Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Waveform-Free Characterization of Intraoperative Cardiac Arrhythmia: Beat-Timing Signatures, Reproducible Burden Phenotypes, and Cross-Surgical-Context Generalization in an Anesthesiologist- Validated Cohort

Muhammet Rojhat KARA

mrkara@kmu.edu.tr

Karamanoğlu Mehmetbey University

Article

Keywords: intraoperative arrhythmia, heart rate variability, RR interval, waveform-free analysis, unsupervised phenotyping, interpretable machine learning

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10629683/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Waveform-Free Characterization of Intraoperative Cardiac Arrhythmia: Beat-Timing Signatures, Reproducible Burden Phenotypes, and Cross-Surgical-Context Generalization in an Anesthesiologist-Validated Cohort

Muhammet Rojhat KARA*

Department of Electrical-Electronics Engineering, Faculty of Engineering, Karamanoğlu Mehmetbey University, KARAMAN, TÜRKİYE

***Correspondence:** mrkara@kmu.edu.tr

Abstract

Cardiac arrhythmia during surgery can demand immediate recognition, yet almost every publicly available electrocardiographic resource is drawn from ambulatory or intensive care recordings rather than the operating room. The first large anesthesiologist-validated intraoperative database therefore invites the reflexive response of training yet another waveform-morphology classifier. This study asks a more economical question: how far can intraoperative arrhythmia be characterized without the raw waveform, from expert beat annotations and routine perioperative metadata alone? Across 482 surgical cases and more than 660,000 validated beats, features built from R-peak times and beat-type labels separated five clinically grouped rhythm classes with a balanced accuracy of 0.869. A transparent linear model matched a higher-capacity random forest, indicating that the signal reflects genuine structure rather than model complexity, and it remained accurate on windows the experts flagged as artifact, where morphology-based methods are most vulnerable. The same features resolved three reproducible arrhythmia-burden phenotypes and transferred between surgical departments; older age was the principal adjusted correlate of high burden. Resolving the ectopy classes still relied on the expert beat-type labels rather than interval timing alone. Beat annotations and metadata thus offer a low-cost, interpretable, and transferable baseline for a setting that previously lacked one.

Keywords: intraoperative arrhythmia; heart rate variability; RR interval; waveform-free analysis; unsupervised phenotyping; interpretable machine learning

1. Introduction

Cardiac arrhythmia during anesthesia and surgery is a distinct clinical entity. The intraoperative environment is arrhythmogenic through a convergence of surgical stimulation, anesthetic agents, autonomic fluctuation, electrolyte and acid-base shifts, and hemodynamic instability, and although most events are transient, some require immediate recognition and intervention to avoid morbidity and mortality [1,2,3]. Perioperative atrial fibrillation in particular is common after noncardiothoracic surgery and is associated with prolonged hospitalization and adverse outcomes [4,5], while the incidence and character of arrhythmia vary substantially with surgical context, being notably higher after thoracic procedures [6]. The clinical importance of the intraoperative setting is therefore established, but the data resources needed to study it computationally have lagged behind.

Publicly available ECG databases have been drawn almost entirely from ambulatory Holter monitoring or intensive care telemetry. The MIT-BIH Arrhythmia Database has anchored beat-level benchmarking for more than two decades [7], and PTB-XL has become the reference clinical twelve-lead corpus [8]. These resources have enabled end-to-end deep neural networks that classify a broad range of arrhythmias from the raw waveform at a level comparable to practicing cardiologists [9,10]. That success, however, has produced a form of benchmark saturation: on canonical morphology databases the marginal gains from ever larger models are small, and the operating conditions of the training data (resting, ambulatory, or critical-care) do not reflect the rapid physiological transitions and characteristic artifacts of the operating room.

The VitalDB Arrhythmia Database addresses this gap. It provides continuous single-lead intraoperative ECG from 482 noncardiac surgical patients with more than 660,000 heartbeats labeled and validated by five anesthesiologists across four beat types and ten rhythm categories [11], and it is built upon the broader VitalDB high-resolution perioperative repository from the same institution [12]. The database is explicitly intended to support the development of waveform-based detection models. A common analytic response is therefore to download the waveform and train yet another morphology classifier. The present study takes a deliberately different position and asks a more basic question: how much of intraoperative arrhythmia is recoverable from the beat annotations alone, without ever touching the raw waveform?

There are strong reasons to expect that the answer is substantial. The temporal sequence of heartbeats carries a large fraction of rhythm information independent of P-wave or QRS morphology. Classical detectors identify atrial fibrillation from RR-interval irregularity alone, using the coefficient of variation and density histograms of RR and successive-difference intervals [13], and modern RR-only classifiers achieve high accuracy on benchmark data [14]. Heartbeat interval features have long complemented morphology in canonical beat classification and established the patient-independent evaluation paradigm that guards against optimistic within-patient leakage [15]. The same principle underlies low-cost consumer screening, where passive detection from pulse timing on commodity wearables identifies atrial fibrillation without a diagnostic-grade waveform [16,17,18]. Beat timing is also the substrate of heart rate variability, whose measurement standards, metrics, norms, and nonlinear extensions (sample entropy and Poincare geometry) are mature and interpretable [19,20,21]. If these timing signatures generalize to the operating room, a waveform-free approach would be inexpensive, transparent, and portable, and it would sidestep the accuracy race on saturated morphology benchmarks.

The intraoperative context also sharpens the practical case for a timing-based representation. Continuous intraoperative ECG is typically acquired from a small number of monitoring leads, most often lead II, under conditions in which electrocautery, patient movement, and positioning introduce transient

baseline and morphological corruption that a morphology classifier must either reject or tolerate. R-peak timing is comparatively resilient to these corruptions, because the fiducial event survives distortions that obscure P-wave and ST-segment detail. A representation built on interval dynamics and beat composition therefore inherits a form of robustness that is specifically valuable in the operating room, and it does so while remaining computationally light enough for bedside deployment. These considerations make the waveform-free question not merely a methodological curiosity but a route to a monitor that is portable across institutions and stable across the signal-quality conditions of real surgery.

In this work, a waveform-free characterization of intraoperative arrhythmia is developed and validated on the VitalDB Arrhythmia Database. The novelty lies not in the timing-based method itself, which builds on established RR-interval approaches to rhythm discrimination [13,14,15], but in its systematic application to a previously unstudied intraoperative distribution and in the construction of a transparent, reproducible baseline for that setting. The contribution is organized around a single claim with three supporting components. First, beat-timing and beat-composition features derived from expert annotations separate the major intraoperative rhythm phenotypes without recourse to waveform morphology. Second, at the case level these features resolve a small number of reproducible arrhythmia-burden phenotypes with distinct clinical and pharmacological profiles. Third, the resulting representation generalizes across surgical contexts and is not degraded when the expert flags the signal as artifact. Two of the features are derived from the expert beat-type label layer rather than from R-peak timing alone, so the approach is described as waveform-free, meaning that no raw waveform is retrieved, rather than as strictly interval-only; this distinction is made explicit throughout, and its consequences for the interpretation of the results are examined directly. The analysis is reported transparently, including where timing alone is insufficient, and the approach is positioned as a complement to morphology-based detection and as a reproducible baseline for a setting that has, until now, lacked one.

2. Materials and Methods

2.1. Dataset and study cohort

The VitalDB Arrhythmia Database v1.0.0, distributed through PhysioNet [11,22] and derived from the VitalDB perioperative repository [12], was used in this study. The database comprises continuous single-lead (lead II) intraoperative ECG from 482 patients undergoing noncardiac surgery at a single tertiary center, with beats and rhythms annotated and adjudicated by five anesthesiologists. The cohort is arrhythmia-enriched by design, because candidate segments were identified by an automated screening step before expert review; consequently the prevalences reported here describe the character of annotated arrhythmia and are not estimates of population incidence. The analysis is thus a characterization study rather than an epidemiological one. The provided annotation files supply, for every beat, the time of the R peak (time_second), the beat type (normal, supraventricular, ventricular, or unknown), the rhythm label, and a Boolean signal-quality flag; the accompanying metadata table supplies demographic, preoperative, surgical, and intraoperative pharmacological variables. All analyses in this study were computed exclusively from these annotation and metadata fields. No raw ECG waveform was retrieved, and no external dataset was used. Cohort characteristics are summarized in Table 1.

Table 1. Clinical and surgical characteristics of the study cohort (n = 482).

Characteristic	Value
Number of cases	482
Age, years, median (IQR)	64 (56-73)

Characteristic	Value
Female sex, n (%)	200 (41.5)
Body mass index, median	22.8
ASA physical status, median	2.0
Preoperative potassium, mmol/L, median	4.1
Department: general surgery, n (%)	361 (74.9)
Department: thoracic surgery, n (%)	107 (22.2)
Department: gynecology, n (%)	8 (1.7)
Department: urology, n (%)	6 (1.2)
Emergency surgery, n (%)	77 (16.0)
Non-sinus beat burden, %, median (IQR)	12.5 (2.3-86.3)
Distinct rhythm classes per case, median	2
Analysis windows extracted, n	20,364

IQR, interquartile range; ASA, American Society of Anesthesiologists. Non-sinus beat burden is the percentage of beats within a case that fall under a non-sinus rhythm label. Department percentages are of the full cohort.

2.2. Clinically grouped rhythm scheme

The ten fine-grained rhythm labels of the database were mapped to a five-class scheme reflecting how intraoperative rhythms are grouped for clinical decision making. Sinus rhythm formed the reference class. Atrial fibrillation and atrial flutter were combined into an atrial fibrillation/flutter class. Patterned atrial ectopy, wandering atrial pacemaker with multifocal atrial tachycardia, and supraventricular tachyarrhythmia were combined into a supraventricular ectopy class. Patterned ventricular ectopy and ventricular tachycardia were combined into a ventricular ectopy class. Sinus node dysfunction and atrioventricular block were combined into a brady/conduction class. This grouping preserves clinically actionable distinctions while providing sufficient support per class for patient-level cross-validation. Ventricular tachycardia is rare within homogeneous 30-beat windows and is therefore evaluated within the ventricular ectopy group rather than as a separate class.

2.3. Analysis units

Two units of analysis were defined. For supervised rhythm classification, the beat sequence of each case was partitioned into non-overlapping windows of 30 consecutive beats. A window was retained only if a single rhythm label accounted for at least 80 percent of its beats, which assigns the window an unambiguous majority label while excluding transitional segments; 20,364 windows met this criterion. For case-level burden phenotyping and determinant analysis, each case contributed one record summarizing its arrhythmia burden and rhythm complexity: the non-sinus beat burden, the atrial fibrillation burden, the number of distinct rhythm classes, the number of rhythm transitions, the Shannon entropy of the rhythm-label distribution, the fraction of expert-flagged artifact beats, and case-level RR summaries.

The 30-beat window and the 80 percent homogeneity criterion were chosen to balance three competing requirements. A window must be long enough for the interval-variability and entropy statistics to be estimable, since short-term heart rate variability indices require on the order of dozens of intervals to stabilize [19,20]; it must be short enough that a single rhythm typically dominates, so that a majority label is meaningful; and the homogeneity threshold must exclude transitional segments in which two rhythms coexist and no single label is representative. A beat-count window rather than a fixed-duration window was used so that the number of intervals contributing to each feature is held constant across rhythms of widely differing rate, which avoids conflating window rate with window length. These choices are deliberately

conservative and standard, and they are stated explicitly so that the effect of alternative windowing can be assessed in future work.

2.4. Waveform-free feature extraction

For each 30-beat window, RR intervals were formed from successive R-peak times and physiologically implausible intervals (below 0.2 s or above 3.0 s) were removed. Fifteen features were computed. Time-domain heart rate variability comprised the mean RR interval, mean heart rate, the standard deviation of RR intervals (SDNN), the root mean square of successive differences (RMSSD), the percentage of successive differences exceeding 50 ms (pNN50), and the coefficient of variation of RR intervals, following established measurement conventions [19,20]. Poincare geometry contributed the short-term and long-term dispersion descriptors (SD1 and SD2) and their ratio. Interval-irregularity descriptors comprised the RR range and the percentages of short (below 0.6 s) and long (above 1.2 s) intervals, which capture the tachy- and brady-directional structure exploited by classical RR-only detectors [13]. Nonlinear regularity was quantified by sample entropy of the RR series with the standard embedding dimension of two and tolerance of 0.2 standard deviations [21]. Finally, beat-type composition contributed the percentages of ventricular and supraventricular beats within the window, computed from the expert beat-type labels. All features are defined in Table 2. Because the beat-type composition features are derived from the expert beat-type annotation layer rather than from RR timing alone, the overall approach is described as waveform-free rather than as strictly RR-interval-only; the provenance and its implications are addressed explicitly in the Discussion.

Table 2. Beat-annotation-derived feature dictionary and overall distribution across the 20,364 analysis windows.

Feature	Description	Mean	SD	Median	IQR
meanRR	Mean RR interval (s)	0.796	0.185	0.778	0.662-0.922
meanHR	Mean heart rate (beats/min)	79.64	19.66	77.12	65.10-90.67
SDNN	SD of RR intervals (s)	0.087	0.089	0.059	0.010-0.145
RMSSD	RMS of successive RR differences (s)	0.122	0.135	0.077	0.007-0.207
pNN50	% successive RR differences > 50 ms	32.20	36.01	10.35	0.00-72.41
CVrr	Coefficient of variation of RR	0.110	0.106	0.078	0.012-0.196
SD1	Poincare short-term variability (s)	0.088	0.097	0.055	0.005-0.149
SD2	Poincare long-term variability (s)	0.081	0.084	0.050	0.012-0.134
SD1_SD2	Ratio of SD1 to SD2	1.043	2.209	0.879	0.505-1.217
RRrange	Range of RR intervals (s)	0.351	0.353	0.272	0.036-0.592
pShortRR	% RR intervals < 0.6 s	17.07	29.02	0.00	0.00-20.00
pLongRR	% RR intervals > 1.2 s	3.98	13.05	0.00	0.00-0.00
SampEn	Sample entropy of the RR series	1.096	0.562	1.099	0.693-1.386
pV_beats	% ventricular beats in the window	1.92	6.94	0.00	0.00-0.00
pS_beats	% supraventricular beats in the window	28.95	43.03	0.00	0.00-90.00

SD, standard deviation; IQR, interquartile range; RMS, root mean square. The last two features are derived from the expert beat-type labels; all others are derived from R-peak timing.

2.5. Supervised rhythm classification

The five-class scheme was classified from the fifteen features using five-fold stratified group cross-validation, with the surgical case as the grouping unit so that no window from a given patient appeared in both training and test folds. This patient-level partitioning implements the inter-patient evaluation paradigm and prevents the optimistic leakage that arises when windows from the same patient are split across folds [15]. An L2-regularized logistic regression on standardized features (balanced class weights) served as the primary, interpretable classifier, and a random forest (300 trees, balanced class weights) was included as a higher-capacity nonlinear comparator to test whether additional model flexibility improved separability.

Balanced class weighting was used in both models because the five rhythm classes are strongly imbalanced, with sinus and atrial fibrillation/flutter windows outnumbering the ectopy and conduction classes by more than an order of magnitude; without reweighting, a classifier can achieve high overall accuracy while ignoring the minority classes that are of greatest clinical interest. For the same reason, performance was summarized by balanced accuracy, which averages per-class recall and is therefore not inflated by the majority classes, and by macro-averaged F1, which weights all classes equally. Both were reported as the mean and standard deviation across folds; per-class precision, recall, and F1 were computed from the pooled out-of-fold predictions. To quantify uncertainty beyond a single split, the patient-level five-fold cross-validation was repeated over ten random seeds, and balanced accuracy and macro F1 were summarized by their mean and a 95 percent percentile confidence interval across the resulting fold estimates. Calibration of the primary linear model was assessed on the pooled out-of-fold predicted probabilities using the multiclass Brier score and the top-label expected calibration error. Evaluating a linear model alongside the random forest was a deliberate design choice: if a linear model performs at least as well as a high-capacity one, the measured separability reflects genuine structure in the feature space rather than the flexibility of the classifier. The per-class confusion structure and permutation importance are reported from the random forest as a conservative nonlinear characterization of where residual error lies; permutation importance was measured as the decrease in balanced accuracy on a held-out patient split when each feature was permuted. To isolate the contribution of the two beat-type-composition features, a strict interval-only ablation was performed by removing the ventricular and supraventricular beat fractions and repeating the classification with the remaining thirteen R-peak-timing features under identical cross-validation and models.

2.6. Unsupervised arrhythmia-burden phenotyping

Case-level phenotypes were derived by k-means clustering on ten standardized burden and complexity descriptors (ventricular and supraventricular beat fractions, non-sinus and atrial fibrillation burden, number of distinct rhythms, number of transitions, rhythm entropy, and case-level mean RR, SDNN, and RMSSD). The number of clusters was examined over k from two to six using the silhouette coefficient, and stability was quantified by the adjusted Rand index between the reference labeling and 25 clusterings of 80 percent subsamples. As reported in the Results, the silhouette coefficient was maximized at $k = 2$; the three-cluster solution was nonetheless retained because it resolves a clinically meaningful separation of an atrial fibrillation/flutter-predominant group from a mixed-ectopy intermediate group, and because that three-cluster solution was itself highly reproducible under resampling. Clusters were ordered by mean non-sinus burden and assigned descriptive names. This choice is a deliberate, disclosed departure from silhouette-optimal partitioning in favor of clinical interpretability; it does not affect the supervised results, which are independent of the clustering.

2.7. Cross-context generalization and artifact robustness

Two evaluations probed whether the waveform-free representation is stable rather than idiosyncratic to a single distribution. Generalization was tested by department-out transfer: a random forest was trained on all windows from one surgical department and evaluated on all windows from the other, in both directions (general surgery and thoracic surgery, the two departments with adequate window counts). Robustness was tested using the expert signal-quality flag: the pooled out-of-fold predictions were stratified into clean and artifact-flagged windows, with a window designated artifact when more than 20 percent of its beats carried the flag. To ensure that the clean-versus-artifact comparison was not confounded by differing class composition, both subsets were restricted to the rhythm classes present among artifact windows before balanced accuracy was computed.

2.8. Determinants of high arrhythmia burden

Determinants of high intraoperative arrhythmia burden were examined by multivariable logistic regression. The binary outcome contrasted cases with above-median against at-or-below-median non-sinus beat burden. Candidate predictors were age, sex, body mass index, ASA physical status, preoperative potassium, emergency status, and cumulative intraoperative doses of ephedrine, phenylephrine, and calcium. Odds ratios with 95 percent confidence intervals are reported. These estimates are explicitly associational. In a retrospective single-center cohort, intraoperative drug exposures are administered in response to evolving hemodynamics, so any drug-burden association is subject to confounding by indication and possible reverse causation and must not be read causally.

2.9. Statistical analysis and software

Continuous features are summarized by median and interquartile range or by mean and standard deviation as indicated. Differences in feature distributions across rhythm classes were tested by the Kruskal-Wallis test. A two-sided p value below 0.05 was considered significant. Analyses were implemented in Python with scikit-learn for classification, clustering, and cross-validation and statsmodels for logistic regression [23]. The complete single-file pipeline is available with this article, and all figures and tables were generated directly from it, so that every reported value is reproducible from the annotation and metadata files alone.

2.10. Ethics declaration

This study is a retrospective secondary analysis of the VitalDB Arrhythmia Database v1.0.0, a fully de-identified dataset that is publicly available through PhysioNet. Ethical approval for the collection of the primary data and the associated handling of informed consent were obtained by the investigators responsible for the source repository, as reported in the primary data descriptor [11]. All methods were carried out in accordance with the relevant guidelines and regulations. Because only publicly available, de-identified data were used, no additional ethical approval was required for the present secondary analysis.

[Author to verify against the primary data descriptor and, where required by the journal, insert the name of the approving institutional review board, the approval or protocol reference number, and the statement regarding informed consent.]

3. Results

3.1. Cohort and arrhythmia-burden landscape

The cohort had a median age of 64 years (IQR 56 to 73) and was 41.5 percent female, with a median ASA physical status of 2 and a median preoperative potassium of 4.1 mmol/L (Table 1). General surgery accounted for 74.9 percent of cases and thoracic surgery for 22.2 percent, with small gynecology and urology strata; 16.0 percent of procedures were emergent. Consistent with the screening-based, arrhythmia-enriched construction of the database, the case-level non-sinus beat burden was highly skewed, with a median of 12.5 percent but an interquartile range spanning 2.3 to 86.3 percent.

The rhythm landscape was dominated by sinus rhythm and atrial fibrillation/flutter at the beat level, with the remaining rhythm classes contributing far fewer beats but appearing across many cases (Figure 1). The case-level view revealed marked heterogeneity: a large group of cases carried little non-sinus burden, a second group occupied an intermediate range, and a distinct group was almost entirely non-sinus, while rhythm complexity (the number of distinct rhythms and of transitions per case) increased with intermediate burden and then collapsed in the predominantly-arrhythmic cases that remained in a single sustained rhythm

(Figure 2). This bimodal-to-trimodal structure at the case level motivated the phenotyping analysis reported below.

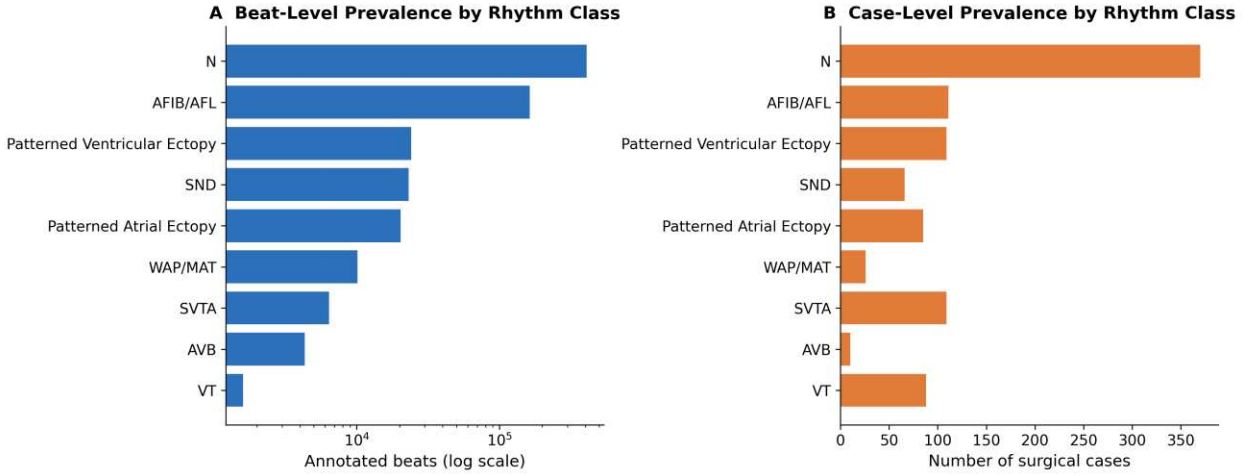

Figure 1. Prevalence of rhythm classes in the cohort. (A) Beat-level prevalence, shown on a logarithmic scale, is dominated by sinus rhythm and atrial fibrillation/flutter. (B) Case-level prevalence shows that less frequent rhythm classes nonetheless appear across many surgical cases. The contrast between the two panels illustrates that beat counts and case counts convey different information about arrhythmia distribution. These are prevalences within an arrhythmia-enriched screening cohort and do not represent population incidence.

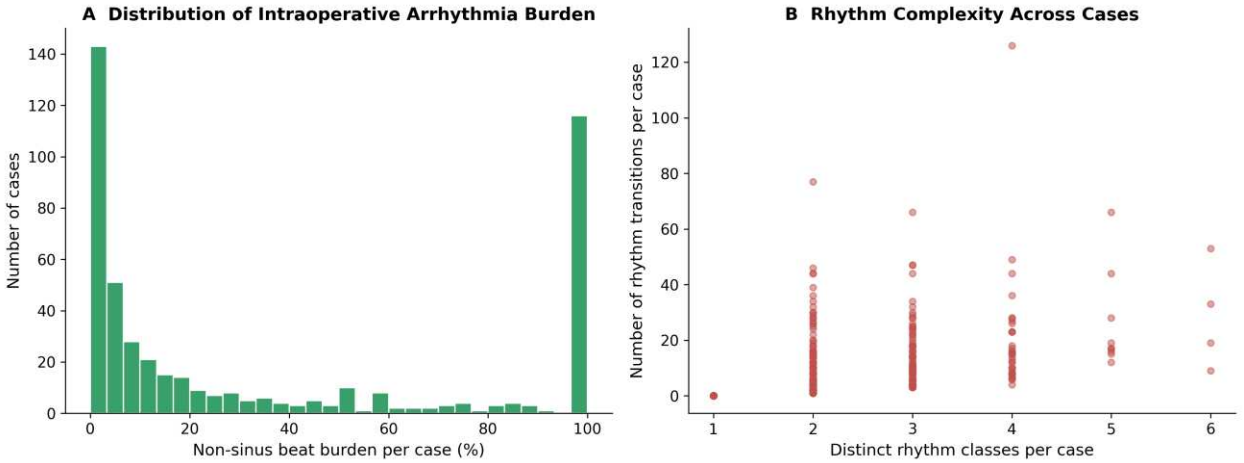

Figure 2. Case-level arrhythmia burden and rhythm complexity. (A) The distribution of non-sinus beat burden per case is strongly bimodal, with concentrations at low burden and near-total burden. (B) Rhythm complexity, expressed as the number of transitions against the number of distinct rhythm classes per case, is greatest for cases of intermediate burden and lowest for cases that remain in a single sustained rhythm.

3.2. Beat-timing features separate rhythm phenotypes

A two-dimensional principal-component projection of the standardized feature space showed clear organization by rhythm class, with sinus windows forming a compact low-variability region and the arrhythmic classes dispersing along axes of increasing interval variability and irregularity (Figure 3). Every feature differed significantly across the five rhythm classes by the Kruskal-Wallis test (all $p < 0.001$; Table 3). The separation was physiologically coherent. Sinus windows had near-zero variability (median SDNN 0.014 s, RMSSD 0.010 s, pNN50 0.0 percent). Atrial fibrillation/flutter windows combined high variability with a defining supraventricular beat composition (median pS_beats 100 percent). Ventricular ectopy

windows carried the highest short-term variability (median RMSSD 0.338 s, SD1 0.243 s) together with a ventricular beat signature (median pV_beats 33.3 percent). Brady/conduction windows were distinguished by long intervals (median mean RR 0.978 s, mean heart rate 61 beats/min, and the only nonzero long-RR fraction). Supraventricular ectopy occupied an intermediate position, which foreshadows its comparatively lower separability under classification.

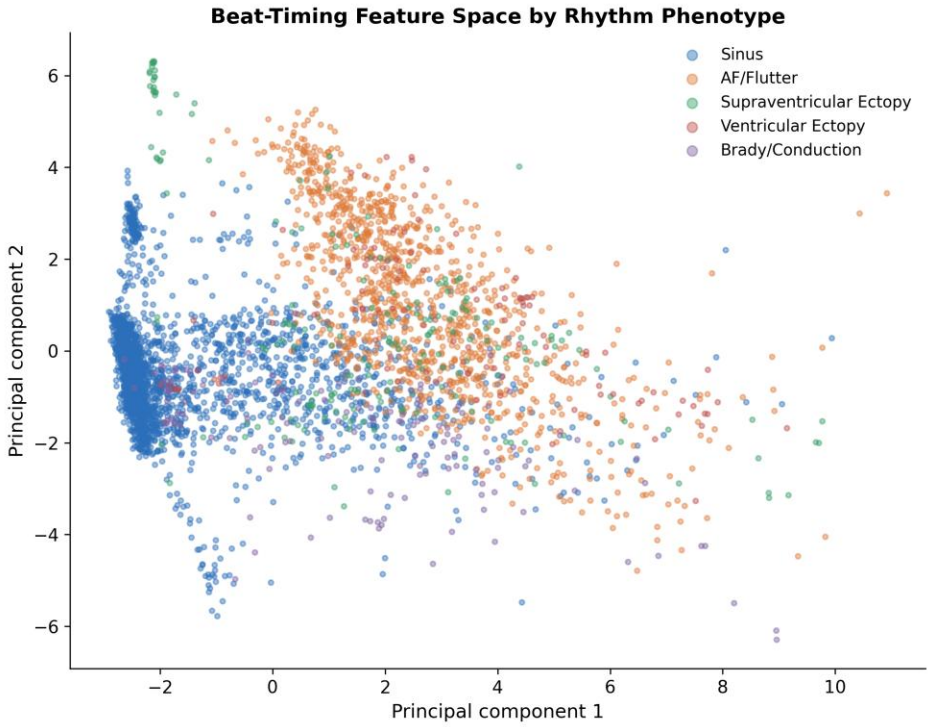

Figure 3. Principal-component projection of the standardized beat-annotation-derived feature space, colored by rhythm class. Sinus windows form a compact low-variability cluster, while the arrhythmic classes separate along axes of interval variability and irregularity. The projection is descriptive; classification (Section 3.3) is performed in the full feature space under patient-level cross-validation.

Table 3. Median feature values by rhythm class and Kruskal-Wallis test across classes.

Feature	Sinus	AF/Flut.	SVE	VE	Brady	H	p
meanRR	0.793	0.703	0.784	0.734	0.978	1707	<0.001
meanHR	75.66	85.32	76.53	81.73	61.37	1707	<0.001
SDNN	0.014	0.150	0.154	0.201	0.142	9147	<0.001
RMSSD	0.010	0.208	0.238	0.338	0.182	8937	<0.001
pNN50	0.00	79.31	72.41	93.10	31.59	13420	<0.001
CVrr	0.016	0.212	0.197	0.251	0.143	10515	<0.001
SD1	0.007	0.150	0.171	0.243	0.131	8936	<0.001
SD2	0.017	0.147	0.114	0.105	0.147	8999	<0.001
SD1_SD2	0.646	1.022	1.433	1.734	0.813	3181	<0.001
RRrange	0.053	0.597	0.486	0.547	0.575	7462	<0.001
pShortRR	0.00	26.67	23.33	30.00	0.00	5739	<0.001
pLongRR	0.00	0.00	0.00	0.00	6.67	2709	<0.001
SampEn	1.099	1.099	0.523	0.322	0.772	2857	<0.001
pV_beats	0.00	0.00	0.00	33.33	0.00	5073	<0.001
pS_beats	0.00	100.0	30.00	0.00	0.00	16167	<0.001

AF/Flut., atrial fibrillation/flutter; SVE, supraventricular ectopy; VE, ventricular ectopy; Brady, brady/conduction; H, Kruskal-Wallis statistic. Values are class medians. The Kruskal-Wallis test evaluates whether feature distributions differ across the five classes.

3.3. Waveform-free multiclass classification

Under patient-level five-fold cross-validation, the transparent linear model, adopted as the primary classifier, separated the five rhythm classes with a balanced accuracy of 0.869 (SD 0.039). A random forest, included as a higher-capacity comparator, reached a balanced accuracy of 0.811 (SD 0.018) and a macro F1 of 0.840 (SD 0.028) and therefore did not improve on the linear model. Repeating the patient-level cross-validation over ten random seeds confirmed the stability of these estimates, yielding a mean balanced accuracy of 0.862 (95 percent confidence interval 0.793 to 0.936) for the linear model and 0.816 (95 percent confidence interval 0.741 to 0.890) for the random forest. This comparison is reported directly because it is informative rather than incidental: the competitiveness of a simple standardized logistic regression indicates that the class structure in the feature space is largely linearly separable, so the discriminative signal is a genuine property of beat timing and composition and not an artifact of a high-capacity model. It also implies that the waveform-free representation can be deployed as a transparent linear rule when interpretability or computational economy is paramount. The per-class and confusion analyses that follow are reported from the random forest, which provides a conservative characterization of the residual error.

Per-class performance was uneven in a clinically interpretable way (Table 4, Figure 4). Sinus rhythm (F1 0.970), atrial fibrillation/flutter (F1 0.991), and ventricular ectopy (F1 0.906) were classified with high fidelity, reflecting their distinctive timing and composition signatures. Supraventricular ectopy was the weakest class (recall 0.554, F1 0.661): windows of patterned atrial ectopy are frequently confused with sinus rhythm because a modest burden of atrial premature beats perturbs interval timing only slightly, a distinction that P-wave morphology would resolve but interval statistics cannot. Brady/conduction rhythms were intermediate (recall 0.642, F1 0.708). Although these minority classes contribute fewer windows, they were drawn from many patients (77, 68, and 64 distinct patients for supraventricular ectopy, ventricular ectopy, and brady/conduction, respectively, against 366 for sinus and 108 for atrial fibrillation/flutter), so their estimates do not rest on a handful of individuals. The confusion structure in Figure 4 confirms that the residual error is concentrated in these two minority classes rather than distributed across the well-separated majority classes, and it bounds the claim of the waveform-free approach precisely: it recovers the dominant intraoperative rhythms with high accuracy while leaving a defined subset of ectopy for which morphology remains advantageous. Beyond discrimination, the predicted probabilities of the primary linear model were well calibrated on the pooled out-of-fold predictions, with a multiclass Brier score of 0.142 and a top-label expected calibration error of 0.018, and the reliability diagram lies close to the identity line across the range of predicted probability (Figure 5).

Table 4. Classification performance by rhythm class from pooled patient-level out-of-fold predictions.

Rhythm class	Precision	Recall	F1-score	Support
Sinus	0.954	0.986	0.970	12,709
AF/Flutter	0.986	0.995	0.991	5,365
Supraventricular Ectopy	0.819	0.554	0.661	929
Ventricular Ectopy	0.937	0.877	0.906	561
Brady/Conduction	0.787	0.642	0.708	800
Macro average	0.897	0.811	0.847	20,364
Balanced accuracy (linear model, primary)			0.869 +/- 0.039	
Balanced accuracy (random forest, comparator)			0.811 +/- 0.018	

Recall is equivalent to sensitivity. Support is the number of windows; the corresponding numbers of distinct patients contributing to each class were 366 (sinus), 108 (atrial fibrillation/flutter), 77 (supraventricular ectopy), 68 (ventricular ectopy), and 64 (brady/conduction). The two balanced-accuracy rows are means across the five patient-level folds. Per-class values are computed from the pooled out-of-fold predictions of the random forest and are reported as a conservative nonlinear characterization of the residual error.

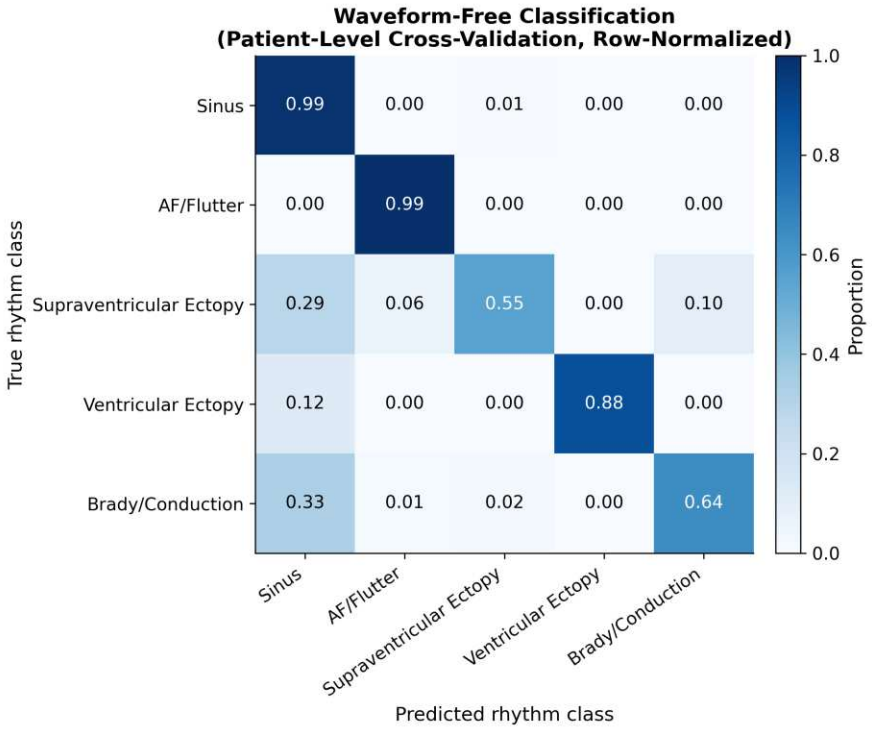

Figure 4. Row-normalized confusion matrix for waveform-free five-class rhythm classification under patient-level cross-validation. The dominant rhythms (sinus and atrial fibrillation/flutter) and ventricular ectopy are recovered with high accuracy, whereas residual error concentrates in supraventricular ectopy, which is most often confused with sinus rhythm.

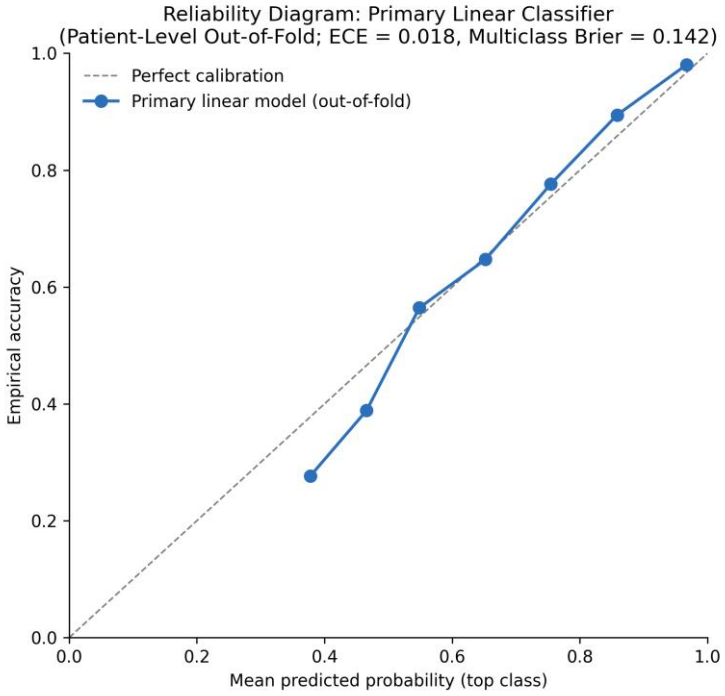

Figure 5. Reliability diagram for the primary linear classifier under patient-level out-of-fold prediction. Each point relates the mean predicted probability of the top class to the empirical accuracy within a probability bin; proximity to

the identity line indicates good calibration. The model is well calibrated across the range of predicted probability (multiclass Brier score 0.142, top-label expected calibration error 0.018).

Permutation importance identified the drivers of the classification (Figure 6). The supraventricular beat fraction, the successive-difference variability measures (pNN50, RMSSD, SD1), and the coefficient of variation contributed most, consistent with the class medians in Table 3 and with the physiology of the arrhythmias in question. Importantly, the leading contribution of the supraventricular beat fraction reflects the correlation, by construction, between the beat-type and rhythm annotation layers, and this correlation is examined as a scope condition in the Discussion.

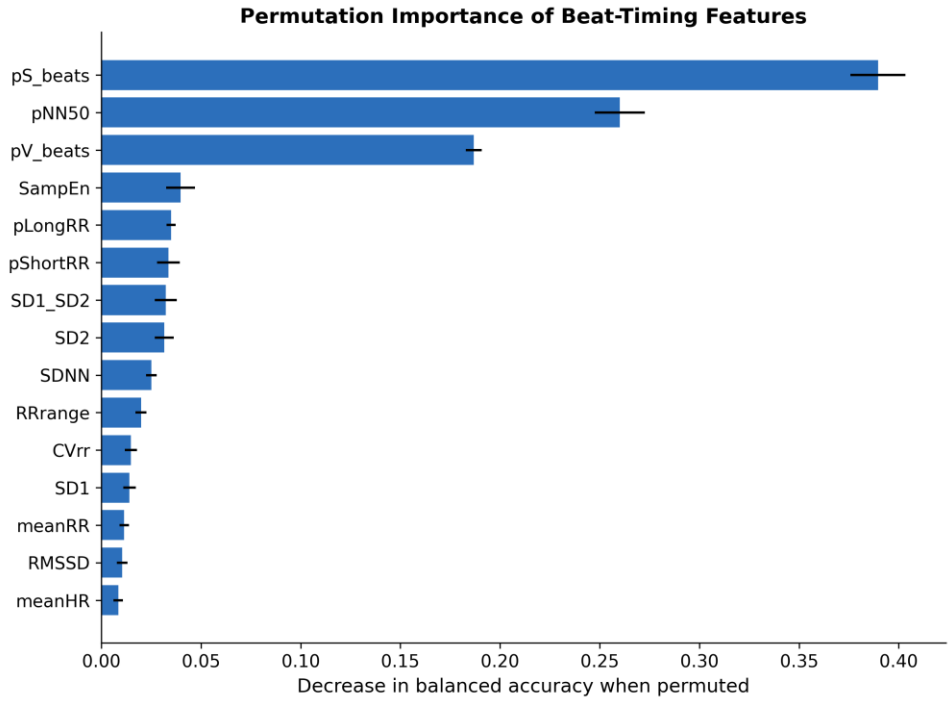

Figure 6. Permutation importance of the beat-annotation-derived features, measured as the decrease in balanced accuracy on a held-out patient split when each feature is permuted. Successive-difference variability, the coefficient of variation, and the supraventricular beat fraction dominate; error bars denote the standard deviation across ten permutation repeats.

To quantify how much of the discrimination depends on the two beat-type-composition features rather than on R-peak timing alone, a strict interval-only model was fitted after removing the ventricular and supraventricular beat fractions (Table 8). Balanced accuracy fell from 0.869 to 0.653 for the linear model and from 0.811 to 0.581 for the random forest. The loss was strongly class-specific. Sinus rhythm and atrial fibrillation/flutter, which are defined by gross interval structure, were essentially retained (random-forest recall 0.979 and 0.965). The classes defined by beat type collapsed: supraventricular ectopy recall fell from 0.554 to 0.223, ventricular ectopy from 0.877 to 0.458, and brady/conduction from 0.642 to 0.278. This ablation delineates the scope of the approach precisely: interval timing alone recovers the rhythms of gross irregularity, whereas resolving the ectopy classes requires the expert beat-type labels, which are themselves morphology-informed.

Table 8. Strict interval-only ablation: classification performance with and without the two beat-type-composition features.

Feature set	Balanced acc. (linear)	Balanced acc. (RF)	SVE recall	VE recall	Brady recall
Full (15 features, waveform-free)	0.869 +/- 0.039	0.811 +/- 0.018	0.554	0.877	0.642
Interval-only (13 features, RR timing)	0.653 +/- 0.044	0.581 +/- 0.033	0.223	0.458	0.278

Balanced accuracy is the mean and standard deviation across the five patient-level folds for the primary linear model and the random-forest comparator. Per-class recall values are from the random forest, consistent with Table 4. The interval-only set removes the ventricular and supraventricular beat fractions. SVE, supraventricular ectopy; VE, ventricular ectopy; RF, random forest.

3.4. Reproducible arrhythmia-burden phenotypes

Unsupervised clustering of case-level burden and complexity descriptors resolved a small number of coherent phenotypes. The silhouette coefficient was highest at two clusters (0.438) and declined for larger k (0.352 at three, 0.290 at four). The three-cluster solution was retained; it separated an atrial fibrillation/flutter-predominant group from a mixed-ectopy intermediate group that the two-cluster solution merges with the low-burden group, and it was adopted because this distinction is clinically meaningful and because the three-cluster partition was highly reproducible under resampling (bootstrap adjusted Rand index 0.82; Figure 7). The three phenotypes were readily interpretable (Table 5). The near-sinus/low-burden phenotype (n = 256) had a median non-sinus burden of 2.6 percent and a median age of 62 years. The mixed-ectopy/intermediate-burden phenotype (n = 121) had a median non-sinus burden of 32.1 percent, the greatest rhythm complexity (median three distinct rhythms and fifteen transitions), and a median age of 66 years. The atrial fibrillation/flutter-predominant phenotype (n = 105) was defined by a non-sinus and atrial fibrillation burden of 100 percent, minimal rhythm complexity (a single sustained rhythm with no transitions), and the oldest median age (69 years).

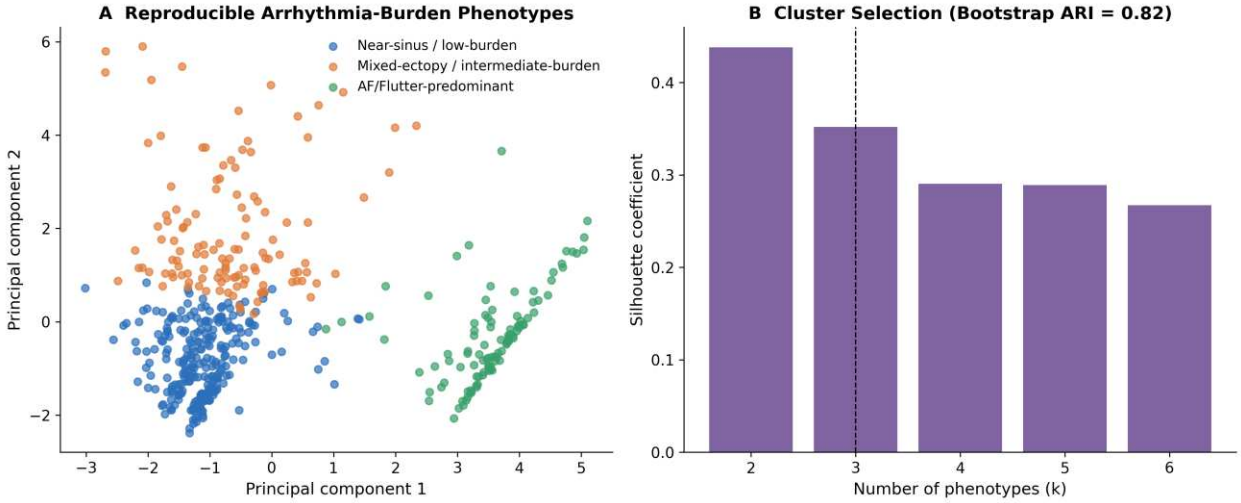

Figure 7. Reproducible arrhythmia-burden phenotypes. (A) Principal-component projection of the case-level burden and complexity space, colored by the three retained phenotypes. (B) Silhouette coefficient by number of clusters; the coefficient is maximized at k = 2, and the retained three-cluster solution (marked) is adopted for clinical interpretability and remains highly reproducible under resampling (bootstrap adjusted Rand index 0.82).

Table 5. Profiles of the three arrhythmia-burden phenotypes (case-level medians).

Phenotype	n	Non-sinus burden (%)	AF burden (%)	Ventricular beats (%)	Distinct rhythms	Transitions	Age (yr)
Near-sinus / low-burden	256	2.6	0	0.19	2	4	62
Mixed-ectopy / intermediate-burden	121	32.1	0	0.69	3	15	66
AF/Flutter-predominant	105	100.0	100	0	1	0	69

Values are per-phenotype medians. AF, atrial fibrillation. Rhythm complexity (distinct rhythms and transitions) peaks in the intermediate-burden phenotype and is minimal in the predominantly-arrhythmic phenotype, which remains in a single sustained rhythm.

The phenotypes carried distinct clinical and pharmacological signatures (Figure 8). The atrial fibrillation/flutter-predominant phenotype was the oldest and displayed an intraoperative pharmacological profile that differed from the low-burden phenotype, while the mixed-ectopy phenotype was intermediate on most axes. These signatures are descriptive summaries of standardized median values across the phenotypes and are intended to characterize, not to explain, the groups; formal adjusted associations are examined in Section 3.6.

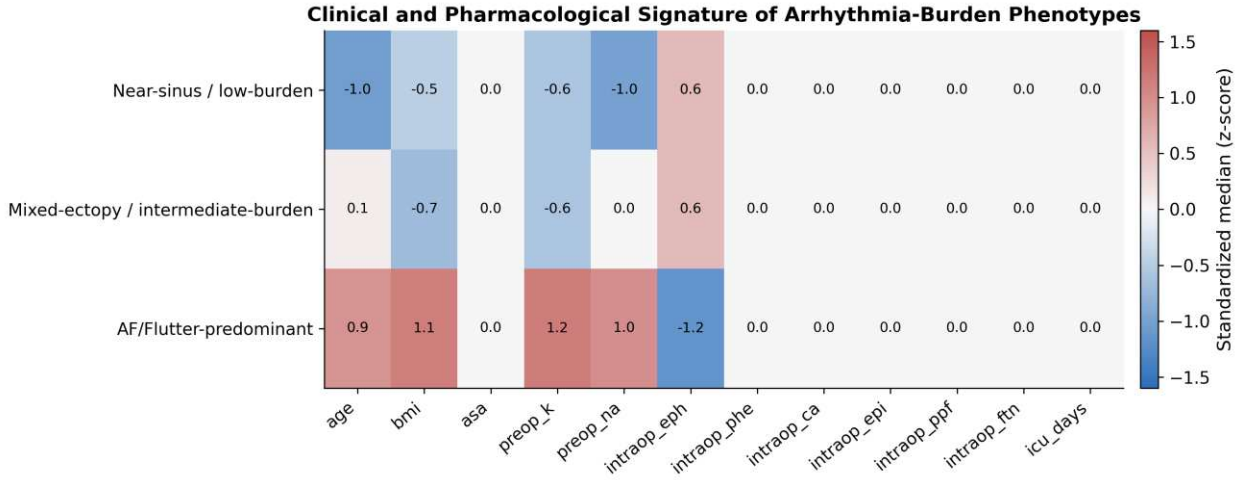

Figure 8. Clinical and pharmacological signature of the arrhythmia-burden phenotypes, shown as standardized median values (z-scores across phenotypes) for demographic, physiological, and intraoperative pharmacological variables. Warmer cells denote higher and cooler cells lower relative values. The heatmap characterizes the phenotypes and is not a causal model.

3.5. Cross-surgical-context generalization and artifact robustness

The waveform-free representation transferred between the two surgical departments of this single-center cohort (Figure 9, Table 6). A classifier trained on general surgery and evaluated on thoracic surgery achieved a balanced accuracy of 0.793 (macro F1 0.823), and the reverse transfer achieved 0.732 (macro F1 0.783), against a within-cohort reference of 0.811. The asymmetry is expected: the thoracic training set was small (4,671 windows) relative to the general test set (15,091 windows), so the thoracic-to-general direction is the more demanding transfer. That both directions remained well above chance for a five-class problem indicates that the timing and composition signatures of intraoperative rhythms are not idiosyncratic to one surgical population.

The representation was also robust to signal-quality degradation. Restricting both subsets to shared rhythm classes, balanced accuracy on expert-flagged artifact windows (0.878, n = 989) was not lower than on clean windows (0.810, n = 19,375). This result is consistent with the mechanism of the approach: R-peak timing is comparatively resilient to the baseline wander and morphological distortion that trigger the quality flag, whereas a morphology classifier would be expected to degrade on precisely these windows. The comparison should be read as evidence of non-degradation rather than of superior performance on artifact, because the artifact subset is small and its shared-class composition is enriched for rhythms, such as grossly irregular atrial fibrillation, that timing features classify most readily.

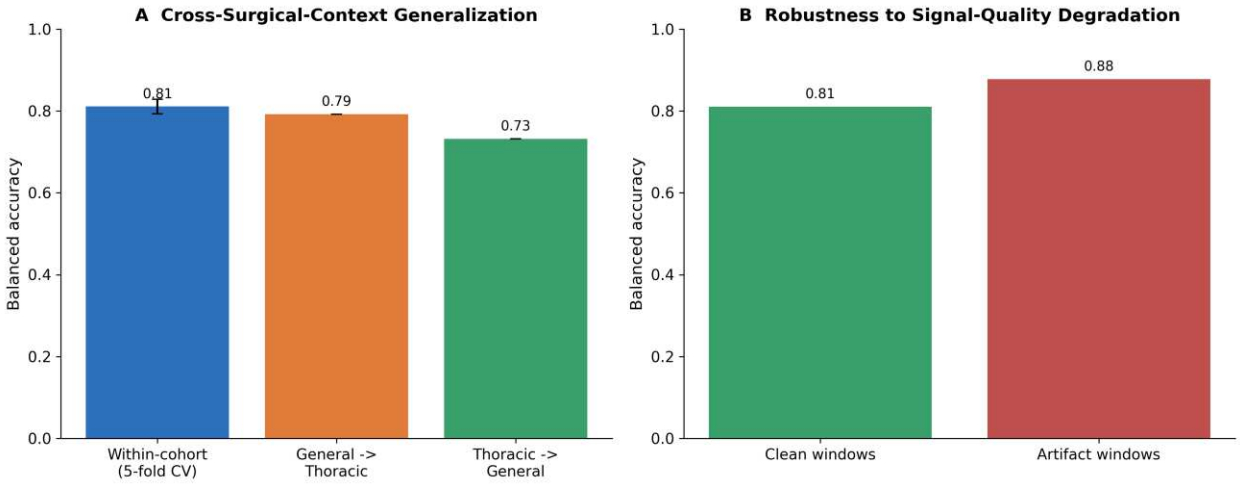

Figure 9. Generalization and robustness of waveform-free classification. (A) Balanced accuracy within the cohort and under department-out transfer between general and thoracic surgery. (B) Balanced accuracy on clean versus expert-flagged artifact windows, restricted to shared rhythm classes. Performance is retained across surgical contexts and is not degraded on artifact-flagged windows.

Table 6. Cross-context generalization and artifact robustness of waveform-free classification.

Evaluation scenario	Balanced accuracy	Macro F1	n windows
Within-cohort (5-fold, patient-level)	0.811	0.840	20,364
Transfer: general to thoracic surgery	0.793	0.823	4,671
Transfer: thoracic to general surgery	0.732	0.783	15,091
Clean-signal windows (shared classes)	0.810	n/a	19,375
Artifact windows (shared classes)	0.878	n/a	989

For transfer rows, n windows is the size of the test department. Clean and artifact rows are restricted to rhythm classes shared between the two subsets so that the comparison is not confounded by class composition.

3.6. Determinants of high arrhythmia burden

In the multivariable logistic model contrasting above-median with at-or-below-median non-sinus burden, older age was the principal adjusted correlate of high arrhythmia burden (odds ratio 1.035 per year, 95 percent confidence interval 1.017 to 1.052, $p < 0.001$; Figure 10, Table 7). Sex, body mass index, ASA physical status, preoperative potassium, and emergency status were not significantly associated with high burden. Two intraoperative drug exposures reached statistical significance but were of small magnitude and inverse direction: ephedrine (odds ratio 0.979 per mg, $p = 0.016$) and calcium (odds ratio 0.999 per mg, $p = 0.006$), the latter with an effect size that is negligible in practical terms because the odds ratio is essentially unity. Because the raw dosing units differ across agents, the drug associations were also expressed per standard deviation of administered dose: ephedrine 0.740 (95 percent confidence interval 0.579 to 0.945) and calcium 0.544 (0.353 to 0.838) per standard deviation, with phenylephrine not significant (1.266, 0.991 to 1.616). This rescaling does not change the interpretation, since these exposures are administered in response to the evolving intraoperative state. These drug associations are most plausibly explained by confounding by indication rather than by any protective effect, because vasoactive and electrolyte therapy is administered in response to the evolving intraoperative state; they are reported for completeness and should not be interpreted causally. The consistent and interpretable finding is that age structures intraoperative arrhythmia burden, in agreement with the age gradient observed across the burden phenotypes in Table 5.

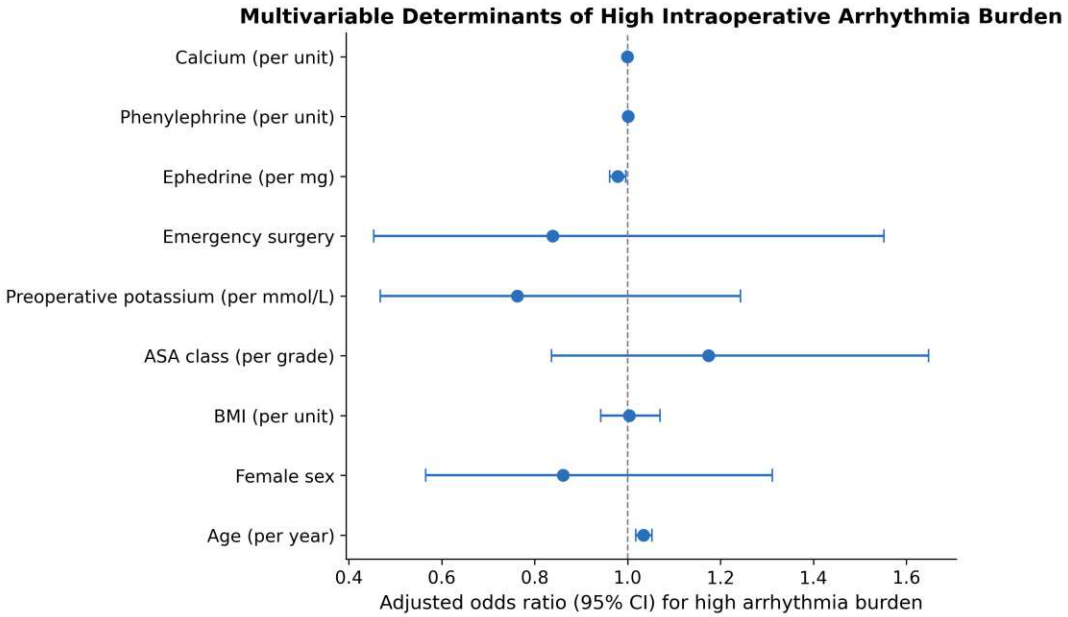

Figure 10. Adjusted determinants of high intraoperative arrhythmia burden from multivariable logistic regression, shown as odds ratios with 95 percent confidence intervals on a logarithmic axis. The vertical reference line denotes an odds ratio of one. Older age is the principal adjusted correlate; the significant drug associations are of small magnitude and are interpreted as associational.

Table 7. Multivariable logistic regression for high (above-median) intraoperative arrhythmia burden.

Predictor	Odds ratio	95% CI	p-value
Age (per year)	1.035	1.017-1.052	<0.001
Female sex	0.861	0.565-1.312	0.486
Body mass index (per unit)	1.004	0.942-1.070	0.907
ASA class (per grade)	1.174	0.836-1.648	0.354
Preoperative potassium (per mmol/L)	0.763	0.468-1.243	0.277
Emergency surgery	0.839	0.454-1.552	0.576
Ephedrine (per mg)	0.979	0.962-0.996	0.016
Phenylephrine (per microgram)	1.001	1.000-1.002	0.059
Calcium (per milligram)	0.999	0.999-1.000	0.006

CI, confidence interval; ASA, American Society of Anesthesiologists. The outcome is above-median non-sinus beat burden. Estimates are associational and subject to confounding by indication for the intraoperative drug exposures. Ephedrine is expressed per milligram, phenylephrine per microgram, and calcium per milligram.

4. Discussion

4.1. Principal findings

Using only the expert annotation layer and routine metadata of the first large intraoperative arrhythmia database, the present analysis showed that intraoperative arrhythmia can be characterized without the raw ECG waveform. Beat-timing and beat-composition features separated the major intraoperative rhythm classes under strict patient-level cross-validation, with a balanced accuracy near 0.87 for the primary transparent linear model, which matched or exceeded a higher-capacity nonlinear comparator, indicating that the discriminative structure is real and largely linear. A strict interval-only ablation localized the source of this discrimination: interval timing alone recovered the rhythms of gross irregularity, whereas the ectopy classes depended on the expert beat-type labels, so the balanced accuracy fell to 0.653 when those labels were withheld. At the case level, the same information resolved three reproducible arrhythmia-burden

phenotypes with distinct age and pharmacological profiles. The representation transferred between surgical departments within the single-center cohort and was not degraded on windows the expert flagged as artifact, and older age emerged as the principal adjusted correlate of high burden. Taken together, these results establish that a substantial part of intraoperative arrhythmia analysis is achievable from beat annotations alone.

4.2. Relation to prior work

The finding that beat timing carries major rhythm information is consistent with a long line of work outside the operating room. Atrial fibrillation has been detected from RR-interval irregularity since classical density-histogram methods [13], modern RR-only classifiers remain accurate [14], and heartbeat interval features have complemented morphology in canonical beat classification [15]. The clinical translation of timing-based detection is visible in consumer wearables that screen for atrial fibrillation from pulse timing without a diagnostic waveform [16,17]. More recent work reinforces this: RR-interval features remain competitive for atrial fibrillation detection with contemporary machine learning [24,25], and deep models operating on beat-to-beat intervals generalize across geography, age, and sex while retaining the noise robustness conferred by the R-peak fiducial [26]. The contribution of the present study is to demonstrate that these principles hold in the intraoperative setting, which is characterized by rapid physiological transitions and distinctive artifacts absent from ambulatory and critical-care corpora such as MIT-BIH and PTB-XL [7,8]. Whereas the prevailing trajectory on those corpora is toward larger morphology models that approach cardiologist-level performance [9,10,27], the marginal value of that trajectory is limited by benchmark saturation. Moving to a waveform-free framing on a genuinely new distribution avoids competing on a saturated axis and instead establishes a transparent, reproducible baseline where none previously existed. The novelty is thus one of setting, dataset, and baseline rather than of algorithm.

A further point of contact with prior methodology concerns evaluation rigor. A recurring source of optimism in ECG classification is within-patient leakage, in which windows or beats from the same individual populate both training and test partitions, so that the model can exploit patient-specific idiosyncrasy rather than generalizable rhythm structure. The inter-patient paradigm was introduced precisely to guard against this [15], and it was adopted throughout by grouping all windows of a case together under stratified group cross-validation. The department-out transfer imposes an even stricter condition, requiring generalization across surgical populations that differ in comorbidity and procedure. That performance degraded only modestly under these constraints, rather than collapsing, is therefore a stronger statement than a high within-patient accuracy would have been, and it is consistent with the view that beat timing encodes rhythm information that is portable rather than patient-specific. The importance of assessing both discrimination and calibration when comparing ECG-based predictors has likewise been emphasized in recent atrial fibrillation prediction studies [28].

The three burden phenotypes align with clinical intuition and with the perioperative literature. An older, atrial fibrillation-predominant group is concordant with the established age dependence and clinical weight of perioperative atrial fibrillation [4], and the separation of a complex mixed-ectopy intermediate group is consistent with the observation that intraoperative arrhythmia varies with surgical context and comorbidity [2,6]. That older age was the dominant adjusted correlate of high burden reinforces this concordance and is unsurprising given the physiology, but it is here derived entirely from annotation-plus-metadata data without any waveform input.

4.3. Clinical and methodological implications

The practical significance of a waveform-free approach follows from what it does not require. Because the analysis depends only on R-peak times, beat-type labels, and metadata, it is inexpensive to compute, does not require transmission or storage of the raw waveform, and yields features that are individually interpretable in physiological terms. This makes it well suited to settings where waveform-grade infrastructure is unavailable, where interpretability is required for clinical acceptance, or where a transparent reference is needed against which more complex models can be judged. The clinical value of annotation-driven and AI-assisted electrocardiography is underscored by recent evidence that ECG-based artificial intelligence can influence hard outcomes at scale [29] and by its expanding role across healthcare [30]. The demonstrated cross-context transfer and artifact resilience are properties that matter specifically in the operating room, where signal quality is variable and where a monitor must behave predictably across surgical populations. The burden-phenotyping component further suggests a route to monitoring prioritization: characterizing which physiological and pharmacological states accompany elevated arrhythmia burden can help direct attention, even though the associations reported here are not causal.

The competitiveness of the linear baseline has an additional, practical consequence. Because a standardized logistic regression matched the random forest on balanced accuracy, the waveform-free layer can be expressed as a compact, auditable linear rule whose coefficients are directly inspectable, rather than as an opaque ensemble. In a domain where clinical acceptance depends on the ability to explain why an alarm fired, a representation that is both accurate and expressible in closed form is a substantive advantage. Much recent effort has been directed at making opaque deep ECG models explainable after the fact [31]; a model that is interpretable by construction sidesteps that difficulty [32], an aim equally emphasized in transparent cardiovascular risk-estimation models [33]. Combined with the fact that every feature has an established physiological interpretation, this makes the approach unusually transparent for a machine-learning system and positions it well for the regulatory and clinical scrutiny that intraoperative monitoring attracts.

The waveform-free approach is positioned as a complement to, not a replacement for, morphology-based detection. The per-class results delineate the boundary precisely. Timing and composition recover the dominant intraoperative rhythms and ventricular ectopy with high fidelity but leave supraventricular ectopy comparatively under-resolved, because a small burden of atrial premature beats perturbs interval statistics only slightly and is best distinguished by P-wave morphology. A practical system would therefore use the waveform-free layer as an efficient, interpretable first stage and reserve waveform analysis for the residual ectopy that timing cannot separate.

4.4. Limitations

Several limitations bound the interpretation. First, the data derive from a single tertiary center and a single lead (lead II), so external and multi-lead generalization remain to be established. The department-out transfer reported here is an internal evaluation across two surgical populations of the same institution and therefore does not substitute for validation on an independent intraoperative cohort. Second, the cohort is arrhythmia-enriched by a screening step, so the prevalences reported here characterize annotated arrhythmia and are not estimates of population incidence; this is a characterization study by construction. Third, the reported determinants are associational, and the intraoperative drug associations in particular are subject to confounding by indication and possible reverse causation and must not be read causally. Fourth, R-peak detection quality is inherited from the database rather than independently verified. Fifth, and importantly for the framing of the method, two of the fifteen features (the ventricular and supraventricular beat fractions) are derived from the expert beat-type annotation layer rather than from RR timing alone.

Because the beat-type and rhythm annotations are correlated by construction, part of the discriminative power, and specifically the leading importance of the supraventricular beat fraction, reflects this correlation. This is why the approach is described as waveform-free rather than as strictly interval-only: it uses no raw waveform, but it does use the expert beat-type labels. This dependency was quantified directly by the interval-only ablation (Section 3.3, Table 8): removing the two beat-type features reduced balanced accuracy from 0.869 to 0.653 and specifically collapsed the ectopy classes, while the rhythms of gross irregularity were retained. The waveform-free performance reported here should therefore be understood as depending, for the ectopy classes, on the expert beat-type labels rather than on interval timing alone. Sixth, the choice of three burden phenotypes was made for clinical interpretability and was not the silhouette-optimal partition, which favored two clusters; this choice has been disclosed, and the three-cluster solution has been shown to be nonetheless highly reproducible. Finally, ventricular tachycardia is rare within homogeneous 30-beat windows and is evaluated within the ventricular ectopy group, so class-specific conclusions about ventricular tachycardia cannot be drawn from these data.

4.5. Future directions

Two further extensions follow naturally. First, external validation on independent intraoperative cohorts and additional leads would test the generalization suggested by the department-out transfer. Second, because the interval-only ablation showed that the ectopy classes depend on the beat-type labels, coupling the waveform-free first stage to a targeted morphology model for the residual ectopy would recover that resolution while preserving the efficiency of the timing-based layer, and could be evaluated prospectively. Both are feasible within the annotation-based framework established here.

Beyond these immediate extensions, the burden-phenotyping component invites longitudinal study. Because the phenotypes are defined from features that are computable online, a natural next step is to evaluate whether phenotype membership, or trajectory across phenotypes during a case, carries prognostic information for intraoperative and early postoperative events. Such an analysis would move the contribution from characterization toward decision support, but it requires linked outcome data and prospective design and is therefore outside the scope of the present descriptive study. It should also be noted that the single-lead, single-center provenance of the data is a constraint that the field as a whole faces, given the scarcity of expert-validated intraoperative recordings; the waveform-free framing is comparatively well suited to multi-institutional replication precisely because it depends only on annotations and metadata that are inexpensive to share, and it could therefore serve as a common baseline as further intraoperative databases become available.

5. Conclusions

Intraoperative cardiac arrhythmia can be quantified, phenotyped, and its determinants examined from beat annotations and routine metadata alone, without retrieving the raw ECG waveform or additional hardware, with the qualification that resolution of the ectopy classes relies on the morphology-informed expert beat-type labels rather than on interval timing alone. On the first large anesthesiologist-validated intraoperative database, waveform-free features separated the major rhythm classes with a balanced accuracy near 0.87 for the primary transparent linear model, which performed at least as well as a higher-capacity nonlinear one, three reproducible arrhythmia-burden phenotypes emerged with coherent clinical profiles, the representation generalized across surgical contexts and tolerated signal-quality degradation, and older age was the principal adjusted correlate of high burden. The approach does not compete with saturated waveform-morphology classifiers but complements them, offering a low-cost, interpretable, and

transferable foundation for intraoperative arrhythmia analytics and a reproducible baseline for a setting that previously lacked one.

Data availability

The VitalDB Arrhythmia Database v1.0.0 is publicly available through PhysioNet [11].

Funding

No funding was received for this study.

Author contributions

M.R.K. is the sole author of this work and was responsible for the conception and design of the study, the analysis and interpretation of the data, the preparation of all figures and tables, and the drafting and critical revision of the manuscript. The author read and approved the final manuscript.

Competing interests

The author declares no competing interests.

References

[1] Kwon CH, Kim SH. Intraoperative management of critical arrhythmia. Korean J Anesthesiol. 2017;70(2):120-126. doi:10.4097/kjae.2017.70.2.120

[2] Ganesh R, Kebede E, Mueller M, Gilman E, Mauck KF. Perioperative cardiac risk reduction in noncardiac surgery. Mayo Clin Proc. 2021;96(8):2260-2276. doi:10.1016/j.mayocp.2021.03.014

[3] Pecha S, Kirchhof P, Reissmann B. Perioperative arrhythmias. Dtsch Arztebl Int. 2023;120(33-34):564-574. doi:10.3238/arztebl.m2023.0052

[4] Subramani Y, El Tohamy O, Jalali D, Nagappa M, Yang H, Fayad A. Incidence, risk factors, and outcomes of perioperative atrial fibrillation following noncardiothoracic surgery: a systematic review and meta-regression analysis of observational studies. Anesthesiol Res Pract. 2021;2021:5527199. doi:10.1155/2021/5527199

[5] Lin MH, Kamel H, Singer DE, Wu YL, Lee M, Ovbiagele B. Perioperative/postoperative atrial fibrillation and risk of subsequent stroke and/or mortality. Stroke. 2019;50(6):1364-1371. doi:10.1161/STROKEAHA.118.023921

[6] Amar D, Zhang H, Roistacher N. The incidence and outcome of ventricular arrhythmias after noncardiac thoracic surgery. Anesth Analg. 2002;95(3):537-543. doi:10.1097/00000539-200209000-00006

[7] Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. IEEE Eng Med Biol Mag. 2001;20(3):45-50. doi:10.1109/51.932724

[8] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, Schaeffter T. PTB-XL, a large publicly available electrocardiography dataset. Sci Data. 2020;7(1):154. doi:10.1038/s41597-020-0495-6

[9] Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med. 2019;25(1):65-69. doi:10.1038/s41591-018-0268-3

[10] Ribeiro AH, Ribeiro MH, Paixao GMM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nat Commun. 2020;11(1):1760. doi:10.1038/s41467-020-15432-4

[11] Eun D, Shim K, Lee H, Lim Y, Lim H, Lee H, Lee HC. VitalDB Arrhythmia Database: an anesthesiologist-validated large-scale intraoperative arrhythmia dataset with beat and rhythm labels. Sci Data. 2026. doi:10.1038/s41597-026-07076-8. (Dataset: PhysioNet. doi:10.13026/axd6-wm13)

[12] Lee HC, Park Y, Yoon SB, Yang SM, Park D, Jung CW. VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients. Sci Data. 2022;9(1):279. doi:10.1038/s41597-022-01411-5

[13] Tateno K, Glass L. Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and delta-RR intervals. Med Biol Eng Comput. 2001;39(6):664-671. doi:10.1007/BF02345439

[14] Duan J, Wang Q, Zhang B, Liu C, Li C, Wang L. Accurate detection of atrial fibrillation events with R-R intervals from ECG signals. PLoS One. 2022;17(8):e0271596. doi:10.1371/journal.pone.0271596

[15] de Chazal P, O'Dwyer M, Reilly RB. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Trans Biomed Eng. 2004;51(7):1196-1206. doi:10.1109/TBME.2004.827359

[16] Tison GH, Sanchez JM, Ballinger B, et al. Passive detection of atrial fibrillation using a commercially available smartwatch. JAMA Cardiol. 2018;3(5):409-416. doi:10.1001/jamacardio.2018.0136

[17] Faust O, Ciaccio EJ, Acharya UR. A review of atrial fibrillation detection methods as a service. Int J Environ Res Public Health. 2020;17(9):3093. doi:10.3390/ijerph17093093

[18] Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med. 2019;381(20):1909-1917. doi:10.1056/NEJMoa1901183

[19] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation. 1996;93(5):1043-1065. doi:10.1161/01.CIR.93.5.1043

[20] Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health. 2017;5:258. doi:10.3389/fpubh.2017.00258

[21] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol. 2000;278(6):H2039-H2049. doi:10.1152/ajpheart.2000.278.6.H2039

[22] Goldberger AL, Amaral LAN, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101(23):e215-e220. doi:10.1161/01.CIR.101.23.e215

[23] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-2830.

[24] Shrikanth Rao SK, Martis RJ. RR interval-based atrial fibrillation detection using traditional and ensemble machine learning algorithms. J Med Signals Sens. 2023;13(3):224-232. doi:10.4103/jmss.jmss_4_22

[25] Yin Z, Liu C, Xie C, Nie Z, Wei J, Zhang W, Liang H. Identification of atrial fibrillation using heart rate variability: a meta-analysis. Front Cardiovasc Med. 2025;12:1581683. doi:10.3389/fcvm.2025.1581683

[26] Biton S, Aldhafeeri M, Marcusohn E, Tsutsui K, Szwagier T, Elias A, Oster J, Sellal JM, Suleiman M, Behar JA. Generalizable and robust deep learning algorithm for atrial fibrillation diagnosis across geography, ages and sexes. npj Digit Med. 2023;6(1):44. doi:10.1038/s41746-023-00791-1

[27] Ansari Y, Mourad O, Qaraqe K, Serpedin E. Deep learning for ECG arrhythmia detection and classification: an overview of progress for period 2017-2023. Front Physiol. 2023;14:1246746. doi:10.3389/fphys.2023.1246746

[28] Baj G, et al. Comparison of discrimination and calibration performance of ECG-based machine learning models for prediction of new-onset atrial fibrillation. BMC Med Res Methodol. 2023;23:169.

[29] Lin CS, et al. AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. Nat Med. 2024;30:1461-1470.

[30] Sau A, Ng FS. The emerging role of artificial intelligence-enabled electrocardiograms in healthcare. BMJ Med. 2023;2:e000193.

[31] van de Leur RR, Bos MN, Taha K, et al. Improving explainability of deep neural network-based electrocardiogram interpretation using variational auto-encoders. Eur Heart J Digit Health. 2022;3(3):390-404.

[32] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1(5):206-215. doi:10.1038/s42256-019-0048-x

[33] Sau A, et al. Artificial intelligence-enabled electrocardiogram for mortality and cardiovascular risk estimation: a model development and validation study. Lancet Digit Health. 2024;6:e791-e802.