



Original Article

Multimodal Machine Learning Framework for Cardiovascular Risk Stratification in Adult Obesity: A Cross-sectional Study



Pedro Ribeiro¹, João Alexandre Lobo Marques², Marconi Pereira Brandão³, Octávio Barbosa Neto³,
Camila Ferreira Leite³ and Pedro Miguel Rodrigues^{1*} 

¹Universidade Católica Portuguesa, Centro de Biotecnologia e Química Fina – Laboratório Associado, Escola Superior de Biotecnologia, Porto, Portugal;

²University of Saint Joseph, Laboratory of Applied Neurosciences, Macao, China; ³Federal University of Ceará, Graduate Program in Cardiovascular Sciences, Fortaleza, Ceará, Brazil

Received: July 16, 2025 | Revised: September 04, 2025 | Accepted: September 28, 2025 | Published online: November 06, 2025

Abstract

Background and objectives: Cardiovascular diseases account for approximately 80% of all deaths caused by known medical conditions, making them the leading cause of mortality worldwide. The present study investigates the use of electrocardiogram (ECG) non-linear features and different topological medical features (heart rate, anthropometry, blood, glucose, and lipid profile, and heart rate variability) to discriminate between different Framingham Cardiovascular Risk Scale status groups in adult obesity using machine learning.

Methods: We conducted a cross-sectional study between November 2023 and May 2024 in Fortaleza, Ceará, Brazil. Based on the Framingham Cardiovascular Risk Scale, patients were categorized into three cardiovascular risk groups: Low (22 participants), Moderate (14 participants), and High (17 participants). From ECG signals at two different positions (ECG_Down and ECG_UP), 27 non-linear features were extracted using multi-band analysis. Additionally, 42 medical features provided by physicians were included. From a pool of 19 machine learning classifiers, models were trained and tested within a nested leave-one-out cross-validation procedure using information solely from ECG, solely from medical features, and combining both (multimodal), respectively, to distinguish between Low vs. Moderate, Low vs. High, Moderate vs. High, and All vs. All.

Results: The multimodal model presented the best results for every comparison group, reaching (1) 88.89% Accuracy and 0.8831 area under the curve (AUC) for Low vs. Moderate; (2) 97.44% Accuracy and 0.9706 AUC for Low vs. High; (3) 93.55% Accuracy and an AUC of 0.9412 for Moderate vs. High; (4) 86.79% Accuracy and 0.9346 AUC for All vs. All.

Conclusions: The multimodal model outperformed single-source models in cardiovascular risk classification. ECG-derived non-linear features, especially from ECG_Down, were key drivers, with medical features adding complementary value. The results support its potential use in clinical triage and diagnosis.

Introduction

Cardiovascular diseases (CVDs) are the primary cause of death worldwide.^{1,2} Until the 1940s, the risk factors associated with

CVD were not known in a context in which these diseases had already caused many casualties among Americans.³ Thus, during times of epidemiological transition, the treatment and prevention of CVD did not have precise directions.⁴ The need for a study to investigate the causes of the increasing burden of CVD was recognized.⁵

The Framingham study was initiated following 5,209 adults free from overt CVD, representing about 19% of Framingham's population.⁶ The main objective was to follow the included participants for the development of cardiac heart failure for 20 years.³ The first results of this cohort emerged four years after the beginning of the study and revealed that high blood pressure, high cholesterol levels, and overweight were associated with the de-

Keywords: Adult obesity; Framingham risk score; Cardiovascular risk; Non-linear features; Multimodal; Machine-learning.

***Correspondence to:** Pedro Miguel Rodrigues, Universidade Católica Portuguesa, Centro de Biotecnologia e Química Fina, Escola Superior de Biotecnologia, Office EBI112, 1327 Rua Diogo Botelho, Porto 4169-005, Portugal. ORCID: <https://orcid.org/0000-0002-5381-6615>. Tel: +351-225580027, E-mail: pmrodrigues@ucp.pt

How to cite this article: Ribeiro P, Marques JAL, Brandão MP, Neto OB, Leite CF, Rodrigues PM. Multimodal Machine Learning Framework for Cardiovascular Risk Stratification in Adult Obesity: A Cross-sectional Study. *Explor Res Hypothesis Med* 2025;11(1):e00037. doi: 10.14218/ERHM.2025.00037.

velopment of new-onset coronary heart disease.⁶ Although being overweight had stood out as a factor related to the development of CVD, obesity was not very prevalent at that time. It was only at the end of the 1970s that its prevalence reached epidemic levels worldwide,⁷ and today it represents one of the most serious public health challenges.⁸

Obesity is a complex disease of multiple aetiologies, with its own pathophysiologies, comorbidities, and disabilities, with direct or indirect influence on CVD, affecting endothelial and myocyte function, as well as enhancing major cardiovascular risk factors like diabetes, hypertension, and hyperlipidaemia.⁹ Thus, the problem lies in the fact that obesity evokes or exacerbates CVD. In addition, parameters indicative of autonomic nervous system (ANS) imbalance are commonly observed in these individuals, with increased sympathetic and reduced parasympathetic nervous activity.¹⁰

The ANS imbalance, or dysautonomia, often progresses to ANS dysfunction and is one of the most overlooked and misdiagnosed conditions.¹¹ The ANS plays a major role in the integrated regulation of food intake, involving satiety signals and energy expenditure; thus, ANS dysregulation might favor body weight gain. Conversely, obesity might trigger alterations in the sympathetic regulation of cardiovascular function, thus favoring the development of cardiovascular complications and events.¹² Neural mechanisms have been involved in the pathogenesis of obesity, particularly sympathovagal imbalance, and the relative prevalence of sympathetic activity has been suggested to play a pivotal role in this complex bi-directional relationship.¹³

Heart rate variability (HRV) is a non-invasive method to evaluate the modulation of the ANS on the electrophysiological sinoatrial node. By describing the oscillations between consecutive electrocardiogram (ECG) R-R intervals, HRV can measure the physiological link between the ANS and the heart.¹⁴ Thus, HRV is a relevant indicator of cardiovascular autonomic dysregulation in individuals with obesity.^{10,15}

From information from the Framingham cohort, cardiovascular risk prediction models for different CVDs have emerged.¹⁶ Variables such as age, serum cholesterol, systolic blood pressure, cigarette smoking, left ventricular hypertrophy on ECG, and glucose intolerance were included in the score.¹⁶ The 10-year risk estimates used in the 1998 score provided a convenient way to classify individuals as having low, intermediate, or high risk for future coronary heart disease.^{3,17}

Measuring the risk of a future adverse event in low-income countries from a simple ECG may be a good ally for therapeutic adjustment and directing behaviors for individuals at risk of developing CVD, especially when considering long-term clinical follow-up. Simplifying the process that defines cardiovascular risk based on isolated information (i.e., ECG data) can facilitate the identification of a larger number of individuals at risk. Recent advances in artificial intelligence strategies incorporating machine learning (ML) are gaining new applications in the clinical context, including disease prognosis,^{18–20} and may be tools to be incorporated into the definition of cardiovascular risk.

We found in state-of-the-art 11 ML studies using Framingham-style clinical variables for cardiovascular risk detection that most ML models rely on classic risk factors (age, sex, systolic blood pressure/diastolic blood pressure,^{21–31} total and high-density lipoprotein cholesterol, smoking, diabetes), sometimes extending to body mass index (BMI), medications, HbA1c, family history, estimated glomerular filtration rate, and heart rate (HR). Only one study (Yang *et al.*²⁹) used a broader feature set (49 electronic medi-

cal record variables). Classifiers are varied but traditional: Support Vector Machine appears in 4/11 studies, Random Forest in 3/11, and single instances of Logistic Regression, XGBoost, Neural Networks, and gradient boosted trees/proportional hazards regression. Validation is predominantly simple: 7/11 use hold-out splits, and 5/11 use cross-validation; no study in the table reports external validation. Cohort sizes and balance vary widely from very small and imbalanced (e.g., Dogan *et al.*²²: 504 vs 20; Navarini *et al.*²⁶: 18 vs 115) to large and balanced (e.g., Sajeev *et al.*²⁸: 23,152 vs 23,152; Quesada *et al.*²³: 5,837 vs 5,837). Performance is heterogeneous. Reported accuracies (n = 9) range from 65.41% to 93.01%, with a mean of 80.21% and a median of 82.35%. Reported area under the curves (AUCs) (n = 9) span 0.6333 to 0.9220, with a mean of 0.764 and a median of 0.751. The highest AUC is from Yang *et al.*²⁹ (stroke, XGBoost, 0.9220) on a moderately sized, hold-out cohort; the highest accuracy (93.01%) is from Dogan *et al.*²² but on a tiny, highly imbalanced dataset (504 low-risk vs 20 high-risk), which likely inflates accuracy. Among larger or balanced cohorts, AUCs cluster in the mid-0.7s to mid-0.8s (Alaa *et al.*²⁴: 0.774; Sajeev *et al.*²⁸: 0.852; Cho *et al.*³⁰: 0.751; Chun *et al.*³¹: 0.836), while Quesada *et al.*²³ is a lower outlier at 0.6333 despite balance. Methodologically, 10/11 studies reduce risk to binary endpoints (Low vs. High or event vs. no event), limiting calibration and clinical interpretability relative to full, continuous risk scoring; several explicitly note unbalanced data (at least 5/11) and small sample sizes (3/11).

This study aimed to develop and validate a multimodal ML model that integrates ECG non-linear parameters with medical features (HR, anthropometry, blood glucose lipid profile, and HRV) to classify cardiovascular risk severity in alignment with the Framingham risk score. Specifically, we sought to quantify the incremental value of multimodal integration by comparing ECG-only and medical features-only models with the combined model, to assess the relative contribution of each feature source (including ECG lead positions) to discrimination, and to make our curated dataset publicly available to support transparency, reproducibility, and further research.

Materials and methods

In this section, we describe each step of the experimental study and the introduction of the database. Figure 1 illustrates the proposed methodology, structured into five main phases: (1) data collection and curation of a database; (2) signal normalization and filtering: to ensure quality and consistency; (3) ECG multi-band decomposition via Discrete Wavelet Transform (DWT) and feature extraction: extraction of 27 non-linear features per 1-second segment across five decomposition levels, followed by dimensionality reduction using six statistical functions; (4) statistical analysis: refinement of the ECG-based feature set; and (5) classification evaluation metrics: integration of the optimized ECG features with 42 additional medical parameters, including HR, anthropometric measures, blood glucose lipid profile (BGLP), and HRV, followed by normalization and processing through a ML pipeline comprising 19 classifiers, resulting in a comprehensive classification report. More details and specific explanations for each phase are provided in the following subsections.

Experimental setup

This work was conducted using a MacBook Pro 14 with an M1 Pro chip (8-core CPU, 14-core GPU) and 16 GB of RAM, utilizing MATLAB[®] and Python coding languages. MATLAB[®], ver-

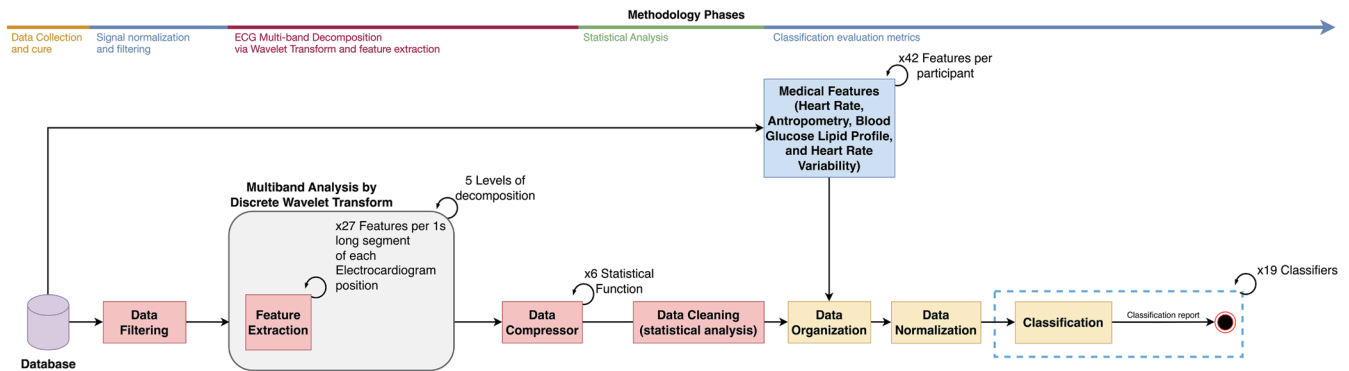


Fig. 1. Methodology workflow diagram.

sion R2023b, was used to extract the non-linear characteristics of the ECG signals, organize them with different topological medical features, and compress and organize the data for ML tasks. Python (version 3.9.12) was used to design, train/test, and obtain discrimination reports from ML models.

Database collection and curation

The database is a cross-sectional study with a quantitative approach and was conducted at the Hospital Universitário Walter Cantídio in Fortaleza, the capital of the State of Ceará, Brazil. The research sample was selected by convenience, and the research was conducted from November 2023 to May 2024 after approval by the institutional Research Ethics Committees (CAAE: 74256823.4.0000.5054 and 74256823.4.3001.5045). The ethical principles recommended by the Declaration of Helsinki and Resolution 466/12 of the Brazilian National Health Council were followed.

Individuals of both sexes aged 30 years or older with a previous nosological diagnosis of obesity, asymptomatic for heart disease, and who underwent laboratory tests (lipid profile and fasting blood glucose) within a maximum period of six months from the interview and data collection were included. Participants in the acute phase of any disease and those with an inability to communicate verbally or cognitive deficits were not included.

A total of 60 patients were initially screened based on the inclusion criteria mentioned above. Of these, five were excluded due to the presence of cardiac disease, discrepancies in diagnostic or comorbidity information, or refusal to provide complete information during the interview, leaving 55 patients eligible for further evaluation. Following anthropometric measurements, vital signs assessment, cardiovascular risk stratification using the Framingham score, and ECG acquisition, two additional patients were excluded due to high interference in HRV measurements. Consequently, the final structured collected database includes data from a total of 53 participants with no missing data (Fig. 2).

ECG recording was performed noninvasively using the PowerLab data capture hardware system, and HRV parameters were consolidated using lead II of the ECG (Labchart Pro version 7.3.4, Brazil) beat by beat in the CM5 position and analyzed by software (MATLAB® 6.1.1.450 Release 12.1.2001). The collection was performed with a sampling frequency of 1,000 Hz.

ECG acquisition was performed at rest in a climate-controlled room in the morning to minimize circadian HR variations.³² For this purpose, all volunteers were previously instructed to abstain from stimulant drugs, caffeine, tobacco, alcohol, ingestion of high-fat foods, and physical activity for at least 24 h beforehand. Re-

cordings were taken from 07:00 to 11:00 a.m. to avoid any hemodynamic effect on HRV. Participants were instructed not to talk during the assessment, thus avoiding interference that could affect the capture of the HR signal.^{33,34} Initially, the participant remained at rest on a stretcher in a supine position (ECG_D) for 5 minutes. They were then asked to perform the active postural maneuver (APM), during which they were instructed to stand up abruptly and remain in an orthostatic position (ECG_UP) without movement for 5 minutes until the end of the measurement.^{35,36}

The use of APM was considered because it is a technique with potential sensitivity for assessing vagal and cardiac sympathetic responses.³⁵ The use of APM causes reflex stimulation of the baroreceptors and contraction of the muscles of the lower limbs, thus changing the individual's position from supine to bipedal, favoring the acquisition of higher delta HRV values.³⁶

In the HRV analyses performed at rest and standing, the stability of the tracings was highlighted, excluding the phase related to the movement during posture change. This is because such movement generates intense instability in the signal due to high cardiocirculatory stress in progress, and therefore, it was not considered for validating the interpretation of the variable information.^{34,37} Thus, for interpretation purposes, the indexes used in the analysis were: standard deviation of RR intervals (hereinafter referred to as SDRR), square root of the mean of the squares of successive differences (hereinafter referred to as RMSSD), percentage of successive RR intervals with differences greater than 50 ms (hereinafter referred to as pRR50), low-frequency power (LF), high-frequency power (HF), low/high-frequency ratio (LF/HF), short-term variability (SD1), long-term variability (SD2), ratio between standard deviations 1 and 2 (SD1/SD2), and ratio between standard deviations 2 and 1 (SD2/SD1).

The volunteers were then assessed and stratified for cardiovascular risk using the Framingham score with information on sex, age, total and fractional cholesterol, blood pressure, smoking, and diabetes.³⁸ Based on the percentage estimate, cardiovascular risk was classified into three categories: Low (<5%), Moderate (5% to 20% for men and 5% to 10% for women), and High (>20% for men and >10% for women).³⁸⁻⁴⁰

The Low cardiovascular risk class included 22 participants, of whom 95.45% (21 out of 22) were females, with an average age of 35.77 years and an average BMI of 42.96 Kg/m². The Moderate cardiovascular risk class included 14 participants, of whom 71.43% (10 out of 14) were women, with an average age of 48.43 years and a mean BMI of 39.26 Kg/m².

The High cardiovascular risk class comprised 17 participants, of which 88.24% (15 out of 17) were women, with an average age

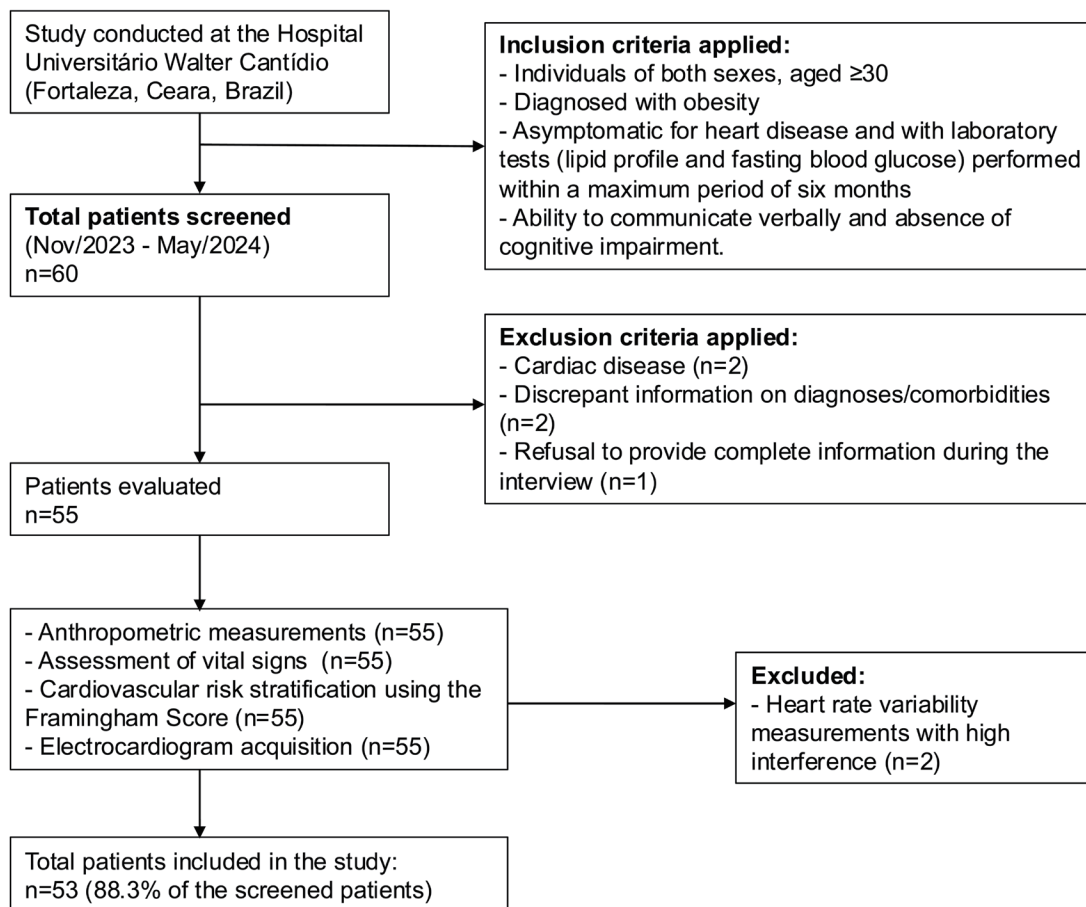


Fig. 2. Flowchart of the participant recruitment and selection process. Participant flow diagram showing the recruitment, inclusion criteria, and exclusions of the cross-sectional study.

of 54.24 years and an average BMI of 45.03 Kg/m². The number of participants results in a minor class imbalance, with a maximum ratio of 60:40, which is generally not considered critical for ML tasks' performance, as suggested by Thabtah *et al.*⁴¹ Table S1 in the Supplement provides additional details regarding the participants' characteristics.

The database is registered at Mendeley Data - DOI: 10.17632/z8mrvy259n.1, and it includes:

- An Excel file (*Framingham Patients Information.xlsx*), which contains each patient's gender, age, Framingham risk score, and a set of medical features grouped into four categories:
 - HR: resting HR, HR during APM, average HR, systolic and diastolic blood pressure, mean arterial pressure, resting double product, estimated double product during APM, and average double product.
 - Anthropometry: BMI, abdominal circumference, waist circumference, and neck circumference.
 - BGLP: LDL cholesterol, HDL cholesterol, and glycemia.
 - HRV: average RR interval, SDRR, RMSSD, pRR50, LF Power, HF Power, LF/HF Power ratio, SD1, SD2, SD1/SD2, and SD2/SD1 ratios.
- A folder named "ECG Signals", containing raw ECG recordings for each patient in.mat format.

Note that in this study, we additionally combined non-linear features extracted from the ECG signals with the clinical data pro-

vided in the Excel file to enhance the analysis of cardiovascular risk.

Signal normalization

The ECG signals, $x(n)$, were loaded into MATLAB[®] and normalized according to the Root Mean Square normalization formula.⁴²

$$x(n) = \frac{x(n)}{\sqrt{\frac{\sum_{n=0}^{N-1} x^2(n)}{N}}}$$

where N represents the signal's length. Then, the average value was removed from the entire signal.

Signal filtering

The signals were sampled at a frequency rate of 1,000 Hz, and an elliptic band-pass filter of order 16 with frequency cut-offs at 1 Hz and 40 Hz,⁴³ a steepness of 0.85, and a stop band attenuation of 60 dB was applied to the signal. Figure 3 presents the 2 seconds per class after signal filtering.

ECG Multi-band decomposition via wavelet transform and feature extraction

The DWT is a robust tool for analyzing finite-energy discrete-time signals. It decomposes a signal into a family of base functions con-

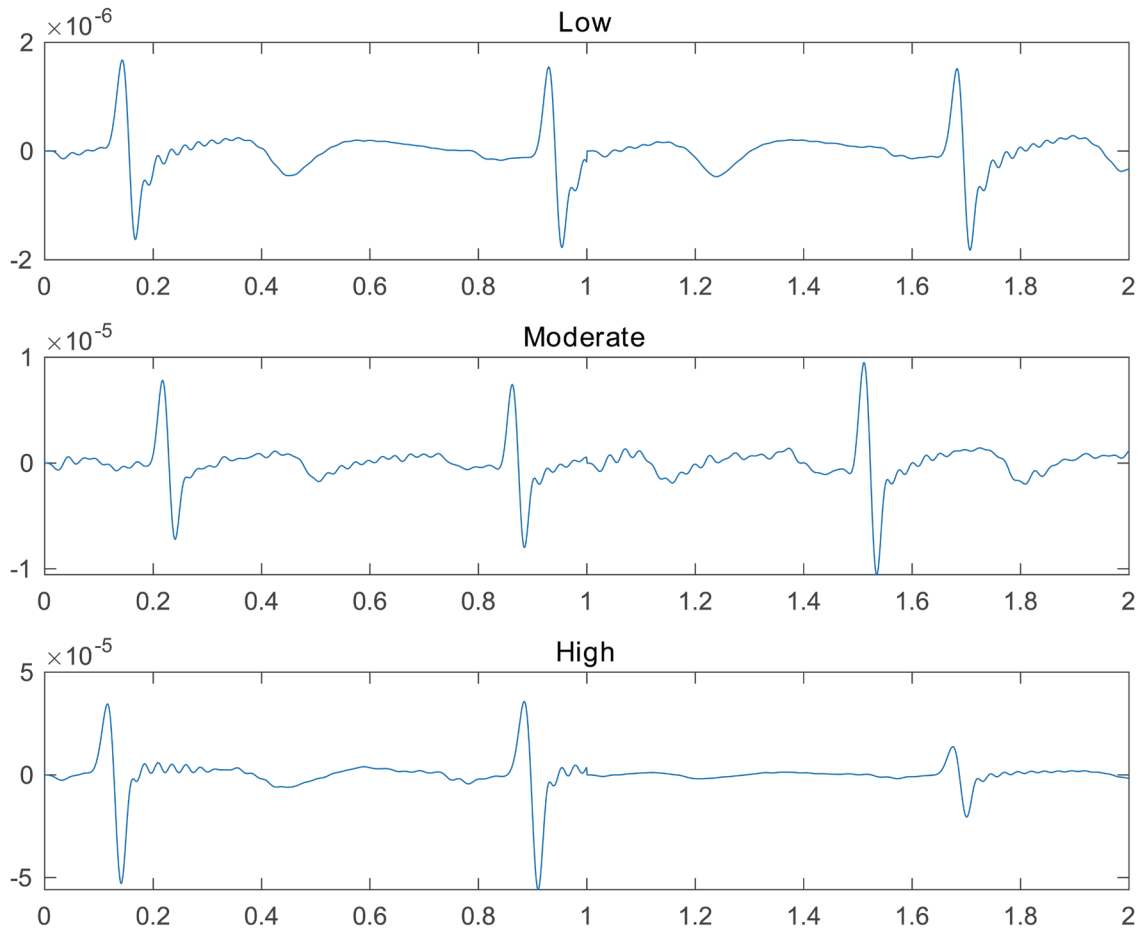


Fig. 3. 2-second electrocardiogram signal example of each cardiovascular risk class after normalization and filtering.

structured from a small set of prototype sequences and their time shifts, enabling compact, time-frequency-localized representations that are well-suited to nonstationary signals. Decomposition and perfect reconstruction are performed with an octave-band, critically decimated filter bank, following the framework introduced by Malvar *et al.*⁴⁴ and extended by Vetterli *et al.*⁴⁵ When focusing on positive frequencies, each sub-band in the transform is confined to a specific range:

$$W_k = \begin{cases} \left[0, \frac{\pi}{2^S}\right], & m = 0, \\ \left[\frac{\pi}{2^{S-m+1}}, \frac{\pi}{2^{S-m}}\right], & m = 1, 2, \dots, S, \end{cases}$$

where S is the number of levels, $S+1$ is the number of sub-bands, and π is the normalized angular frequency equivalent to half the sampling rate. The DWT utilizes an analysis scale function, $\tilde{\phi}_1(n)$, and an analysis wavelet function, $\tilde{\psi}_1(n)$, defined as:

$$\tilde{\phi}_1(n) = h_{LP}(n)$$

and

$$\tilde{\psi}_1(n) = h_{HP}(n),$$

where $h_{LP}(n)$ and $h_{HP}(n)$ are the impulse responses of the low-pass and high-pass analysis filters, respectively.

The following recursion formulas are defined:

$$\begin{aligned} \tilde{\phi}_{i+1}(n) &= \tilde{\phi}_i(n/2) * \tilde{\phi}_1(n), \\ \tilde{\psi}_{i+1}(n) &= \tilde{\phi}_i(n) * \tilde{\psi}_1(n/2^i), \end{aligned}$$

where the symbol “*” denotes the convolution operation. The analysis filter for the m^{th} sub-band is expressed as:

$$h_m(n) = \begin{cases} \tilde{\phi}_S(n), & m = 0 \\ \tilde{\psi}_{S+1-m}(n) & m = 1, 2, \dots, S. \end{cases}$$

The m^{th} sub-band signal is computed as:

$$x_m(n) = \begin{cases} \sum_{k=-\infty}^{\infty} x(k)h_m(2^S n - k), & m = 0 \\ \sum_{k=-\infty}^{\infty} x(k)h_m(2^{S-m+1} n - k) & m = 1, 2, \dots, S. \end{cases}$$

In this study, the DWT was used to decompose each ECG signal into sub-bands ($x_m(n)$) up to level five ($S = 5$). The Symlet7 wavelet was selected for its strong performance in ECG analysis up to five decomposition levels.^{19,25} To maintain consistency with the original sampling rate, the sub-band signals $x_m(n)$ were resampled to the native rate using a wavelet interpolation method. For example, Figure 4 provides a visual representation of the multiscale analysis of a 1-second ECG segment performed using the DWT.

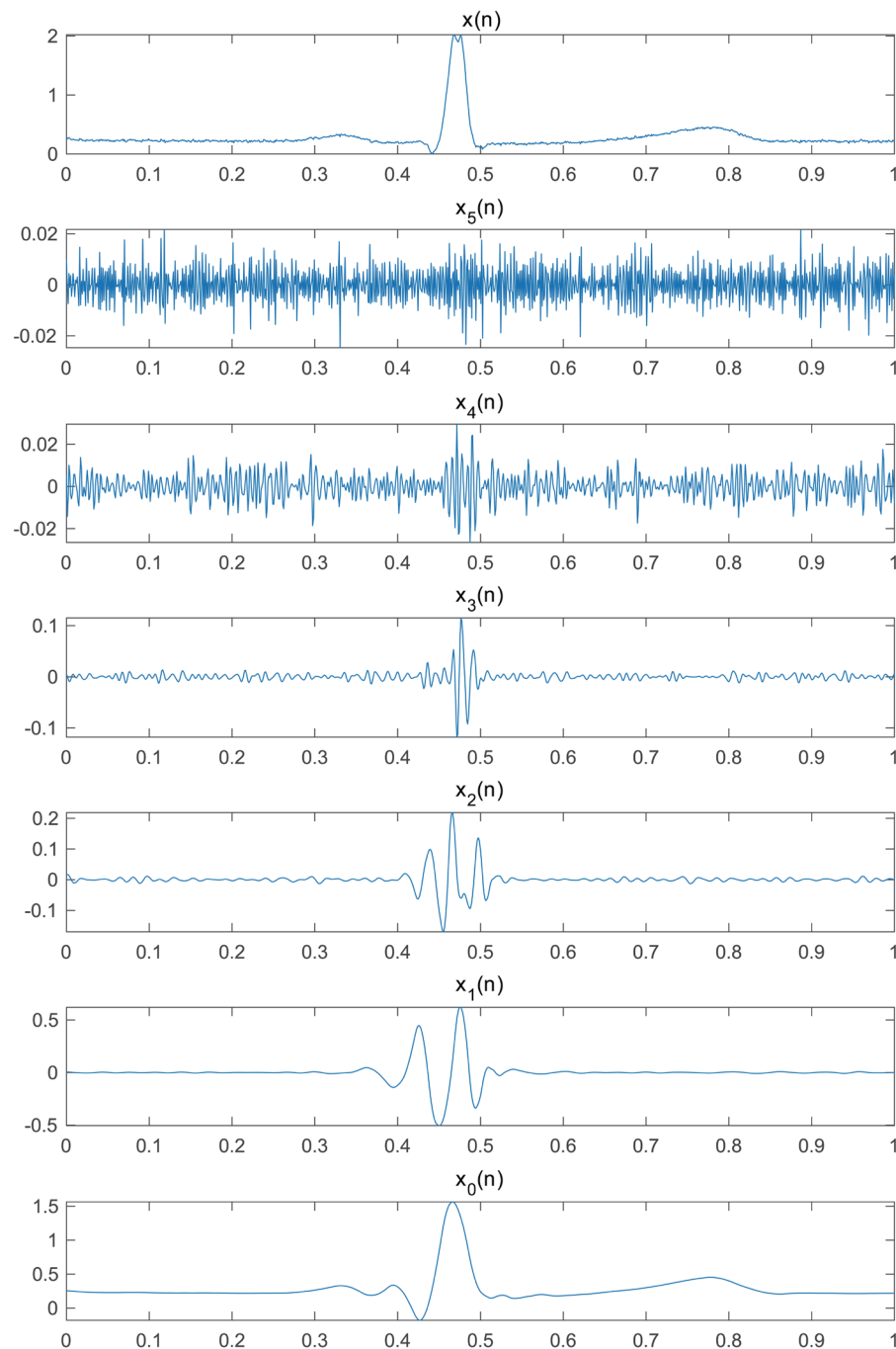


Fig. 4. Multiscale analysis of a 1s electrocardiogram segment using the discrete wavelet transform.

Feature extraction

After that, 27 non-linear features (see Table 1 for more information^{46–68}) were collected from the signal's full-band and sub-bands every 1 s within a sliding rectangular non-overlapping window analysis. Then, the resulting time series of features per feature and band were compressed over time, respectively, by six distinct statistical functions: average (Avg), standard deviation (Std), 95th percentile (P95), variance (Var), median (Med), and kurtosis (Kur).¹⁹

At the end of the feature extraction from ECG analysis, we obtained a total of 1,944 non-linear features (27 non-linear features extracted from $[\times] 6$ bands (ECG full-band + sub-bands), compacted by $[\times] 6$ data compressors in $[\times] 2$ different positions).

Statistical analysis

To optimize the feature selection process, we performed a total of 1,944 statistical analyses with the Kruskal-Wallis statistical test using MATLAB® Statistics and Machine Learning Toolbox, which

Table 1. Extracted features with corresponding equations and definitions

Feature	Definition	Equation
Approximate Entropy (AE)	AE assesses the likelihood that similar patterns within the data remain consistent when additional data points are included. A lower AE suggests more regular or predictable data, while a higher AE indicates greater complexity or irregularity	$AE(m, r) = \lim_{N \rightarrow \infty} \frac{\Theta^m(r) - \Theta^{m+1}(r)}{N - m + 1}$ <p>Θ is the Heaviside step function and can be defined as</p> $\Theta^m(r) = \frac{\sum_j \ln [C_j^m(r)]}{N - m + 1}$ <p>where r denotes a predetermined tolerance, N is the number of data points in the time series, and m is the dimension⁴⁶</p>
Correlation Dimension (CD)	CD measures self-similarity, with higher values indicating a high degree of complexity and less similarity	$CD = \lim_{M \rightarrow \infty} \frac{2 \sum_{i=1}^{M-k} \sum_{j=i+k}^M \Theta(l X_i - X_j)}{M^2}$ <p>where $\Theta(x)$ is the Heaviside step function, X_i and X_j are position vectors on the attractor, l is the distance under consideration, k is the summation offset, and M is the number of reconstructed vectors from $x(n)$⁴⁶</p>
Detrended Fluctuation Analysis (DFA)	DFA measures power scaling observed through R/S analysis	$DFA(n) = \sqrt{\frac{\sum_{k=1}^n [y(k) - \bar{y}_n(k)]^2}{n}}$ <p>where N is the length, $y_n(k)$ is the local trend, and $y(k)$ is defined as</p> $y(k) = \sum_{i=1}^k [x(i) - \bar{x}]$ <p>with $x(i)$ as the inter-beat interval and \bar{x} as its average⁴⁷</p>
Energy (En)	En represents the system capacity for performing work ⁴⁸	$En = \sum_{n=0}^{N-1} x(n) ^2$
Higuchi Fractal Dimension (H)	H estimates the fractal dimension of a time series ⁴⁹	$H = \frac{\ln(L(k))}{\ln\left(\frac{1}{k}\right)}$ <p>where k is the number of composed sub-series, and $L(k)$ is the average curve size</p>
Hurst Exponent (EH)	EH quantifies how chaotic or unpredictable a time series is	$K_q(\tau) \sim \left(\frac{\tau}{v}\right)^{qEH(q)}$ <p>with</p> $K_q(\tau) = \frac{ X(t+\tau) - X(t) ^q}{ X(t) ^q}$ <p>where q is the order moments of the distribution increments, v is the time resolution, τ is the incorporation time delay of the attractor, and t is the period of a given time series signal $X(t)$⁵⁰</p>

(continued)

Table 1. (continued)

Feature	Definition	Equation
Katz Fractal Dimension (<i>K</i>)	<i>K</i> estimates the fractal dimensions through a waveform analysis of a time series ⁵⁰	$K = \frac{\log(n)}{\log(n) + \log \left(\frac{\max_n \left(\sqrt{(n-1)^2 + (x(n) - x(1))^2} \right)}{\sum_{n=2}^N \sqrt{1 + (x(n-1) - x(n))^2}} \right)}$
Lyapunov Exponent (<i>ELy</i>)	<i>ELy</i> evaluates the system's predictability and sensitivity to change	$ELy(x_0) = \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \ln f'(x_k - 1) }{n}$ <p>where <i>f'</i> is the derivative of <i>f</i>⁵¹</p>
Logarithmic Entropy (<i>LE</i>)	<i>LE</i> quantifies the average amount of information (in bits) needed to represent each event in the probability distribution. Higher logarithmic entropy values indicate greater unpredictability or randomness, while lower values suggest more certainty or order ⁴⁸	$LE = \sum_{n=1}^N \log_2 x(n) ^2$
Shannon Entropy (<i>SE</i>)	<i>SE</i> are measured in bits when the base-2 logarithm (\log_2) is used, quantifying the average bits required to represent each outcome in a probability distribution. Higher entropy values indicate greater uncertainty, unpredictability, or randomness, while lower values suggest more order or certainty ⁴⁸	$SE = - \sum_{n=1}^N x(n) ^2 \log_2 x(n) ^2$
Sample Entropy (<i>SampEn</i>)	where <i>m</i> is the length of the vector, <i>r</i> is the tolerance, <i>N</i> is the total size of the vector, and <i>n</i> is a vector portion that is being analysed ⁵²	$SampEn(m, r, N) = - \log \left(\frac{n^{m+1}}{n^m} \right)$
Fuzzy Entropy (<i>FuzzEn</i>)	where <i>r</i> is the resolution and <i>m</i> is the dimension ⁵³	$FuzzEn(x, m, r, n_e, d) = - \ln \left(\frac{\psi^{m+1,d}(n_e, r)}{\psi^{m,d}(n_e, r)} \right)$ <p>where $\psi^{m,d}(n_e, r)$ is defined as</p> $\psi^{m,d}(n_e, r) = \frac{\sum_{\lambda=1}^{N-md} \sum_{\lambda+1 \leq j \leq N} \exp(\Delta_{\lambda,j}^{n_e})}{(N-md)(N-md-1)}$
Kolmogorov Entropy (<i>K2En</i>)	where $C_d(\epsilon)$ is the correlation integral ⁵⁴	$K2En = \lim_{\epsilon \rightarrow 0} \lim_{d \rightarrow \infty} \frac{\ln \frac{C_d(\epsilon)}{C_{d+1}(\epsilon)}}{\tau}$
Permutation Entropy (<i>PermEn</i>)	where <i>P</i> is the probability distribution for each symbol ⁵⁵	$PermEn(m) = - \sum_{y=1}^k P_y \ln P_y$
Conditional Entropy (<i>CondEn</i>)	where <i>E</i> is the Shannon Entropy, and <i>L</i> is the dimension ⁵⁶	$CondEn \left(\frac{L}{L-1} \right) = E(L) - E(L-1)$

(continued)

Table 1. (continued)

Feature	Definition	Equation
Distribution Entropy (<i>DistEn</i>)	where p is the probability of each bin ⁵⁷	$DistEn = -\frac{\sum_{i=1}^M p_i \log_2(p_i)}{\log_2(M)}$
Dispersion Entropy (<i>DispEn</i>)	where $\pi_{v_0^{(1)}, \dots, v_{m-1}^{(m)}}$ is the dispersion pattern and c^m is the potential dispersion patterns ⁵⁸	$DispEn = -\sum_{\pi_{v_0^{(1)}, \dots, v_{m-1}^{(m)}}} p(\pi_{v_0^{(1)}, \dots, v_{m-1}^{(m)}}) \ln(p(\pi_{v_0^{(1)}, \dots, v_{m-1}^{(m)}}))$ <p>where $p(\pi_{v_0^{(1)}, \dots, v_{m-1}^{(m)}})$ can be defined as</p> $p(\pi_{v_0^{(1)}, \dots, v_{m-1}^{(m)}}) = \frac{\#\{i i \leq N - (m-1)d, Z_i^{m,c}\}}{N - (m-1)}$
Spectral Entropy (<i>SpecEn</i>)	where $S(f)$ is the power spectrum and f_n is the upper limit Frequency ⁵⁹	$SpecEn = -\sum_{f=0}^{f_n} S(f) \log S(f)$
Symbolic Dynamic Entropy (<i>SyDyEn</i>)	where $p(l)$ is the probability of occurrence and $c(l)$ is the frequency of occurrence ⁶⁰	$SyDyEn = -\sum_{l=1}^{4^m} p(l) \log p(l)$ <p>and $p(l)$ is defined as</p> $p(l) = \frac{c(l)}{N - m + 1}$
Increment Entropy (<i>IncrEn</i>)	where $p(w_n)$ is the frequency of each unique word (w_n), q is the quantifying precision, and m is the dimension ⁶¹	$IncrEn = -\frac{\sum_{n=1}^{2q+1} p(w_n) \log p(w_n)}{m-1}$ <p>and $p(w_n)$ is defined as</p> $p(w_n) = \frac{q(w_n)}{N - m}$
Cosine Similarity Entropy (<i>CosEn</i>)	where $B_{(c)}^{(m)}(r_{CSE})$ is the global probability of occurrences of similar patterns and r_{CSE} is the tolerance ⁶²	$CosEn = -\left[B_{(c)}^{(m)}(r_{CSE}) \log_2 B_{(c)}^{(m)}(r_{CSE}) + (1 - B_{(c)}^{(m)}(r_{CSE})) \log_2 (1 - B_{(c)}^{(m)}(r_{CSE})) \right]$
Phase Entropy (<i>PhasEn</i>)	where the $p(i)$ is the probability distribution and k is the maximum sector of φ ⁶³	$PhasEn = -\frac{\sum_{i=1}^k p(i) \log p(i)}{\log k}$

(continued)

Table 1. (continued)

Feature	Definition	Equation
Slope Entropy (<i>SlopEn</i>)	<i>SlopEn</i> is defined as the computation of a sub-sequence of symbols where <i>d</i> is the difference between two consecutive samples, and γ and δ are thresholds ⁶⁴	$\left\{ \begin{array}{l} d > \gamma \rightarrow 2 \\ d \leq \gamma \ \& \ d > \delta \rightarrow 1 \\ d \leq \delta \rightarrow 0 \\ d < -\delta \ \& \ d > -\gamma \rightarrow -1 \\ d < -\gamma \rightarrow -2 \end{array} \right.$
Bubble Entropy (<i>BubbEn</i>)	where H_{swaps}^m is the conditional Rényi entropy ⁶⁵	$BubbEn = \frac{H_{swaps}^{m+1} - H_{swaps}^m}{\log \frac{m+1}{m-1}}$
Gridded Distribution Entropy (<i>GridEn</i>)	where <i>j</i> is the <i>j</i> th grid, the p_j is the probability of each grid, <i>b</i> is the number of points in the <i>j</i> th grid, and <i>N</i> is the length of the RR intervals ⁶⁶	$GridEn = -\sum_{j=1}^{n^2} p_j \log p_j$ $p_j = \frac{b}{N - m}$ <p>where p_j is defined as</p>
Entropy of Entropy (<i>EnofEn</i>)	where <i>i</i> is the level index, $y_i^{(t)}$ is the sequence, and N/τ is the number of representative states for each original time series ⁶⁷	$EnofEn = -\sum_{i=1}^{d_s} p_i \ln p_i$ $p_i = \frac{\text{Number of } y_j^{(t)} \ln i}{N / \tau}$ <p>and p_i is defined as</p>
Attention Entropy (<i>AttrEn</i>)	where $a_{i,j}$ are the attention weight probabilities between the positions <i>i</i> and <i>j</i> and $a'_{i,j}$ is the averaged weights ⁶⁸	$AttrEn = -\sum_{j=0}^{d_s} a_{i,j} \log a_{i,j}$ $a_{i,j} = \frac{e^{a'_{i,j}}}{\sum_j e^{a'_{i,j}}}$ <p>and $a_{i,j}$ is defined as</p>

Table 2. Number of features per origin

	# of features
Non-linear features	
ECG_D	39
ECG_UP	21
Medical features	
Heart rate	9
Anthropometry	4
Blood glucose lipid Profile	3
Heart rate Variability	26

ECG, electrocardiogram.

evolved all the group's parameter distributions of features. Only those analyses showing significant differences ($p < 0.05$) between class distributions were selected, and the others were discarded. For multiclass distribution, we applied the Bonferroni correction. The percentage of statistical analyses with significant differences was approximately 3.1% (60 out of 1,944). Additionally, we combined the selected ECG features with the medical features (defined in Section 2.2), initially provided by health professionals, for use in the ML discrimination stage. Table 2 shows the number of features per origin after filtering.

Table 3. 19 Scikit-learn machine learning classifiers configuration

Classifier – Scikit-learn class (Abbreviation)	Hyperparameters
AdaBoostClassifier (AdaBoost)	n_estimators = 50, learning_rate = 1.0, algorithm = 'SAMME.R' + Default
BaggingClassifier (BaggC)	n_estimators = 10, max_samples = 1.0, bootstrap = true + Default
DecisionTreeClassifier (DeTreeC)	max_depth = 5, criterion = 'gini', splitter = 'best', min_samples_split = 2 + Default
ExtraTreesClassifier (ExTreeC)	n_estimators = 300, criterion = 'gini', max_features = 'auto', bootstrap=false + Default
GaussianNB (GauNB)	priors = none, var_smoothing = 1e-9, store_covariance = false + Default
GaussianProcess Classifier (GauPro)	1.0 * RBF (1.0), optimizer = 'fmin_l_bfgs_b', max_iter_predict = 100, copy_X_train = true + Default
GradientBoosting Classifier (GradBoost)	loss = 'log loss', learning_rate = 0.1, n_estimators = 100 + Default
KNearestNeighbors Classifier (KNN)	n_neighbors = 5, weights = 'uniform', algorithm = 'auto' + Default
LinearDiscriminant Analysis (LinDis)	solver = 'svd', shrinkage = none, priors = none + Default
LinearSVC (LinSVC)	penalty = 'l2', loss = 'squared_hinge', dual = True + Default
LogisticRegression (LogReg)	solver = "lbfgs", penalty = 'l2', C = 1.0, max_iter = 100 + Default
LogisticRegressionCV (LogRegCV)	cv = 3, penalty = 'l2', solver = 'lbfgs', max_iter = 100 + Default
MLPClassifier (MLP)	$\alpha = 1$, max_iter = 1,000, hidden_layer_sizes = 100, activation = 'relu', solver = 'adam' + Default
OneVsRestClassifier (OvsR)	estimator = LinearSVC(random_state = 0), n_jobs = none, verbose = false + Default
QuadraticDiscriminantAnalysis (QuadDis)	reg_param = 0.0, priors = none, store_covariance = false + Default
RandomForest Classifier (RF)	max_depth = 5, n_estimators = 300, criterion = 'gini', bootstrap = true, min_samples_split = 2 + Default
SGDClassifier (SGD)	max_iter = 100, tol = 0.001, penalty = 'l2', loss = 'hinge', alpha = 0.0001 + Default
SGDClassifierMod (SGDCMod)	penalty = 'l2', loss = 'hinge', alpha = 0.0001 + Default
Support-vector Machines (SVC)	$\gamma = \text{"auto"}$, kernel = 'Radial Basis Function', C = 1.0, probability = false + Default

RBF, radial basis function.

Classification evaluation metrics

To train and test the ML models using a nested leave-out cross-validation procedure, we utilized data derived from three sources: ECG-only, medical features-only, and a combination of both as a multimodal model. Cross-validation techniques allow the entire dataset to be used in classification while preventing data leakage between training and testing sets. This method is beneficial for deriving classification conclusions from small datasets.⁶⁹

For all models, data were pre-filtered using the analysis of variance F-value feature selector, which selects feature hierarchy based on their F-statistic power. This approach maximizes model performance by iteratively feeding the model with different feature combinations to identify the optimal set for each discriminative task (Low vs. Moderate, Low vs. High, Moderate vs. High, and All vs. All). Specifically:

1. For ECG-only models, the best combination of features ranged from 1 to 60.
2. For medical features-only models, the best combination of features ranged from 1 to 42.
3. For multimodal models, the best combination of features ranged from 1 to 102.

This approach resulted in three optimized models, each responsible for one of the four discrimination tasks previously identified. These models were selected from a pool of 19 pre-designed Scikit-learn ML classifiers,⁷⁰ as shown in Table 3, with the hyperparameter tuning step selected within Nested leave-one-out cross-validation (LOOCV) due to the constraints of our dataset size. Given

Table 4. Best classification results for the electrocardiogram (ECG)-only models per comparison group

Comparison group	# of features	Feature combination (category percentage)	Classifier	Acc (%)	Rec (%)	Prec (%)	Spec (%)	F1 – Score (%)	Negative predictive value (%)	Area under the curve
Low vs. Moderate	28	Electrocardiogram_D (100%) & Electrocardiogram_UP (0%)	DeTreeC	80.56	86.36	82.61	71.43	84.44	76.92	0.7890
Low vs. High	18	Electrocardiogram_D (\approx 83.3%) & Electrocardiogram_UP (\approx 16.7%)	SGD	97.44	100	95.65	94.12	97.78	100	0.9706
Moderate vs. High	23	Electrocardiogram_D (\approx 65.2%) & Electrocardiogram_UP (\approx 34.8%)	MLP	93.55	85.71	100	100	92.31	89.47	0.9286
All vs. All	47	Electrocardiogram_D (\approx 68.1%) & Electrocardiogram_UP (\approx 31.9%)	LinSVC	86.79	86.79	86.92	93.06	86.86	93.47	0.9346

the sample size, we therefore relied on rigorous internal validation, using Nested LOOCV (a strategy widely used in the literature for small-sample studies) with all model selection and tuning confined to the training folds,^{71–75} and we report performance aggregated across folds along with calibration and sensitivity analyses. While this does not replace external validation, it provides an unbiased estimate of out-of-sample performance.⁷⁶ The selection of the best results was based on their AUC performance for distinguishing between comparison groups.

AUC quantifies a binary classifier’s ability to distinguish between positive and negative instances across all decision thresholds by comparing the true positive rate (TPR) to the false positive rate (FPR). AUC ranges from 0 to 1, where 1 indicates a perfect classifier and 0.5 indicates performance equivalent to random guessing. Because it summarizes performance across thresholds, AUC provides a single, threshold-independent measure that is especially useful for model comparison under class imbalance.⁷⁷ In addition to AUC, we report the following metrics: Accuracy (*Acc*), Precision (*Prec*; positive predictive value), Recall (*Rec*; sensitivity), Specificity (*Spec*), *F1 – Score*, and negative predictive value (NPV). Let *TP*, *TN*, *FP*, and *FN* denote true positives, true negatives, false positives, and false negatives, respectively.

Acc is the proportion of correctly classified instances among all instances.⁷⁸

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%,$$

Prec, the proportion of predicted positives that are truly positive,⁷⁹ can be defined as

$$Prec = \frac{TP}{TP + FP} \times 100\%,$$

NPV shows the proportion of predicted negatives that are truly negative and can be defined as⁸⁰

$$NPV = \frac{TN}{TN + FN} \times 100\%,$$

Rec represents the proportion of actual positives that are correctly identified and is defined as⁷⁹

$$Rec = \frac{TP}{TP + FN} \times 100\%,$$

Spec is the proportion of actual negatives that are correctly identified.⁸¹

$$Spec = \frac{TN}{TN + FP} \times 100\%$$

F1 – Score is the harmonic mean of *Prec* and *Rec*, balancing both types of error,⁸² and it is defined as

$$F1 - Score = \frac{2 \times Prec \times Rec}{Prec + Rec} \times 100\%.$$

Results

We divided these results into four subsections: (1) the performance of the ECG-only model, trained and tested exclusively with ECG features; (2) the performance of the medical features-only model, trained and tested on medical features; (3) the performance of the multimodal models, trained with both ECG and medical features; and (4) a comparison between the performance of the three classes.

ECG-only models’ performance

Table 4 shows the classification report for the ECG-only models. Models trained exclusively on ECG-derived features achieved strong discrimination across all comparison groups. For Low vs. Moderate, the best model using 100% of features from the ECG_D position reached an *Acc* of 80.36%, *Rec* of 86.36%, *Prec* of 82.61%, *Spec* of 71.43%, *F1 – Score* of 84.44%, NPV of 76.92%, and an AUC of 0.7890. For Low vs. High, the best model, trained with 83.3% of features from ECG_D and the remainder from ECG_UP, attained an *Acc* of 97.44%, *Rec* of 100%, *Prec* of 95.65%, *Spec* of 94.12%, *F1 – Score* of 97.78%, NPV of 100%, and an AUC of 0.9706. For Moderate vs. High, performance included an *Acc* of 93.55%, *Rec* of 85.71%, *Prec* of 100%, *Spec* of 100%, *F1 – Score* of 92.31%, NPV of 89.47%, and an AUC of 0.9286, with 65.2% of discriminative information coming from ECG_D. In the All vs. All comparison, the model achieved an *Acc* of 86.79%, *Rec* of 86.92%, *Prec* of 86.92%, *Spec* of 93.06%, *F1 – Score* of 86.86%, NPV of 93.47%, and an AUC of 0.9346. Collectively, these results indicate that the ECG_D position contributed the most informative features for maximizing discrimination, with ECG_UP playing a secondary role, particularly in the Moderate vs. High comparison.

Medical features models’ performance

Table 5 illustrates the classification results for the medical fea-

Table 5. Best classification results for the medical features-only models per comparison group

Comparison group	# of features	Feature combination (category percentage)	Classifier	Acc (%)	Rec (%)	Prec (%)	Spec (%)	F1 – Score (%)	Negative predictive value (%)	Area under the curve
Low vs. Moderate	7	Heart rate (\approx 28.6%) & Anthropometry (0%) & Blood glucose lipid profile (\approx 14.3%) & Heart rate variability (\approx 57.1%)	QuadDis	83.33	86.36	86.36	78.57	86.36	78.57	0.8247
Low vs. High	26	Heart rate (\approx 27%) & Anthropometry (\approx 11.5%) & Blood glucose lipid profile (0%) & Heart rate variability (\approx 61.5%)	QuadDis	87.18	86.36	90.48	88.24	88.37	83.33	0.8730
Moderate vs. High	29	Heart rate (0%) & Anthropometry (\approx 13.8%) & Blood glucose lipid profile (\approx 6.9%) & Heart rate variability (\approx 79.3%)	LinDis	80.65	92.86	72.22	70.59	81.25	92.31	0.8172
All vs. All	5	Heart rate (0%) & Anthropometry (0%) & Blood glucose lipid profile (20%) & Heart rate variability (80%)	SGDMod	60.38	60.38	60.03	77.66	60.20	80.17	0.6626

tures-based models. When trained exclusively on medical parameters (HR, anthropometry, BGLP, and HRV), models showed competitive but generally lower performance than ECG-based models, except in one case. For Low vs. Moderate, *Acc* was 83.33%, *Rec* 86.36%, *Prec* 86.36%, *Spec* 78.57%, *F1 – Score* 86.36%, *NPV* 78.57%, and *AUC* 0.8247. For Low vs. High, *Acc* reached 87.18%, with *Rec* 86.36%, *Prec* 90.48%, *Spec* 88.24%, *F1 – Score* 88.24%, *NPV* 83.33%, and *AUC* 0.8730. For Moderate vs. High, *Acc* was 80.65%, *Rec* 92.86%, *Prec* 72.22%, *Spec* 70.59%, *F1 – Score* 81.25%, *NPV* 92.31%, and *AUC* 0.8172. In the All vs. All comparison, *Acc* was 60.38%, *Rec* 60.38%, *Prec* 60.03%, *Spec* 77.66%, *F1 – Score* 60.20%, *NPV* 80.17%, and *AUC* 0.6626. HRV features were the most influential among the medical parameters, contributing 57.1%, 61.5%, 79.3%, and 80% of the discriminative information for Low vs. Moderate, Low vs. High, Moderate vs. High, and All vs. All, respectively. Comparing ECG-based and medical features-only models, the medical features-only model was 3.45% more accurate for Low vs. Moderate but was less accurate by 10.78% to 26.42% for Low vs. High, Moderate vs. High, and All vs. All, indicating stronger ECG-driven discrimination, particularly in binary tasks involving the High class and in multiclass settings.

Multimodal models' performance

Table 6 corresponds to the best classification results for the multimodal models (ECG features + medical features). Combining ECG and medical features yielded the best overall results in selected comparisons. For Low vs. Moderate, the multimodal model achieved an *Acc* of 88.89%, *Rec* 90.91%, *Prec* 90.91%, *Spec* 85.71%, *F1 – Score* 90.91%, *NPV* 85.71%, and *AUC* 0.8831, improving accuracy by 10.34% over the ECG-only model. For Low vs. High, the combined model matched the ECG-only performance with an *Acc* of 97.44%, *Rec* 100%, *Prec* 95.65%, *Spec* 94.12%, *F1 – Score* 97.78%, *NPV* 100%, and *AUC* 0.9706. For Moderate vs. High, *Acc* was 93.55%, *Rec* 100%, *Prec* 87.50%, *Spec* 88.24%, *F1 – Score* 93.33%, *NPV* 100%, and *AUC* 0.9412. For All vs. All, performance was comparable to ECG-only: *Acc* 86.79%, *Rec* 86.79%, *Prec* 86.92%, *Spec* 93.06%, *F1 – Score* 86.86%, *NPV* 93.47%, and *AUC* 0.9346. These patterns reflect that feature selec-

tion often discarded medical variables in the Low vs. High and All vs. All settings, effectively aligning the combined model's behavior with ECG-only models in those comparisons.

Comparison between different types of models' performance

Across models, Figure 5 summarizes AUC behavior between the approaches presented in Tables 4–6, for each comparison group: adding ECG to medical features improved AUC in every group (by 7.1% to 41.1%), while adding medical features to ECG improved AUC for Low vs. Moderate and Moderate vs. High but not for Low vs. High or All vs. All due to the feature selection process. Confusion matrices in Figure 6 show a small number of misclassifications: in Low vs. Moderate, two participants per class were misclassified; in Low vs. High, one High participant was predicted as Low; in Moderate vs. High, two High participants were misclassified; and in All vs. All, most errors occurred between Low and Moderate. Receiver-operating characteristic curves (Fig. 7) approached near-perfect classification with AUCs of 0.8831, 0.9706, 0.9412, and 0.9346 for Low vs. Moderate, Low vs. High, Moderate vs. High, and All vs. All, respectively.

Discussion

The results demonstrate that ECG-derived information, particularly from the ECG_D position, is the dominant discriminative signal. Compared with medical features-only models, the ECG-only model achieved higher *Acc* in Low vs. High (97.44% vs. 87.18%; *AUC* +11.2% relative), Moderate vs. High (93.55% vs. 80.65%; *AUC* +13.6%), and All vs. All (86.79% vs. 60.38%; *AUC* +41.1%), with the sole exception of Low vs. Moderate, where the medical features-only model slightly surpassed the ECG-only model in *Acc* (83.33% vs. 80.36%; *AUC* +4.5% relative). Multimodal integration delivered its largest gains for adjacent strata, improving Low vs. Moderate *Acc* by 10.6% relative over ECG-only (88.89% vs. 80.36%) and raising *AUC* by 11.9% vs. ECG-only and 7.1% vs. medical features-only. For Moderate vs. High, *AUC* increased by 1.4% vs. ECG-only and 15.2% vs. medical features-only, with accuracy unchanged. In Low vs. High and All vs. All, multimodal

Table 6. Best classification results for the multimodal models per comparison group

Comparison group	# of features	Feature combination (category percentage)	Classifier	Acc (%)	Rec (%)	Prec (%)	Spec (%)	F1 – Score (%)	Negative predictive value (%)	Area under the curve
Low vs. Moderate	33	Electrocardiogram_D (0%) & Electrocardiogram_UP (~51.5%) & Heart Rate (0%) & Anthropometry (~3%) & Blood Glucose lipid profile (~3%) & Heart rate variability (~42.5%)	LinDis	88.89	90.91	90.91	85.71	90.91	85.71	0.8831
Low vs. High	18	Electrocardiogram_D (~83.3%) & Electrocardiogram_UP (~16.7%) & Heart rate (0%) & Anthropometry (0%) & Blood glucose lipid profile (0%) & Heart rate variability (0%)	SGD	97.44	100	95.65	94.12	97.78	100	0.9706
Moderate vs. High	31	Electrocardiogram_D (~83.3%) & Electrocardiogram_UP (~16.7%) & Heart Rate (0%) & Anthropometry (0%) & Blood glucose lipid profile (0%) & Heart rate variability (0%)	LinDis	93.55	100	87.50	88.24	93.33	100	0.9412
All vs. All	47	Electrocardiogram_D (~68.1%) & Electrocardiogram_UP (~31.9%) & Heart rate (0%) & Anthropometry (0%) & Blood glucose lipid profile (0%) & Heart rate variability (0%)	LinSVC	86.79	86.79	86.92	93.06	86.86	93.47	0.9346

performance matched ECG-only (0% change in *Acc* and AUC) because feature selection discarded medical variables, suggesting that ECG alone captured the dominant signal or that selection favored parsimony over complementarity. The multimodal findings underscore both the potential and the limitations of feature selection in combined pipelines. In Low vs. Moderate and Moderate

vs. High, combining modalities improved discrimination (AUC +11.9% and +1.4% versus ECG-only; +7.1% and +15.2% versus medical features-only), with a notable *Acc* gain only for Low vs. Moderate (+10.3%), consistent with the idea that adjacent categories benefit from complementary features. In contrast, for Low vs. High and All vs. All, the feature selection procedure eliminated

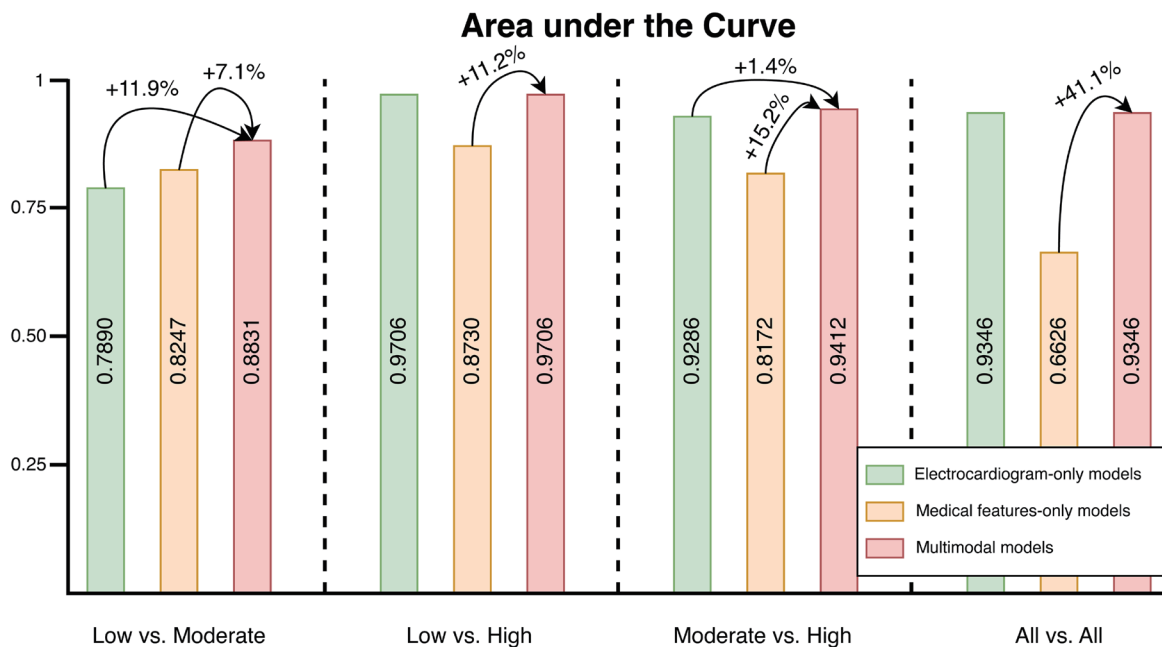


Fig. 5. Area under the curve (AUC) comparison results, between the electrocardiogram-only models, the medical features-only models, and the multimodal models, per comparison groups.

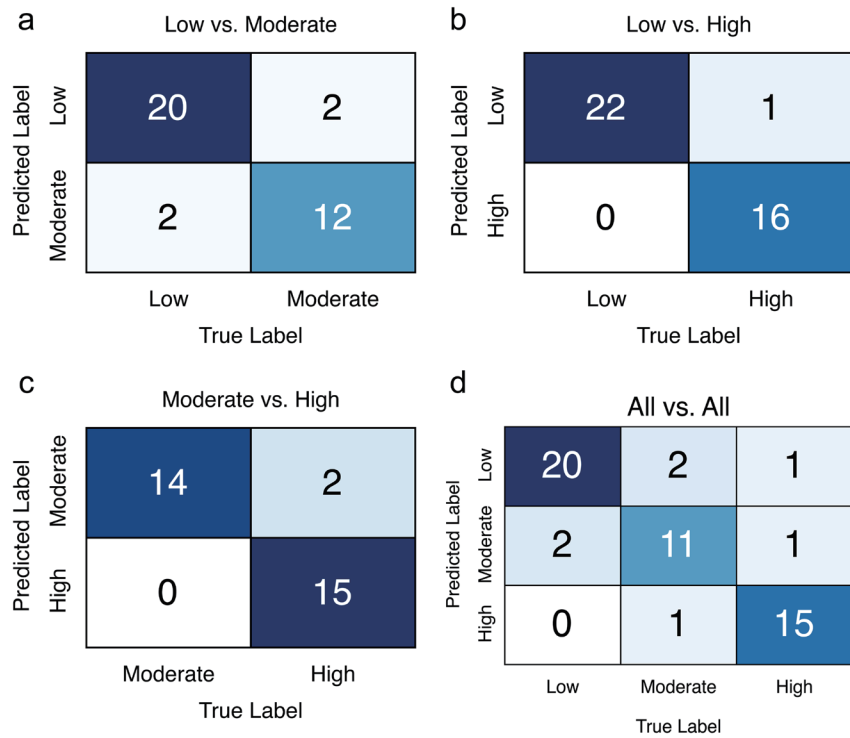


Fig. 6. Confusion matrices' best prediction results per study group comparison. (a) Low vs. Moderate; (b) Low vs. High; (c) Moderate vs. High; (d) All vs. All.

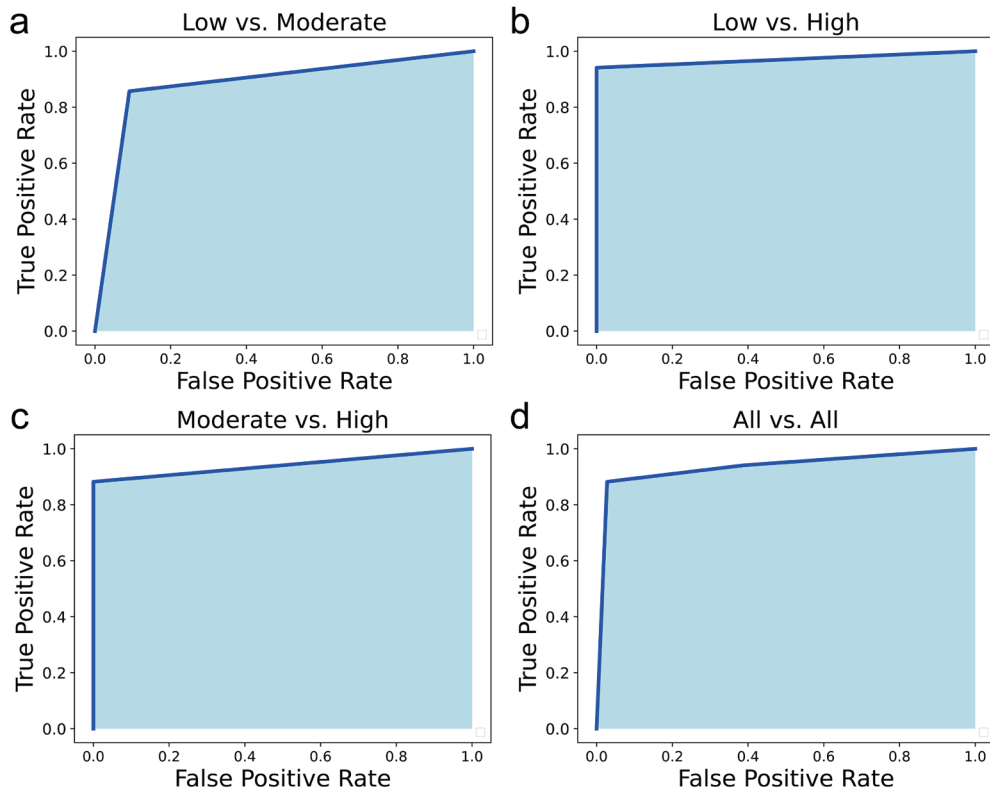


Fig. 7. Best prediction results for the receiver-operating characteristic curves per study group comparison. (a) Low vs. Moderate; (b) Low vs. High; (c) Moderate vs. High; (d) All vs. All.

medical variables, yielding no change versus ECG-only models (0% for both AUC and *Acc*), although AUC still improved substantially over medical features-only baselines (+11.2% and +41.1%, respectively), indicating either that ECG features alone sufficiently captured strong cardiovascular risk evidence or that the selection algorithm favored parsimony over potential complementarity. This behavior highlights an important methodological consideration: aggressive feature selection can optimize performance while inadvertently suppressing clinically relevant complementarity. Alternative selection strategies or model families designed to leverage multimodal complementarity may further improve performance in settings where ECG already dominates. Error patterns align with clinical plausibility. Misclassifications most frequently occurred between Low and Moderate, reflecting the expected continuum in risk phenotypes. High *Spec* and *Prec* in the High class suggest a favorable profile for ruling in higher risk, while strong *Rec* and NPV in several settings support safe rule-out. Nonetheless, even small numbers of misclassifications in the High class warrant attention; model calibration and threshold selection should consider clinical consequences and decision costs.

Table 6 indicates that LinDis, LinSVC, and SGD classifiers achieved the best performance. All three are linear classifiers: LinDis (linear discriminant, LDA-style) fits class-conditional densities and applies Bayes' rule to produce a linear decision boundary; LinSVC is a support vector machine with a linear kernel; and SGD-Classifier with loss = 'hinge' optimizes the linear SVM (hinge) objective using stochastic gradient descent, yielding a linear large-margin classifier trained via SGD. The superiority of these linear models likely reflects a favorable bias-variance trade-off. On small or noisy datasets, non-linear kernels (e.g., polynomial or radial basis function) can capture spurious patterns unless heavily regularized and carefully tuned, which degrades generalization. Linear models impose a simpler hypothesis class that emphasizes dominant trends over noise and are straightforward to regularize (e.g., via C/α or early stopping), resulting in more stable out-of-sample performance.⁸³ The similar results from LinSVC and SGD further suggest that the margin-based linear decision boundary, rather than solver-specific nuances, drives the gains.⁸³

It should be noted that, although we report in Tables 4–6 the best-performing model for each classification method and comparison based on AUC, we have included Figure S1 to provide a comprehensive overview of the AUC distributions across all combinations of feature sets and comparisons per classifier. These boxplots clearly illustrate that (1) performance differences between classifiers are relatively minor, (2) the results are consistent across models, indicating no evidence of cherry-picking, and (3) the choice of classifier does not significantly affect overall performance, thereby reinforcing the robustness and reliability of our findings.

In comparison with the state-of-the-art (Table 7),^{21–31} prior works predominantly reported binary tasks, with 66.67% using unbalanced datasets and 41.67% using cross-validation. Among 11 studies, five used an equivalent Low vs. High risk formulation.^{22,23,25,27,28} Relative to these, our best overall results (Table 6) showed *Acc* gains of 4.43% to 32.03%, though direct comparison must be interpreted cautiously due to differences in databases and validation protocols. Nevertheless, external comparison with published studies is encouraging but must be interpreted with caution. Reported *Acc* improvements over prior work in Low vs. High tasks suggest that our approach, particularly with ECG_D features and, where beneficial, multimodal integration, is competitive with the state-of-the-art. However, differences in datasets, class balance, and validation protocols can substantially affect headline metrics. A robust external

validation, harmonized evaluation protocols, and transparent reporting will be essential to confirm generalizability.

Overall, the evidence supports ECG as the primary source of discriminative information for cardiovascular risk detection, with ECG_D emerging as the most informative position and ECG_UP contributing to specific pairwise comparisons. Medical parameters (especially HRV) add value when discriminating adjacent risk levels, and their utility can be amplified by careful feature selection and model design choices that preserve multimodal complementarity.

Study limitations

While the findings are promising, several limitations warrant caution. First, the dataset is relatively small, meaning that even a few misclassifications can significantly affect the reported metrics. Without confidence intervals, the true uncertainty may be greater than the point estimates suggest. Second, generalizability may be constrained by the specific acquisition protocol, electrode positions (ECG_D and ECG_UP), device characteristics, and preprocessing choices. These factors, along with the operational definitions of risk groups, may not transfer to other populations, settings, or hardware. Moreover, medical parameters (particularly HRV) are sensitive to transient influences such as autonomic state, medications, and recording conditions, introducing potential confounding if not fully controlled.

Methodological choices may also introduce optimism. The observation that the feature selection step frequently discarded medical variables for feeding the multimodal model suggests that complementary multimodal information may have been underutilized; alternative integration strategies might yield different results. Finally, we did not evaluate model calibration or clinical utility (e.g., decision-curve analysis), and no external or prospective validation was conducted.

Future directions

Future efforts should prioritize updating the database to ensure the generalization of the results. This will also enable the use of a hold-out validation process for classification, which can provide a more straightforward evaluation of model performance on unseen data compared to cross-validation.

Additionally, it should prioritize the use of large and more balanced multi-center cohorts, harmonized acquisition protocols, rigorously nested model selection with calibration assessment, alternative feature selection and integration strategies, and external validation to substantiate robustness and clinical applicability.

Conclusions

In a cohort of 53 patients, we extracted 27 non-linear ECG features from two positions and 42 physician-curated clinical features and, using nested LOOCV with analysis of variance F-value selection, trained models to discriminate Framingham risk strata. Across Low vs. Moderate, Low vs. High, Moderate vs. High, and All vs. All tasks, the multimodal model consistently outperformed ECG-only and medical features-only models, achieving 86–97% *Acc* with AUCs up to 0.97. ECG-derived non-linear features, especially from the ECG_D position, were the principal drivers of discrimination, while medical features provided complementary gains, indicating the proposed multimodal approach is a promising tool to support clinical triage.

Table 7. State-of-the-art literature report on cardiovascular risk detection systems based on the Framingham risk scale, with information about the database, the comparison groups, the features extracted, used classifiers, limitations, and Accuracy.

Ref	Year	Source	Features extracted	Comparison group (number of participants)	Classifier	Limitations	Validation	Accuracy	Area under the curve
Unnikrishnan <i>et al.</i> ²¹	2016	Medical Features	Age, body mass index, current smoker, gender, total cholesterol, systolic blood pressure, high density lipoprotein cholesterol, diabetes, medication for hypertension, retinopathy, diastolic blood pressure	No-Cardiovascular Diseases (382) vs. Cardiovascular Diseases (128)	Support Vector Machine	Limited assessment of risk score, just Cardiovascular Diseases vs. No-Cardiovascular Diseases. Unbalanced dataset. Small dataset	Hold-on	82.35%	0.71
Dogan <i>et al.</i> ²²	2018	Medical Features	Age, gender, total cholesterol, High density lipoprotein cholesterol, systolic blood pressure, diastolic blood pressure, hemoglobin A1c, and smoking status	Low-Risk (504) vs. High-Risk (20)	Ensemble of Random Forest	Limited assessment of risk score, just Low vs. High. Unbalanced dataset	Hold-on	93.01%	–
Quesada <i>et al.</i> ²³	2019	Medical Features	Age, sex, total cholesterol, systolic blood pressure, tobacco use, diastolic blood pressure, high density lipoprotein cholesterol, and the presence of diabetes	Low-Risk (5,837) vs. High-Risk (5,837)	Random Forest	Limited assessment of risk score, just Low vs. High	Hold-on	80.9%	0.6333
Alaa <i>et al.</i> ²⁴	2019	Medical Features	Gender, age, systolic blood pressure, treatment for hypertension, smoking status, history of diabetes, and body mass index	5-year risk of Cardiovascular Diseases (4,801) vs. No-5-year risk of Cardiovascular Diseases (4,801)	AutoPrognosis	Limited assessment of risk score, just Cardiovascular Diseases vs. No Cardiovascular Diseases	Cross-validation	–	0.774
Chen <i>et al.</i> ²⁵	2020	Medical Features	Age, sex, systolic blood pressure, diastolic blood pressure, total cholesterol, high density lipoprotein cholesterol, diabetes status, smoking status	Low-Risk (1,036) vs. High-Risk (983)	Support Vector Machine	Limited assessment of risk score, just Low vs. High	Cross-validation	85.11%	–
Navarini <i>et al.</i> ²⁶	2020	Medical Features	Age, sex, systolic blood pressure, total cholesterol, smoking status, and hypertension treatment	Cardiovascular Diseases (18) vs. No-Cardiovascular Diseases (115)	Random Forest	Limited assessment of risk score, just Cardiovascular Diseases vs. No Cardiovascular Diseases. Unbalanced dataset. Small dataset	Cross-validation	65.41%	0.7297

(continued)

Table 7. (continued)

Ref	Year	Source	Features extracted	Comparison group (number of participants)	Classifier	Limitations	Validation	Accuracy	Area under the curve
Jamthikar <i>et al.</i> ²⁷	2020	Medical Features	Age, Sex, glycated hemoglobin, low density lipoprotein cholesterol, high density lipoprotein cholesterol, total cholesterol, triglyceride, systolic blood pressure, diastolic blood, pressure, hypertension, estimated glomerular filtration rate, and family history	Low-Risk (22) vs. High-Risk (180)	Support Vector Machine	Limited assessment of risk score, just Low vs. High. Unbalanced dataset. Small dataset	Cross-validation	65.41%	0.67
Sajeev <i>et al.</i> ²⁸	2021	Medical Features	Age, sex, total cholesterol, High density lipoprotein cholesterol, systolic blood pressure, hypertension medication, diabetes, and smoking status	Low-Risk (23,152) vs. High-Risk (23,152)	Logistic Regression	Limited assessment of risk score, just Low vs. High	Cross-validation	–	0.852
Yang <i>et al.</i> ²⁹	2021	Medical Features	49 Medical records' features	Stroke (2,648) vs. No-Stroke (3,337)	XGBoost	Limited assessment of risk score, just Stroke vs. No-Stroke	Hold-on	84.78%	0.9220
Cho <i>et al.</i> ³⁰	2021	Medical Features	Age, sex, systolic blood pressure, total cholesterol, high density lipoprotein cholesterol, smoking status, history of diabetes, and antihypertensive medication use	5-year risk of Cardiovascular Diseases (1,862) vs. No-5-year risk of Cardiovascular Diseases (50,327)	Neural Network	Limited assessment of risk score, just Cardio Vascular Diseases vs. No Cardio Vascular Diseases. Unbalanced dataset	Hold-on	–	0.751
Chun <i>et al.</i> ³¹	2021	Medical Features	Age, smoking status, coronary heart disease, diabetes, blood pressure-lowering treatment, systolic blood pressure-untreated, and systolic blood pressure-treated	Stroke (532) vs. No-Stroke (6,185)	Gradient Boosted Trees and proportional hazards regression	Limited assessment of risk score, just Stroke vs. No Stroke. Unbalanced dataset	Hold-on	80%	0.836
Present study	2025	Electrocardiogram + Medical features	102 features	Low (22) vs. Moderate (14)	LinDis	Small dataset.	Cross-validation	88.89%	0.8831
				Low (22) vs. High (17)	SGD			97.44%	0.9706
				Moderate (14) vs. High (17)	LinDis			93.55%	0.9412
				Low (22) vs. Moderate (14) vs. High (17)	LinSVC			86.79%	0.9346

Acknowledgments

This work was supported by National Funds from FCT - Fundação para a Ciência e a Tecnologia through project UIDB/50016/2020.

Funding

No funding was received.

Conflict of interest

The authors have no conflicts of interest related to this publication.

Author contributions

Conceptualization, drafting of manuscript, investigation, manuscript editing (PR, JALM, PMR), validation (CL, MB, ON, JALM, PMR), critical revision of the manuscript for important intellectual content (PR, CL, MB, ON, JALM, PMR), study supervision, and funding acquisition (JALM, PMR). All authors have approved the final version and publication of the manuscript.

Ethical statement

The research was approved by the Institutional Research Ethics Committees (CAAE: 74256823.4.0000.5054 and 74256823.4.30.01.5045). The ethical principles recommended by the Declaration of Helsinki (as revised in 2024) and Resolution 466/12 of the Brazilian National Health Council were followed. All patients provided consent beforehand.

Data sharing statement

The dataset used to support the findings of this study has been deposited in Mendeley Data (doi:10.17632/z8mrvy259n.1). The dataset is currently under embargo and will be publicly available on February 11, 2026.

References

- [1] Roth GA, Mensah GA, Fuster V. The Global Burden of Cardiovascular Diseases and Risks: A Compass for Global Action. *J Am Coll Cardiol* 2020;76(25):2980–2981. doi:10.1016/j.jacc.2020.11.021, PMID:33309174.
- [2] Tsoo CW, Aday AW, Almarzooq ZI, Anderson CAM, Arora P, Avery CL, *et al*. Heart Disease and Stroke Statistics-2023 Update: A Report From the American Heart Association. *Circulation* 2023;147(8):e93–e621. doi:10.1161/CIR.0000000000001123, PMID:36695182.
- [3] Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *Lancet* 2014;383(9921):999–1008. doi:10.1016/S0140-6736(13)61752-3, PMID:24084292.
- [4] Teo KK, Rafiq T. Cardiovascular Risk Factors and Prevention: A Perspective From Developing Countries. *Can J Cardiol* 2021;37(5):733–743. doi:10.1016/j.cjca.2021.02.009, PMID:33610690.
- [5] Mendis S. The contribution of the Framingham Heart Study to the prevention of cardiovascular disease: a global perspective. *Prog Cardiovasc Dis* 2010;53(1):10–14. doi:10.1016/j.pcad.2010.01.001, PMID:20620420.
- [6] Andersson C, Naylor M, Tsoo CW, Levy D, Vasan RS. Framingham Heart Study: JACC Focus Seminar, 1/8. *J Am Coll Cardiol* 2021;77(21):2680–2692. doi:10.1016/j.jacc.2021.01.059, PMID:34045026.
- [7] Conway B, Rene A. Obesity as a disease: no lightweight matter. *Obes Rev* 2004;5(3):145–151. doi:10.1111/j.1467-789X.2004.00144.x, PMID:15245383.
- [8] Koliaki C, Dalamaga M, Liatis S. Update on the Obesity Epidemic: After the Sudden Rise, Is the Upward Trajectory Beginning to Flatten? *Curr Obes Rep* 2023;12(4):514–527. doi:10.1007/s13679-023-00527-y, PMID:37779155.
- [9] Welsh A, Hammad M, Piña IL, Kulinski J. Obesity and cardiovascular health. *Eur J Prev Cardiol* 2024;31(8):1026–1035. doi:10.1093/eur-jpc/zwae025, PMID:38243826.
- [10] Yadav RL, Yadav PK, Yadav LK, Agrawal K, Sah SK, Islam MN. Association between obesity and heart rate variability indices: an intuition toward cardiac autonomic alteration - a risk of CVD. *Diabetes Metab Syndr Obes* 2017;10:57–64. doi:10.2147/DMSO.S123935, PMID:28255249.
- [11] Novak P. Autonomic Disorders. *Am J Med* 2019;132(4):420–436. doi:10.1016/j.amjmed.2018.09.027, PMID:30308186.
- [12] Guarino D, Nannipieri M, Iervasi G, Taddei S, Bruno RM. The Role of the Autonomic Nervous System in the Pathophysiology of Obesity. *Front Physiol* 2017;8:665. doi:10.3389/fphys.2017.00665, PMID:28966594.
- [13] Zapparoli L, Devoto F, Giannini G, Zonca S, Gallo F, Paulesu E. Neural structural abnormalities behind altered brain activation in obesity: Evidence from meta-analyses of brain activation and morphometric data. *Neuroimage Clin* 2022;36:103179. doi:10.1016/j.nicl.2022.103179, PMID:36088842.
- [14] Forte G, Favieri F, Casagrande M. Heart Rate Variability and Cognitive Function: A Systematic Review. *Front Neurosci* 2019;13:710. doi:10.3389/fnins.2019.00710, PMID:31354419.
- [15] Williams SM, Eleftheriadou A, Alam U, Cuthbertson DJ, Wilding JPH. Cardiac Autonomic Neuropathy in Obesity, the Metabolic Syndrome and Prediabetes: A Narrative Review. *Diabetes Ther* 2019;10(6):1995–2021. doi:10.1007/s13300-019-00693-0, PMID:31552598.
- [16] Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham Study. *Am J Cardiol* 1976;38(1):46–51. doi:10.1016/0002-9149(76)90061-8, PMID:132862.
- [17] Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97(18):1837–1847. doi:10.1161/01.cir.97.18.1837, PMID:9603539.
- [18] Fan X, Li Y, He Q, Wang M, Lan X, Zhang K, *et al*. Predictive Value of Machine Learning for Recurrence of Atrial Fibrillation after Catheter Ablation: A Systematic Review and Meta-Analysis. *Rev Cardiovasc Med* 2023;24(11):315. doi:10.31083/j.rcm2411315, PMID:39076446.
- [19] Ribeiro P, Marques JAL, Pordeus D, Zacarias L, Leite CF, Sobreira-Neto MA, *et al*. Machine learning-based cardiac activity non-linear analysis for discriminating COVID-19 patients with different degrees of severity. *Biomed Signal Process Control* 2024;87(Pt A):105558. doi:10.1016/j.bspc.2023.105558.
- [20] Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, *et al*. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med* 2020;46(3):383–400. doi:10.1007/s00134-019-05872-y, PMID:31965266.
- [21] Unnikrishnan P, Kumar DK, Poosapadi Arjunan S, Kumar H, Mitchell P, Kawasaki R. Development of Health Parameter Model for Risk Prediction of CVD Using SVM. *Comput Math Methods Med* 2016;2016:3016245. doi:10.1155/2016/3016245, PMID:27594895.
- [22] Dogan MV, Beach SRH, Simons RL, Lendasse A, Penaluna B, Philibert RA. Blood-Based Biomarkers for Predicting the Risk for Five-Year Incident Coronary Heart Disease in the Framingham Heart Study via Machine Learning. *Genes (Basel)* 2018;9(12):641. doi:10.3390/genes9120641, PMID:30567402.
- [23] Quesada JA, Lopez-Pineda A, Gil-Guillén VF, Durazo-Arvizu R, Orozco-Beltrán D, López-Domenech A, *et al*. Machine learning to predict cardiovascular risk. *Int J Clin Pract* 2019;73(10):e13389. doi:10.1111/ijcp.13389, PMID:31264310.
- [24] Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF, van der Schaar M. Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *PLoS One* 2019;14(5):e0213653. doi:10.1371/journal.pone.0213653, PMID:31091238.
- [25] Chen YS, Cheng CH, Chen SF, Jhuang JY. Identification of the Framing-

- ham Risk Score by an Entropy-Based Rule Model for Cardiovascular Disease. *Entropy (Basel)* 2020;22(12):1406. doi:10.3390/e22121406, PMID:33322122.
- [26] Navarini L, Caso F, Costa L, Currado D, Stola L, Perrotta F, *et al*. Cardiovascular Risk Prediction in Ankylosing Spondylitis: From Traditional Scores to Machine Learning Assessment. *Rheumatol Ther* 2020;7(4):867–882. doi:10.1007/s40744-020-00233-4, PMID:32939675.
- [27] Jamthikar A, Gupta D, Saba L, Khanna NN, Araki T, Viskovic K, *et al*. Cardiovascular/stroke risk predictive calculators: a comparison between statistical and machine learning models. *Cardiovasc Diagn Ther* 2020;10(4):919–938. doi:10.21037/cdt.2020.01.07, PMID:32968651.
- [28] Sajeev S, Champion S, Beleigoli A, Chew D, Reed RL, Magliano DJ, *et al*. Predicting Australian Adults at High Risk of Cardiovascular Disease Mortality Using Standard Risk Factors and Machine Learning. *Int J Environ Res Public Health* 2021;18(6):3187. doi:10.3390/ijerph18063187, PMID:33808743.
- [29] Yang Y, Zheng J, Du Z, Li Y, Cai Y. Accurate Prediction of Stroke for Hypertensive Patients Based on Medical Big Data and Machine Learning Algorithms: Retrospective Study. *JMIR Med Inform* 2021;9(11):e30277. doi:10.2196/30277, PMID:34757322.
- [30] Cho SY, Kim SH, Kang SH, Lee KJ, Choi D, Kang S, *et al*. Pre-existing and machine learning-based models for cardiovascular risk prediction. *Sci Rep* 2021;11(1):8886. doi:10.1038/s41598-021-88257-w, PMID:33903629.
- [31] Chun M, Clarke R, Cairns BJ, Clifton D, Bennett D, Chen Y, *et al*. Stroke risk prediction using machine learning: a prospective cohort study of 0.5 million Chinese adults. *J Am Med Assoc* 2021;28(8):1719–1727. doi:10.1093/jamia/ocab068, PMID:33969418.
- [32] Souza Neto EG, Peixoto JVC, Rank Filho C, Petterle RR, Fogaça RTH, Wolska BM, *et al*. Effects of High-Intensity Interval Training and Continuous Training on Exercise Capacity, Heart Rate Variability and Isolated Hearts in Diabetic Rats. *Arq Bras Cardiol* 2023;120(1):e20220396. doi:10.36660/abc.20220396, PMID:36629606.
- [33] Ernst G. Hidden Signals-The History and Methods of Heart Rate Variability. *Front Public Health* 2017;5:265. doi:10.3389/fpubh.2017.00265, PMID:29085816.
- [34] Vanderlei LC, Pastre CM, Hoshi RA, Carvalho TD, Godoy MF. Basic notions of heart rate variability and its clinical applicability. *Rev Bras Cir Cardiovasc* 2009;24(2):205–217. doi:10.1590/s0102-76382009000200018, PMID:19768301.
- [35] Dídimo N dos S, Cruz LAS, Do Carmo GS, Da Rocha RB, Cardoso VS. Parasympathetic autonomic modulation of heart rate and glycemic profile in type 2 Diabetese patients (in Portuguese). *Braz J Dev* 2022;8(10):67735–67749. doi:10.34117/bjdv8n10-189.
- [36] Paschoal MA, Volanti VM, Pires CS, Fernandes FC. Heart rate variability in different age groups (in Portuguese). *Braz J Phys Ther* 2006;10(4):413–419. doi:10.1590/S1413-35552006000400009.
- [37] Nunan D, Donovan G, Jakovljevic DG, Hodges LD, Sandercock GR, Brodie DA. Validity and reliability of short-term heart-rate variability from the Polar S810. *Med Sci Sports Exerc* 2009;41(1):243–250. doi:10.1249/MSS.0b013e318184a4b1, PMID:19092682.
- [38] Simão AF, Precoma DB, Andrade JP, Correa Filho H, Saraiva JFK, Oliveira GMM, Murro ALB, *et al*. First Brazilian Guideline on Cardiovascular Prevention (in Portuguese). *Arq Bras Cardiol* 2013;101(6 Suppl 2):1–63.
- [39] Mensogera AL, Sundarakumar JS, Diwakar L, Issac TG, SANSCO Study Team. Relationship between Framingham Cardiovascular Risk Score and cognitive performance among ageing rural Indian participants: a cross-sectional analysis. *BMJ Open* 2023;13(11):e074977. doi:10.1136/bmjopen-2023-074977, PMID:37949620.
- [40] Chen X, Tu Q, Wang D, Liu J, Qin Y, Zhang Y, *et al*. Effectiveness of China-PAR and Framingham risk score in assessment of 10-year cardiovascular disease risk in Chinese hypertensive patients. *Public Health* 2023;220:127–134. doi:10.1016/j.puhe.2023.05.007, PMID:37315498.
- [41] Thabtah F, Hammoud S, Kamalov F, Gonsalves A. Data imbalance in classification: Experimental evaluation. *Inf Sci* 2020;513:429–441. doi:10.1016/j.ins.2019.11.004.
- [42] Lux RL, Sower CT, Allen N, Etheridge SP, Tristani-Firouzi M, Saarel EV. The application of root mean square electrocardiography (RMS ECG) for the detection of acquired and congenital long QT syndrome. *PLoS One* 2014;9(1):e85689. doi:10.1371/journal.pone.0085689, PMID:24454918.
- [43] Kalidas V, Tamil L. Real-time QRS detector using Stationary Wavelet Transform for Automated ECG Analysis. 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE); 2017 Oct 23–25. Washington, DC, USA, Piscataway, NJ, USA: IEEE; 2017:457–461. doi:10.1109/BIBE.2017.00-12.
- [44] Malvar H. *Signal Processing with Lapped Transforms*. Norwood, Massachusetts: Artech House; 1992.
- [45] Vetterli M, Kovachovska J. *Wavelets and Subband Coding*. Englewood Cliffs, New Jersey: Prentice Hall; 1995.
- [46] Caesarendra W, Kosasih B, Tieu K, Moodie CAS. An application of nonlinear feature extraction - A case study for low speed slewing bearing condition monitoring and prognosis. 2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics; 2013 Jul 09–12. Wollongong, NSW, Australia, Piscataway, NJ, USA: IEEE; 2013:1713–1718. doi:10.1109/AIM.2013.6584344.
- [47] Hardstone R, Poil SS, Schiavone G, Jansen R, Nikulin VV, Mansvelder HD, *et al*. Detrended fluctuation analysis: a scale-free view on neuronal oscillations. *Front Physiol* 2012;3:450. doi:10.3389/fphys.2012.00450, PMID:23226132.
- [48] Sundararajan D. *Discrete wavelet transform a signal processing approach*. 1st ed. New Jersey: John Wiley & Sons; 2015. doi:10.1002/9781119113119.
- [49] Silva M, Ribeiro P, Bispo BC, Rodrigues PM. Detection of Alzheimer's Disease through Nonlinear Parameters of Speech Signals (in Portuguese). Proceedings of the 41st Brazilian Symposium on Telecommunications and Signal Processing; 2023 Oct 8–11. São José dos Campos, SP, Brazil, São Paulo: Brazilian Society of Telecommunications; 2023:1–5. doi:10.14209/sbrt.2023.1570923829.
- [50] Garcia A, Garcia C, Villasenor-Pineda L, Montoya O. *Biosignal Processing and Classification Using Computational Learning and Intelligence Principles, Algorithms, and Applications*. London: Academic Press; 2022.
- [51] Silva G, Batista P, Rodrigues PM. COVID-19 activity screening by a smart-data-driven multi-band voice analysis. *J Voice* 2025;39(3):602–611. doi:10.1016/j.jvoice.2022.11.008, PMID:36464573.
- [52] Richman JS, Lake DE, Moorman JR. Sample entropy. *Methods Enzymol* 2004;384:172–184. doi:10.1016/S0076-6879(04)84011-4, PMID:15081687.
- [53] Azami H, Li P, Arnold SE, Escudero J, Humeau-Heurtier A. Fuzzy entropy metrics for the analysis of biomedical signals: Assessment and comparison. *IEEE Access* 2019;7:104833–104847. doi:10.1109/ACCESS.2019.2930625.
- [54] Gao L, Wang J, Chen L. Event-related desynchronization and synchronization quantification in motor-related EEG by Kolmogorov entropy. *J Neural Eng* 2013;10(3):036023. doi:10.1088/1741-2560/10/3/036023, PMID:23676901.
- [55] Bian C, Qin C, Ma QD, Shen Q. Modified permutation-entropy analysis of heartbeat dynamics. *Phys Rev E Stat Nonlin Soft Matter Phys* 2012;85(2 Pt 1):021906. doi:10.1103/PhysRevE.85.021906, PMID:22463243.
- [56] Porta A, Baselli G, Liberati D, Montano N, Cogliati C, Gnecchi-Ruscone T, *et al*. Measuring regularity by means of a corrected conditional entropy in sympathetic outflow. *Biol Cybern* 1998;78(1):71–78. doi:10.1007/s004220050414, PMID:9485587.
- [57] Li P, Liu C, Li K, Zheng D, Liu C, Hou Y. Assessing the complexity of short-term heartbeat interval series by distribution entropy. *Med Biol Eng Comput* 2015;53(1):77–87. doi:10.1007/s11517-014-1216-0, PMID:25351477.
- [58] Rostaghi M, Azami H. Dispersion entropy: A measure for time-series analysis. *IEEE Signal Process Lett* 2016;23(5):610–614. doi:10.1109/LSP.2016.2542881.
- [59] Inouye T, Shinosaki K, Sakamoto H, Toi S, Ukai S, Iyama A, *et al*. Quantification of EEG irregularity by use of the entropy of the power spectrum. *Electroencephalogr Clin Neurophysiol* 1991;79(3):204–210. doi:10.1016/0013-4694(91)90138-t, PMID:1714811.
- [60] Wang J, Li T, Xie R, Wang XM, Cao YY. Fault feature extraction for multiple electrical faults of aviation electro-mechanical actuator based on symbolic dynamics entropy. 2015 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC);

- 2015 Sep 19-22. Ningbo, China, Piscataway, NJ, USA: IEEE; 2015:1–6. doi:10.1109/ICSPCC.2015.7338844.
- [61] Liu X, Wang X, Zhou X, Jiang A. Appropriate use of the increment entropy for electrophysiological time series. *Comput Biol Med* 2018;95:13–23. doi:10.1016/j.combiomed.2018.01.009, PMID:29433037.
- [62] Chanwimalueang T, Mandic DP. Cosine similarity entropy: Self-correlation-based complexity analysis of dynamical systems. *Entropy* 2017;19(12):652. doi:10.3390/e19120652.
- [63] Rohila A, Sharma A. Phase entropy: a new complexity measure for heart rate variability. *Physiol Meas* 2019;40(10):105006. doi:10.1088/1361-6579/ab499e, PMID:31574498.
- [64] Cuesta-Frau D, Schneider J, Bakštein E, Vostatek P, Spaniel F, Novák D. Classification of Actigraphy Records from Bipolar Disorder Patients Using Slope Entropy: A Feasibility Study. *Entropy (Basel)* 2020;22(11):1243. doi:10.3390/e22111243, PMID:33287011.
- [65] Manis G, Aktaruzzaman M, Sassi R. Bubble Entropy: An Entropy Almost Free of Parameters. *IEEE Trans Biomed Eng* 2017;64(11):2711–2718. doi:10.1109/TBME.2017.2664105, PMID:28182552.
- [66] Yan C, Li P, Liu C, Wang X, Yin C, Yao L. Novel gridded descriptors of poincaré plot for analyzing heartbeat interval time-series. *Comput Biol Med* 2019;109:280–289. doi:10.1016/j.combiomed.2019.04.015, PMID:31100581.
- [67] Hsu CF, Wei SY, Huang HP, Hsu L, Chi S, Peng CK. Entropy of entropy: Measurement of dynamical complexity for biological systems. *Entropy* 2017;19(10):550. doi:10.3390/e19100550.
- [68] Zhang J, Zhao Y, Li H, Zong C. Attention With Sparsity Regularization for Neural Machine Translation and Summarization. *IEEE/ACM Trans. Audio Speech Lang Process* 2019;27(3):507–518. doi:10.1109/TASLP.2018.2883740.
- [69] Morin K, Davis JL. Cross-validation: What is it and how is it used in regression? *Commun Stat Theory Methods* 2016;46(11):5238–5251. doi:10.1080/03610926.2015.1099672.
- [70] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, *et al*. Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011;12:2825–2830.
- [71] Allgaier J, Pryss R. Cross-Validation Visualized: A Narrative Guide to Advanced Methods. *Mach Learn Knowl Extr* 2024;6(2):1378–1388. doi:10.3390/make6020065.
- [72] Sushentsev N, Rundo L, Abrego L, Li Z, Nazarenko T, Warren AY, *et al*. Time series radiomics for the prediction of prostate cancer progression in patients on active surveillance. *Eur Radiol* 2023;33(6):3792–3800. doi:10.1007/s00330-023-09438-x, PMID:36749370.
- [73] Geroldinger A, Lusa L, Nold M, Heinze G. Leave-one-out cross-validation, penalization, and differential bias of some prediction model performance measures—a simulation study. *Diagn Progn Res* 2023;7(1):9. doi:10.1186/s41512-023-00146-0, PMID:37127679.
- [74] Bradshaw TJ, Huemann Z, Hu J, Rahmim A. A Guide to Cross-Validation for Artificial Intelligence in Medical Imaging. *Radiol Artif Intell* 2023;5(4):e220232. doi:10.1148/ryai.220232, PMID:37529208.
- [75] Yates LA, Aandahl Z, Richards SA, Brook BW. Cross validation for model selection: A review with examples from ecology. *Ecol Monogr* 2023;93(1):e1557. doi:10.1002/ecm.1557.
- [76] Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 2006;7:91. doi:10.1186/1471-2105-7-91, PMID:16504092.
- [77] Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J Anesthesiol* 2022;75(1):25–36. doi:10.4097/kja.21209, PMID:35124947.
- [78] Dođan O. Data Linkage Methods for Big Data Management in Industry 4.0. In: Öner S, Yüregir O (eds). *Optimizing Big Data Management and Industrial Systems With Intelligent Techniques*. Optimizing Big Data Management and Industrial Systems With Intelligent Techniques. Hershey, PA: IGI Global; 2019:108–127. doi:10.4018/978-1-5225-5137-9.ch005.
- [79] Ting KM, Sammut C, Webb GI. Precision and Recall. *Encyclopedia of Machine Learning and Data Mining*. Boston, MA: Springer; 2017:990–991. doi:10.1007/978-1-4899-7687-1_659.
- [80] Safari S, Baratloo A, Elfil M, Negida A. Evidence Based Emergency Medicine Part 2: Positive and negative predictive values of diagnostic tests. *Emerg (Tehran)* 2015;3(3):87–88. PMID:26495390.
- [81] Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol* 2019;19(1):64. doi:10.1186/s12874-019-0681-4, PMID:30890124.
- [82] Goutte C, Gaussier E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In: Losada DE, Fernández-Luna JM (eds). *Advances in Information Retrieval*. 27th European Conference on IR Research, ECIR 2005; 2025 March 21-23. Santiago de Compostela, Spain, Berlin, Heidelberg: Springer; 2005:345–359. doi:10.1007/978-3-540-31865-1_25.
- [83] Mehta P, Wang CH, Day AGR, Richardson C, Bukov M, Fisher CK, *et al*. A high-bias, low-variance introduction to Machine Learning for physicists. *Phys Rep* 2019;810:1–124. doi:10.1016/j.physrep.2019.03.001, PMID:31404441.