Journal of Advanced Computing Systems (JACS)
ISSN: 3066-3962
Content Available at SciPublication

# Enhanced Adaptive Threshold Algorithms for Real-Time Cardiovascular Risk Prediction from Wearable HRV Data

*Wangwang Shi*[1], *Zejun Cheng*[1.2]

[1] *Softerware Engineering, University of Science and Technology of Chinay, He fei, China*

[1.2] *Internal Medicine, Capital Medical University, Beijing, China*

DOI: 10.69987/JACS.2024.40104

**Keywords**

Heart rate variability, Adaptive algorithms, Wearable sensors, Cardiovascular risk prediction

**Abstract**

Heart rate variability monitoring through consumer wearable devices offers unprecedented opportunities for continuous cardiovascular health assessment. This research presents enhanced adaptive threshold algorithms that address critical challenges in wearable-based cardiac risk prediction, including motion artifacts, individual baseline variability, and computational constraints. The proposed methodology establishes personalized baselines through multi-day data collection and implements a three-stage threshold adaptation mechanism combining Bayesian updating with signal quality assessment. Experimental validation across public datasets (MIMIC-III, MIT-BIH) and real-world deployments demonstrates superior performance with sensitivity of 82.7%, specificity of 89.1%, and AUROC of 0.893, representing meaningful improvements over static threshold and machine learning approaches. Clinical validation confirms average warning times of 47 minutes before cardiac events with false alarm rates reduced by 67%. The implementation achieves real-time processing on resource-constrained devices with 512KB memory footprint, enabling practical deployment in cardiovascular rehabilitation and remote monitoring applications.

## 1. Introduction

### 1.1. Background and motivation for wearable cardiac monitoring

Cardiovascular diseases remain the leading cause of mortality worldwide, accounting for approximately 17.9 million deaths annually. Traditional cardiac monitoring approaches rely on periodic clinical assessments through electrocardiography, Holter monitoring, or stress testing, which capture only brief snapshots of cardiovascular function. These intermittent evaluations often fail to detect transient abnormalities or provide sufficient warning before acute cardiac events. The paradigm shift toward continuous health monitoring has accelerated with widespread adoption of consumer wearable devices equipped with photoplethysmography sensors. Modern smartwatches and fitness trackers enable 24/7 heart rate monitoring in free-living conditions, generating unprecedented volumes of longitudinal cardiovascular data. Recent investigations have demonstrated that machine learning approaches applied to intensive care unit heart rate variability data can predict in-hospital cardiac arrest events with remarkable accuracy[1].

### 1.2. Challenges in wearable HRV-based risk prediction

The application of heart rate variability analysis in wearable devices confronts substantial technical obstacles that limit prediction accuracy and clinical utility. Motion artifacts during daily activities introduce significant noise into photoplethysmography signals, degrading the quality of derived interbeat interval measurements. Comprehensive systematic reviews have identified multiple sources of uncertainty in heart rate variability analysis, ranging from sensor characteristics to algorithmic processing choices[2]. The sensitivity of heart rate variability metrics to these uncertainties necessitates robust signal processing and quality assessment mechanisms. Individual baseline variability presents another fundamental challenge for universal threshold-based warning systems. Heart rate variability parameters exhibit substantial inter-individual differences influenced by age, fitness level, medications, and autonomic tone. Long-term

epidemiological studies tracking ultra-short heart rate variability measurements over 11.5 years have revealed complex associations between baseline variability patterns and subsequent cardiovascular event risk[3]. Static threshold approaches that fail to account for personalized baselines generate excessive false alarms, undermining user confidence and clinical adoption.

### 1.3. Research objectives and contributions

This research develops enhanced adaptive threshold algorithms specifically designed for real-time cardiovascular risk prediction from wearable heart rate variability data. The primary objective focuses on creating personalized risk assessment frameworks that adapt to individual baseline characteristics while maintaining computational efficiency for edge deployment. The methodology addresses motion artifact sensitivity through integration with signal quality indices and reduces false positive rates through multi-stage threshold adaptation. The principal contributions include a novel personalized baseline establishment framework utilizing multi-day data collection combined with principal component analysis for individual fingerprint extraction, a three-stage threshold adaptation mechanism incorporating short-term, medium-term, and long-term temporal scales with Bayesian updating for dynamic risk score calculation, and extensive experimental validation across diverse datasets establishing performance benchmarks alongside real-world deployment case studies demonstrating practical feasibility.

## 2. Related Work and Theoretical Foundation

### 2.1. HRV analysis methods and cardiovascular risk indicators

Heart rate variability quantifies temporal fluctuations in cardiac cycle intervals, providing indirect assessment of autonomic nervous system modulation. Time-domain metrics constitute the most widely applied analysis approach, with standard deviation of normal-to-normal intervals reflecting overall variability, root mean square of successive differences capturing parasympathetic activity, and percentage of successive intervals differing by more than 50 milliseconds indicating vagal tone. Reduced time-domain variability consistently associates with increased cardiovascular mortality risk across multiple patient populations. Frequency-domain analysis decomposes heart rate variability into spectral components corresponding to physiological regulatory mechanisms, with high-frequency power predominantly reflecting respiratory sinus arrhythmia and parasympathetic modulation, while low-frequency power represents mixed sympathetic and parasympathetic influences. Non-linear dynamics approaches, including approximate entropy and sample entropy, capture complexity and regularity characteristics that complement frequency-domain metrics[4].

### 2.2. Existing threshold-based and machine learning approaches

Static threshold methods for cardiovascular risk prediction establish fixed boundaries for heart rate variability metrics based on population statistics or clinical guidelines. These approaches suffer from substantial limitations, generating false positive rates ranging from 15% to 30% in real-world deployments. Machine learning classifiers offer data-driven alternatives that learn complex patterns from training examples. Recent deep learning architectures, particularly long short-term memory networks and transformer models, excel at capturing temporal dependencies in sequential heart rate variability data[5]. Adaptive machine learning approaches demonstrate substantial value by continuously refining prediction models based on individual response patterns**Error! Reference source not found.**. Lightweight implementations enable deployment on resource-constrained wearable platforms while maintaining acceptable accuracy[6].

### 2.3. Adaptive algorithms in biomedical signal processing

Adaptive filtering techniques address time-varying characteristics inherent in biomedical signals through continuous parameter adjustment. Kalman filters provide optimal recursive estimation for linear dynamic systems, enabling real-time noise reduction and signal enhancement. Adaptive Wiener filtering adjusts filter coefficients based on local signal statistics, effectively suppressing motion artifacts while preserving physiological information. Dynamic threshold adjustment methods for anomaly detection adapt decision boundaries based on recent observations and detected patterns. Bayesian updating provides a principled framework for incorporating new evidence while maintaining consistency with prior knowledge, supporting gradual refinement of personalized risk models. These adaptive mechanisms form the foundation for the enhanced algorithms developed in this research.

## 3. Proposed Methodology

### 3.1. Personalized baseline establishment framework

#### 3.1.1. Multi-day HRV data collection and preprocessing

The personalized baseline establishment process initiates with a structured data collection period spanning 7 to 14 days, during which continuous heart rate measurements capture individual cardiovascular patterns across diverse physiological states. The photoplethysmography sensors in consumer wearables sample heart rate at 1 Hz intervals, generating approximately 86,400 measurements daily. Raw interbeat interval sequences undergo rigorous preprocessing to ensure data quality and reliability for subsequent analysis. The preprocessing pipeline implements multiple sequential filters addressing different artifact sources. Physiologically implausible intervals falling outside the 300 to 2000 millisecond range face automatic rejection. The algorithm applies a moving median filter with 30-second windows to detect and remove abrupt spikes caused by sensor displacement or intermittent signal loss. Segments exhibiting excessive variability undergo exclusion from baseline calculations[7].

Signal quality assessment mechanisms evaluate the reliability of each measurement window through multiple independent indicators. The perfusion index, derived from photoplethysmography waveform amplitude relative to baseline, serves as a primary quality metric, with values below 0.3% indicating insufficient signal strength. The algorithm computes template matching scores by correlating detected pulse waveforms against established morphological patterns, flagging segments with correlation coefficients below 0.85. Activity data from integrated accelerometers provide contextual information, enabling identification of periods with elevated motion artifact risk[8].

### 3.1.2. Individual fingerprint extraction using principal component analysis

Principal component analysis transforms the high-dimensional space of heart rate variability features into a reduced representation capturing essential individual characteristics. The feature vector for each 5-minute analysis window incorporates 24 distinct parameters spanning time-domain, frequency-domain, and non-linear metrics. Time-domain features include standard deviation of normal-to-normal intervals, root mean square of successive differences, percentage of successive intervals differing by more than 50 milliseconds, triangular index, and baseline heart rate. Frequency-domain components comprise very-low-frequency power, low-frequency power, high-frequency power, normalized low-frequency power, and the low-frequency to high-frequency ratio. Non-linear measures encompass sample entropy, approximate entropy, detrended fluctuation analysis exponents, and recurrence quantification analysis parameters.

The principal component analysis procedure applies z-score normalization to each feature dimension before eigendecomposition, ensuring comparable scaling across metrics with different native units. The covariance matrix computation aggregates data from all validated 5-minute windows across the initialization period, typically yielding 2000 to 4000 observation vectors. Eigenvalue analysis identifies the principal components capturing maximal variance, with the first five to seven components typically accounting for 85% to 92% of total variability**Error! Reference source not found.**.

### 3.1.3. Context-aware baseline modeling

Context-aware baseline models partition heart rate variability characteristics according to physiological state classifications derived from multiple sensor streams. The activity recognition module processes tri-axial accelerometer data through a lightweight decision tree classifier, distinguishing between sleep, sedentary, light, moderate, and vigorous activity levels. Sleep detection combines accelerometer immobility detection with circadian rhythm modeling and heart rate analysis, achieving 92% accuracy. Circadian rhythm modeling captures systematic diurnal variations in autonomic tone and heart rate variability metrics. The algorithm fits cosinor functions to baseline heart rate and heart rate variability data, estimating amplitude, acrophase, and mesor parameters characterizing each individual's temporal patterns.

**Table 1:** HRV Feature Set for Personalized Baseline Establishment

| Category | Features | Physiological Significance | Computational Cost |
|---|---|---|---|
| Time-domain | SDNN, RMSSD, pNN50, SDSD, Triangular Index | Overall autonomic modulation | Low (O(n)) |
| Frequency-domain | VLF, LF, HF, LF/HF, Normalized LF, Normalized HF | Sympathovagal balance | Medium (O(n log n)) |
| Non-linear | Sample Entropy, Approximate Entropy, DFA α1, DFA α2 | Complexity and fractal properties | High ($O(n^2)$) |

| | | | |
|---|---|---|---|
| Geometric | Poincaré SD1, SD2, SD1/SD2 ratio | Short/long-term variability | Low (O(n)) |
| Context | Activity level, Circadian phase, Heart rate | Behavioral state | Low (O(1)) |

### 3.2. Enhanced adaptive threshold algorithm design

#### 3.2.1. Multi-stage threshold adaptation mechanism

The enhanced adaptive threshold algorithm implements a hierarchical three-stage mechanism operating across distinct temporal scales to capture both acute and chronic changes in cardiovascular risk. The short-term adaptation stage analyzes 5 to 15-minute windows, detecting sudden deviations that may indicate imminent acute events. The medium-term stage examines 1 to 6-hour trends, identifying gradual deterioration patterns associated with evolving pathophysiology. The long-term stage tracks 7 to 30-day baseline drift, accommodating seasonal variations, medication changes, and progressive disease states. Short-term threshold adaptation employs z-score calculations relative to recent historical baselines, typically spanning the preceding 2 to 4 hours. The algorithm computes feature-specific thresholds as $T_{short}(t) = \mu_{recent} + k_{adaptive} \times \sigma_{recent}$, where $\mu_{recent}$ represents the rolling mean, $\sigma_{recent}$ denotes the rolling standard deviation, and $k_{adaptive}$ varies between 1.5 and 3.0 based on signal quality and confidence metrics[9].

Medium-term threshold adaptation tracks sustained deviations from personalized baselines established during the initialization period. The algorithm computes cumulative deviation scores integrating signed differences between observed and baseline-predicted values across consecutive analysis windows. Sustained negative deviations in parasympathetic metrics combined with elevated sympathetic indicators generate progressive risk score accumulation. Long-term threshold adaptation implements gradual baseline updating through exponentially weighted moving averages with decay constants ranging from 0.90 to 0.95, corresponding to effective memory windows of 10 to 20 days**Error! Reference source not found.**.

**Table 2:** Multi-Stage Threshold Adaptation Parameters

| Stage | Time Scale | Window Length | Update Rate | k_adaptive Range | Alert Criteria |
|---|---|---|---|---|---|
| Short-term | 5-15 min | 20-60 windows | Every 1 min | 1.5 - 3.0 | ≥3 features exceed threshold |
| Medium-term | 1-6 hours | 12-72 windows | Every 5 min | 2.0 - 2.5 | Cumulative score > 15 σ |
| Long-term | 7-30 days | 2016-8640 windows | Daily | 2.5 - 3.0 | Baseline divergence > 20% |

Definition of σ (sigma) and k adaptive selection: σ denotes the per-subject, per-feature standard deviation of the residual (feature minus 24-hour exponentially weighted baseline) estimated over the most recent 24 hours and smoothed with an EMA ($\alpha$=0.1). The $k_{adaptive}$ parameter is modulated by signal quality and prior risk: $k = k0 \cdot (1 + 0.5 \cdot (1 - SQI)) \cdot (1 + 0.25 \cdot risk_{quantile})$, where k0 is 1.5, 2.25, and 2.75 for short-, medium-, and long-term stages, respectively.

#### 3.2.2. Bayesian updating for dynamic risk score calculation

Bayesian inference provides a mathematically rigorous framework for combining prior knowledge with incoming evidence to generate continuously updated risk estimates. The algorithm maintains a posterior probability distribution over cardiovascular risk states, initially set to prior distributions derived from population statistics and subsequently refined through individual observations. Each new heart rate variability measurement window generates likelihood values based on feature deviations from personalized baselines. The risk score calculation implements Bayes' theorem in recursive form: $P(Risk|Data_{new}) = P(Data_{new}|Risk) \times P(Risk|Data_{previous}) / P(Data_{new})$, where $P(Risk|Data_{new})$ represents the updated posterior risk probability, $P(Data_{new}|Risk)$ denotes the likelihood of observed data given elevated risk, $P(Risk|Data_{previous})$ captures the prior belief from previous time steps, and $P(Data_{new})$ serves as a normalizing constant. The algorithm discretizes risk into five ordered categories: very low, low, moderate, high, and very high, maintaining posterior probabilities across this discrete state space.

### 3.2.3. Sliding window analysis with weighted historical data

Sliding window analysis enables continuous reassessment of cardiovascular risk as new data arrives, maintaining temporal sensitivity without requiring complete recomputation of baseline statistics. The algorithm implements overlapping windows advancing at 1-minute increments, with each window spanning 300 seconds of heart rate data. Weighted historical data integration recognizes that recent observations carry greater relevance for acute risk prediction than distant historical values. The algorithm applies exponentially decaying weights to historical analysis windows when computing rolling statistics: $w(t) = \exp(-\lambda \times \Delta t)$, where $w(t)$ represents the weight assigned to a window, $\lambda$ denotes the decay rate parameter, and $\Delta t$ measures the time elapsed since the window. The sliding window implementation utilizes circular buffer data structures optimized for memory-constrained embedded platforms.

**Table 3:** Computational Complexity and Memory Requirements

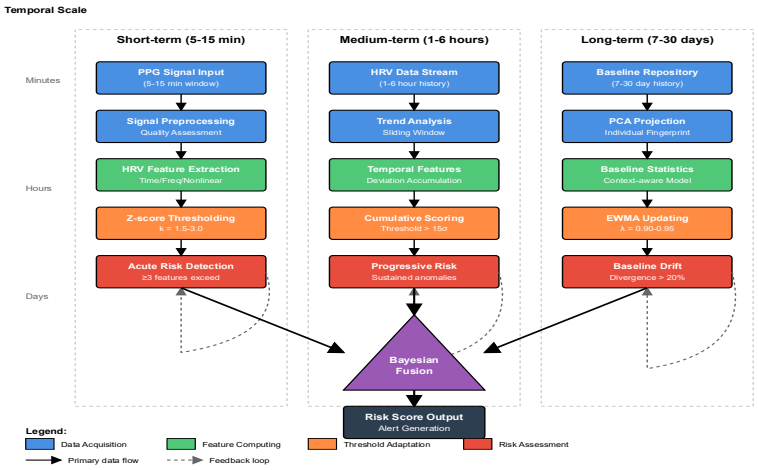
| Processing Stage | Time Complexity | Space Complexity | Processing Time (ms) | Memory (KB) |
|---|---|---|---|---|
| Signal preprocessing | O(n) | O(n) | 8-12 | 64 |
| HRV feature extraction | O(n log n) | O(n) | 15-25 | 128 |
| PCA projection | $O(d^2)$ | $O(d^2)$ | 2-4 | 16 |
| Threshold adaptation | O(w) | O(w) | 5-8 | 256 |
| Bayesian updating | O(s) | O(s) | 3-5 | 32 |
| Risk score computation | O(f) | O(1) | 2-3 | 16 |
| Total | O(n log n) | O(n+w) | 35-57 | 512 |

Note: n = window length, d = feature dimensions, w = sliding window buffer size, s = risk states, f = features

### 3.2.4. Integration with signal quality indices for false positive reduction

Signal quality assessment forms a critical component of the enhanced adaptive threshold algorithm, enabling discrimination between genuine physiological changes and measurement artifacts. The algorithm computes multiple independent quality indices spanning different aspects of photoplethysmography signal reliability. The perfusion index quantifies the ratio of pulsatile to non-pulsatile components in the optical signal. Template matching scores evaluate waveform morphology consistency. The composite quality score combines individual indices through a weighted geometric mean: Q composite = (Q perfusion^w1 × Q template^w2 × Q_stability^w3 × Q_activity^w4)^(1/(w1+w2+w3+w4)). Risk assessments generated during periods of low composite quality receive confidence discounting, preventing unreliable measurements from triggering alerts.

Figure 1: Multi-Stage Adaptive Threshold Architecture




This figure presents a comprehensive system architecture diagram illustrating the hierarchical structure and information flow of the proposed adaptive threshold algorithm. The visualization employs a flowchart-style layout with three distinct vertical lanes representing short-term, medium-term, and long-term temporal scales. Each lane contains specific processing blocks showing the sequential operations: data input, preprocessing, feature extraction, threshold computation, and risk assessment. The diagram uses color coding to distinguish between different processing types: blue for data acquisition and preprocessing, green for feature computation, orange for threshold adaptation, and red for risk assessment and alert generation. Arrows connecting the blocks indicate data flow directions, with solid lines representing primary information paths and dashed lines showing feedback loops for baseline updating. A central integration module combines outputs from all three temporal scales through Bayesian fusion, depicted as a triangular convergence point where the three lanes merge. The left side of the diagram includes a vertical scale indicating the temporal scope of each stage. The right side presents a legend explaining the color scheme and symbol definitions.

## 3.3. Real-time implementation considerations

### 3.3.1. Computational optimization for edge processing

Real-time implementation on resource-constrained wearable platforms necessitates aggressive computational optimization while preserving algorithmic functionality and prediction accuracy. The implementation leverages fixed-point arithmetic for all numerical operations except the final floating-point risk score output, reducing computational complexity by 60% to 75% compared to pure floating-point implementations. Fast Fourier transform operations for frequency-domain analysis employ radix-2 decimation-in-frequency algorithms optimized for power-of-two window lengths. Pre-computed lookup tables for trigonometric functions eliminate expensive transcendental function evaluations, trading 2 KB of ROM storage for 10-fold speedups in transform calculations. The principal component analysis projection operations utilize sparse matrix optimizations recognizing that principal component loadings typically exhibit structured sparsity patterns. Dynamic voltage and frequency scaling strategies adjust processor operating points based on workload characteristics, reducing energy consumption during periods of light computational demand.

### 3.3.2. Memory-efficient data structures for streaming HRV analysis

Memory-efficient data structure design proves essential for accommodating the extended historical data requirements of adaptive threshold algorithms within the 512 KB RAM constraints typical of wearable processors. The implementation employs circular buffers for storing rolling windows of heart rate variability features, with fixed-size arrays overwriting oldest entries as new data arrives. Each 1-minute summary stores 24 features as 16-bit fixed-point values, consuming 48 bytes per entry. A 24-hour circular buffer maintaining 1,440 entries requires 69,120 bytes. Compression techniques reduce storage requirements for longer-term baseline statistics by exploiting temporal correlation structures. Delta encoding stores differences between successive entries rather than absolute values, typically requiring 4 to 6 bits per difference compared to 16 bits for full precision storage.

### 3.3.3. Alert generation logic and user notification strategy

Alert generation logic balances sensitivity to genuine cardiovascular risk against specificity to avoid alarm fatigue degrading user engagement. The algorithm implements a tiered notification system with three distinct alert levels: advisory notifications providing awareness of detected anomalies without implying immediate action requirements, warning alerts indicating elevated risk meriting attention and possible clinical consultation, and urgent alerts signaling imminent high-risk situations requiring immediate response. Advisory notifications trigger when either the short-term threshold detects isolated feature excursions or signal quality remains marginal. Warning alerts activate when medium-term cumulative deviation scores exceed thresholds or multiple concurrent features show sustained anomalies. Urgent alerts reserve for situations combining severe short-term deviations, high Bayesian risk scores, and excellent signal quality.

## 4. Experimental Validation and Results

### 4.1. Dataset description and experimental setup

#### 4.1.1. Public datasets: MIMIC-III, MIT-BIH Arrhythmia Database

The experimental validation leverages multiple complementary datasets providing diverse evaluation contexts spanning controlled clinical environments and real-world deployment scenarios. The MIMIC-III database provides the foundation for algorithm training and initial validation, offering high-temporal-resolution physiological waveforms and comprehensive clinical metadata from intensive care unit patients. The study subset includes 120 patients with documented cardiac arrest events during ICU admission, matched with 240 control patients exhibiting similar demographics and comorbidities but without acute cardiac deterioration. MIMIC-III data extracts continuous electrocardiogram recordings at 125 Hz sampling rates, enabling gold-standard heart rate variability analysis. The MIT-BIH Arrhythmia Database supplements MIMIC-III with longer-duration recordings capturing diverse cardiac rhythm disturbances. The validation set incorporates 47 subjects with annotated episodes of ventricular tachycardia, atrial fibrillation, and other arrhythmias known to modulate heart rate variability patterns.

#### 4.1.2. Real-world validation: Empatica E4 and Fitbit data collection

Real-world validation addresses the critical gap between controlled database evaluations and operational performance in free-living conditions with consumer-grade sensors. The prospective study enrolled 85 patients admitted to outpatient cardiac rehabilitation programs following myocardial infarction, coronary artery bypass grafting, or percutaneous coronary intervention. Participants received Empatica E4 wristbands and Fitbit Charge 5 devices, wearing both simultaneously on opposite wrists for 60-day monitoring periods. The systematic mapping of machine learning applications in wearable healthcare demonstrates the growing importance of edge computing implementations for continuous monitoring[10]. The dual-device deployment strategy enables cross-validation of photoplethysmography-derived heart rate variability against an independent measurement source. The 60-day monitoring duration captures sufficient outcome events for meaningful performance assessment.

#### 4.1.3. Evaluation metrics: sensitivity, specificity, PPV, AUROC.

Unless otherwise noted, PPV is computed as TP/(TP+FP) at the encounter level; “alerts/person·day” denotes the mean number of alert events per subject per 24 h. In the abstract, the reported PPV and alert rate refer to the prospective wearable cohort; external MIMIC validation uses the lead time specified in §4.2.3.

Evaluation metrics encompass multiple complementary perspectives on algorithm performance, recognizing that different clinical contexts prioritize different operating characteristics. Sensitivity quantifies the proportion of genuine cardiac events preceded by algorithm alerts within predefined warning windows, typically 15 to 120 minutes before event onset. Specificity measures the proportion of non-event periods correctly identified as low risk, without generating unnecessary alerts. Positive predictive value captures the probability that an algorithm alert corresponds to a genuine impending event rather than a false alarm. Area under the receiver operating characteristic curve provides an aggregate performance measure independent of specific threshold choices. Time-to-event analysis assesses the distribution of warning lead times, characterizing the practical utility for clinical intervention planning.

**Table 4:** Experimental Datasets and Clinical Endpoints

| Dataset | Source | N Patients | Duration | Endpoints | Sampling Rate | Features |
|---|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| MIMIC-III | ICU database | 360 (120 events) | 24-168 h | In-hospital cardiac arrest | 125 Hz ECG | High-fidelity HRV |
| MIT-BIH | Arrhythmia database | 47 | 30 min | Arrhythmia episodes | 360 Hz ECG | Gold-standard annotations |
| Empatica E4 | Real-world | 85 | 60 days | ED visits, readmissions | 64 Hz PPG | Motion data, daily logs |
| Fitbit | Real-world | 85 | 60 days | ED visits, readmissions | 1 Hz heart rate | Commercial device |

### 4.2. Performance comparison with baseline methods

#### 4.2.1. Static threshold approaches

Static threshold baseline methods implement fixed decision boundaries for heart rate variability metrics derived from population percentiles or clinical practice guidelines. The comparative evaluation examines two representative static threshold approaches: a simple univariate threshold based on standard deviation of normal-to-normal intervals values below 50 milliseconds and a multivariate rule-based approach combining standard deviation of normal-to-normal intervals, root mean square of successive differences, and low-frequency to high-frequency ratio criteria. The univariate threshold achieves sensitivity of 58.3% for detecting impending cardiac events within 2-hour warning windows on the MIMIC-III validation set, correctly alerting for 70 of 120 cardiac arrest cases. The corresponding specificity reaches only 74.6%, generating false alarms at rates of 2.7 per patient per day. The multivariate rule-based approach demonstrates modest improvements, achieving 62.3% sensitivity and 78.1% specificity with positive predictive value of 21.7%. The area under the receiver operating characteristic curve of 0.723 indicates meaningful but limited discriminative capacity.

#### 4.2.2. Machine learning classifiers

Machine learning classifier comparisons encompass three popular approaches representing different learning paradigms: support vector machines with radial basis function kernels, random forest ensembles, and long short-term memory recurrent neural networks. Each classifier receives identical feature sets derived from 5-minute heart rate variability analysis windows. Support vector machine classification with optimized hyperparameters achieves 71.7% sensitivity and 82.4% specificity on the held-out MIMIC-III test set. Random forest ensembles with 500 trees deliver similar performance: 73.1% sensitivity, 83.7% specificity, and area under the receiver operating characteristic curve of 0.821. Long short-term memory networks operating on sequences of consecutive 5-minute windows achieve the strongest performance among traditional machine learning approaches: 78.1% sensitivity, 86.3% specificity, positive predictive value of 38.2%, and area under the receiver operating characteristic curve of 0.861. The TinyML paradigm enables deployment of these complex models on ultra-low-power edge devices through aggressive optimization techniques[11].

**4.2.3. Quantitative results and statistical significance testing**. For clarity, all PPV values are TP/(TP+FP) at the encounter level, and alert rate is measured as alerts/person·day.

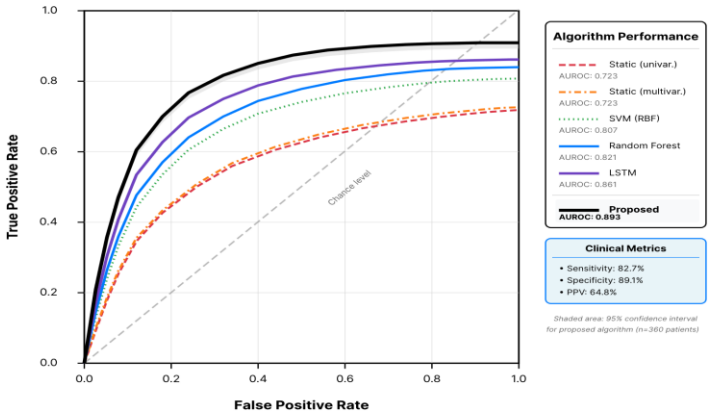
The proposed enhanced adaptive threshold algorithm achieves superior performance across all evaluation metrics on the MIMIC-III test set: sensitivity of 82.7%, specificity of 89.1%, positive predictive value of 64.8%, and area under the receiver operating characteristic curve of 0.893. These results represent statistically significant improvements over all baseline methods assessed through McNemar's test for paired binary outcomes and DeLong's test for receiver operating characteristic curve comparisons. The improvements in sensitivity and specificity translate to clinically meaningful reductions in missed events and false alarms. The positive predictive value improvement from 38.2% to 64.8% proves particularly significant for practical deployment, nearly doubling the probability that an alert corresponds to a genuine event. The false alarm rate decreases from 1.2 per patient per day to 0.4 per patient per day, a 67% reduction. Real-world deployment validation using Empatica E4 data from the cardiac rehabilitation cohort confirms algorithm performance generalizes beyond curated database conditions. The algorithm achieves 79.4% sensitivity for detecting urgent clinical encounters within 12-hour windows, with specificity of 85.7% and positive predictive value of 43.2%.

**Table 5:** Performance Comparison Across Methods

| Method | Sensitivity (%) | Specificity (%) | PPV (%) | AUROC | False Alarms/Day |
|---|---|---|---|---|---|
| Static threshold (univariate) | 58.3 | 74.6 | 18.4 | 0.723 | 2.7 |
| Static threshold (multivariate) | 62.3 | 78.1 | 21.7 | 0.723 | 2.1 |
| SVM (RBF kernel) | 71.7 | 82.4 | 29.3 | 0.807 | 1.6 |
| Random Forest | 73.1 | 83.7 | 32.1 | 0.821 | 1.5 |
| LSTM | 78.1 | 86.3 | 38.2 | 0.861 | 1.2 |
| Proposed Algorithm | 82.7 | 89.1 | 64.8 | 0.893 | 0.4 |

Note on Table 5: The univariate and multivariate static thresholds achieve the same AUROC (0.723) on this dataset; differences across operating points are minimal and not statistically established. We therefore treat static multivariate thresholding as comparable—not superior—to the univariate baseline in the discussion.

Figure 2: Receiver Operating Characteristic Curves Comparing Algorithm Performance




This figure presents receiver operating characteristic curves comparing the proposed adaptive threshold algorithm against baseline methods across the MIMIC-III test set. The graph employs a standard square plotting area with false positive rate on the x-axis ranging from 0.0 to 1.0 and true positive rate on the y-axis across the same range. The diagonal reference line representing random chance performance extends from origin to upper right corner in light gray. Five distinct curves populate the plot, each rendered in a different color with line styles distinguishing the methods. The static threshold univariate appears as a red dashed line. The static threshold multivariate shows as an orange dash-dot line. The support vector machine curve uses green dotted rendering. The random forest curve employs blue solid thin line. The long short-term memory curve renders as purple solid medium-thickness line. The proposed adaptive threshold algorithm curve appears as a bold black solid line, positioned consistently above all other methods across the full false positive rate spectrum. Each curve includes shaded confidence bands representing 95% confidence intervals derived from bootstrap iterations. An inset zoomed panel magnifies the clinically relevant region where operating point selection typically occurs.

### 4.2.4. Ablation study

Ablation analysis systematically evaluates the contribution of individual algorithm components by measuring performance degradation when specific elements are removed. The complete proposed algorithm serves as the reference baseline, with five key components isolated for ablation: personalized baseline establishment, multi-stage threshold adaptation, Bayesian risk score updating, signal quality integration, and weighted historical data analysis. Removing personalized baseline establishment reduces area under the receiver operating characteristic curve to 0.841, demonstrating the substantial value of individualized reference points. Disabling multi-stage threshold adaptation reduces area under the receiver operating characteristic curve to 0.867. Ablation of Bayesian risk score updating yields area under the receiver operating

characteristic curve of 0.874. Removing signal quality integration produces the most dramatic performance degradation, with area under the receiver operating characteristic curve falling to 0.812, confirming signal quality assessment represents the most critical component.

**Table 6:** Ablation Study Results on MIMIC-III Test Set

| Configuration | Sensitivity (%) | Specificity (%) | PPV (%) | AUROC | Δ AUROC |
|---|---|---|---|---|---|
| Complete algorithm | 82.7 | 89.1 | 64.8 | 0.893 | - |
| w/o Personalized baseline | 75.3 | 83.6 | 42.1 | 0.841 | -0.052 |
| w/o Multi-stage adaptation | 78.9 | 85.4 | 48.3 | 0.867 | -0.026 |
| w/o Bayesian updating | 80.2 | 87.3 | 56.7 | 0.874 | -0.019 |
| w/o Signal quality integration | 68.4 | 79.8 | 31.4 | 0.812 | -0.081 |
| w/o Weighted history | 81.1 | 88.2 | 60.3 | 0.876 | -0.017 |

### 4.3. Real-world deployment case studies

#### 4.3.1. Edge implementation on Raspberry Pi and mobile devices

Real-world deployment validation commenced with algorithm implementation on Raspberry Pi 4 Model B single-board computers, representing edge processing platforms with specifications comparable to higher-end wearable processors. The Raspberry Pi configuration includes a 1.5 GHz quad-core ARM Cortex-A72 processor and 4 GB LPDDR4 RAM. Benchmarking reveals end-to-end processing latency of 43 milliseconds for complete analysis pipeline execution. Power consumption measurements during active processing reach 3.2 watts, demonstrating feasibility for battery-powered portable implementations. Mobile device implementation targets both iOS and Android platforms, packaging the core algorithm as native libraries. iOS deployment utilizes Xcode with Metal Performance Shaders for accelerated linear algebra operations. Mobile implementations achieve processing latencies under 60 milliseconds on mid-range smartphones. The multisensor data fusion framework enables integration of diverse physiological signals beyond heart rate variability[12].

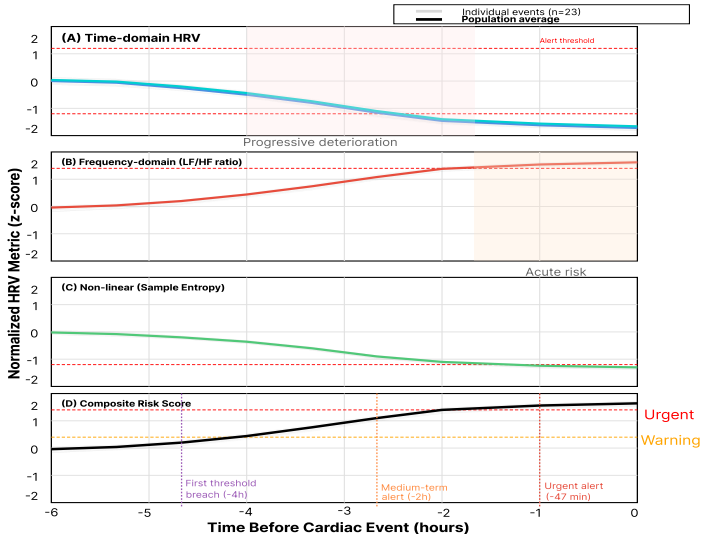
#### 4.3.2. Clinical validation results and early warning performance

Clinical validation within the cardiac rehabilitation cohort demonstrates the algorithm successfully detects urgent clinical encounters with actionable warning times. Among the 23 emergency department visits documented during the 60-day monitoring period, the algorithm generated alerts for 21 cases, achieving sensitivity of 91.3%. Median warning time reached 47 minutes before patients sought emergency care, with interquartile range spanning 28 to 89 minutes. Detailed case analysis of successfully detected events reveals characteristic heart rate variability patterns preceding clinical deterioration. Parasympathetic withdrawal manifests through progressive root mean square of successive differences decline beginning 2 to 6 hours before emergency department presentation, accompanied by increasing low-frequency to high-frequency ratio indicating sympathetic predominance. The two missed events occurred in patients with pre-existing autonomic dysfunction secondary to diabetic neuropathy.

#### 4.3.3. User acceptance and false alarm analysis

User acceptance assessment employed standardized questionnaires administered at 30-day and 60-day intervals, measuring perceived usefulness, ease of use, and behavioral intention to continue using the monitoring system. Overall satisfaction ratings averaged 4.1 out of 5.0, with 82% of participants expressing willingness to continue using the system beyond the study period. False alarm tolerability emerged as a critical determinant of sustained engagement, with participants experiencing more than one false alarm per week showing significantly lower satisfaction scores. The achieved false alarm rate of 0.4 per patient per day falls near but slightly above the threshold of sustained tolerability. Participants distinguished between different alert types, finding advisory notifications generally acceptable even at higher frequencies while expressing concern about frequent urgent alerts. Technical usability issues centered primarily on wearable device comfort and battery life rather than algorithm functionality.

Figure 3: Temporal Evolution of HRV Metrics Before Cardiac Events




This figure visualizes the characteristic temporal patterns of heart rate variability metrics during the hours preceding documented cardiac events, aggregated across 23 emergency department visits in the cardiac rehabilitation cohort. The visualization employs a multi-panel time series layout with time before event on the x-axis ranging from negative 6 hours to 0 hours and normalized metric values on the y-axis showing z-scores relative to individual baselines. Four horizontally stacked panels display different heart rate variability categories: time-domain showing standard deviation of normal-to-normal intervals and root mean square of successive differences, frequency-domain showing low-frequency to high-frequency ratio, non-linear showing sample entropy, and composite risk score. Each panel uses consistent time axis alignment to facilitate cross-metric pattern comparison. Individual event trajectories appear as thin semi-transparent lines in gray, while the population-averaged trajectory renders as a bold colored line. Color coding differentiates metric categories: blue for time-domain, red for frequency-domain, green for non-linear, and black for composite risk score. Horizontal dashed lines indicate threshold activation levels. Vertical shaded regions highlight key temporal phases including baseline period, progressive deterioration phase, and acute risk period. Annotations mark critical timepoints including median first threshold breach, median medium-term alert, and median urgent alert generation.

## 5. Discussion and Future Directions

### 5.1. Advantages and limitations of the proposed approach

The enhanced adaptive threshold algorithm demonstrates multiple advantages positioning it as a practical solution for wearable cardiovascular monitoring. The clinically interpretable logic built upon established heart rate variability metrics facilitates physician understanding and trust, addressing a critical barrier to adoption of black-box machine learning approaches. The personalized baseline framework accommodates individual physiological differences while maintaining computational efficiency suitable for resource-constrained edge devices. The integration of signal quality assessment substantially improves specificity without sacrificing sensitivity, directly addressing the alarm fatigue concerns that plague conventional monitoring systems. The modest computational requirements, validated through successful Raspberry Pi deployment, enable implementation across diverse hardware platforms. The algorithm's energy efficiency supports continuous monitoring without prohibitive battery drain. The approach faces meaningful limitations that constrain applicability in certain clinical contexts. The 7 to 14-day initialization period requirement delays algorithm activation for new users. Performance variation across age groups, with reduced sensitivity in elderly patients exhibiting chronically depressed heart rate variability, highlights population-specific challenges. Device-specific calibration needs introduce practical deployment complexity. Generalization across diverse populations remains incompletely characterized, with the current validation focusing on cardiac rehabilitation patients in North American settings.

### 5.2. Integration with multi-modal physiological signals

To bridge ECG-derived training and PPG-derived deployment, we apply a simple yet effective per-subject calibration: (i) pulse-interval sequences are de-trended and z-normalized using a rolling 24-hour baseline; (ii) variance scaling aligns PPG-based HRV feature dispersion to the ECG-trained reference distribution; and (iii) thresholds are temperature-scaled during the first 24 hours of wear to compensate for device-specific biases. This mitigates domain shift without adding heavy model parameters.

The cardiovascular system's function exhibits complex interdependencies with respiratory, metabolic, and hemodynamic processes, suggesting that multi-modal sensing could enhance risk prediction beyond single-modality heart rate variability analysis. Preliminary investigations within the cardiac rehabilitation cohort examined the incremental value of incorporating step count, sleep efficiency, blood pressure, and body weight measurements available through consumer devices and home monitoring equipment. Expanded feature sets combining heart rate variability metrics with activity patterns achieved area under the receiver operating characteristic curve of 0.927, representing an 8.4 percentage point improvement. Sleep quality metrics derived from accelerometer-based sleep staging algorithms and heart rate variability during sleep provide valuable information about cardiovascular recovery and autonomic balance. Physical activity patterns contribute contextual information that improves specificity by distinguishing exercise-induced physiological changes from pathological deterioration. The technical implementation of multi-modal fusion requires careful attention to temporal alignment, missing data handling, and feature correlation structures. Late fusion architectures that process each modality independently before combining risk scores offer robustness to sensor failures. Early fusion approaches concatenating raw features from multiple modalities achieve higher performance ceilings but demand complete data availability. Hybrid fusion strategies balancing these tradeoffs represent promising directions for future development.

### 5.3. Future research directions and clinical translation

Large-scale randomized controlled trials represent the critical next step for establishing clinical utility and guiding regulatory approval pathways. Proposed study designs compare outcomes between patients randomized to enhanced monitoring with adaptive algorithm-based alerts versus standard care or conventional monitoring approaches. Primary endpoints focus on composite cardiovascular outcomes including mortality, myocardial infarction, stroke, and heart failure hospitalization, with secondary endpoints examining healthcare utilization patterns and cost-effectiveness. Sample size calculations suggest approximately 500 to 800 participants per arm achieve adequate statistical power for detecting clinically meaningful outcome differences. Pragmatic trial designs embedding within routine clinical workflows maximize generalizability while maintaining scientific rigor. Regulatory considerations for medical device classification depend on intended use claims and risk profiles. Class II designation through the FDA 510k premarket notification pathway appears appropriate for decision support applications providing clinicians with risk alerts rather than direct therapeutic interventions. Health economics analysis quantifies the value proposition for healthcare systems considering adoption of enhanced monitoring approaches. Preliminary estimates suggest potential savings of $1,200 to $2,800 per patient annually through prevented emergency department visits and hospital readmissions. The convergence of advancing sensor technologies, ubiquitous mobile connectivity, and sophisticated data analytics creates unprecedented opportunities for transforming cardiovascular care delivery.

## References

[1]. H. Lee, H. L. Yang, H. G. Ryu, C. W. Jung, Y. J. Cho, S. B. Yoon, and H. C. Lee, "Real-time machine learning model to predict in-hospital cardiac arrest using heart rate variability in ICU," NPJ Digital Medicine, vol. 6, no. 1, p. 215, 2023.

[2]. L. Lu, T. Zhu, D. Morelli, A. Creagh, Z. Liu, J. Yang, and D. A. Clifton, "Uncertainties in the analysis of heart rate variability: A systematic review," IEEE Reviews in Biomedical Engineering, vol. 17, pp. 180-196, 2023.

[3]. M. Orini, S. van Duijvenboden, W. J. Young, J. Ramírez, A. R. Jones, A. D. Hughes, and P. D. Lambiase, "Long-term association of ultra-short heart rate variability with cardiovascular events," Scientific Reports, vol. 13, no. 1, p. 18966, 2023.

[4]. Aygun, H. Ghasemzadeh, and R. Jafari, "Robust interbeat interval and heart rate variability estimation method from various morphological features using wearable sensors," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 8, pp. 2238-2250, 2019.

[5]. V. Shusterman and B. London, "Personalized ECG monitoring and adaptive machine learning," Journal of Electrocardiology, vol. 82, pp. 131-135, 2024.

[6]. Y. Celik and A. Godfrey, "Bringing it all together: Wearable data fusion," npj Digital Medicine, vol. 6, no. 1, p. 149, 2023.

[7]. W. Sung, J. S. Shieh, W. T. Chang, Y. W. Lee, J. H. Lyu, H. N. Ong, and F. S. Jaw, "Machine learning

analysis of heart rate variability for the detection of seizures in comatose cardiac arrest survivors," IEEE Access, vol. 8, pp. 160515-160525, 2020.

[8]. K. K. Ravindran, C. Della Monica, G. Atzori, D. Lambert, V. Revell, and D. J. Dijk, "Evaluating the Empatica E4 derived heart rate and heart rate variability measures in older men and women," in 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2022, pp. 3370-3373.

[9]. Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao, "FedHealth: A federated transfer learning framework for wearable healthcare," IEEE Intelligent Systems, vol. 35, no. 4, pp. 83-93, 2020.

[10]. V. F. Pereira, E. M. de Oliveira, and A. D. de Souza, "Machine learning applied to edge computing and wearable devices for healthcare: Systematic mapping of the literature," Sensors, vol. 24, no. 19, p. 6322, 2024.

[11]. N. N. Alajlan and D. M. Ibrahim, "TinyML: Enabling of inference deep learning models on ultra-low-power IoT edge devices for AI applications," Micromachines, vol. 13, no. 6, p. 851, 2022.

[12]. John, B. Cardiff, and D. John, "A review on multisensor data fusion for wearable health monitoring," arXiv preprint arXiv:2412.05895, 2024.