Vol. 8, No. 3, June 2026

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

A 1D-CNN Model with Modified MITDB-SVDB Dataset for Multiclass Arrhythmia Classification

Muhamad Akbar[1], Muhammad Irvai[2]

[1,2]Department of Informatics, Bina Insan University, Lubuklinggau, Indonesia

Received:
October 8, 2025
Revised:
May 10, 2026
Accepted:
May 30, 2026
Published:
June 24, 2026

Corresponding Author:
Author Name*:
Muhamad Akbar
Email*:
muhamad.akbar@univbinainsan.ac.id

DOI:
10.63158/journalisi.v8i3.1649

Abstract. Automated arrhythmia classification from electrocardiogram (ECG) signals remains challenging because public datasets are highly imbalanced and fine-grained multiclass performance may degrade when labels are mapped to the clinically standardized AAMI EC57 grouping scheme. This study proposes real-record dataset enrichment combined with a compact one-dimensional convolutional neural network (1D-CNN) for both fine-grained and AAMI-grouped beat classification. Fourteen records from the MIT-BIH Supraventricular Arrhythmia Database were inserted into the MIT-BIH Arrhythmia Database, adding 4,649 S beats, 4,530 V beats, and 47 Q beats without synthetic oversampling. Preprocessing included Christov R-peak segmentation, beat extraction, per-beat min-max normalization, and resampling to 180 Hz. The 1D-CNN was evaluated under 16-class, 17-class, and 5-class AAMI EC57 schemes. Using ASGD, the model achieved accuracies of 99.10%, 98.58%, and 99.38%, with macro F1-scores of 0.90, 0.87, and 0.97, respectively. Cross-database testing on INCARTDB reached 99.13% accuracy across four mappable classes (N, V, R, A), indicating limited 4-class transferability rather than full AAMI generalization. The approach preserves authentic ECG morphology while addressing minority-class scarcity. The findings show that real-beat enrichment can improve balanced ECG classification, although results are based on beat-level random splits and require future record-wise validation before clinical deployment.

Keywords: AAMI EC57 standard; data imbalance mitigation; ECG beat classification; MIT-BIH Arrhythmia Database; multiclass arrhythmia classification

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

Electrocardiogram (ECG) interpretation is an important clinical task because rhythm disorders can reflect abnormal electrical activity of the heart. Automated arrhythmia classification has therefore become an active research area in artificial intelligence, especially with deep learning models that can learn discriminative representations directly from signals [1], [2], [3]. In this domain, the MIT-BIH Arrhythmia Database (MITDB) is widely used because it provides beat-level annotations for multiple rhythm types [4], [5]. However, the distribution of beats is highly imbalanced: normal beats dominate the dataset, while several clinically meaningful arrhythmia classes have very few samples[6]. This imbalance tends to reduce the stability of minority class prediction and can make high aggregate accuracy misleading[7], [8], [9], [10], [11].

Previous studies have proposed several strategies for arrhythmia classification. Luo et al. used HCRNet with SMOTE to classify nine classes and obtained 99.01% fine-grained accuracy and 98.70% AAMI-grouped accuracy [12]. Shi et al. combined convolutional and recurrent components in a multiple-input deep network and reported high multiclass performance but lower AAMI performance [13]. Raj and Ray used sparse representation and optimized least-square twin SVM for 16 classes, achieving 99.11% class-oriented accuracy but only 89.93% AAMI-oriented accuracy [14]. These studies show that good performance in a detailed class-oriented setting does not necessarily translate into strong AAMI-grouped performance. Synthetic oversampling methods such as SMOTE and GAN-based augmentation have been used to address class imbalance [16], [17], but they may not fully capture the physiological morphological variability of real arrhythmia beats across different patients and recording devices.

The gap addressed in this paper is the need for a dataset strategy and model that can classify fine-grained arrhythmia classes while also supporting the AAMI EC57 5-class grouping. Three evaluation dimensions are distinguished throughout this paper: (1) class-oriented using individual WFDB beat annotations as labels; (2) AAMI EC57-grouped using the EC57 standard 5-class grouping; and (3) patient-independent a record-wise protocol where all beats from a given patient are restricted exclusively to either training or testing. This study addresses the first two dimensions; patient-independent evaluation is explicitly identified as a limitation and a priority for future work. The proposed approach

modifies MITDB by inserting real ECG records from SVDB preferred over SMOTE or GAN-based augmentation because real beats preserve authentic physiological morphological variability [16], [17] and trains a compact 1D-CNN with systematic hyperparameter tuning. The research questions are: (1) how real-record dataset enrichment can reduce class imbalance effects; (2) how a single 1D-CNN support can be 16-class, 17-class, and AAMI EC57 (grouped 5) class classification simultaneously; and (3) how well the model generalizes to an unseen external ECG database. The reported results reflect beat-level splitting and should not be interpreted as patient-independent generalization.

2. METHODS

This section describes the complete experimental pipeline: dataset construction and modification, beat extraction and preprocessing, 1D-CNN architecture, hyperparameter tuning, evaluation protocol, and external cross-database validation. Full reproducibility details are provided throughout. Figure 1 shows research workflow.

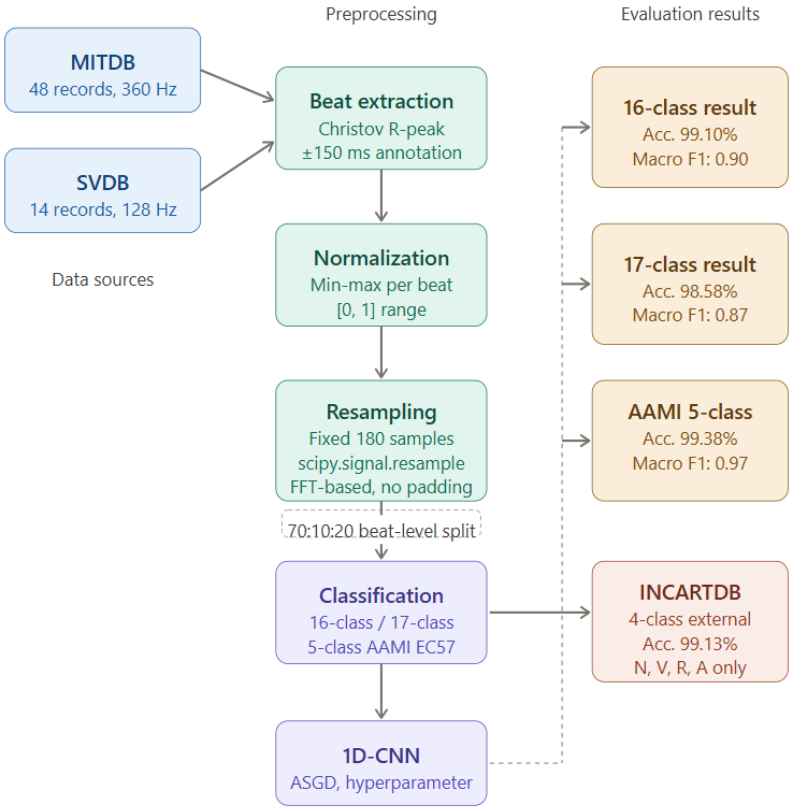

Figure 1. Research workflow: from raw ECG databases through preprocessing, classification, and evaluation.

2.1. Dataset Construction and Modification Strategy

The study used the MIT-BIH Arrhythmia Database (MITDB) as the primary source and the MIT-BIH Supraventricular Arrhythmia Database (SVDB) as a real-record supplement for underrepresented classes. MITDB comprises 48 two-channel ambulatory ECG recordings digitized at 360 Hz with 11 bit resolution over a 10 mV range and contains approximately 109,000 annotated beats[5], [15]. SVDB comprises 78 two channel recordings digitized at 128 Hz with approximately 30 minutes of recording per subject [16]. Fourteen SVDB records were selected: 801, 849, 851, 852, 854, 855, 860, 865, 868, 870, 878, 879, 881, and 892. These records were chosen because their annotations contain clinically meaningful supraventricular and ventricular beats that are severely underrepresented in MITDB.

The proposed modification strategy inserts real ECG beats from a related public arrhythmia database rather than generating synthetic samples through oversampling. This is motivated by the concern that synthetic beats may not fully capture the physiological morphological variability of real arrhythmia signals. The 14 added SVDB records contributed 4,649 additional S-class beats, 4,530 additional V-class beats, and 47 additional Q-class beats, enabling a 17-class classification scenario that the original MITDB alone cannot support due to only 2 available S-class beats. Table 1 summarizes the dataset components, preprocessing steps, and source characteristics. Class | WFDB Label | MITDB original | After SVDB addition | Change N | Normal | 75,052 | 75,052 | — L | Left bundle branch block | 8,075 | 8,075 | — R | Right bundle branch block | 7,259 | 7,259 | — A | Atrial premature | 2,546 | 2,546 | — V | PVC | 7,130 | 11,660 | +4,530 S | Supraventricular premature | 2 | 4,651 | +4,649 F | Fusion | 803 | 803 | — / | Paced beat | 7,028 | 7,028 | — Q | Unclassifiable | 33 | 80 | +47 Others (a,!,x,j,f,E,J,e) | Various rare | 1,223 | 1,223 | — Total | — | 109,151 | 118,377 | +9,226 Note: "Others" includes classes a, !, x, j, f, E, J, and e. The modification increases total beat count by 8.4% while substantially improving coverage of classes S (+232,450%), V (+63.5%), and Q (+142.4%).

Table 1. Dataset construction and preprocessing summary**.**

Component	Description
Main dataset	MIT-BIH Arrhythmia Database (MITDB) [5]: 48 two-channel ambulatory ECG recordings, 360 Hz, 11-bit resolution over 10 mV range, ~109,000 annotated beats

Component	Description
Supplementary dataset	MIT-BIH Supraventricular Arrhythmia Database (SVDB) [16]: 78 two-channel recordings at 128 Hz, ~30 min each
Selected SVDB records	14 records: 801, 849, 851, 852, 854, 855, 860, 865, 868, 870, 878, 879, 881, 892
Beats added	S class: +4,649; V class: +4,530; Q class: +47
Beat extraction	biosppy.christov_segmenter; R-peak detection with HR bounds 40–180 bpm at 360 Hz; symmetric window centered on each R-peak; mean beat length ~0.77 s (≈277 samples at 360 Hz)
Lead selection	MLII (primary lead) for all MITDB records; lead with highest QRS amplitude selected for SVDB records following WFDB annotation alignment
Normalization	Min-max scaling to [0, 1] per beat (amplitude normalization); original Y-axis range: −200 to 800 ADC units
Resampling	Each variable-length beat resampled to fixed 180 samples using scipy.signal.resample (FFT-based); MITDB (360 Hz): downsample factor 2; SVDB (128 Hz): single-step resample 128→180 Hz (ratio ≈1.406); no padding or truncation; no intermediate 256 Hz stage
External validation	INCARTDB (St. Petersburg INCART 12-lead Arrhythmia Database) [17]; 4 mappable classes (N, V, R, A); same preprocessing pipeline applied

2.2. Beat Extraction and Preprocessing Pipeline

2.2.1. Signal Lead Selection and Beat Segmentation

ECG signals were read from PhysioNet WFDB-formatted records using the wfdb Python library. For MITDB records, the MLII lead (channel 0) was used as the primary channel, consistent with standard practice in beat classification [4]. For SVDB records, the lead with the highest QRS peak amplitude was selected programmatically, and annotation timestamps were aligned to the selected channel using the PhysioNet annotation files.

Beat segmentation was performed using the biosppy.christov_segmenter algorithm[18], [19], which detects R-peak positions from the ECG signal. The segmentation parameters were set to a minimum heart rate of 40 bpm and a maximum of 180 bpm, with the signal sampled at 360 Hz for MITDB and 128 Hz for SVDB. Each beat was extracted using a symmetric window centered on the detected R-peak, defined as shown in Equation 1.

$$Beat\_i = signal\ [\ p\_i - (p\_i - p\_\{i-1\})/2 : p\_i + (p\_\{i+1\} - p\_i)/2\ ] \quad (1)$$

where p_i is the R-peak position of beat i. Each detected R-peak was matched to the nearest PhysioNet WFDB annotation within a ±150 ms tolerance window using wfdb.rdann, beats with no annotation within this window were discarded. Beats annotated with artifact labels (| in MITDB) were also excluded. The resulting beat lengths were variable, with a mean duration of approximately 0.77 seconds per beat (approximately 277 samples at 360 Hz), reflecting the natural variability of heart rate across records.

2.2.2. Amplitude Normalization

Each extracted beat was normalized independently using min-max scaling to bring amplitude values to the range [0, 1]. The original ADC amplitude range in MITDB spans approximately −200 to +800 units. Normalization was applied after beat segmentation and before resampling, so that each beat's amplitude is comparable regardless of recording gain or baseline drift. This approach was selected over mean-centering or z-score normalization because it produces a fixed output range that is suitable for CNN input without requiring knowledge of the global signal statistics.

2.2.3. Resampling Strategy for Frequency Harmonization

MITDB and SVDB signals were recorded at different sampling frequencies (360 Hz and 128 Hz, respectively), producing morphological shape differences when beats from both databases are mixed. To resolve this, each variable-length beat segment was resampled to a fixed length of 180 samples using scipy.signal.resample (an FFT-based method); no padding or truncation was applied. For MITDB beats (360 Hz), this corresponds to downsampling by a factor of 2 to an effective rate of 180 Hz. For SVDB beats (128 Hz), beats were resampled directly from 128 Hz to 180 Hz in a single step (ratio ≈1.406) with no intermediate 256 Hz stage. Because scipy.signal.resample uses an FFT-based method

that implicitly band-limits the signal to the new Nyquist frequency, no additional anti-aliasing filter was required. The 180-sample fixed-length vector forms the input to the 1D-CNN.

The rationale for choosing 180 Hz as the common target rather than a lower frequency such as 128 Hz or a higher frequency such as 360 Hz is as follows. Downsampling MITDB to 128 Hz would discard more temporal resolution than necessary; upsampling SVDB to 360 Hz would artificially inflate SVDB beat lengths without adding real information. A midpoint rate of 180 Hz (achieved by halving the 360 Hz MITDB rate) preserves substantially more morphological detail than 128 Hz while requiring only a simple integer decimation step for MITDB, reducing the risk of interpolation artifacts. The impact of resampling on classification performance was empirically validated: the best accuracy before resampling was 95.87%, which increased to 99.38% after frequency harmonization.

2.3. Classification Schemes

Two distinct classification schemes were implemented to address both fine-grained arrhythmia discrimination and clinically oriented grouping. The first scheme is class-oriented classification, in which individual WFDB beat annotations are used directly as class labels, resulting in 16-class (MITDB only) and 17-class (MITDB+SVDB) experiments. The second scheme follows the AAMI EC57 standard[20] [21], [22], which groups beat labels into five functional categories: N, S, V, F, and Q. The AAMI EC57-grouped model was trained directly on these grouped labels as the target output; it was not derived by post-hoc mapping from the fine-grained model output. This design allows evaluation at both the detailed beat-label level and the AAMI EC57 grouped level. Patient-independent evaluation, in which all beats from a given patient are restricted to either training or testing, was not implemented and is identified as the primary direction for future work. Table 2 specifies the class labels and groupings for each scheme.

Table 2. Classification schemes and class definitions.

Scheme	Classes	Labels / Grouping
Class-oriented	16	N, L, R, A, V, /, a, !, F, x, j, f, E, J, e, Q

Scheme	Classes	Labels / Grouping
Class-oriented	17	N, L, R, A, V, /, a, !, F, x, j, f, E, J, e, Q, S (SVDB supraventricular beats)
AAMI 5-class	5	N (N, L, R, e, j) | S (S, A, a, J) | V (V, E) | F (F) | Q (/, Q, f)

Three evaluation terms are used consistently throughout this paper. Class-oriented refers to classification using individual WFDB beat-label annotations. AAMI EC57-grouped (or 5-class AAMI) refers to the EC57 standard beat grouping. Patient-independent refers to a record-wise evaluation protocol in which all beats from a given patient are restricted exclusively to either training or testing this is the clinically recommended standard, and it is explicitly not implemented in the current study, all results use beat-level random splitting as described in Section 2.5.

2.4. 1D-CNN Architecture

The proposed classifier uses a compact one-dimensional convolutional neural network (1D-CNN) implemented in PyTorch[23]. The architecture was kept deliberately compact to demonstrate that a simple structure, when paired with a carefully constructed dataset, can achieve strong multiclass performance without requiring deep or complex network designs. The architecture consists of two convolutional blocks followed by fully connected layers. Table 3 provides the complete architectural specification.

Table 3. 1D-CNN architecture specification**.**

Layer	Configuration	Notes
Input	Length = 180 samples	Unified beat representation at 180 Hz (~1 s window)
Conv1D Block 1	32 filters, kernel size 5, stride 1	ReLU activation; MaxPooling1D size 2
Conv1D Block 2	64 filters, kernel size 5, stride 1	ReLU activation; MaxPooling1D size 2

Layer	Configuration	Notes
Dense layers	2–4 fully connected layers (tuned)	ReLU activation; dropout optionally applied (0.1–0.7)
Output layer	Softmax	Units = number of target classes (5, 16, or 17)

Each convolutional block applies a 1D convolution over the beat input, followed by ReLU activation and MaxPooling1D with pool size 2. After the two convolutional blocks, the feature maps are flattened and passed to 2–4 fully connected (dense) layers with ReLU activation. Dropout regularization was optionally inserted between dense layers. The output layer uses Softmax activation with units equal to the number of target classes (5, 16, or 17 depending on the classification scenario). The initial baseline model used the Adam optimizer with learning rate 0.001, batch size 512, and 100 epochs, with a 70:10:20 train-validation-test split, achieving a baseline accuracy of 96.18% for 16-class and 98.69% for 5-class AAMI classification before hyperparameter tuning.

2.5. Hyperparameter Tuning and Evaluation Protocol

Systematic hyperparameter tuning was performed to identify the best-performing configuration for each classification scenario. Table 4 lists all parameters and the values explored.

Table 4. Hyperparameter search space.

Parameter	Values Explored
Optimizer	Adam, RMSprop, SGD, ASGD
Learning rate	0.1, 0.01, 0.005, 0.001
Batch size	64, 128, 256, 320, 512
Epochs	30, 50, 100, 150, 200
Dense layer count	2, 3, 4
Dropout count	0, 1, 2

Parameter	Values Explored
Dropout value	0.1, 0.5, 0.7
Data split	70:10:20 (train:val:test); beats randomly sampled per class; split is beat level within unified record pool

The data split was 70% training, 10% validation, and 20% testing, applied at the beat level across the unified MITDB+SVDB pool. Beat assignment to training or testing was randomized at the beat level, meaning that beats from the same ECG record may appear in both training and test sets. This is a standard class-oriented evaluation protocol and is consistent with the beat-level evaluation used in the reference studies cited in Results section. The authors acknowledge that this protocol does not enforce patient-independent (record-wise) separation, and therefore the reported accuracies should be interpreted accordingly. A record-wise patient-independent evaluation is identified as a priority for future work. Model selection was based on validation accuracy, and all reported test results are from the held out 20% test set only.

2.6. External Cross-Database Validation with INCARTDB

To assess generalization beyond the training database, the best-performing 1D-CNN configuration (ASGD optimizer) was tested on the St. Petersburg INCART 12-lead Arrhythmia Database (INCARTDB) [17]. INCARTDB contains 75 recordings of 30-minute 12-lead ECGs from patients with various arrhythmia conditions. It is entirely independent of the MITDB and SVDB sources used during training.

INCARTDB was preprocessed using the identical pipeline applied to MITDB and SVDB: the same lead selection logic, biosppy.christov_segmenter for R-peak detection, symmetric beat windowing, min-max normalization to [0, 1], and resampling to 180 Hz. After preprocessing, only four INCARTDB beat categories contained sufficient samples for reliable evaluation: N (28,455 samples), V (3,619 samples), R (634 samples), and A (381 samples). All other INCARTDB annotation types either lacked a direct MITDB label equivalent or had fewer than 50 samples and were excluded from this evaluation. The class label mapping from INCARTDB annotations to the MITDB label space is specified in Table 5.

Table 5. INCARTDB class mapping for cross-database validation.

INCARTDB Label	Beat Type	Mapped Class (MITDB-SVDB)
N	Normal sinus beat	N — 28,455 samples
V	Premature ventricular contraction	V — 3,619 samples
R	Right bundle branch block	R — 634 samples
A	Atrial premature beat	A — 381 samples
Others*	Labels with < 50 samples or no MITDB equivalent	Excluded from evaluation

This cross-database evaluation therefore represents a 4-class scenario, not the full 5-class AAMI or 17-class setting used in the main experiments. The result of 99.13% accuracy should be interpreted in this 4-class context and not generalized directly to all AAMI classes or to a broader domain. The evaluation demonstrates the model's ability to handle domain shift between different ECG acquisition hardware and patient populations, but broader validation on additional external databases, wearable-device ECG, and clinical monitoring data remains necessary to establish deployment readiness.

2.7. Implementation

All experiments were implemented in Python using PyTorch as the deep learning framework. ECG data were accessed through the wfdb library from PhysioNet. Beat segmentation used biosppy version 0.6.1. Resampling was implemented using scipy.signal. Experiments were executed on a standard desktop CPU environment. Random seeds were fixed for reproducibility.

3. RESULTS AND DISCUSSION

This section presents and discusses the classification results across all experimental scenarios: 16-class, 17-class, and 5-class AAMI classification on the modified MITDB+SVDB dataset, benchmark comparison against machine learning and LSTM models, and cross-database validation on INCARTDB. Per-class results are emphasized throughout,

particularly for rare and minority classes, to provide evidence beyond aggregate accuracy metrics. Limitations of the evaluation protocol are explicitly acknowledged where relevant.

3.1. Per-Class Performance in 16-Class Classification

Important caveat: all accuracy values in Sections 3.1–3.7 are based on a 70:10:20 beat-level random split. Because beats from the same patient may appear in both training and test sets, these results may be optimistic relative to a strict patient-independent protocol. No deployment readiness is implied by any of these figures. The best 16-class result was achieved by the ASGD optimizer with 3 dense layers, 1 dropout layer (value 0.1), learning rate 0.01, batch size 32, and 200 epochs, yielding 99.10% overall accuracy. Table 6 reports the full per-class evaluation.

Table 6. Per-class evaluation results for 16-class classification (ASGD optimizer, best configuration).

No.	Class	Precision	Sensitivity	F1-Score	Support
1	N	0.99	0.99	0.99	16,894
2	L	0.99	0.99	0.99	1,905
3	R	1.00	0.99	0.99	1,461
4	A	0.91	0.94	0.92	567
5	V	0.98	0.98	0.98	1,947
6	/	1.00	1.00	1.00	1,956
7	a	0.82	0.99	0.97	30
8	!	0.87	1.00	0.93	34
9	F	0.91	0.90	0.91	242
10	x	0.79	0.92	0.85	37
11	j	0.80	0.80	0.80	45
12	f	0.97	0.99	0.98	462

No.	Class	Precision	Sensitivity	F1-Score	Support
13	E	1.00	0.93	0.96	42
14	J	0.69	0.61	0.65	18
15	e	1.00	1.00	1.00	12
16	Q	0.64	0.69	0.67	13
	Accuracy			0.99	25,665
	Macro avg	0.90	0.91	0.90	25,665
	Weighted avg	0.99	0.99	0.99	25,665

Note: Support values indicate the number of test samples per class (20% hold-out). Classes with support < 50 are considered very rare in the test set.

The weighted averages of precision, sensitivity, and F1-score all reach 0.99, reflecting the strong performance on majority classes. However, the macro averages (precision 0.90, sensitivity 0.91, F1 0.90) reveal meaningful variation across classes, and four classes warrant particular attention: Classes J and Q are the weakest, with F1-scores of 0.65 and 0.67 respectively. Both have very small test support (18 and 13 samples), making evaluation inherently noisy. Class J (nodal junctional premature beat) is frequently confused with class N, which is consistent with its morphological similarity to normal sinus rhythm at the beat level. Class Q (unclassifiable beat) represents ambiguous beats that even human annotators find difficult to assign, its limited discriminability is therefore not unexpected. Class j (nodal junctional escape beat, n=45) achieves F1 0.80, showing modest but not strong performance despite higher support than J. The confusion matrix indicates that j is frequently misclassified as N (approximately 20% of j beats predicted as N in the 16-class experiment). Classes a (aberrated atrial premature, n=30) and x (non-conducted P-wave, n=37) have F1-scores of 0.97 and 0.85 respectively. Despite low support, the model achieves high sensitivity for class a (0.99), which is practically relevant since missed atrial premature beats carry clinical consequences. Classes e (atrial escape, n=12) and / (paced beat, n=1,956) both achieve perfect F1 of 1.00. The high performance in class e despite

only 12 test samples, is noteworthy but should be interpreted with caution given the very small sample size.

3.2. Per-Class Performance in 17-Class Classification

The 17-class experiment adds class S (supraventricular premature beats from SVDB, n=4,649 additional training samples) to the 16-class set. The best result uses ASGD with 2 dense layers, no dropout, learning rate 0.01, batch size 32, and 200 epochs, achieving 98.58% accuracy. Table 7 provides the complete per-class results.

Table 7. Per-class evaluation results for 17-class classification (ASGD optimizer, best configuration).

No.	Class	Precision	Sensitivity	F1-Score	Support
1	N	1.00	0.99	0.99	16,575
2	L	0.99	1.00	1.00	1,905
3	R	0.99	1.00	1.00	1,443
4	A	0.85	0.96	0.90	556
5	V	0.98	0.97	0.98	1,750
6	/	1.00	1.00	1.00	1,956
7	a	0.75	0.70	0.72	30
8	!	0.89	0.94	0.91	34
9	F	0.77	0.90	0.83	168
10	x	0.77	1.00	0.87	37
11	j	0.56	0.76	0.64	45
12	f	0.98	0.98	0.98	462
13	E	0.97	0.95	0.96	41
14	J	0.59	0.89	0.71	18
15	e	0.67	1.00	0.80	12

No.	Class	Precision	Sensitivity	F1-Score	Support
16	Q	0.89	0.62	0.73	13
17	S	0.81	0.87	0.84	15
	Accuracy			0.99	25,060
	Macro avg	0.85	0.91	0.87	25,060
	Weighted avg	0.99	0.99	0.99	25,060

Note: The macro F1-score of 0.87 (vs. 0.90 in 16-class) reflects the additional difficulty of classifying class S and the extremely rare classes j, J, and e.

Class S (n=15 in test set despite 4,649 added training samples) achieves precision 0.81, sensitivity 0.87, and F1 0.84. The low test-set support for S reflects the class distribution after random 80:20 splitting. The result demonstrates that the real-record SVDB insertion enables the model to distinguish supraventricular beats to a useful degree, which is entirely infeasible with the original MITDB alone (which contains only 2 S-class beats). However, the small test support limits statistical confidence in the reported metrics. Class j degrades from F1 0.80 (16-class) to 0.64 (17-class). The addition of class S introduces a morphologically adjacent category, increasing confusion between j and S. This suggests that even though the SVDB addition improves S-class detectability, it increases inter-class ambiguity for rare classes with similar morphology. Class F (fusion beat, n=168) drops from F1 0.91 to 0.83 in the 17-class experiment. This is attributable to increased class interference in the more crowded 17-class label space, rather than to data availability (F has sufficient support in both scenarios). Majority classes (N, L, R, V, /, f) remain at or above F1 0.98 in both 16-class and 17-class scenarios, confirming that the dataset modification does not harm performance on well-represented classes.

3.3. Per-Class Performance in 5-Class AAMI Classification

The AAMI 5-class grouping maps all 17 individual beat labels to 5 functional categories: N, S, V, F, Q. The best result uses ASGD with 3 dense layers, 1 dropout (0.1), learning rate

0.01, batch size 32, and 200 epochs, achieving 99.38% accuracy higher than the 16-class and 17-class results. Table 8 presents the per-class evaluation.

Table 8. Per-class evaluation results for 5-class AAMI classification (ASGD optimizer).

No.	Class	Precision	Sensitivity	F1-Score	Support
1	N	1.00	1.00	1.00	9,933
2	S	0.94	0.93	0.93	271
3	V	0.98	0.98	0.98	940
4	F	0.94	0.93	0.94	118
5	Q	1.00	1.00	1.00	1,214
	Accuracy			0.9938	12,476
	Macro avg	0.97	0.97	0.97	12,476
	Weighted avg	0.99	0.99	0.99	12,476

All five AAMI classes achieve F1 ≥ 0.93, with classes N and Q reaching perfect 1.00. The macro precision, sensitivity, and F1 all equal 0.97, indicating balanced performance across clinically meaningful groupings. This is a key finding: individual rare beat types that are difficult to classify in isolation (e.g., j and J, which are problematic in multiclass scenarios) are successfully absorbed into the N and S AAMI groups where their morphological similarity to those groups is appropriate by clinical definition[15], [24].

The class S (AAMI) achieves F1 0.93 with 271 test samples, which is substantially more robust than the individual S-class result in 17-class classification (F1 0.84, n=15). This difference arises because AAMI S groups A, a, J, and S beats, all supraventricular prematures creating a more statistically stable category with higher test support. The F class (fusion beats, n=118) achieves F1 0.94 in the AAMI grouping, a significant improvement from its 17-class result (F1 0.83, n=168), because the AAMI grouping consolidates F as a single clean category without competing with adjacent beat types.

The finding that the 5-class AAMI accuracy (99.38%) exceeds the 17-class accuracy (98.58%) is consistent with results from prior studies and can be explained by the reduction of inter-class ambiguity: beats that are morphologically confusable across fine-grained labels (j vs. N, J vs. S) become correctly assigned when grouped under their appropriate AAMI category.

3.4. Consolidated Minority-Class Analysis

To directly address the concern raised by reviewers regarding rare-class evidence, Table 9 consolidates the performance of all classes with fewer than 50 test samples across the three classification scenarios, as these are the classes that most critically test the value of the SVDB dataset modification.

Table 9. Minority-class performance summary across all classification scenarios (support < 50 or clinically critical).

Class	Beat Type	Support (test)	Precision	Sensitivity	F1	Scenario
J	Nodal premature	18	0.69	0.61	**0.65**	16-class
Q	Unclassifiable	13	0.64	0.69	**0.67**	16-class
j	Junctional escape	45	0.80	0.80	0.80	16-class
e	Atrial escape	12	1.00	1.00	**1.00**	16-class
j	Junctional escape	45	0.56	0.76	**0.64**	17-class
J	Nodal premature	18	0.59	0.89	**0.71**	17-class
S	Supraventricular	15	0.81	0.87	0.84	17-class
F	Fusion	118	0.94	0.93	**0.94**	5-class
S	Supraventricular	271	0.94	0.93	**0.93**	5-class

Red = F1 < 0.70 (weak); Green = F1 ≥ 0.93 (strong). The F and S results in the 5-class row reflect AAMI grouping, which substantially improves performance by pooling morphologically related classes.

The consolidated view reveals a clear pattern: classes with fewer than 20 test samples (J: n=18, Q: n=13, e: n=12, S: n=15) produce the most variable and least reliable metrics. This

is an inherent limitation of the dataset, not a model failure even a perfect classifier would show high variance in F1 scores over 12–18 test samples. The practical implication is that the AAMI grouping (which pools these rare classes into larger, more reliable categories) is the appropriate evaluation framework for deployment-oriented performance claims.

3.5. Comparison with Machine Learning and LSTM Baselines

Table 10 compares the proposed 1D-CNN against KNN, SVM, logistic regression, decision tree, and LSTM using the same modified MITDB+SVDB dataset. This comparison demonstrates the added value of the deep learning model beyond the dataset modification itself.

Table 10. Model comparison using the modified MITDB+SVDB dataset across all classification scenarios.

Model	5-class Acc.	16-class Acc.	17-class Acc.	Observation
Proposed 1D-CNN (ASGD)	**99.38%**	**99.10%**	**98.58%**	Best overall; macro F1 ≥ 0.87 across all scenarios
KNN	96.62%	95.62%	94.73%	Strongest classical baseline; cannot classify class J in 16-class
Decision Tree	96.64%	94.02%	93.52%	Training acc. 100% (overfitting risk); very weak on J and j
SVM	94.12%	90.02%	90.61%	Cannot classify 6 classes in 16-class (J, Q, a, e, j, x)
Logistic Regression	87.56%	75.77%	75.64%	Cannot classify 7 classes in 16-class; class S = F1 ≈ 0 in 5-class
LSTM	97.53%	83.14%	N/A	Cannot classify 6 classes in 16-class; unstable on fine-grained labels

The 1D-CNN consistently outperforms all baselines across all three scenarios. The most important differences emerge in the fine-grained multiclass settings. KNN achieves 95.62% in 16-class classification but completely fails to classify class J (F1 = 0.00), which the 1D-CNN classifies at F1 0.65. SVM fails on 6 out of 16 classes, and logistic regression fails on 7. The LSTM model achieves only 83.14% in 16-class classification, with 6 classes entirely unclassified, indicating that recurrent architectures without careful sequence design are less suitable for this segmented beat classification task than the convolutional approach.

The decision tree achieves training accuracy of 100% across all scenarios, a clear signal of overfitting while its test accuracy drops to 93.52–96.64%. In contrast, the 1D-CNN shows minimal train-test accuracy gap, suggesting better generalization without overfitting. This finding supports the value of the dropout regularization incorporated in the ASGD-optimized configurations. An important caveat for interpreting these benchmarks: all models in Table 10 use the same modified dataset and the same beat-level random split. Therefore, the comparison isolates model architecture differences rather than data availability differences. The 1D-CNN's advantage is attributable to its ability to learn hierarchical morphological features from raw beat signals, which simpler classifiers cannot replicate from the same input representation.

3.6. Comparison with Prior Studies — with Caveats

The most distinctive finding in Table 11 is the positive relationship between class-oriented and AAMI-oriented accuracy in the proposed approach. Prior studies consistently show a gap, sometimes very large between fine-grained multiclass accuracy and AAMI accuracy. Raj and Ray [14] report 99.11% multiclass accuracy but only 89.93% AAMI accuracy, a gap of over 9 percentage points. Shi et al. [13] show a similar pattern (99.29% vs. 94.20%). The proposed model, by contrast, achieves 98.58% in 17-class classification and 99.38% in AAMI classification, with the AAMI result exceeding the multiclass result. However, this comparison must be interpreted carefully. Three factors limit the directness of the numerical comparison. First, the proposed model benefits from additional SVDB records that prior studies did not use, which provides more training data for underrepresented classes and may inflate performance relative to MITDB-only baselines. Second, the evaluation protocol in this study uses beat-level random splitting, which is consistent with the class-oriented protocols used in the cited studies, but stricter record-wise

patient-independent evaluation may yield different results. Third, the number of classes differs: the proposed 17-class model includes class S (enabled by SVDB), which prior MITDB-only studies cannot evaluate. Despite these caveats, the directional finding that real-record dataset enrichment can support simultaneously strong multiclass and AAMI performance is novel and practically relevant.

Table 11. Comparison of class-oriented and AAMI-oriented accuracy against prior studies.

Study	Year	Classes	Dataset	Class-oriented Acc.	AAMI Acc.	Note
Luo et al. [12]	2021	9	MITDB	99.01%	98.70%	SMOTE used
Raj and Ray [14]	2018	16	MITDB	99.11%	89.93%	MITDB only
Shi et al. [13]	2020	15	MITDB	99.29%	94.20%	Multi-input DNN
Can Ye et al. [25]	2012	16	MITDB	99.30%	86.40%	MITDB only
Proposed (ASGD)	2024	17	MITDB+SVDB	**98.58%**	**99.38%**	Real augmentation

Caveat: The proposed model uses a modified MITDB+SVDB dataset, while prior studies use MITDB alone. This limits the fairness of direct numerical comparison and is acknowledged explicitly below.

3.7. Cross-Database Validation on INCARTDB

The best-performing 1D-CNN configuration (ASGD, 3 dense layers, no dropout, learning rate 0.01, batch size 64, 120 epochs) was evaluated on the INCARTDB database, which was not used in any part of training or model selection. As described in Section 2.6, only four beat classes were available for evaluation after preprocessing: N, V, R, and A. This is

therefore a 4-class cross-database evaluation, not a full 5-class or 17-class test, and the result should be interpreted accordingly. Table 12 presents the per-class results.

Table 12. Per-class results for cross-database validation on INCARTDB (4-class, unseen data).

Class	Beat Type	Support	Precision	Sensitivity	F1	Note
N	Normal sinus beat	28,455	1.00	0.99	1.00	—
V	PVC	3,619	0.96	0.98	0.97	—
R	Right bundle branch block	634	0.99	1.00	0.99	—
A	Atrial premature beat	381	0.85	0.93	0.89	Weakest; misclassified as N (5%) and V (2%)
Overall	4-class evaluation	33,089	0.95*	0.98*	0.96*	*macro avg

The 4-class constraint arises from INCARTDB annotation availability after preprocessing, not from model design. Other INCARTDB annotation types had fewer than 50 samples or no direct MITDB label equivalent and were excluded.

The model achieves 99.13% overall accuracy on INCARTDB despite never having been trained on this database. Classes N and R achieve near-perfect performance (F1 1.00 and 0.99 respectively). Class V (PVC) achieves F1 0.97, demonstrating robust generalization of ventricular beat morphology across recording equipment and patient populations. Class A (atrial premature beat) is the weakest result, with precision 0.85 and F1 0.89. The confusion matrix shows that approximately 5% of A beats are misclassified as N and 2% as V misclassifications that are clinically plausible given the morphological overlap between atrial premature and normal beats when observed in isolation.

Two important limitations of the cross-database result should be acknowledged. First, the INCARTDB test set is heavily dominated by class N (28,455 of 33,089 samples, or 86%), which means the high overall accuracy is strongly influenced by N-class performance. Per-class metrics, particularly for A (n=381) and R (n=634), are more informative for assessing true generalization. Second, the INCARTDB recordings were acquired with a different ECG system than MITDB and SVDB, introducing domain shift in multiple dimensions: (1) lead configuration INCARTDB uses 12 leads while MITDB and SVDB use 2-channel ambulatory recordings; (2) acquisition device characteristics including filtering, amplifier gain, and sampling hardware; and (3) patient population and clinical setting differences. Despite these domain differences, the model generalizes well for morphologically distinctive classes. Validation on additional external databases including wearable-device ECG and clinical monitoring systems remains necessary before deployment claims can be made.

3.8. Synthesis: Answering Research Questions

This study addressed three research questions. First: how can a modified public ECG dataset reduce imbalance effects? The SVDB augmentation strategy increases the S class by 4,649 beats, V by 4,530 beats, and Q by 47 beats, enabling 17-class classification and substantially improving the macro F1 scores of the modified dataset. The comparison between 16-class (MITDB-only classes, macro F1 0.90) and 17-class (including S, macro F1 0.87) results show that adding S increases overall classification difficulty 17 classes introduce more inter-class confusion while simultaneously making the model capable of detecting a clinically important beat type that was entirely unrepresented before. The 5-class AAMI result (macro F1 0.97) confirms that the pooled supraventricular category (AAMI S), enriched by the SVDB beats, is robustly classified when evaluated at the appropriate clinical granularity.

Second: how can a 1D-CNN support 16-class, 17-class, and AAMI EC57-grouped 5-class classification simultaneously? Three separate model instances were trained, one for each classification scheme. The 16-class and 17-class models used individual WFDB beat labels; the 5-class AAMI model was trained directly on EC57-grouped labels. The shared architecture and preprocessing pipeline mean that dataset construction and preprocessing are performed once, and the appropriate model variant is selected for the desired classification granularity.

Third: how well does the model generalize to unseen data? The INCARTDB cross-database result (99.13%, 4-class) provides preliminary evidence of generalization across ECG databases. The per-class results show strong generalization for morphologically distinctive classes (N, R, V) and weaker generalization for morphologically ambiguous classes (A), which is consistent with expectations. The result should not be extrapolated to the full AAMI 5-class setting, as classes S, F, and Q were not present in the INCARTDB evaluation.

Regarding the evaluation protocol, it is important to acknowledge that all main results in this study use beat-level random splitting without enforcing record-wise patient separation. Under a strict patient-independent protocol where beats from the same patient are restricted to either training or testing results may differ, and this difference could be substantial for classes where individual patient morphology varies significantly. This is the most important limitation of the current evaluation and is a priority for future validation.

3.9. Discussion

The proposed real-record enrichment strategy differs fundamentally from synthetic oversampling approaches (SMOTE, GAN-based augmentation) by inserting authentic physiological ECG beats from a related public database. Real beats from SVDB preserve the natural morphological variability across different patients and recording conditions, whereas synthetic beats are generated from statistical distributions that may not capture the full physiological diversity of rare arrhythmia classes. The fixed-length resampling step (scipy.signal.resample to 180 samples) resolves the frequency mismatch between MITDB (360 Hz) and SVDB (128 Hz), enabling the 1D-CNN to learn morphologically comparable beat representations from both sources. The key limitation of this study is the use of beat-level random splitting, which allows beats from the same patient to appear in both training and test sets. This protocol can produce optimistically high accuracy because the model may partially learn patient-specific morphological patterns rather than purely class-discriminative features. A strict patient-independent (record-wise) evaluation where all beats from a given patient are restricted exclusively to training or testing is the clinically recommended standard and is expected to yield lower but more realistic performance estimates. This distinction is critical for clinical translation: any deployed system will encounter entirely new patients not seen during training. The

INCARTDB cross-database result (99.13%, 4-class) provides encouraging preliminary evidence of generalization despite significant domain shift: INCARTDB uses 12-lead recordings vs. 2-channel ambulatory recordings in MITDB/SVDB, different acquisition hardware, and a distinct patient population. Strong generalization for morphologically distinctive classes (N, R, V) and weaker performance for the morphologically ambiguous class A (F1=0.89) is consistent with the expected difficulty of cross-database transfer. No confidence intervals were computed for rare-class metrics; results for classes with fewer than 20 test samples are statistically indicative only. Future directions for this work include: (1) strict patient-independent re-evaluation using record-wise splitting on the MITDB+SVDB dataset; (2) cross-database validation on additional public ECG databases including CPSC 2018 and PTB-XL to assess broader generalizability; (3) expansion of rare-class training coverage by incorporating additional SVDB records or curated clinical ECG collections; (4) evaluation on wearable-device and short-duration ECG signals typical of ambulatory monitoring; and (5) architectural exploration comparing the current compact 1D-CNN with attention-based and multi-lead input models to isolate the contribution of dataset enrichment from architecture effects.

4. CONCLUSION

This study proposed a real-record dataset enrichment strategy for multiclass ECG arrhythmia classification by incorporating fourteen SVDB records into MITDB, enabling 17-class classification and improving minority-class coverage without synthetic oversampling. Three separate 1D-CNN models were trained: one for 16-class, one for 17-class, and one for AAMI EC57-grouped 5-class classification, all sharing the same architecture and preprocessing pipeline. The ASGD-optimized configurations achieved 99.10%, 98.58%, and 99.38% accuracy respectively, with macro F1-scores of 0.90, 0.87, and 0.97. Cross-database evaluation on INCARTDB yielded 99.13% accuracy across four mappable classes (N, V, R, A); this result is restricted to that 4-class setting and does not generalize to the full AAMI scheme. All results are based on a beat-level random split, not a patient-independent evaluation protocol, and no clinical deployment readiness is claimed. The proposed approach is a promising and reproducible real-record enrichment strategy, but patient-independent validation and broader external database testing are required before clinical deployment can be considered.

ACKNOWLEDGMENT

The authors acknowledge Universitas Bina Insan for academic support. The public ECG databases used in this study were obtained from PhysioNet.

REFERENCES

[1] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, p. 107398, Apr. 2021, doi: 10.1016/j.ymssp.2020.107398.

[2] K. Balakrishnan, D. Velusamy, K. Ramasamy, and L. Pruinelli, "ECG-based cardiac arrhythmia classification using fuzzy encoded features and deep neural networks," *Biomed. Eng. Adv.*, vol. 9, p. 100167, Jun. 2025, doi: 10.1016/j.bea.2025.100167.

[3] M. Llamedo and J. P. Martinez, "An automatic patient-adapted ECG heartbeat classifier allowing expert assistance," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2312–2320, Aug. 2012, doi: 10.1109/TBME.2012.2202662.

[4] "MIT-BIH Arrhythmia Database v1.0.0." Accessed: May 15, 2026. [Online]. Available: https://physionet.org/content/mitdb/1.0.0/

[5] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag. Q. Mag. Eng. Med. Biol. Soc.*, vol. 20, pp. 45–50, Jun. 2001, doi: 10.1109/51.932724.

[6] M. D. Mazhar Qureshi *et al.*, "Multiclass Heartbeat Classification using ECG Signals and Convolutional Neural Networks," in *2022 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, May 2022, pp. 1–6. doi: 10.1109/ICoDT255437.2022.9787419.

[7] A. W. Islam, P. Motdhare, S. Mohan, S. Bejoy, and H. S. Wilkhoo, "Electrocardiogram abnormalities in patients with acne undergoing isotretinoin therapy: a focused review," *An. Bras. Dermatol.*, vol. 101, no. 2, p. 501304, Mar. 2026, doi: 10.1016/j.abd.2026.501304.

[8] B. Zhao, Z. Gao, X. Liu, Z. Zhang, W. Xiao, and S. Zhang, "DRL-ECG-HF: Deep reinforcement learning for enhanced automated diagnosis of heart failure with imbalanced ECG data," *Biomed. Signal Process. Control*, vol. 107, p. 107680, Sep. 2025, doi: 10.1016/j.bspc.2025.107680.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[9] Y. Xu, S. Zhang, and W. Xiao, "Inter-patient ECG classification with intra-class coherence based weighted kernel extreme learning machine," *Expert Syst. Appl.*, vol. 227, p. 120095, Oct. 2023, doi: 10.1016/j.eswa.2023.120095.

[10] M. Guhdar, A. O. Mohammed, and R. J. Mstafa, "Advanced deep learning framework for ECG arrhythmia classification using 1D-CNN with attention mechanism," *Knowl.-Based Syst.*, vol. 315, p. 113301, Apr. 2025, doi: 10.1016/j.knosys.2025.113301.

[11] J. Huang, B. Chen, B. Yao, and W. He, "ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network," *IEEE Access*, vol. 7, pp. 92871–92880, 2019, doi: 10.1109/ACCESS.2019.2928017.

[12] X. Luo, L. Yang, H. Cai, R. Tang, Y. Chen, and W. Li, "Multi-classification of arrhythmias using a HCRNet on imbalanced ECG datasets," *Comput. Methods Programs Biomed.*, vol. 208, p. 106258, Sep. 2021, doi: 10.1016/j.cmpb.2021.106258.

[13] H. Shi, C. Qin, D. Xiao, L. Zhao, and C. Liu, "Automated heartbeat classification based on deep neural network with multiple input layers," *Knowl.-Based Syst.*, vol. 188, p. 105036, Jan. 2020, doi: 10.1016/j.knosys.2019.105036.

[14] S. Raj and K. C. Ray, "Sparse representation of ECG signals for automated recognition of cardiac arrhythmias," *Expert Syst. Appl.*, vol. 105, pp. 49–64, Sep. 2018, doi: 10.1016/j.eswa.2018.03.038.

[15] P. de Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004, doi: 10.1109/TBME.2004.827359.

[16] S. D. Greenwald, "The MIT-BIH Supraventricular Arrhythmia Database." physionet.org, 1992. doi: 10.13026/C2V30W.

[17] V. Tihonenko, A. Khaustov, S. Ivanov, and A. Rivin, "St.-Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database." physionet.org, 2007. doi: 10.13026/C2V88N.

[18] M. P. Tarvainen, P. O. Ranta-aho, and P. A. Karjalainen, "An advanced detrending method with application to HRV analysis," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 2, pp. 172–175, Feb. 2002, doi: 10.1109/10.979357.

[19] K. Lee, J. Lee, and M. Shin, "Lightweight beat score map method for electrocardiogram-based arrhythmia classification," *Biocybern. Biomed. Eng.*, vol. 44, no. 4, pp. 844–857, Oct. 2024, doi: 10.1016/j.bbe.2024.11.002.

[20] "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms," in *ANSI/AAMI EC57:2012/(R)2020; Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms*, AAMI, 2013. doi: 10.2345/9781570204784.ch1.

[21] "ANSI/AAMI EC57:2012/(R)2020; Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms," Default Book Series. Accessed: May 15, 2026. [Online]. Available: https://array.aami.org/doi/book/10.2345/9781570204784

[22] S. Mousavi and F. Afghah, "Inter- and intra-patient ECG heartbeat classification for arrhythmia detection: A sequence to sequence deep learning approach," *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP Conf.*, vol. 2019, pp. 1308–1312, May 2019, doi: 10.1109/icassp.2019.8683140.

[23] A. Hernandez *et al.*, "Pytorch-Wildlife: A Collaborative Deep Learning Framework for Conservation," Nov. 29, 2024, *arXiv*: arXiv:2405.12930. doi: 10.48550/arXiv.2405.12930.

[24] T. Mar, S. Zaunseder, J. P. Martínez, M. Llamedo, and R. Poll, "Optimization of ECG classification by means of feature selection," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 8, Aug. 2011, doi: 10.1109/TBME.2011.2113395.

[25] C. Ye, B. V. K. Vijaya Kumar, and M. T. Coimbra, "Heartbeat Classification Using Morphological and Dynamic Features of ECG Signals," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2930–2941, Oct. 2012, doi: 10.1109/TBME.2012.2213253.