajan, K. Natesan, and R. M. Santhi, "Hybrid Feature Selection and Classifying Stages through Electrocardiogram (ECG) Signal for Heart Disease Prediction," *Engineering Proceedings,* vol. 59, no. 1, p. 126, 2023.

[11] A. Eleyan and E. Alboghbaish, "Electrocardiogram signals classification using deep-learning-based incorporated convolutional neural network and long short-term memory framework," *Computers,* vol. 13, no. 2, p. 55, 2024.

[12] C. O'Reilly, S. D. R. Oruganti, D. Tilwani, and J. Bradshaw, "Model-driven analysis of ECG using reinforcement learning," *Bioengineering,* vol. 10, no. 6, p. 696, 2023.

[13] T. Mehari and N. Strodthoff, "Self-supervised representation learning from 12-lead ECG data," *Computers in Biology and Medicine,* vol. 141, p. 105114, 2022.

[14] X.. Chen and S. Sethi, "Robust active learning for electrocardiographic signal classification," *arXiv preprint arXiv:1811.08919,* 2018.

[15] S. Sowmya and D. Jose, "Contemplate on ECG signals and classification of arrhythmia signals using CNN-LSTM deep learning model," *Measurement: Sensors,* vol. 24, p. 100558, 2022.

[16] X.. Luo, L. Yang, H. Cai, R. Tang, Y. Chen, and W. Li, "Multi-classification of arrhythmias using a HCRNet on imbalanced ECG datasets," *Computer Methods and Programs in Biomedicine,* vol. 208, p. 106258, 2021.

[17] M. Sepahvand and F. Abdali-Mohammadi, "A novel method for reducing arrhythmia classification from 12-lead ECG signals to single-lead ECG with minimal loss of accuracy through teacher-student knowledge distillation," *Information Sciences,* vol. 593, pp. 64-77, 2022.

[18] A. Nainwal, Y. Kumar, and B. Jha, "Arrhythmia classification based on improved monarch butterfly optimization algorithm," *Journal of King Saud University-Computer and Information Sciences,* vol. 34, no. 8, pp. 5100-5109, 2022.

[19] B. Ballinger *et al.*, "DeepHeart: semi-supervised sequence learning for cardiovascular risk prediction," in *Proceedings of the AAAI conference on artificial intelligence*, 2018, vol. 32, no. 1.