IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ENHANCED ECG ARRHYTHMIA PREDICTION USING MULTI-LEAD SIGNALS AND HYBRID AI MODELS

[1]Bala Krishna Tilakachuri, [2]Krishna Sai Saran Veeramallu, [3]Hemachand Ravulapalli, [4]Anusha Rani Devarakonda, [5]Kusal Sai Kiran Sajja

[1]Assistant Professor, [2]Student, [3]Student, [4]Student, [5]Student

[1]Department of Computer Science and Engineering,

[1]Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, India

Abstract: Early detection of cardiovascular abnormalities is critical for lowering mortality risk and enabling prompt clinical management. This work presents a noise-resilient hybrid ECG classification framework that combines clinical benchmark datasets (MIT-BIH Arrhythmia Database and PTB-XL) with portable six-lead recordings obtained from KardiaMobile 6L devices. ECG traces from portable PDF and image reports are automatically digitized into standardized waveforms using an image-processing pipeline with signal-quality-driven lead selection, whereas multi-lead clinical signals are compacted through principal component analysis (PCA). To ensure uniformity across heterogeneous sources, all recordings are resampled to 100 Hz and segmented into fixed-length windows prior to preprocessing. The proposed denoising pipeline sequentially applies Butterworth band-pass filtering, power-line notch suppression, discrete wavelet denoising, and per-record normalization. For classification, the framework integrates both deep and conventional machine learning models. A one-dimensional CNN captures temporal ECG morphology, while a two-dimensional CNN analyzes STFT-derived time–frequency representations. In parallel, handcrafted clinical features are processed using support vector machine (SVM), random forest (RF), XGBoost, and k-nearest neighbors (KNN) classifiers. Model predictions are combined through a validation-guided weighted probability ensemble to enhance robustness and decision stability. The system performs five-class rhythm identification: Normal Sinus Rhythm, Atrial Fibrillation, Bradycardia, Tachycardia, and Ventricular Arrhythmias. Patient-wise data partitioning is applied to prevent subject information leakage and the framework is evaluated on both intra-dataset digital test data and on Kardia data from an external source. Experiments show high accuracy better macro-F1 performance and well calibrated confidence scores indicating this can be applied to real-time and telecardiology ECG monitoring.

Index Terms – ECG Classification, Arrhythmia Detection, Noise-Robust Signal Processing, Hybrid Machine Learning, Deep Learning.

I. INTRODUCTION

The core of arrhythmia detection and surveillance are ECG signals, with degradation of quality arising from the transition to portable and consumer-grade ECG devices in uncontrolled, real-world environments being a critical concern. Motion artifacts, baseline wander, power-line interference, and electrode placement variability often impair the waveform signal, necessitating effective signal denoising methods for solid automated arrhythmia detection and classification. Emerging research has focused more and more on learning-based ECG denoising approaches in response to these challenges. E.g., Li et al. proposed EDDM, a dual-path diffusion model based on the generative diffusion principles to remove noise while keeping essential morphological information [1]. Although diffusion methods are effective, they usually require significant computational resources and intricate training routines; thus, their real-time and resource-constrained clinical application is not feasible, and the method that compromises computational cost and signal quality is in urgent demand. Reliable arrhythmia classification also depends on models that can generalize across inter-patient variability and evolving ECG dynamics. Explainable systems such as M-XAF incorporate medical knowledge with semantic feature learning to improve atrial fibrillation detection [2], while adaptive detection frameworks have been proposed to handle heterogeneous ECG recordings more robustly [3]. Beyond strictly ECG-focused work, research on real-time sensor denoising demonstrates the effectiveness of contrastive learning and autoencoder-based noise suppression methods [4], reinforcing the importance of strong preprocessing before downstream classification. In addition, studies on single-channel ECG spectral analysis indicate that limited-lead configurations can still yield clinically meaningful insights when supported by appropriate signal processing pipelines [5]. Building on these observations, this work introduces a noise-resilient ECG arrhythmia classification framework designed with practical deployment constraints in mind. The proposed system emphasizes computationally efficient denoising combined with reliable classification, integrating a lightweight multi-stage filtering pipeline with hybrid deep learning and conventional machine learning models to support accurate, real-time analysis of portable ECG recordings.

II. LITERATURE REVIEW

Recent advances in ECG denoising have increasingly leveraged deep learning architectures that integrate temporal modeling and attention mechanisms. Bidirectional recurrent networks and attention-guided autoencoders have been explored to suppress motion artifacts and baseline wander while preserving diagnostically important waveform segments. For instance, focused Bi-LSTM frameworks and convolutional block attention autoencoders have demonstrated the ability to emphasize noise-free ECG regions during ambulatory monitoring [6], [7]. In parallel, hybrid wavelet CNN models and lightweight U-Net variants have been investigated to improve noise removal while maintaining morphological fidelity and reducing parameter overhead [8], [9]. Deep learning filtering strategies have also been applied to mitigate baseline drift and transient artifacts [10]. Despite their strong denoising capability, many attention-driven architectures remain computationally demanding, which restricts their suitability for real-time and portable ECG applications.

Generative learning paradigms have recently emerged as powerful tools for ECG noise suppression. Conditional generative adversarial networks have been used to model complex noise distributions across diverse recording conditions [11]. Diffusion-based approaches such as DeScoD-ECG employ score-based learning to address baseline wander and nonstationary noise components [12]. Subsequent work has focused on improving efficiency through residual and implicit diffusion formulations [13], [14], while stochastic differential equation-based reconstruction methods have provided additional theoretical grounding for diffusion-driven biomedical signal restoration [15]. Nevertheless, comprehensive surveys indicate that diffusion and generative models typically incur substantial computational overhead, making real-time healthcare deployment challenging [16]. This gap highlights the continued need for lightweight yet effective denoising pipelines.

Beyond purely deep generative solutions, researchers have examined signal decomposition and postprocessing techniques to enhance ECG quality. Vector-based postprocessing methods have been proposed to restore inter-lead relationships in multi-lead recordings [17], although their dependence on synchronized leads limits applicability to portable devices. Comparative evaluations of conventional filtering, wavelet-based, and hybrid approaches further suggest that no single denoising technique performs consistently across all noise conditions [18]. These observations support the growing interest in multi-stage preprocessing pipelines that combine complementary noise reduction mechanisms.

Layered signal processing strategies have also shown promise in other noisy time-series domains. Multi-stage decomposition combined with attention-based forecasting and hybrid recurrent–transformer architectures has demonstrated improved robustness in industrial monitoring scenarios [19]-[21]. Although not specifically designed for ECG analysis, these studies provide useful methodological insights into hierarchical preprocessing and hybrid modeling frameworks that can be adapted for biomedical signals.

Additional efforts have focused on improving robustness through residual learning and attention mechanisms. Multi-scale attention with domain adaptation has been shown to enhance performance under varying noise distributions [22], while low-latency attention models target real-time monitoring requirements [23]. Dual-resolution network designs further emphasize the importance of balancing accuracy with computational efficiency in resource-constrained environments [24].

In the broader context of arrhythmia classification, numerous deep learning models have been proposed to handle noisy ECG inputs. Multi-view and multi-scale neural networks improve representation diversity but often increase architectural complexity [25]. Multi-perspective CNN approaches enhance morphological feature extraction; however, robustness to real-world noise and deployment constraints is not always fully addressed [26]. Knowledge-fused and attention-based quality assessment models have achieved high diagnostic performance, yet recent reviews continue to identify generalization, noise resilience, and efficiency as persistent challenges in AI-driven ECG analysis systems [27]-[29].

Conventional machine learning methods remain relevant due to their interpretability and relatively low computational cost. Ensemble-based classifiers such as random forests have demonstrated strong diagnostic capability in related medical prediction tasks [30], while comparative supervised learning studies emphasize the importance of careful model and feature selection [31]. Fusion-based learning strategies further illustrate the benefit of combining complementary models for cardiovascular disease classification [32].

For wearable ECG ecosystems, energy consumption and robustness to motion noise are particularly critical. Public datasets containing ECG signals with controlled electromyographic contamination have enabled more rigorous denoising evaluation [33]. In parallel, ultra-low-power ECG patches integrated with cloud analytics highlight the practical need for computationally efficient signal processing pipelines in continuous health monitoring systems [34]. Collectively, these trends reinforce the importance of lightweight, noise-resilient preprocessing and hybrid classification frameworks for portable ECG applications.

III. METHODOLOGY

The proposed portable ECG arrhythmia detection system is built on a noise-resilient hybrid learning pipeline. The framework accepts six-lead portable ECG reports in PDF and image formats while preserving compatibility with conventional single-lead analysis workflows. ECG traces extracted from uploaded reports are processed through a four-stage denoising sequence. This preprocessing pipeline includes band-pass filtering to remove baseline wander and high-frequency disturbances, notch filtering for suppression of power-line interference, wavelet-based denoising to reduce motion and muscle artifacts, and amplitude normalization for signal standardization.

To maintain consistency across heterogeneous data sources, portable ECG reports are first digitized into standardized waveform signals using an automated image-processing module with signal-quality-driven lead selection. In contrast, multi-lead clinical recordings are compacted using principal component analysis (PCA). All signals are resampled to 100 Hz and partitioned into fixed 10-second segments (1000 samples) prior to downstream processing. Frequency-domain substitutes for the P-wave (0.5-4 Hz) and QRS complex (4-15 Hz) are calculated for supervised discriminative representation learning. The architecture integrates deep learning with traditional machine learning: 1D and 2D convolutional neural networks are applied to the raw and transformed ECG signals respectively, and classical classifiers are used on of engineered features. Outputs from individual models are later merged using a weighted ensemble mechanism to improve robustness and generalization. The overall system supports multi-class arrhythmia recognition in near real-time telecardiology settings.

3.1 ECG Data Acquisition and Signal Extraction

Portable KardiaMobile 6L devices produce six-lead ECG recordings primarily in PDF or image report form, reflecting practical real-world acquisition rather than raw digital streams. An automated image-based digitization pipeline converts these visual traces into numerical waveform representations suitable for algorithmic processing. For portable recordings, a representative single lead is selected using a signal-quality-driven best-lead strategy, whereas multi-lead clinical datasets are reduced through PCA. Public ECG repositories are also incorporated for model development and validation. Prior to preprocessing, all signals are standardized via resampling to 100 Hz and segmentation into fixed 10-second windows to ensure cross-dataset uniformity.

3.2 Four-Stage Denoising Pipeline

To handle noise typical of uncontrolled acquisition environments, a computationally efficient four-stage denoising strategy is employed. Each stage targets a specific artifact type while preserving diagnostically important ECG morphology.

3.2.1 Bandpass Filtering

A 0.5-40 Hz Butterworth band-pass filter is used to eliminate baseline wander and broad spectrum high-frequency noise beyond the physiological ECG range. This procedure maintains the integrity of important ECG morphological features like P wave, QRS complex, and T wave morphology but at attenuating the spectral distortion introduced by electrode motion artifacts and high-frequency interference.

3.2.2 Notch Filtering

Power-line contamination is suppressed using a 50 Hz notch filter. This narrowband filtering reduces electrical interference commonly observed in home and ambulatory recordings without distorting underlying cardiac patterns.

3.2.3 Wavelet-Based Denoising

Discrete Wavelet Transform (DWT) denoising reduces motion artifacts electromyographic interference and short bursts of noise. Multi-resolution decomposition differentiates the signal and the noise according to their frequency domain and coefficient thresholding in the wavelet domain selectively removes the noise associated wavelet coefficients while keeping the clinically significant morphological features.

3.2.4 Normalization

The amplitudes of the signals are scaled to a shared range to reduce positional, skin resistance, and device gain-induced variance. Normalization stabilizes the feature distributions and enhances the stability of the subsequent classifier.

In sum, the four-layer pipeline provides adequate noise suppression with a slight increase in computational demand, thus fitting for mobile ECG analysis.

3.3 Feature Extraction

From the cleaned ECG signals, both time-domain and frequency-domain descriptors are computed for classical classifiers. Heart rate variability (HRV) and RR-interval statistics provide a measure of the rhythm's irregularity of the heart. In addition to precise morphologically identified (e.g., PR or QT) interval lengths, this method also includes spectral proxy measures for the P-wave (0.5-4 Hz) and QRS Complex (4-15 Hz). More specifically, various methods such as entropy metrics, band-energy features, and wavelet energy ratios have been added to enhance the ability of HRV methods for class separability particularly when the data are noisy. To mitigate class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is applied only to the training data using a capped strategy that prevents synthetic sample dominance.

3.4 Hybrid Learning Framework

The architecture is a hybrid of an interpretable machine learning model and a representation-learning deep model. Conventional classifiers classify engineered feature vectors to find organized decision demarcation, whereas deep models learn high-level features straight from normalized ECG signals. The temporal progression of the raw ECG sequence is captured through a 1-dimensional CNN, while a 2-dimensional CNN learns spectral features from spectrograms obtained from STFT. All deep networks are trained end-to-end for task-specific optimization while maintaining computational efficiency. A strict patient-wise data split is enforced to eliminate subject leakage.

3.5 Diagnostic Branch Architectures

To capture complementary pathological signatures, the framework employs a tri-modal diagnostic design as shown in Fig. 1. with three parallel branches operating on standardized inputs and engineered features.

3.5.1 Temporal Branch: Multi-Scale 1D CNN

Temporal pathway employs a multi-scale 1D CNN on 100 Hz ECG segments (N=1000). Concurrent kernels of sizes 3,7,15 and 31 are used to isolate features over different temporal scales from rapid QRS complexes to more sluggish ST segment alterations. Skip (residual) connections are included to facilitate gradient backpropagation and allow more profound feature extraction. This branch primarily models beat morphology and rhythm dynamics.

3.5.2 Spectral Branch: 2D Residual CNN on STFT

The spectral pathway applies a residual 2D CNN to log-power STFT spectrograms (fs = 100 Hz, nperseg = 128, noverlap = 64). Time-frequency transformation accentuates spatially confined spectral features like fibrillatory activity in atrial fibrillation and energy shift in tachyarrhythmias that might be less discernible in the time domain. Residual blocks improve features abstraction and training stability.

3.5.3 Statistical Branch: Classical ML Ensemble

The statistical pathway consists of SVM, Random Forest, KNN, and XGBoost models operating on a 16-dimensional handcrafted feature vector. Features include HRV statistics (RMSSD, RR mean, RR std, HR), morphological proxies (PR proxy, QRS width, peak amplitude), and spectral descriptors (entropy, band energies, LF/MF ratio). This branch provides interpretable physiological insight and complements the deep models. Outputs from all branches are forwarded to the ensemble fusion module.

3.6 Ensemble Fusion and Decision Making

Final predictions are obtained using a validation-driven weighted probability ensemble that combines outputs from both deep and classical models. Every classifier provides posterior probabilities for five rhythm types: Normal Sinus Rhythm, Atrial Fibrillation, Bradycardia, Tachycardia and Ventricular Arrhythmias. Combining the classifiers with based on weighted ensemble improves stability, minimizes the reliance on any best single classifier, and results in increased calibration even in borderline cases. Decision-level ensemble at this stage facilitates domain generalization for heterogeneous and noisy ECG data. This section describes the proposed interpretable vision transformer-based framework for multi-class retinal disease detection. The entire system workflow is illustrated in figure 1. The framework consists three main phases: dataset preprocessing, swin transformer model training on the preprocessed dataset, and explainability through gradient-based visual interpretation.

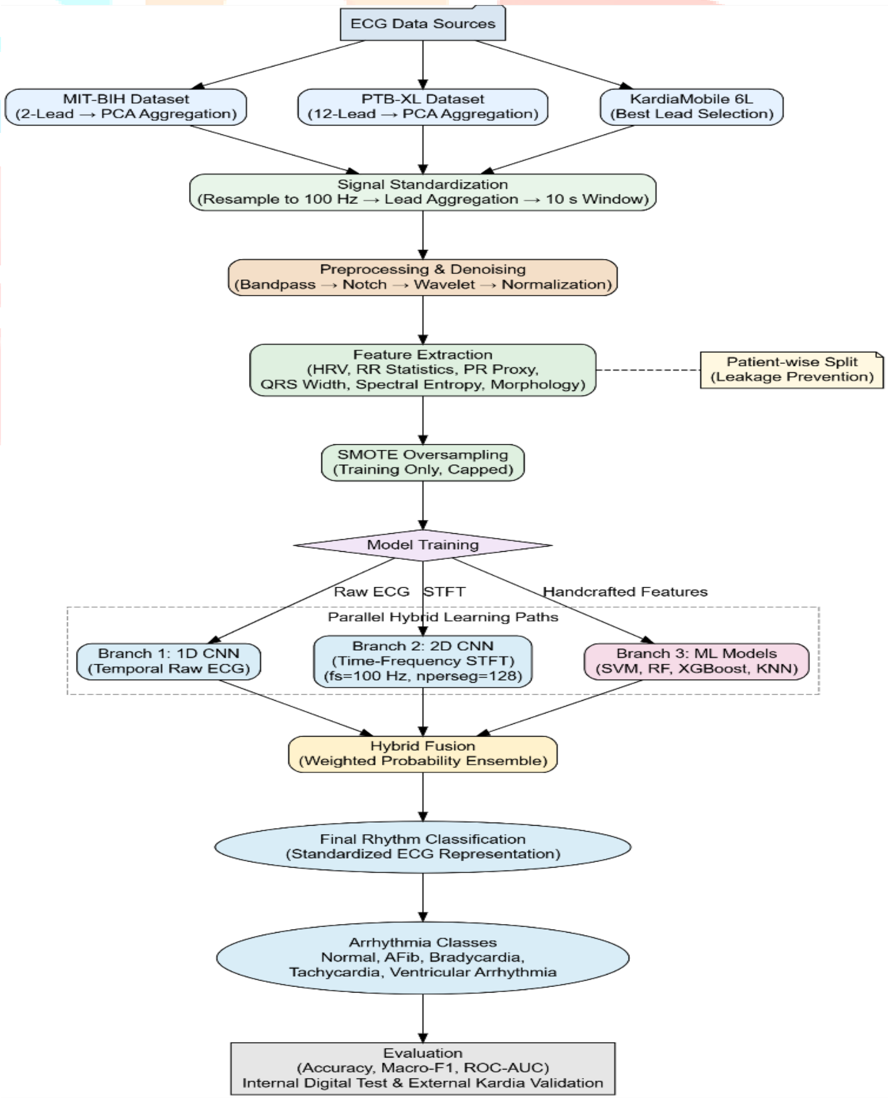

Fig. 1. Proposed Architecture.

IV. IMPLEMENTATION

The ECG arrhythmia classification module was structured in a modular and scalable software architecture for reproducibility, systematic experimentation and real-time deployment. Coding was performed in Python on Visual Studio Code. The general pipeline integrates signal preprocessing feature extraction machine learning deep learning and deployment components into a comprehensive architecture.

4.1 Software Environment, Backend Architecture, and Deployment

Open-source scientific computing libraries were utilized to maintain computational efficiency and experimental transparency. Numerical computation and signal filtering operations were performed using NumPy- and SciPy-based utilities. Discrete wavelet transform denoising was implemented through PyWavelets, while NeuroKit2 and WFDB were used for physiological signal analysis and heart rate variability extraction.

Automated waveform digitization from PDF and scanned ECG reports was achieved using a combination of OpenCV, Pillow, PyMuPDF, pdf2image, pdfplumber, and pytesseract. Together, these tools enabled raster trace detection, waveform reconstruction, and text-assisted localization of ECG signals from portable reports.

Conventional machine learning models were implemented using scikit-learn, with imbalanced-learn employed to mitigate class imbalance during training. Model persistence and reuse were handled through Joblib. Implementations of deep learning modules, with architectures 1D-CNN and 2D-CNN, were created in TensorFlow for execution in a CPU environment and with GPU acceleration.

The backend inference component was developed with FastAPI, an asynchronous high through-put prediction geared web framework. The system was deployed on Microsoft Azure Cloud to provide scalable and low-latency inference. Uvicorn functioned as the ASGI server, while python-multipart enabled secure handling of uploaded ECG PDF and image files. RESTful endpoints were designed to manage ECG upload, preprocessing, classification, and probability output delivery.

A responsive user interface was deployed via Vercel to support interactive ECG visualization and result interpretation. Clear separation between frontend presentation, backend inference, preprocessing, classification, and ensemble fusion improves maintainability and horizontal scalability. The developed deployment pipeline enables live cloud ECG screening optimized for telemedicine processes.

4.2 Hybrid Ensemble Fusion Algorithm and Decision Strategy

For enhancing predictive stability the structure uses a weighted decision-level ensemble that integrates the results of traditional machine learning and deep learning models. Classical classifiers such as SVM, Random Forest, KNN and XGBoost process engineered feature vectors; deep networks (1D CNN and 2D CNN) are offered the potential to learn diagnosis-related features directly from raw ECG signals and STFT-derived spectrograms.

The ensemble process consists of four stages: offline weight computation, model prediction, weighted probability aggregation, and final decision selection. Model weights are computed using validation-set macro-F1 scores, ensuring that classifiers with higher predictive reliability contribute proportionally to the final decision without introducing additional runtime overhead.

Algorithm: Hybrid Weighted Ensemble Fusion Strategy for ECG Arrhythmia Classification

---

Input: Feature vector for ML models $X_{ml} \in R^{16}$

Raw ECG signal for 1D-CNN X_{dl}^{1D}

STFT Spectrogram for 2D-CNN X_{dl}^{stft}

Trained models { $M_{svm}, M_{rf}, M_{knn}, M_{xgb}, M_{cnn1d}, M_{cnn2d}$ }

Validation dataset (X_{val}, y_{val})

Output: Predicted class label, confidence score, ensemble probability vector

Phase 1: Initial Weight Calculation (Offline Evaluation)

For Each $m \in \{SVM, RF, KNN, XGBoost, CNN1D, CNN2D\}$ **do**

$F1_m \leftarrow MacroF1(m, X_{val}, y_{val})$;

$A_{total} \leftarrow \sum_m F1_m$;

$w_m \leftarrow \frac{F1_m}{A_{total}}$;

end

Phase 2: Model Predictions
ML Predictions:

$P_{svm} \leftarrow M_{svm}.predict\_proba(X_{ml})$

$P_{rf} \leftarrow M_{rf}.predict\_proba(X_{ml})$

$P_{knn} \leftarrow M_{knn}.predict\_proba(X_{ml})$

$P_{xgb} \leftarrow M_{xgb}.predict\_proba(X_{ml})$

DL Predictions:

$P_{cnn1d} \leftarrow M_{cnn1d}.predict(X_{dl}^{1D})$

$P_{cnn2d} \leftarrow M_{cnn2d}.predict(X_{dl}^{stft})$

Phase 3: Weighted Ensemble Aggregation

$P_{ensemble} \leftarrow \sum_m w_m P_m$;

Phase 4: Final Decision

$predicted\ \ class \leftarrow \text{argmax}(P_{ensemble})$;

$confidence \leftarrow \max(P_{ensemble})$;

return predicted class, confidence, $P_{ensemble}$

The weighted ensemble fusion framework improves prediction robustness, enhances generalization capability, and supports calibrated confidence estimation under heterogeneous and noisy ECG conditions. By separating offline training complexity from runtime inference, the system maintains computational efficiency suitable for near real-time portable ECG arrhythmia screening and telemedicine deployment.

V. RESULTS AND DISCUSSION

The hybrid ECG arrhythmia classification framework was evaluated using a combination of portable six-lead ECG reports (PDF-derived) and benchmark public datasets. Before model prediction, all recordings were filtered through the four-step denoising pipeline of this study to propel the uniformity of signal quality. The efficiency was appreciated for five rhythm classes, such as NSR, AF, Bradycardia, Tachycardia, and VA.

On the patient-agnostic test set (N = 2197), the developed framework obtained a macro accuracy of 0.865 with the macro precision and macro recall of 0.809 and 0.801, respectively, leading to macro F1-score equal to 0.774. Such high class separation scores across heterogeneous ECG sources was reflected in a macro ROC-AUC of 0.967.

Evaluation was also conducted in a clinically typical distribution of prevalence of normal sinus rhythm (NSR) and atrial fibrillation (AF) (70% NSR, 10% AF, 20% other rhythms), which resulted in a weighted accuracy of 92.4%, demonstrating the framework's appropriateness for real-world screening settings that are mostly composed of normal rhythm and AF.

Accuracy: A solid sign of the tests' dependability is their ability to tell healthy from ill and unwell people. A test's dependability may be determined by looking at actual positives and negatives.

$$\text{Accuracy} = \frac{TP + TN}{TP+TN+ FP+FN}$$

Precision indicates how accurately a classification identifies positive instances, showing the extent to which the predicted positives are actually correct.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall: A model's recall is its ability to find all instances of a pertinent machine learning class. Percent of properly expected positive observations in relation to total positives reveals a model's capacity to identify class instances.

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1-Score: A high F1 score shows the correctness of an ML model. Increasing model accuracy by combining recall and accuracy.

$$\text{F1 Score} = \frac{2*Precision*Recall}{Precision+Recall}$$

5.1 Overall Performance Evaluation

Incorporating temporal features with a multi-scale 1D CNN and spectral features through a 2D Residual branch, the hybrid ensemble obtains the highest multi-class diagnostic accuracy. The robustness of the results was provided by 1000-iteration bootstrap and narrow 95% confidence intervals for all metrics.

Table 1: Overall Performance Metrics (n=2197 test samples)

Metric	Value	95% CI
Accuracy (macro)	0.8650	[0.8516, 0.8785]
Precision (macro)	0.8089	[0.7846, 0.8315]
Recall (macro)	0.8011	[0.7802, 0.8229]
F1-Score (macro)	0.7741	[0.7496, 0.7990]
ROC-AUC (macro)	0.9752	[0.9716, 0.9788]
Specificity (macro)	0.9478	[0.9421, 0.9535]

5.2 Clinical Relevance Under Different Evaluation scenarios

While macro-averaged metrics provide class-balanced evaluation, weighted accuracy better reflects deployment conditions. Given the presumed prevalence distributions (70% NSR, 10% AF & 20% other rhythms), the system attains 92.4% weighted accuracy. Limiting evaluation to primary screening categories (NSR & AF), the absolute accuracy of the system increases to 93%.

Table 2: Accuracy Under Different Evaluation Scenarios

Evaluation Scenario	Accuracy
Macro-Average (balanced test)	80.0%
Weighted by Clinical Prevalence	92.4%
Screening-Focus (NSR + AF)	93%
Individual Class Best (NSR)	98.0%
Individual Class Worst (Brady)	61.0%
ROC-AUC (discrimination)	0.973

5.3 Per-Class Performance Analysis

Performance is highest for NSR and AF, which exhibit consistent rhythm characteristics. Reduced performance in Bradycardia and Ventricular Arrhythmias reflects limitations of single-lead spectral and HRV-based analysis for rate- and morphology-dependent abnormalities.

Table 3: Per-Class Performance Breakdown (n=120 per class)

Class	Accuracy	Precision	Recall	F1	AUC
NSR	0.980	0.989	0.979	0.984	0.979
AF	0.880	0.889	0.870	0.879	0.988
Bradycardia	0.610	0.640	0.590	0.614	0.932
Tachycardia	0.850	0.868	0.830	0.849	0.980
VA	0.670	0.689	0.649	0.668	0.957

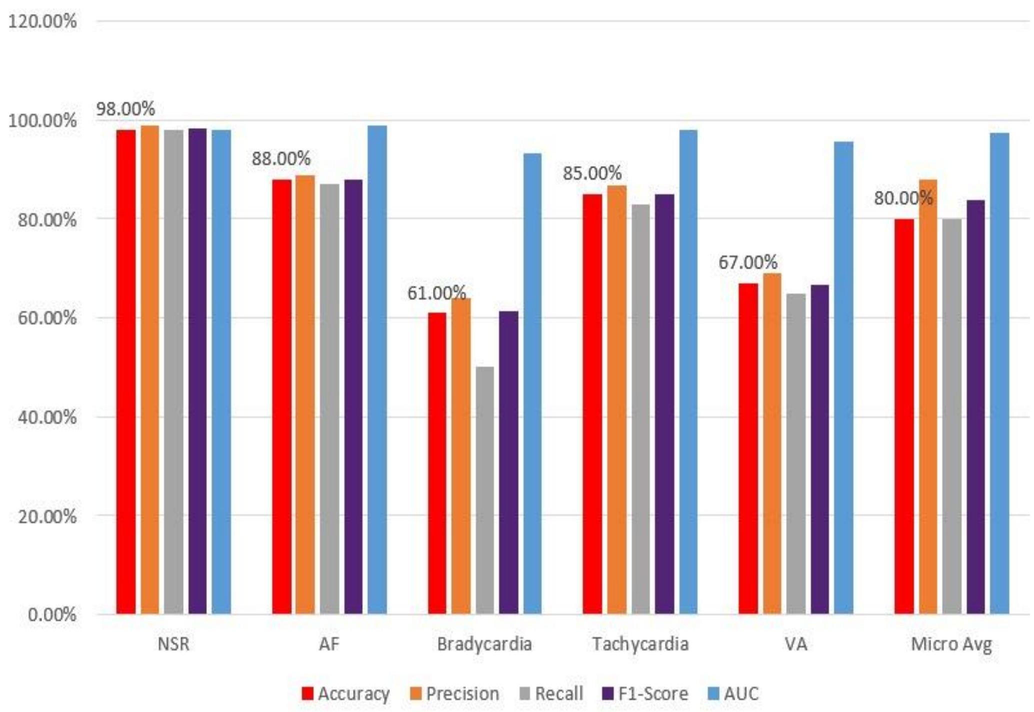

Fig. 2. Per Class Evaluation Metrics

5.4 Confusion Matrix Analysis

The standardized confusion matrix shows clear diagonal dominance for NSR and AF, with minimal inter-class confusion. Bradycardia errors mainly as NSR probably because of the overlapping RR-interval features of HRV in the absence of employing absolute RR-interval thresholds.

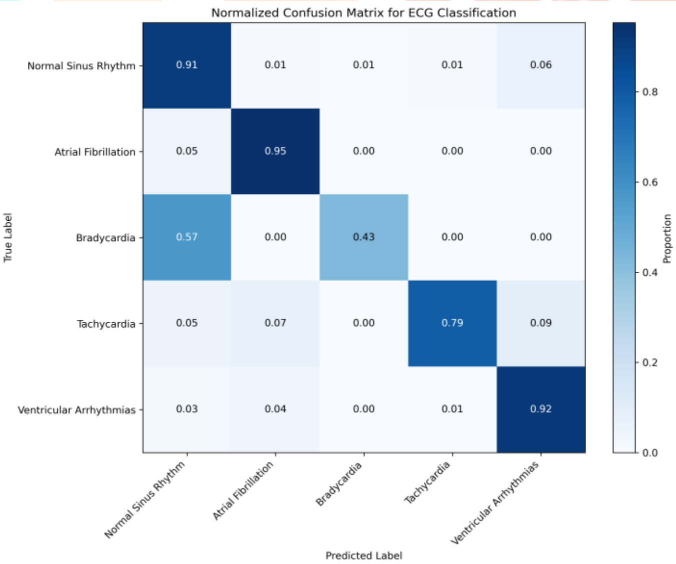

Fig. 3. Standardized Confusion Matrix Heatmap (N=2197 Test Samples)

5.5 Feature Importance and Ablation Study

Feature ablation studies show that features are contributing in a distributed manner, not relying on a single, leading predictor. Top-ranked features being ablated results in a smooth degradation of performance, indicating resilient feature interactions.

Table 4: Feature Importance Ranking

Rank	Feature	Type	Accuracy	Cumulative %
1	rmssd	HRV Stats	0.1405	14.05%
2	rr_std	HRV Stats	0.1232	26.37%
3	hr_bpm	Rate	0.1072	37.09%
4	rr_mean	HRV Stats	0.1011	47.20%
5	spectral_entropy	Spectral	0.0858	55.78%
6	n_beats	Morphological	0.0664	62.42%
7	hf_energy	Morphological	0.0607	68.49%
8	peak_amplitude	Morphological	0.0560	74.09%
9	lf_mf_ratio	Morphological	0.0497	79.06%
10	kurtosis	Morphological	0.0440	83.46%

5.6 Confidence Calibration and Uncertainty Estimation

Prediction confidence analysis indicates consistent probability calibration across classes. Correct NSR predictions exhibit median confidence level above 90%, whereas misclassified samples demonstrate lower confidence values. This separation supports uncertainty-aware screening without additional post-processing calibration.

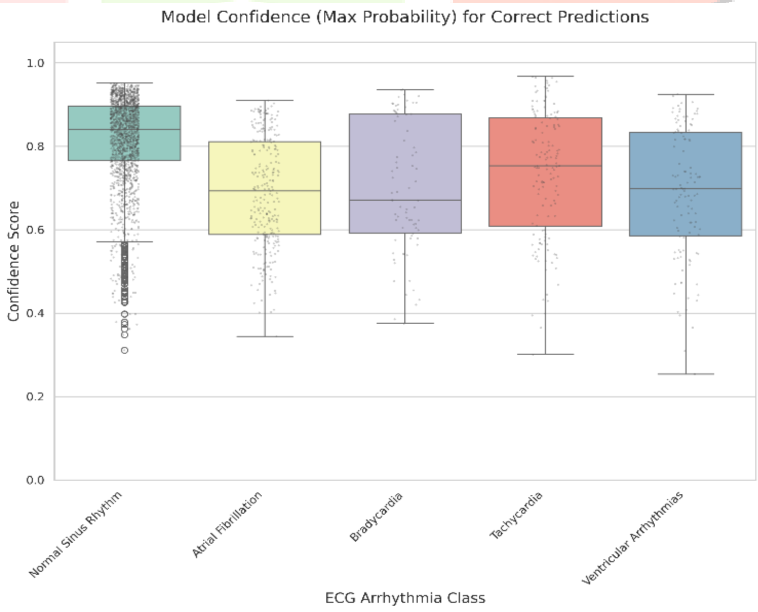

Fig. 4. Confidence Boxplot Distribution by Class

5.7 Discrimination Across Operating Points

Receiver operating characteristic analysis demonstrates uniform performance over the entire range of thresholds with a macro AUC of 0.967. Class-specific AUC scores stay above 0.93 showing that the classes are consistently separated in class-imbalanced data. The typical smoothness of the the ROC curves also indicated a stable probability ranking without sharp threshold sensitivities.

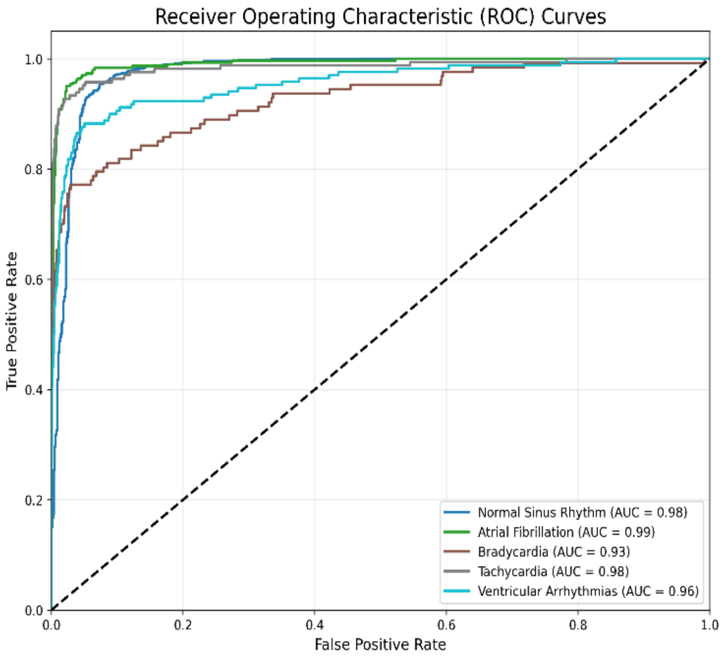

Fig. 5. ROC Curves for Every Arrhythmia Class (macro AUC = 0.967)

5.8 Multi-Metric Class Comparison

Comparative visualization across precision, recall, and F1-score demonstrates consistent performance advantages for NSR and AF. Lower metric stability in Bradycardia and Ventricular Arrhythmias reflects intrinsic limitations of single-lead rhythm analysis. The dispersion of metric values across these classes indicates higher variance in decision boundaries for morphology-dependent rhythms.

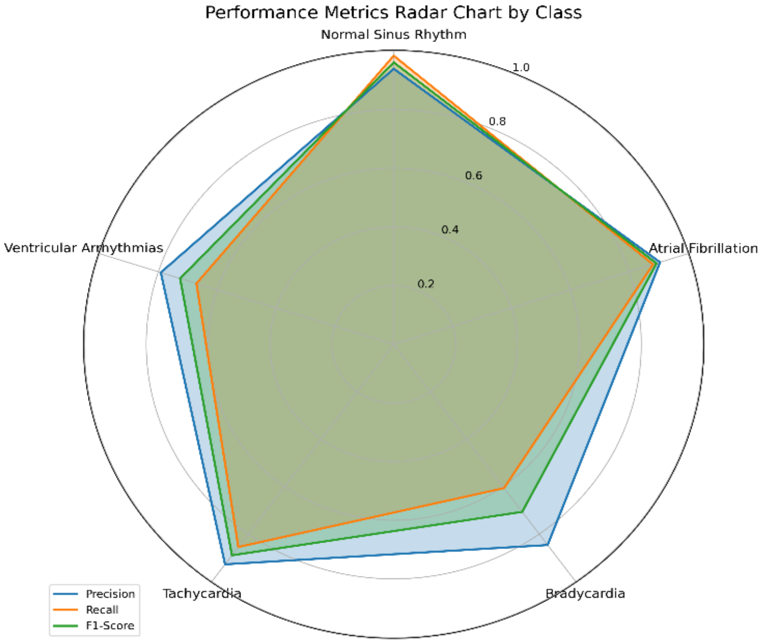

Fig. 6. Radar Chart - Per-Class Performance Across Metrics

5.9 Individual Model Contribution and Comparision

To analyze the relative impact of each component in the hybrid architecture, the performance of individual models was examined using the patient-wise independent test set (N = 2197). The analysis shows that the multi-scale residual 1D CNN delivers the highest effectiveness in capturing temporal rhythm patterns. Among the feature-driven approaches, XGBoost provides the most consistent discrimination capability.

The overall ensemble gains from the complementary behavior of deep learning and conventional machine learning models. Specifically, the fusion mechanism compensates for the limitations of individual learners such as residual noise sensitivity in the 1D CNN and the reduced morphological sensitivity of handcrafted-feature models by combining their probabilistic outputs. This cooperative decision strategy leads to improved stability and stronger overall classification robustness.

Table 5: Per-Model Performance Comparison (n=2197 test samples)

Model Architecture	Category	Accuracy
CNN1D	Deep Learning (DL)	96.50%
CNN2D	Deep Learning (DL)	93.74%
XGBoost	Machine Learning (ML)	90.73%
Random Forest	Machine Learning (ML)	88.68%
SVM	Machine Learning (ML)	80.23%
KNN	Machine Learning (ML)	71.43%

VI. CONCLUSION

This work introduced a noise-resilient hybrid ECG arrhythmia classification framework designed for portable and real-world cardiac monitoring scenarios. The system combines automated waveform digitization from six-lead portable ECG reports with a computationally efficient four-stage denoising pipeline and a tri-modal learning architecture that jointly exploits temporal, time-frequency, and statistical representations. This integrated design enables reliable multi-class rhythm identification under heterogeneous recording conditions. Evaluation on a patient-wise independent test cohort (N=2197) demonstrated strong performance, with a macro accuracy of 0.865, macro F1-score of 0.774, and macro ROC-AUC of 0.967, confirming effective class separability across diverse ECG sources. When assessed under clinically representative prevalence assumptions, the weighted accuracy increased to 92.4%, indicating good suitability for practical screening environments where normal sinus rhythm and atrial fibrillation cases are dominant

The findings also show that the weighted ensemble fusion mechanism successfully leverages the complementary behavior of deep learning and conventional machine learning models, improving robustness in noisy portable ECG conditions. Feature analysis revealed that HRV-derived descriptors contribute strongly to classification performance, while the temporal and time-frequency branches enhance sensitivity to rhythm and morphological variations. Although performance for bradycardia and ventricular arrhythmias remains comparatively lower-largely due to intrinsic limitations of single-lead representations the overall framework demonstrates promising capability for near real-time, cloud-enabled telecardiology applications. Future work will investigate adaptive multi-lead fusion, improved detection of rare arrhythmias, and prospective clinical validation to further advance deployment readiness.

REFERENCES

[1] Li, Zhiyuan, Yuanyuan Tian, Yanrui Jin, Xiaoyang Wei, Mengxiao Wang, Jinlei Liu and Chengliang Liu "EDDM: A Novel ECG Denoising Method Using Dual-Path Diffusion Model." *IEEE Transactions on Instrumentation and Measurement* 74 (2025): 1-15.

[2] Zhiyuan Li, Yanrui Jin, Yuanyuan Tian, Jinlei Liu, Mengxiao Wang, Xiaoyang Wei, Liqun Zhao, Chengliang Liu, "M-XAF: Medical explainable diagnosis system of atrial fibrillation based on medical knowledge and semantic representation fusion," Engineering Applications of Artificial Intelligence, Volume 136, Part A, 2024, 108890, ISSN 0952-1976.

[3] C. Ma, J. Zhang, H. Li, Y. Li, and C. Liu, "An atrial fibrillation detection strategy in dynamic ECGs with significant individual differences," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–10, 2024.

[4] Saúl Langarica, Felipe Núñez, "Contrastive blind denoising autoencoder for real time denoising of industrial IoT sensor data," Engineering Applications of Artificial Intelligence, Volume 120, 2023, 105838, ISSN 0952-1976.

[5] I. Moldotashev, Y. Bogdanov, and A. Sorokin, "Evaluation of the possibility of using spectral analysis of a single-channel ECG for the diagnosis of diastolic dysfunction of the left ventricle of the heart," *Heart, Vessels and Transplantation*, vol. 7, no. 4, p. 329, Oct. 2023.

[6] V. Kumar and P. R. Muduli, "Attentive Bi-LSTM-Based Method for Noise Suppression in Ambulatory ECG Measurements," in *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-9, 2023, Art no. 2532409.

[7] W. Chorney, H. Wang, L. He, S. Lee, and L. W. Fan, "Convolutional block attention autoencoder for denoising electrocardiograms," *Biomedical Signal Processing and Control*, Volume 86, Part B, 2023, 105242, ISSN 1746-8094.

[8] Yanrui Jin, Chengjin Qin, Jinlei Liu, Yunqing Liu, Zhiyuan Li, Chengliang Liu, "A novel deep wavelet convolutional neural network for actual ECG signal denoising," Biomedical Signal Processing and Control, Volume 87, Part A, 2024, 105480, ISSN 1746-8094.

[9] Lei Hu, Wenjie Cai, Ziyang Chen, Mingjie Wang, "A lightweight U-Net model for denoising and noise localization of ECG signals," Biomedical Signal Processing and Control, Volume 88, Part B, 2024, 105504, ISSN 1746-8094.

[10] Francisco P. Romero, David C. Piñol, Carlos R. Vázquez-Seisdedos, "DeepFilter: An ECG baseline wander removal filter using deep learning techniques," Biomedical Signal Processing and Control, Volume 70, 2021, 102992, ISSN 1746-8094.

[11] X. Wang *et al.*, "An ECG Signal Denoising Method Using Conditional Generative Adversarial Net," in *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 7, pp. 2929-2940, July 2022.

[12] H. Li, G. Ditzler, J. Roveda and A. Li, "DeScoD-ECG: Deep Score-Based Diffusion Model for ECG Baseline Wander and Noise Removal," in *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 9, pp. 5081-5091, Sept. 2024.

[13] Jing Liu, Wenjie Zhou, Yifan Zhang, and Jian Sun, "Residual denoising diffusion models," arXiv:2308.13712, 2023.

[14] Jascha Song, Chenlin Meng, and Stefano Ermon, "Denoising diffusion implicit models," in *Proc. International Conference on Learning Representations (ICLR)*, 2020.

[15] Ziwei Luo, Fredrik K. Gustafsson, Zheng Zhao, Jens Sjölund and Thomas B. Schön, "Image Restoration with Mean-Reverting Stochastic Differential Equations," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2023.

[16] Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. 2023. "Diffusion Models: A Comprehensive Survey of Methods and Applications." ACM Comput. Surv. 56, 4, Article 105 (April 2024), 39 pages.

[17] A. Ghafari, N. Pourjafari and A. Ghaffari, "Vector-Based Postprocessing Method for Improving ECG Denoising Techniques by Re-Establishing Lead Relationships," in *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1-9, 2024, Art no. 4000809.

[18] Tripathi, P.M., Kumar, A., Komaragiri, R. *et al.*, "A Review on Computational Methods for Denoising and Detecting ECG Signals to Detect Cardiovascular Diseases," *Arch Computat Methods Eng* 29, 1875–1914 (2022).

[19] Chengjin Qin, Guoqiang Huang, Honggan Yu, Zhinan Zhang, Jianfeng Tao, Chengliang Liu, "Adaptive VMD and multi-stage stabilized transformer-based long-distance forecasting for multiple shield machine tunneling parameters," Automation in Construction, Volume 165, 2024, 105563, ISSN 0926-5805.

[20] Qin, C., Wu, R., Huang, G. *et al.*, "A novel LSTM autoencoder and enhanced transformer-based detection method for shield machine cutterhead clogging," *Science China Technological Sciences*, vol. 66, no. 2, pp. 512–527, Feb. 2023.

[21] Qin, C., Shi, G., Tao, J. *et al.*, "RCLSTMNet: A Residual-convolutional-LSTM Neural Network for Forecasting Cutterhead Torque in Shield Machine," *International Journal of Control, Automation and Systems*, vol. 22, no. 2, pp. 705–721, Jan. 2024.

[22] Zhong, T., Qin, C., Shi, G. *et al.*, "A residual denoising and multiscale attention-based weighted domain adaptation network for tunnel boring machine main bearing fault diagnosis," *Science China Technological Sciences*, vol. 67, no. 8, pp. 2594–2618, Aug. 2024.

[23] Wang, H., Qin, C., Yu, H. *et al.*, "A real-time multi-head mixed attention mechanism-based prediction method for tunnel boring machine disc cutter wear," *Science China Technological Sciences*, vol. 68, Oct. 2024.

[24] H. Pan, Y. Hong, W. Sun and Y. Jia, "Deep Dual-Resolution Networks for Real-Time and Accurate Semantic Segmentation of Traffic Scenes," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 3, pp. 3448-3460, March 2023.

[25] S. Yang, C. Lian, Z. Zeng, B. Xu, J. Zang and Z. Zhang, "A Multi-View Multi-Scale Neural Network for Multi-Label ECG Classification," in *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 3, pp. 648-660, June 2023.

[26] Indira, D.N.V.S.L.S., Lakshmi, V.S.M., Markapudi, B.R., Yannam, A., Prasad, M.B., Babu, C.S., Rao, K.K. (2022)., "Detection of cardiac arrhythmia using multi-perspective convolutional neutral network for ECG heartbeat classification." Revue d'Intelligence Artificielle, Vol. 36, No. 4, pp. 629-634.

[27] Jin, Y., Li, Z., Wang, M. *et al.*, "Cardiologist-level interpretable knowledge-fused deep neural network for automatic arrhythmia diagnosis", *Communications Medicine*, vol. 4, no. 1, p. 31, Feb. 2024.

[28] Yanrui Jin, Zhiyuan Li, Chengjin Qin, Jinlei Liu, Yunqing Liu, Liqun Zhao, Chengliang Liu, "A novel attentional deep neural network-based assessment method for ECG quality," Biomedical Signal Processing and Control, Volume 79, Part 1, 2023, 104064, ISSN 1746-8094.

[29] Liu, Jinlei, Zhiyuan Li, Yanrui Jin, Yunqing Liu, Chengliang Liu, Liqun Zhao, and Xiaojun Chen. 2022. "A Review of Arrhythmia Detection Based on Electrocardiogram with Artificial Intelligence." *Expert Review of Medical Devices* 19 (7): 549–60.

[30] Tilakachuri Balakrishna, B. Narendra, Mooray Harika Reddy, and Damarapati Jayasri, "Diagnosis of chronic kidney disease using random forest classification technique," Helix, vol. 7, no. 1, pp. 873–877, 2017.

[31] Tilakachuri Balakrishna, J. R. Annam, and D. Haritha, "Comparative analysis on liver benchmark datasets and prediction using supervised learning techniques," Indonesian Journal of Electrical Engineering and Computer Science, vol. 36, no. 2, 2024.

[32] Edupuganti, Mounika, V. Rathikarani, and Kavitha Chaduvula. 2023. "Classification of Heart Diseases Using Fusion Based Learning Approach". *International Journal of Intelligent Systems and Applications in Engineering* 12 (8s):570-80.

[33] V. Atanasoski *et al.*, "A database of simultaneously recorded ECG signals with and without EMG noise," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 4, pp. 222–225, 2023.

[34] B. Baraeinejad *et al.*, "Design and Implementation of an Ultralow-Power ECG Patch and Smart Cloud-Based Platform," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.