FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 3 | Pages 796-806 | e-ISSN: 2347-470X

Morphological and RR-Interval Features for ECG Arrhythmia Classification: A Comparative Study of Intra-Patient and Inter-Patient Evaluation

Shams M. Qahtan[1], and Hadeel N. Abdullah[2*]

[1]*Department of Electrical Engineering, University of Technology, Iraq; eng7shams@gmail.com*
[2]*Department of Electrical Engineering, University of Technology, Iraq; hadeel.n.abdullah@uotechnology.edu.iq*

***Correspondence:** *Hadeel N. Abdullah; hadeel.n.abdullah@uotechnology.edu.iq*

ABSTRACT- Heartbeat classification on the MIT-BIH Arrhythmia Database is commonly evaluated with beat-level cross-validation, which places beats from the same subject in both training and test sets. In this work, we evaluate a 61-dimensional heartbeat representation under both that protocol and record-level separation, and compare the two. Each beat is described by 60 signal samples spanning 166.7 ms around the R-peak and one RR interval. Under beat-level five-fold cross-validation, a Random Forest reaches 98.42% accuracy and 95.44% macro-F1 across five classes. Under record-level separation, with the configuration selected entirely within DS1 and DS2 withheld from model selection, the same representation reaches 80.80% accuracy and 42.00% macro-F1, with sensitivity of 0.9864 for Normal, 0.9139 for PVC, 0.2004 for RBBB, 0.0002 for LBBB and 0.0023 for APC. We then separate the causes of these failures experimentally. RBBB sensitivity rises to 0.9178 on one held-out subject when the number of training subjects carrying the class is increased, whereas LBBB sensitivity remains below 0.001 under every configuration and feature subset tested. Across the five classifiers evaluated under DS1/DS2, LBBB remains poorly recognized and RBBB sensitivity remains below 0.40, whereas APC performance varies substantially, showing that the poor transfer is not specific to the Random Forest and that the weak classes are affected by different factors. The comparison illustrates the substantial difference between beat-level and record-level evaluation under the standard DS1/DS2 setting, and shows that per-class failures under record-level separation arise from different causes and require different remedies.

Keywords: ECG, Inter-Patient Evaluation, Arrhythmia Classification, Random Forest, RR Interval.

ARTICLE INFORMATION
Author(s): Shams M. Qahtan, and Hadeel N. Abdullah;
Received: 31/05/26; **Accepted**: 03/09/26; **Published**: 30/09/26;
E- ISSN: 2347-470X;
Paper Id: IJEER 3105B22;
Citation: 10.37391/ijeer.140313
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-14/ijeer-140313.html

Publisher's Note: FOREX Publication stays neutral with regard to jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

Electrocardiography is the primary non-invasive window onto cardiac electrical activity, and manual review of multi-hour ambulatory recordings is slow and inconsistent, which motivates automated beat-level classification. The MIT-BIH Arrhythmia Database [1, 2], comprising 48 half-hour recordings from 47 subjects sampled at 360 Hz with expert beat annotations, has served as the standard benchmark for this task.

Early systems paired morphological samples around the R-peak with inter-beat timing and conventional classifiers [3–10], a body of work that also addresses related concerns such as signal denoising [5] and feature selection for high-dimensional biosignals [9]; Random Forests have been applied widely for their resistance to overfitting and their feature attribution [11–13], while deep architectures report very high accuracies at a computational cost not always compatible with continuous monitoring [14–20].

A methodological issue cuts across much of this literature. Many published results are obtained under an intra-patient protocol, in which beats are pooled across recordings and then split at random. Consecutive beats from one subject are highly correlated, so such a split places near-duplicate beats on both sides of the division, and the resulting accuracy does not estimate performance on a subject the model has never seen. The standard DS1/DS2 division of de Chazal et al. [21] is widely used for inter-patient evaluation. A systematic review of 122 studies published between 2017 and 2024 found that only 30.3% adopted an inter-patient split, while the overall citation trend remained dominated by studies that did not follow the inter-patient methodology [22]. The representation proposed here is therefore evaluated under both protocols. Combining morphology with RR-interval information is not itself new; the contribution lies in the controlled evaluation of a specific 61-dimensional representation under matched protocols, with training-only model selection, class-specific failure analysis and computational characterization. The contributions are stated so that each maps to a specific experiment.

i. A 61-dimensional representation combining a 166.7 ms morphological window with one RR feature, with a

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 3 | Pages 796-806 | e-ISSN: 2347-470X

controlled ablation quantifying each part under both protocols (*Section 4.5*).

ii. A model-selection procedure confined to the training records, using record-grouped cross-validation within DS1 and a single evaluation on DS2, compared against a non-nested search over the full database; the two differ by 0.21 accuracy points and 0.62 macro-F1 points (*Section 4.2*).

iii. A standard de Chazal DS1/DS2 inter-patient evaluation as the primary result, with confusion matrix, per-class sensitivity and specificity, support, and per-class subject coverage (*Section 4.3*). No strong inter-patient generalization is claimed for LBBB, RBBB or APC.

iv. A diagnostic experiment separating three causes of per-class failure. RBBB sensitivity responds to the number of training subjects carrying the class, rising from 0.5419 to 0.9178 on one held-out subject; LBBB sensitivity does not respond to subject count, and remains below 0.001 across the Random Forest hyperparameter grid and the feature-set ablations; APC is constrained by class frequency, and class weighting reduced its sensitivity rather than improving it (*Sections 4.7 and 4.8*).

v. Feature attribution by two methods whose disagreement is quantified, with the interpretability claim restricted to global, model-level attribution (*Section 4.9*).

vi. Measured computational cost on a desktop CPU, with repeated timing (*Section 4.10*).

2. MATHEMATICAL BACKGROUND

2.1. ECG Signal and Normalization

An ECG recording is a sequence x(n), n = 1 … N. To reduce amplitude variability between recordings, each recording is standardized by a z-score transform [23]:

$$x_{norm}(n) = \frac{x(n) - \mu}{\sigma + \varepsilon} \quad (1)$$

where μ and σ are computed over the entire recording, before beat extraction, and $\varepsilon = 10^{-6}$. The normalization unit is therefore the record, not the individual window. This choice was fixed before any inter-patient result was computed, on the grounds that record-level normalization preserves the amplitude of a beat relative to its own recording, which per-beat standardization discards. *Table 1* reports both variants as a post-hoc sensitivity analysis rather than as a selection experiment: the record-level unit was not chosen on the basis of the figures shown there.

On information leakage, μ and σ come from a single recording and are applied only to that recording. The transform is unsupervised and shares no statistic between recordings, so nothing crosses from one subject to another and the DS1/DS2 separation holds. It is, however, an offline operation: it requires the complete recording before any beat within it is normalized, and does not correspond to prospective streaming, where the statistics would have to be estimated from a preceding window. This is stated among the limitations in *section 4.12*.

Table 1. Post-hoc comparison of normalization units; the record-level unit was fixed in advance

Normalization unit	Intra accuracy	Intra macro-F1	Inter accuracy	Inter macro-F1
Record-level (adopted)	98.42%	95.44%	80.80%	42.00%
Per-beat	98.90%	96.67%	79.40%	35.15%

2.2. Segmentation and feature construction

For consecutive R-peaks R_i and R_{i-1}, the RR interval in seconds, the morphological vector formed from the 30 samples preceding and the 30 samples following the annotated R-peak, and the final descriptor that appends the interval to that vector are defined respectively as:

$$RR_i = \frac{R_i - R_{i-1}}{f_s} \quad (2)$$

$$F_{morph} = [f_1, f_2, \dots, f_{60}] \quad (3)$$

$$\boldsymbol{F_{final}} = [f_1, f_2, \dots, f_{60}, RR] \quad (4)$$

where f_s=360 Hz is the sampling frequency, R_i and R_{i-1} are the sample positions of the current and preceding R-peaks, f_j denotes a normalized ECG amplitude sample, and RR is the interval of *equation (2)* appended without further scaling. The resulting descriptor has 61 elements.

At 360 Hz, 60 samples correspond to 166.7 ms in total, that is ±83.3 ms about the R-peak. This span captures the QRS complex and the immediately adjacent morphology, but does not reliably contain the *P* or *T* wave, still less so at elevated heart rates. Rhythm information is therefore supplied by the RR feature rather than by extending the window, a choice consistent with the importance analysis of *section 4.9*, where the most informative sample positions lie after the R-peak.

A record is retained only if channel 0 is labelled MLII, and *equation (1)* is applied to that channel in full. The annotation file is then traversed in temporal order. An annotation counts as a heartbeat if its symbol belongs to the MIT-BIH beat set {N, L, R, e, j, A, a, J, S, V, E, F, /, f, Q, ?}; all other symbols, such as rhythm changes, signal-quality markers and ventricular flutter waves, are skipped. The RR reference position advances at every heartbeat, not only at beats of the five target classes, since otherwise the interval would span any intervening fusion, paced or unclassifiable beat and would be inflated. The first heartbeat of each record has no predecessor and is discarded (42 beats), as is any target beat whose R-peak lies within 30 samples of either end of the recording (3 beats). For each retained beat, the descriptor of *equation (4)* is formed.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 3 | Pages 796-806 | e-ISSN: 2347-470X

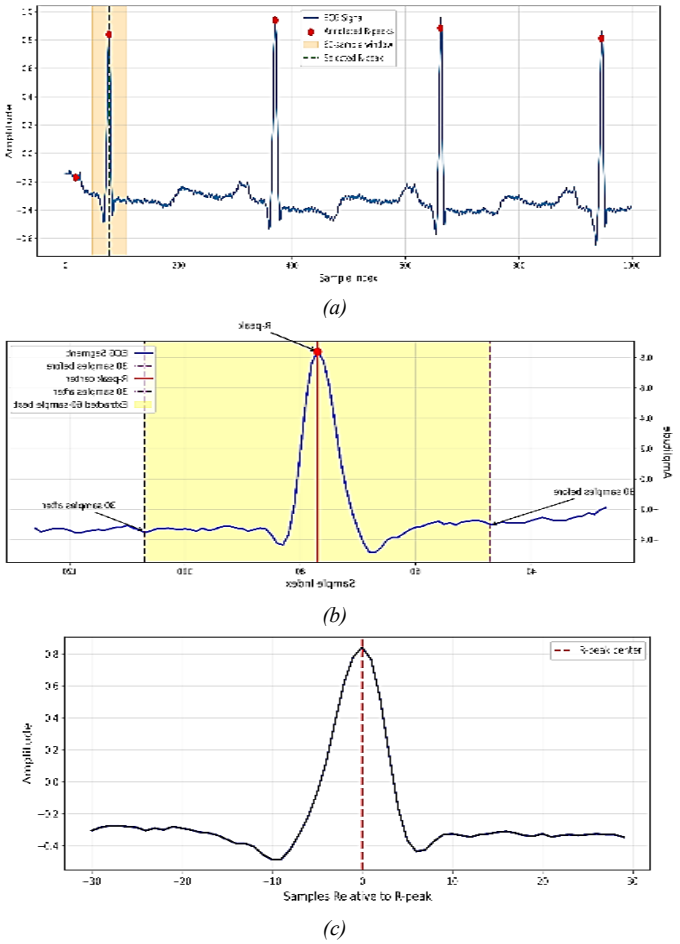

Figure 1. Record 100: (a) annotated R-peaks; (b) magnified beat**;** (c) the extracted 60-sample vector

2.3. Random Forest classifier

A Random Forest is an ensemble of trees grown on bootstrap samples with a random feature subset at each split. The output is the majority vote [24]:

$$\hat{y} = mode\{h_1(x), h_2(x), \dots, h_T(x)\} \qquad (5)$$

where $h_t(x)$ is the class predicted by the t-th tree for an input x, T is the number of trees, and $\hat{y}$ is the predicted class label.

It is used here because it tolerates correlated features and does not require additional feature-wise scaling of the 61-dimensional vector, while providing direct global feature attribution. It is not assumed to be the best classifier for this representation: *Section 4.6* evaluates five baseline classifiers alongside the Random Forest on identical features under the intra-patient protocol; four of these are also evaluated under DS1/DS2. Under record-level separation, logistic regression obtains a higher macro-F1 than the Random Forest.

3. PROPOSED SYSTEM AND WORKFLOW

The system comprises signal acquisition, record-level normalization, beat segmentation, construction of the 61-dimensional vector, and Random Forest classification (*figure 2*). The same extraction code produces the data for both protocols, so the only difference between the experiments is how records are assigned.

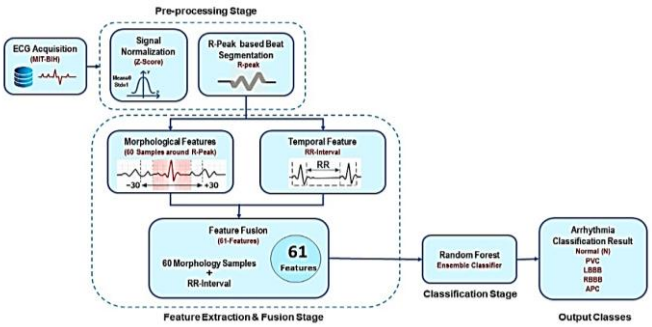

Figure 2. Block diagram of the proposed system

Algorithm 1 sets out the full procedure. Model selection is confined to DS1, and DS2 does not contribute to it; the same fixed record-wise preprocessing and feature extraction is applied independently to both sets, with DS2 used only for the final evaluation.

Algorithm 1. End-to-end experimental pipeline

Input : R = 48 MIT-BIH records; beat symbol set B; W = 30; fs = 360 Hz;
grid G of (n_estimators, max_depth)
Output: DS2 performance metrics
1 X ← ∅
2 for each record r in R do
3 if channel_0(r) ≠ MLII then continue
4 s ← (channel_0(r) − mean) / (std + 1e−6) // Eq. (1), record-level
5 prev ← null
6 for each annotation (sym, pos) in r, in temporal order do
7 if sym ∉ B then continue // non-beat marker
8 if prev = null then prev ← pos; continue
9 if sym is a target class and W < pos < len(s) − W then
10 X ← X ∪ {(concat(s[pos−W:pos+W], (pos−prev)/fs),
11 label(sym), r)} // Eqs. (2)–(4)
12 prev ← pos // advances at every beat
13 end for; end for
14 DS1, DS2 ← split(X by record) // de Chazal et al. [21]
15 best ← argmax over cfg in G of
16 macro_F1(GroupKFold_CV(DS1, cfg, groups = record id, k = 5))
17 // DS2 unused here
18 M ← RandomForest(best).fit(DS1)
19 return evaluate(M, DS2) // one-time final evaluation on DS2

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 3 | Pages 796-806 | e-ISSN: 2347-470X

The only fitted model quantities are the forest structure, learned from DS1 beats alone, and the configuration selected at lines 15–17, also using DS1 alone. The normalization at line 4 uses only the statistics of the record being normalized and no label information, so it creates no dependence between DS1 and DS2.

3.1 Scope: Annotation-Assisted Classification

Segmentation uses the expert R-peak annotations distributed with the database, which bounds what the results can demonstrate. Expert annotations provide an idealised fiducial point and bypass QRS detection, which in a deployed system carries its own failure modes: missed beats, false detections and localization jitter, any of which would shift the extraction window and perturb both the morphological samples and the RR interval. What is evaluated here is therefore offline, annotation-assisted classification of correctly segmented beats, and no claim of end-to-end real-time operation is made in this paper. Dias et al. [6] quantified this dependence by adding jitter of up to 50 ms to the annotated R-peak positions: N and V sensitivity fell by less than one point, while the supraventricular class lost 7.8 points of positive predictivity. Integrating a QRS detection front-end and quantifying its effect on classification is stated as required future work.

3.2. Hyperparameter Selection

The classifier configuration was selected within the training records. DS1 was partitioned into five folds by GroupKFold on the record identifier, so that all beats from a record fall in the same fold, and a grid over n_estimators ∈ {50, 100, 150, 200} and max_depth ∈ {10, 20, 30, None} was evaluated on macro-F1 rather than accuracy, since the classes are severely imbalanced. DS2 did not contribute to hyperparameter selection. The selected configuration is 100 trees at maximum depth 30 (*Table 2*).

Two observations follow from the grid. First, macro-F1 varies by only 0.62 points across the sixteen configurations, from 27.04% to 27.66%, so the choice of configuration is not what determines performance here. Second, LBBB sensitivity is 0.0003 in every configuration without exception, which places that failure inside DS1 itself, independently of the DS2 records on which it is later measured. The macro-F1 values in *table 2* are internal selection scores computed within DS1 and are not comparable with the DS2 results reported in *section 4.3*.

Table 2. Model-selection grid on DS1 only: macro-F1 (%) under record-grouped five-fold cross-validation

n_estimators	depth 10	depth 20	depth 30	depth None
50	27.25	27.15	27.26	27.33
100	27.30	27.46	**27.66**	27.59
150	27.21	27.47	27.61	27.56
200	27.04	27.46	27.53	27.51

Adopted: 100 trees, depth 30. Macro-F1 within DS1 is far below the value obtained on DS2 because record-grouped folds leave some folds without training examples of LBBB or RBBB; the DS1 figure is a selection criterion, not a performance estimate.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Dataset and beat accounting

Lead selection. Channel 0 is used for every record and its label verified programmatically; a record is retained only if that label is MLII. Records 102, 104 and 114 carry V5 on channel 0 and are excluded. Record 114 does contain MLII, but on channel 1; it is excluded rather than remapped so that a single uniform lead position is used throughout, since mixing channel positions within one feature set would introduce the very morphological inconsistency the rule prevents. Forty-five records are retained, of which 43 enter the inter-patient experiment, the DS1/DS2 division of de Chazal et al. [21] additionally excluding the paced records 107 and 217. The figure of 47 given in the previous version was the subject count, applied in error to records.

Beat accounting. The earlier statement that the dataset comprised approximately 83% of annotated beats specified neither a denominator nor an exclusion criterion, and conflated two distinct ratios. Table 3 replaces it with a complete audit trail, every stage of which is reproducible from the rules of Section 2.2. Of the 97,921 beats belonging to the five target classes, 97,876 are retained, or 99.95%. Of the 103,199 true heartbeats present in the 45 records, 97,876 enter the study, or 94.84%, the remainder being beat types outside the five-class problem. No target beat is excluded on grounds of signal quality, and the first-beat and boundary exclusions are independent of the target class label.

Table 3. Beat accounting and class distribution.

Stage	Beats
Annotations across all 48 records	112,647 (109,494 beats, 3,153 markers)
True heartbeats in the 45 MLII records	103,199
—of the five target classes	97,921
—other beat types (/, f, F, j, a, E, J, Q, e, S)	5,278
Excluded: first beat of a record (42) or within 30 samples of a boundary (3)	45
Retained for classification	97,876

Class distribution of the retained beats: Normal 72,936 (74.52%), PVC 7,079 (7.23%), LBBB 8,072 (8.25%), RBBB 7,253 (7.41%), APC 2,536 (2.59%). The Normal share is the majority-class baseline against which accuracy figures should be read; under the inter-patient split this baseline is 74.37%, computed on DS2 alone.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 3 | Pages 796-806 | e-ISSN: 2347-470X

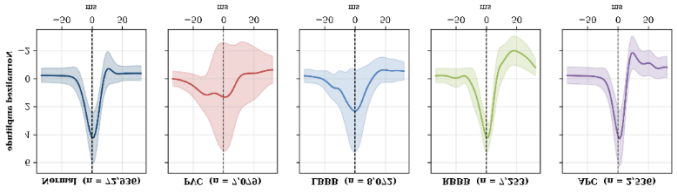

Figure 3. Mean beat morphology per class; shaded regions show ±1 SD

figure 3 shows the class morphologies: Normal a narrow sharp QRS; PVC a broad low-amplitude complex with wide dispersion; LBBB a wide complex of reduced peak amplitude; RBBB a sharp peak followed by a pronounced negative deflection; and APC a waveform closely resembling Normal, anticipating the confusion pattern of *section 4.8*.

4.2. Protocols and Metrics

Inter-patient (primary): Records are divided at the record level into the DS1 training set and DS2 test set specified by de Chazal et al. [21]; the four paced records (102, 104, 107, 217) are excluded by that protocol. Record 114 is already excluded by the lead criterion, so DS1 contributes 21 records and 48,434 beats and DS2 contributes 22 records and 48,977 beats. The remaining 465 beats belong to records 107 and 217 and enter neither set. No record appears in both.

Intra-patient (additional): Beats from all 45 records are pooled and split by StratifiedKFold with five folds, shuffle enabled, random_state = 42, stratified on the class label. Folds are formed over beats rather than records, so beats from one recording appear in several folds; this protocol is retained only for comparability with the published literature.

Model Selection: The configuration is selected within DS1 alone, by the procedure of *section 3.2*, and DS2 is used only after configuration selection for the primary evaluation. The previous version of this manuscript selected hyperparameters by cross-validation over the pooled dataset, so records subsequently used for testing contributed to the choice. Both are reported here for direct comparison: the corrected procedure gives 80.80% accuracy and 42.00% macro-F1 on DS2, against 81.01% and 42.62% under the earlier one. The effect of the selection leakage on this dataset is therefore 0.21 accuracy points and 0.62 macro-F1 points.

Metrics: Precision, recall (equivalently sensitivity), F1 and specificity are computed per class directly from the confusion matrix of the run reported; macro averages are unweighted means over the five classes and weighted averages are weighted by support. Class-wise values are reported as decimals and aggregate values as percentages throughout, and every table reconciles with its confusion matrix.

4.3. Inter-Patient Results (Primary Experiment)

Training on DS1 and testing on DS2 gives 80.80% accuracy, 42.00% macro-F1, 74.03% weighted-F1, macro-precision 0.5441 and macro-recall 0.4207 (*Table 4, Figure 4).*

Table 4. Per-class results under both protocols. Recall is equivalent to sensitivity.

Class	Inter-patient (DS1/DS2)				Intra-patient (5-fold)			
	Precision	Recall	F1	Support	Precision	Recall	F1	Support
Normal	0.8042	0.9864	0.8860	36,426	0.9832	0.9978	0.9904	72,936
PVC	0.8490	0.9139	0.8803	3,218	0.9838	0.9501	0.9667	7,079
LBBB	0.0250	0.0002	0.0005	4,124	0.9925	0.9730	0.9827	8,072
RBBB	0.9134	0.2004	0.3287	3,473	0.9891	0.9720	0.9805	7,253
APC	0.1290	0.0023	0.0045	1,736	0.9727	0.7579	0.8520	2,536
Macro avg	0.5441	0.4207	0.4200	48,977	0.9842	0.9302	0.9544	97,876
Weighted avg	0.7253	0.8080	0.7403	48,977	0.9841	0.9842	0.9837	97,876

Specificity per class: inter-patient 0.3028, 0.9886, 0.9991, 0.9985, 0.9994; intra-patient 0.9500, 0.9988, 0.9993, 0.9991, 0.9994 (Normal, PVC, LBBB, RBBB, APC).

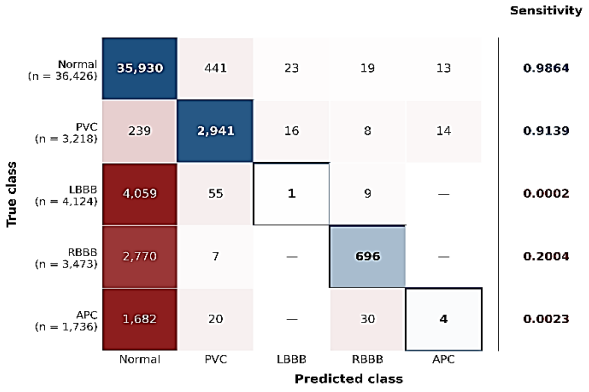

Figure 4. Inter-patient confusion matrix (DS1 train / DS2 test).

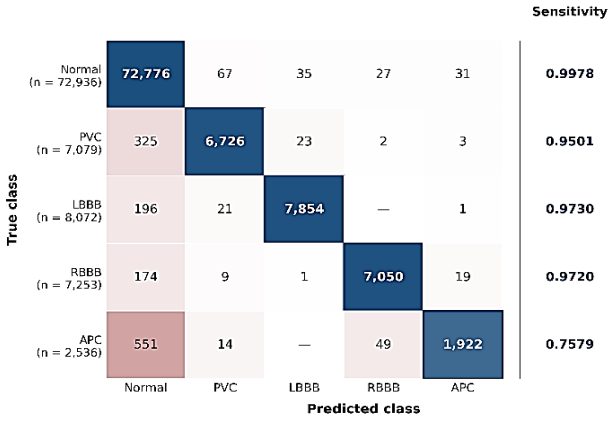

Figure 5. Intra-patient confusion matrix, pooled over five folds

Cell values are beat counts; shading is row-normalized; blue marks correct classifications and red marks errors.

Performance differs sharply across the five classes. Only two records contribute LBBB beats to each split, namely 109 and 207 in DS1 and 111 and 214 in DS2, and only three contribute RBBB beats, namely 118, 124 and 207 in DS1 and 212, 231

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 3 | Pages 796-806 | e-ISSN: 2347-470X

and 232 in DS2. Normal, PVC and APC draw on 17 to 19, 16 and 13 records respectively. Limited subject coverage is the obvious candidate explanation, and a comparable observation appears in [6], where a single record (232) contributes 75% of the supraventricular beats in DS2 and the authors attribute poor generalization to that concentration. Coverage is not, however, sufficient to explain the behaviour of all three weak classes here, and *section 4.7* tests it directly.

The APC class is a different case. Thirteen records contribute APC beats to each split, so subject coverage does not account for its failure. DS1 contains only 800 APC beats against 36,266 Normal beats, a ratio of roughly 45 to 1. Its difficulty is therefore one of class frequency, compounded by close morphological similarity to Normal beats, which is visible in *figure 3.*

4.4. Intra-Patient Results (Additional Experiment)

Under the intra-patient protocol, the same system reaches 98.42% accuracy and 95.44% macro-F1, with a weighted-F1 of 98.37%, a macro-precision of 0.9842 and a macro-recall of 0.9302 (*Table 4*, *Figure 5*). These figures are reported for comparability with the published literature. They must not be read as estimates of performance on unseen subjects. The 53.44-point macro-F1 gap between the two evaluation settings illustrates the substantial difference between beat-level and record-level evaluation, although the inter-patient protocol additionally excludes records 107 and 217.

Fold-wise stability is given in *table 5*. Accuracy is highly stable at 98.42% ± 0.05 and macro-F1 at 95.44% ± 0.18, but APC sensitivity varies by an order of magnitude more relative to its own value, which is why fold-level accuracy alone is an insufficient stability check on imbalanced data.

Table 5. Fold-wise intra-patient results. SD is the population standard deviation over five folds

Fold	1	2	3	4	5	Mean	SD
Accuracy (%)	98.38	98.45	98.48	98.43	98.35	98.42	0.05
Macro-F1 (%)	95.23	95.55	95.66	95.54	95.23	95.44	0.18

Per-class sensitivity is stable for the four larger classes, with standard deviations of 0.0003 for Normal, 0.0047 for PVC, 0.0038 for LBBB and 0.0061 for RBBB, and least stable for APC at 0.0126 about a mean of 0.7579. Fold test sets contain 19,575–19,576 beats, with per-class counts within one beat of each other across folds.

4.5. Ablation: Morphology, RR, and Their Combination

Table 6 tests the complementarity premise with three configurations under both protocols, everything else fixed.

Table 6. Ablation of the feature set under both protocols

Configuration	Intra-patient				Inter-patient		
	Accuracy	Macro-F1	APC rec.	Accuracy	Macro-F1	LBBB rec.	RBBB rec.
Morphology only (60)	97.45%	90.82%	0.5564	78.82%	36.81%	0.0002	0.1137
RR only (1)	75.99%	28.20%	0.0000	71.84%	26.02%	0.0000	0.0107
Morphology RR (61)	98.42%	95.44%	0.7579	80.80%	42.00%	0.0002	0.2004
Majority-class baseline	74.52%	—	—	74.37%	—	—	—

The table supports two conclusions that pull in opposite directions. On its own the RR interval is not discriminative: it reaches 75.99% intra-patient against a majority-class baseline of 74.52%, and 71.84% inter-patient against a baseline of 74.37%, that is below the baseline, with three of the five classes essentially unrecognized. Combined with the morphological samples it is another matter. Macro-F1 rises by 4.62 points intra-patient and 5.19 points inter-patient, and the gain falls where morphology is weakest, with APC sensitivity moving from 0.5564 to 0.7579 and inter-patient RBBB sensitivity from 0.1137 to 0.2004. One feature, 1.6% of the vector, produces this. Deleting the morphological block instead costs 67.24 macro-F1 points intra-patient.

4.6. Controlled Baseline Comparison

To test whether performance is specific to the Random Forest, five baseline classifiers were evaluated alongside it under the intra-patient protocol on identical features and seed. Four of these were also evaluated under DS1/DS2; the RBF-SVM was omitted from the inter-patient comparison for the reason stated below (*Tables 7 and 8*).

Table 7. Baseline classifiers under the intra-patient protocol, identical features and folds

Classifier	Accuracy	Macro-F1	APC recall
Random Forest (proposed)	98.42%	95.44%	0.7579
k-nearest neighbours (k = 5)	98.14%	93.40%	0.6526
SVM, RBF kernel *	97.92%	93.67%	0.6622
Decision tree	97.02%	92.24%	0.7772
Logistic regression	91.71%	82.60%	0.6002
Gaussian naive Bayes	77.16%	57.80%	0.3517

"An RBF kernel on 97,876 × 61 was not tractable within available resources; this row used a stratified subsample of 20,000 beats (20.4%) under the same fold protocol and is indicative rather than strictly comparable."

Under the intra-patient protocol, Random Forest leads on both metrics, though the margin over k-nearest neighbours is 0.28 accuracy points. The separation appears in macro-F1 and originates in the minority class, where APC sensitivity is 0.7579 against 0.6526. A single decision tree reaches higher APC sensitivity while performing worse on every other measure, which suggests that majority voting across many trees leans toward the frequent classes.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 3 | Pages 796-806 | e-ISSN: 2347-470X

Table 8. The same classifiers under the inter-patient protocol, fitted on DS1 and evaluated once on DS2. The RBF-SVM is omitted, since a subsampled row would not be comparable under this protocol

Classifier	Accuracy	Macro-F1	LBBB Rec.	RBBB Rec.	APC Rec.
Random Forest (proposed)	80.80%	42.00%	0.0002	0.2004	0.0023
Decision tree	79.64%	42.90%	0.0056	0.2793	0.0565
Logistic regression	78.50%	45.16%	0.0112	0.3988	0.0035
k-nearest neighbours (k = 5)	76.82%	32.19%	0.0002	0.0438	0.0046
Gaussian naive Bayes	70.80%	34.73%	0.0323	0.1094	0.5017

The ordering does not carry over. Random Forest has the highest accuracy under DS1/DS2 but not the highest macro-F1, logistic regression obtaining 45.16% against 42.00%. Gaussian naive Bayes reaches an APC sensitivity of 0.5017 by predicting the class liberally, at the lowest overall accuracy in the table. The weak inter-patient performance is therefore not specific to the Random Forest: all classifiers remain poor for LBBB and none exceeds 0.40 RBBB sensitivity, although classifier choice materially affects APC. The justification for Random Forest rests on its intra-patient behaviour and its direct feature attribution, not on superiority under the primary protocol.

4.7. Subject Coverage or Representation? A Diagnostic Experiment

Section 4.3 showed that LBBB, RBBB and APC transfer poorly under the DS1/DS2 protocol, and identified limited subject coverage as the candidate explanation for the first two. Coverage and a representational limit produce identical numbers in *table 4*, so they were separated experimentally.

DS1 contains LBBB beats from two records (109, 207) and RBBB beats from three (118, 124, 207); DS2 contains LBBB from two (111, 214) and RBBB from three (212, 231, 232). In the experiment reported here, the number of training subjects carrying a class was increased by moving one or two records from DS2 into training, and sensitivity was then measured on a single remaining held-out subject. This procedure deliberately breaks the DS1/DS2 protocol and is therefore reported separately; it is a diagnostic and not a performance estimate, and none of the figures in *sections 4.3 to 4.6* depend on it.

Table 9. Subject-coverage diagnostic. Sensitivity on a single held-out subject before and after increasing the number of training subjects carrying the class; the number of such subjects is given in parentheses

Class	Held-out record	Sensitivity, baseline → increased coverage
RBBB	231	0.5419 (3) → **0.9178** (5)
RBBB	232	0.0354 (3) → **0.4798** (5)
RBBB	212	0.0016 (3) → 0.0082 (5)
LBBB	111	0.0005 (2) → 0.0009 (3)
LBBB	214	0.0000 (2) → 0.0000 (3)

The two classes respond differently to the same intervention. RBBB sensitivity rises substantially on two of the three test subjects, which is the behaviour expected when the constraint is how many individuals the classifier has seen. LBBB sensitivity does not move on either test subject.

Three further results place the LBBB failure outside subject coverage. It is 0.0003 in all sixteen configurations of the DS1 selection grid (*table 2*), so it is present within DS1 alone and is not a property of the particular DS2 records. It is at or below 0.0002 in all three ablation configurations, morphology only, RR only and combined (*table 6*), so it is not a consequence of the feature combination. And no classifier in *table 8* exceeds 0.0323, so it is not a property of the Random Forest. The 61-dimensional single-lead representation used here does not transfer LBBB morphology across subjects.

This is a limitation of the representation and not of the task. Shi et al. [25], separating the same five MIT-BIH classes under an inter-patient division of their own construction, report 27.9% LBBB sensitivity from a 14-dimensional region-based representation, the lowest of their five classes but three orders of magnitude above the value obtained here.

Record 212 does not follow the RBBB pattern, remaining below 0.01 in both conditions. Additional coverage helps two of the three RBBB test subjects and not the third, so the effect is not uniform across subjects within a class.

4.8. Class Imbalance and APC Error Analysis

Two families of remedy were tested under both protocols, applied to training data only; the held-out fold and DS2 are never resampled (*table 10*).

Table 10. Class-imbalance handling. Undersampling rows are mean ± SD over ten random seeds

Strategy	Intra-patient		Inter-patient			
	Accuracy	Macro-F1	Accuracy	Macro-F1	RBBB	APC
None (proposed)	98.42%	95.44%	80.80%	42.00%	0.2004	0.0023
class_weight = balanced	98.50%	95.50%	79.93%	36.47%	0.0181	0.0000
Random oversampling	98.71%	96.14%	79.91%	36.37%	0.0115	0.0012
Undersampling 1×	94.96%	88.97%	77.89 ± 0.83	45.40 ± 1.01	0.4206	0.1000
Undersampling 2×	—	—	80.01 ± 0.44	45.93 ± 0.52	0.4099	0.0653
Undersampling 4×	—	—	81.03 ± 0.32	45.62 ± 0.70	0.3906	0.0395
Undersampling 8×	—	—	81.08 ± 0.39	44.45 ± 1.41	0.3485	0.0157
Undersampling 16×	—	—	80.78 ± 0.30	42.57 ± 1.15	0.2546	0.0048

Ratios denote the maximum retained count per class as a multiple of the smallest class in DS1, which contains 800 APC beats. Classes exceeding the specified cap are randomly undersampled, whereas classes below the cap are retained unchanged.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 3 | Pages 796-806 | e-ISSN: 2347-470X

Undersampling works differently, reducing the dominance of the more frequent classes rather than duplicating minority-class observations. A twofold cap, limiting each class to 1,600 training beats, gives the best macro-F1 in the sweep at 45.93 ± 0.52%, a gain of 3.93 points over the undersampled model. It is reported as a post-hoc experiment rather than adopted, since it would require different training procedures for the two protocols and the 1× setting reduces intra-patient macro-F1 to 88.97%. LBBB stays at or below 0.0022 under every method and ratio tested, so no imbalance remedy addresses that class.

APC error analysis. Under the intra-patient protocol, 551 of the 2,536 APC beats, or 21.73%, are labelled Normal, with a further 0.55% assigned to PVC and 1.93% to RBBB. The RR distributions qualify the explanation offered in the previous version of this manuscript. Normal beats have a median RR of 0.7472 s, with a fifth to ninety-fifth percentile range of 0.5333 to 1.1861 s, and APC beats a median of 0.6861 s over the range 0.3806 to 0.7583 s. Some 68.22% of APC beats therefore fall inside the central band of the Normal distribution, so the single RR feature does not separate the two classes for most APC beats. Published systems using RR intervals normalized against local and global moving means report substantially stronger supraventricular results under this protocol [26], which identifies the unnormalized interval used here as one target for revision.

4.9. Feature Importance and the Scope of Interpretability

Interpretability is claimed here in a restricted sense. The fitted forest admits global, model-level attribution, which characterises the model as a whole. It does not explain why an individual beat received a particular label, and it is not evidence of clinical relevance. Per-beat attribution and its clinical validation lie outside the scope of this work.

Two methods are reported because they disagree, and the disagreement is itself the finding. Mean decrease in impurity is computed by the forest on its training data and is known to favour continuous features of high cardinality; permutation importance, the drop in macro-F1 when a single feature is shuffled, is not subject to that bias. The first attributes 5.97% of total importance to the RR interval and 94.03% to the morphological block, the second 58.35% and 41.65% respectively. Both rank the interval first among all 61 features, so describing it as the most important individual feature is supported, but neither share should be quoted on its own.

The disagreement arises from the correlation structure: impurity spreads morphological importance across sixty correlated positions, while permutation finds each morphological sample compensated by its neighbours whereas the interval has no substitute. The ablation of *section 4.5* arbitrates, since it removes a feature block rather than perturbing one coordinate: deleting the interval costs 4.62 macro-F1 points and deleting the morphological block costs 67.24. The decision is morphology-dominated, with the interval a small non-redundant addition. Figure 6 maps morphological importance onto time; both methods concentrate it after the R-peak, though impurity peaks between +47ms and +58ms and permutation between +22ms and +50ms. This is a temporal location only. The earlier version associated these positions with the ST segment and the T wave; that association is not supported, since at typical heart rates they fall within the terminal QRS rather than in repolarization, and it has been removed. Because adjacent samples are correlated, individual per-sample rankings should not be over-interpreted.

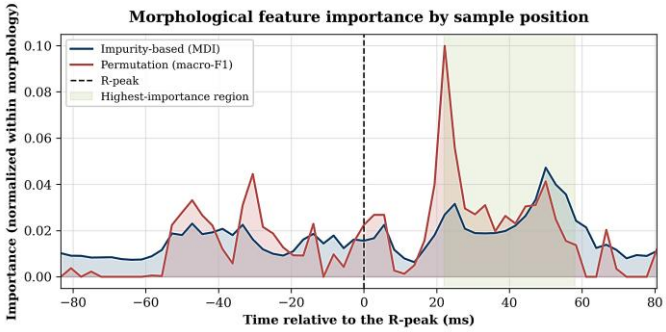

Figure 6. Morphological feature importance by sample position relative to the R-peak, by both methods

4.10. Computational Cost

All timings are medians of seven repeats on an Intel 12-core desktop CPU under Windows 11. Single-beat inference is measured with parallelism disabled, since for one row it is 2.4 times slower than a single thread. These measurements characterize a forest of the selected configuration fitted on all 97,876 beats. The final inter-patient model is instead fitted on the 48,434 DS1 beats; its model size and training time were not separately characterized. Of the 3.2066 ms single-beat latency, 3.1955 ms is a fixed per-call framework overhead and 0.0111 ms is the classification itself, so 99.7% of the measured latency is overhead rather than computation; the single-beat processing margin is computed from the conservative single-beat value. The depth cap was retained on size grounds as well: an unbounded forest of the same width occupies 4.38 MB more without improving macro-F1 within DS1.

Table 11. Measured computational characteristics

Quantity	Measured value
Trees / depth / total nodes	100 / 30 (cap reached by every tree) / 367,102
Nodes per tree	3,233–4,157 (median 3,698)
Serialized model size	36.45 MB (7.87 MB compressed)
Training time, 97,876 beats	13.66 s
Feature extraction per beat	0.0209 ms
Inference: single beat / batched 1000	3.2066 ms / 0.0111 ms per beat
Total per beat, single-beat processing	3.227 ms, i.e. 0.65% of the 500 ms interval at 120 bpm (155× margin)

On embedded deployment, no low-power hardware was tested in this work, and a serialized model of 36.45 MB does not fit

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 3 | Pages 796-806 | e-ISSN: 2347-470X

the flash memory of a typical microcontroller. Energy consumption per beat is not reported either: the measurements were obtained on a general-purpose desktop CPU, where energy figures do not transfer to the low-power processors used in ambulatory monitoring. Embedded operation is therefore described as a future validation target rather than a demonstrated result, and quantitative embedded claims have been removed from this manuscript. What the measurements do support is that classification of a segmented beat consumes less than 1% of the available interval at 120 bpm on a desktop processor.

4.11 Comparison with Previous Studies

Table 12 places this work alongside representative studies, subject to an explicit caveat: the rows are not directly comparable. The studies differ in validation protocol, class definition, number of classes and evaluation metric, and any of these can move reported accuracy by more than the differences between the methods themselves. The protocol column is the most consequential, since intra-patient and inter-patient figures on MIT-BIH do not measure the same quantity; the 53.44-point macro-F1 gap between Sections 4.3 and 4.4 illustrates this. No claim of superiority is made on the basis of raw accuracy.

Table 12. Comparison with prior work; protocols differ and rows are not comparable

Ref.	Protocol	Classes	Feature Construction	Dim.	Classifier	Reported Performance	Computational Cost
[3]	Inter-patient DS1/DS2	4 AAMI	Wavelet, HOS, RR intervals, morphological features	Not stated	Ensemble of SVMs	94.4% accuracy, 65.26% mean sensitivity	Not reported
[4]	Beat-level, 10-fold	15 MIT-BIH	7 parametric morphology features + VMP-QRS visual-pattern feature	8	kNN (k = 31)	97.7% accuracy	Not reported
[6]	Inter-patient DS1/DS2	3 AAMI	4 log-RR, 4 morphological statistics, 2 HOS	10	Linear discriminant	Se: N 94.5, S 92.5, V 88.6 *	2.6 ms/beat
[8]	Beat-level, 80/20 holdout	4 AAMI	LBP, HOS, CM, HBF, DWT, RR intervals	131	MPA-optimized Random Forest	99.67% accuracy	Not reported
[11]	Patient-specific	3 AAMI	pre/post-RR, PR, RT and QRS intervals + P/QRS/T abstract vectors	15 (reported)	Weighted Random Forest	F1: SVEB 81.05, VEB 97.07 *	68 bits/beat feature storage
[25]	Inter-patient, author-defined split	5 MIT-BIH	Region-based P/QRS/T features + RR ratio; DWT/PCA in QRS region	14	Ensemble of KNN, SVM, DT, RF + secondary SVM	74.5% accuracy; Se: LBBB 27.9, RBBB 79.6, APC 81.8	Not reported
[26]	Inter-patient DS1/DS2	3 AAMI	Derivative of 64 samples; normalized RR	64 + 4	Matched-filter 1D CNN	98.18% accuracy, 91.90% mean sensitivity	15 KB, <1 ms/beat †
This work	Inter-patient DS1/DS2; beat-level 5-fold	5 MIT-BIH	60 samples + 1 single unnormalized RR	61	Random Forest	80.80% (macro-F1 42.00%) inter-patient; 98.42% (95.44%) beat-level	36.45 MB, 3.227 ms/beat

"Accuracy not reported in the source." Model size and inference time were reported for variants evaluated on embedded hardware; the accuracy figure is from the best-performing variant in the study.

One distinction matters particularly: in the AAMI grouping, class N absorbs LBBB and RBBB, so studies using AAMI classes do not face the discrimination problem examined here. Only [25] separates the same five MIT-BIH classes under record-level separation, and does so under an author-defined split constructed to balance the classes across training and test sets, which DS1/DS2 does not.

Subject counts are reported only where explicitly available from the cited study; the protocol column identifies the reported evaluation design, and results obtained under different protocols are not treated as directly comparable.

4.12. Limitations

Inter-patient generalization is limited to two of five classes, and the system is not suitable for use on LBBB, RBBB or APC. For RBBB, limited subject coverage appears to be a major contributor, though not uniformly across subjects; for LBBB the constraint is the representation itself. Segmentation depends on expert annotations, so end-to-end performance is unknown. The standard de Chazal partition is retained for comparability, but records 201 and 202 originate from the same subject and fall in DS1 and DS2 respectively, a known exception to complete subject-level separation that has been noted in the literature [22] and avoided in some later divisions [25]. Record-level normalization requires the complete recording and is therefore offline. The ±83.3ms window excludes the P and T waves. Timing was measured on a desktop processor only, energy was not measured, the 36.45 MB model exceeds typical microcontroller memory, the SVM baseline used a 20.4% subsample, and no cross-database evaluation was performed, so generalization beyond MIT-BIH is not demonstrated.

5. CONCLUSIONS

This work evaluated a 61-dimensional heartbeat representation, 60 morphological samples spanning ±83.3 ms about the R-peak plus one RR interval, with a Random Forest classifier, under both the standard de Chazal DS1/DS2 inter-patient protocol and beat-level cross-validation, on 97,876 beats from 45 MLII records of the MIT-BIH Arrhythmia Database.

Under record-level separation the system attains 80.80% accuracy and 42.00% macro-F1, with the configuration selected within DS1 alone. Normal and PVC retain high sensitivity on DS2, at 0.9864 and 0.9139, whereas LBBB, RBBB and APC do not, at 0.0002, 0.2004 and 0.0023, and the causes differ. RBBB sensitivity improves substantially on two of three held-out subjects when more training subjects carry the class, which identifies limited subject coverage as a major contributor rather than a uniform explanation. LBBB sensitivity does not respond to subject count, to any configuration in the selection grid, or to any feature subset, and remains poorly recognized across all five classifiers tested, so this representation does not transfer LBBB morphology reliably across subjects. APC is constrained by class frequency, at roughly 45 to 1 against Normal in DS1; class weighting reduced its sensitivity, while undersampling improved inter-patient macro-F1 at a cost to intra-patient performance.

Under the beat-level protocol the same system reaches 98.42% accuracy and 95.44% macro-F1. The 53.44-point macro-F1 gap illustrates the magnitude of the difference between the two evaluation settings rather than providing a pure estimate of protocol effect. The ablation shows the feature types to be complementary: the RR interval alone only slightly exceeds the majority-class baseline under the intra-patient protocol and falls below it under the inter-patient protocol, yet adding it raises macro-F1 by 4.62 and 5.19 points. Single-beat classification requires 3.227 ms on a desktop processor, and no embedded hardware was tested.

Future work follows from these limitations: a QRS detection front-end and quantification of its effect; cross-database evaluation; an RR construction normalized against local and global moving means, reported as substantially more effective for supraventricular classes; and implementation on low-power hardware.

Author contributions: Conceptualization, methodology, software and original draft: Shams M. Qahtan. Validation: both authors. Review, editing and supervision: Hadeel N. Abdullah.

Conflicts of interest: "The authors declare no conflict of interest".

Data and code: the MIT-BIH Arrhythmia Database is publicly available from PhysioNet [1, 2]; the processing scripts are available from the corresponding author on request, each writing a timestamped log that traces every reported value to the run producing it.

Ethical approval: this study used a publicly available, de-identified database and involved no human participants or animals. The material is the authors' original work and has not been previously published.

Funding: None

REFERENCES

[1] M. Costa, G.B. Moody, I. Henry, and A.L. Goldberger, "PhysioNet: an NIH research resource for complex signals," Journal of electrocardiology, Vol. 36, pp.139-144, 2003. DOI: https://doi.org/10.1016/j.jelectrocard.2003.09.038

[2] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," IEEE engineering in medicine and biology magazine, Vol. 20, Issue. 3, pp. 45–50, 2001. DOI: https://doi.org/10.1109/51.932724

[3] S. K. Pandey, R. R. Janghel, and V. Vani, "Patient specific machine learning models for ECG signal classification," Procedia Computer Science, Vol. 167, pp. 2181–2190, 2020. DOI: https://doi.org/10.1016/j.procs.2020.03.269

[4] H. Yang and Z. Wei, "Arrhythmia recognition and classification using combined parametric and visual pattern features of ECG morphology," IEEE Access, Vol. 8, pp. 47103–47117, 2020. DOI: 10.1109/ACCESS.2020.2979256

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 14, Issue 3 | Pages 796-806 | e-ISSN: 2347-470X

[5] F. W. Abdullah and H. N. Abdullah, "ECG signal denoising: an integrated filtering approach for improved signal quality", Kufa Journal of Engineering, Vol. 17, No. 1, pp. 291–307, 2026. DOI: http://doi.org/10.30572/2018/KJE/170116

[6] Dias, F.M., H.L. Monteiro, T.W. Cabral, R. Naji, M. Kuehni, and E.J.D.S Luz, "Arrhythmia classification from single-lead ECG using the inter-patient paradigm," Computer Methods and Programs in Biomedicine, Vol. 202, p. 105948, 2021. DOI: 10.1016/j.cmpb.2021.105948

[7] S. Aziz, S. Ahmed, and M.-S. Alouini, "ECG-based machine-learning algorithms for heartbeat classification," Scientific reports, Vol. 11, No. 1, p. 18738, 2021. DOI: 10.1038/s41598-021-97118-5

[8] Hassaballah, M., Y.M. Wazery, I.E. Ibrahim, and A. Farag, "ECG heartbeat classification using machine learning and metaheuristic optimization," Bioengineering, Vol. 10, No. 4, p. 429, 2023. DOI: 10.3390/bioengineering10040429

[9] T. Subba and T. Chingtham, "Comparative analysis of machine learning algorithms with advanced feature extraction for ECG signal classification" , IEEE Access, Vol. 12, pp. 57727–57740, 2024. DOI: 10.1109/ACCESS.2024.3387041

[10] Darmawahyuni, A., S. Nurmaini, B. Tutuko, M.N. Rachmatullah, F. Firdaus, A.I. Sapitri, A. Islami, J. Marcelino, R. Isdwanta, and M.I. Perwira, "An Improved electrocardiogram arrhythmia classification performance with feature optimization," BMC medical informatics and decision making, Vol. 24, No. 1, p. 412, 2024. DOI: 10.1186/s12911-024-02822-7

[11] Kung, B.H., P.Y. Hu, C.C. Huang, C.C. Lee, C.Y. Yao, and C.H. Kuan, "An efficient ECG classification system using resource-saving architecture and random forest," IEEE Journal of Biomedical and Health Informatics, Vol. 25, No. 6, pp. 1904–1914, 2021. DOI: https://doi.org/10.1109/JBHI.2020.3035191

[12] A.F. Ahmed, H.N. Abdullah, B.M. Albaker, "A Robust and Efficient Scheme for ECG Signal Classification Based on Digital Signal Processing, Feature Engineering Selection, and Random Forest Classifier", Radioelectronics, Nanosystems. Information Technologies, Vol. 16, No. 1, pp. 53-66, 2024. DOI: 10.17725/j.rensit.2024.16.053

[13] M. Sraitih, Y. Jabrane, and A. H. El Hassani, "An Automated system for ECG arrhythmia detection using machine learning techniques," Journal of Clinical Medicine, Vol. 10, No. 22, p. 5450, 2021. DOI: https://doi.org/10.3390/jcm10225450

[14] Ullah, A., S.M. Anwar, M. Bilal, and R.M. Mehmood, "Classification of arrhythmia by using deep learning with 2-D ECG spectral image representation," Remote sensing, Vol. 12, No. 10, p. 1685, 2020. DOI: https://doi.org/10.3390/rs12101685

[15] M. F. Safdar, P. Pałka, R.M. Nowak, and A. Al Faresi, "A novel data augmentation approach for enhancement of ECG signal classification," Biomedical Signal Processing and Control, Vol. 86, p. 105114, 2023. DOI: 10.1016/j.bspc.2023.105114

[16] Zabihi, F., F. Safara, and B. Ahadzadeh, "An electrocardiogram signal classification using a hybrid machine learning and deep learning approach". Healthcare Analytics, Vol. 6, p.100366, 2024. DOI: 10.1016/j.health.2024.100366

[17] U. Gupta N. Paluru, D. Nankani, K. Kulkarni, and N. Awasthi, "A comprehensive review on efficient artificial intelligence models for classification of abnormal cardiac rhythms using electrocardiograms," Heliyon, Vol. 10, No. 5, e26787, 2024. 10.1016/j.heliyon.2024.e26787

[18] S. Reznichenko, J. Whitaker, Z. Ni, and S. Zhou, "Comparing ECG lead subsets for heart arrhythmia/ECG pattern classification: Convolutional neural networks and random forest." CJC open, Vol. 7, No. 2, pp. 176–186, 2025. DOI: https://doi.org/10.1016/j.cjco.2024.10.012

[19] A. Kirkbas and A. Kizilkaya, "Automated ECG arrhythmia classification using feature images with Common Matrix approach-based classifier," Sensors, Vol. 25, No. 4, p. 1220, 2025. DOI: https://doi.org/10.3390/s25041220

[20] N. S. Kathayat and A. P. Renold, "Hybrid deep learning model for scalogram-based ECG classification of cardiovascular diseases", IEEE Access, Vol. 13, pp. 159628–159638, 2025. DOI: https://doi.org/10.1109/ACCESS.2025.3605279

[21] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," IEEE Transactions on Biomedical Engineering, Vol. 51, No. 7, pp. 1196–1206, 2004. DOI: https://doi.org/10.1109/TBME.2004.827359

[22] G. A. L. Silva, P. H. L. Silva, G. J. P. Moreira, V. L. de S. Freitas, J. C. Gertrudes, and E. J. da S. Luz, "A systematic review of ECG arrhythmia classification: embedded feasibility, adherence to standards, and fair evaluation," Research on Biomedical Engineering, Vol. 42, No. 3, Art. no. 42, 2026. doi: http://doi.org/10.1007/s42600-026-00485-z

[23] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.

[24] L. Breiman, "Random forests," Machine Learning, Vol. 45, No. 1, pp. 5–32, 2001. doi: https://doi.org/10.1023/A:1010933404324

[25] H. Shi, H. Wang, F. Zhang, Y. Huang, L. Zhao, and C. Liu, "Inter-patient heartbeat classification based on region feature extraction and ensemble classifier," Biomedical Signal Processing and Control, Vol. 51, pp. 97–105, 2019. doi: https://doi.org/10.1016/j.bspc.2019.02.012

[26] M. M. Farag, "A tiny matched filter-based CNN for inter-patient ECG classification and arrhythmia detection at the edge," Sensors, vol. 23, no. 3, Art. no. 1365, 2023. doi: https://doi.org/10.3390/s23031365

© 2026 by the Shams M. Qahtan, and Hadeel N. Abdullah. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).