ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

# Stratifying autonomic nervous system regulation patterns in healthy men: A machine learning approach

**Wollner Materko**[1,2*]

[1]Department of Health, Faculty of of Health Sciences, Federal University of Amapá, Macapá, Amapá, Brazil

[2]Department of Education, Faculty of Physical Education, Federal University of Amapá, Macapá, Amapá, Brazil

## Abstract

Heart rate variability (HRV) is a critical non-invasive marker of autonomic nervous system regulation and plays an essential role in cardiovascular health. Individual differences in autonomic function necessitate the development of personalized health strategies. This study aimed to develop and validate a method that integrates principal component analysis (PCA) and K-means clustering to identify distinct patterns of autonomic regulation in healthy men using HRV data. A total of 80 young, healthy men (22.0 ± 2.8 years old, 65.2 ± 6.9 kg, and 171.0 ± 6.5 cm) were recruited, and their HRV data were analyzed using time-domain and frequency-domain parameters. PCA was applied to reduce the dimensionality of the HRV data, while K-means clustering was employed to identify distinct autonomic profiles. Silhouette index values were 0.397 for one cluster, 0.481 for two clusters, and 0.556 for three clusters, indicating that the three-cluster solution provided the best fit. Three statistically distinct and physiologically meaningful clusters were identified. Cluster 3 ($n$ = 19) demonstrated significantly higher HRV parameters than cluster 1 ($n$ = 33) and cluster 2 ($n$ = 28) ($p$ = 0.001). *Post hoc* analysis further confirms that cluster 1 differed significantly from both cluster 2 and cluster 3 ($p$ = 0.001). Based on HRV characteristics, the clusters were characterized as "high vagal tone," "intermediate vagal tone," and "low vagal tone." The "high vagal tone" cluster exhibited the strongest parasympathetic activity, while the "low vagal tone" cluster showed evidence of sympathetic predominance. This study demonstrates a robust approach for stratifying autonomic profiles, highlighting the potential of machine learning in advancing personalized cardiovascular health assessment.

***Keywords:*** Heart rate variability; Autonomic nervous system; Machine learning; Principal component analysis; K-means clustering

***Corresponding author:**
Wollner Materko
(wollner.materko@gmail.com)

**Citation:** Materko W. Stratifying autonomic nervous system regulation patterns in healthy men: A machine learning approach. *Artif Intell Health*. 2025;2(4):103-113. doi: 10.36922/AIH025050006

**Received:** January 29, 2025

**Revised:** June 24, 2025

**Accepted:** June 30, 2025

**Published online:** July 28, 2025



## 1. Introduction

Cardiovascular diseases (CVDs) represent a significant global health burden, accounting for an estimated 17.9 million deaths annually, according to the World Health Organization.[1] Heart rate monitoring serves as an effective method for detecting cardiac irregularities, such as arrhythmias. Continuous heart rate analysis facilitates early

diagnosis and prevention of CVDs, ultimately reducing morbidity and mortality.[2]

Heart rate variability (HRV) measures the beat-to-beat fluctuations in R-R intervals (RRI) as recorded by an electrocardiogram (ECG). As a non-invasive metric, HRV provides quantitative insight into the activity of the autonomic nervous system (ANS) by reflecting its modulation of cardiac function.[3] Elevated HRV is generally associated with enhanced cardiovascular adaptability and greater resilience to stress, whereas reduced HRV may reflect underlying autonomic dysfunction.[4] Due to its sensitivity, HRV has emerged as a valuable biomarker for assessing autonomic regulation and cardiac function, playing a vital role in the early detection and prevention of CVDs.[5]

However, the practical application of HRV monitoring, particularly in field settings such as athletic training, is often limited by the cost, complexity, and bulk of conventional ECG equipment.[6] Wearable technologies, such as the Polar® Heart Rate Monitor Interface Board, provide an affordable, high-quality alternative for acquiring beat-to-beat RRI data in real-world environments.[7] These devices enable athletes and coaches to access objective physiological feedback, thereby enhancing performance monitoring and supporting personalized, data-driven training adjustments.[8]

The stratification of cardiovascular autonomic function, as assessed by HRV in healthy men, provides valuable insights into ANS regulation and classification of cardiovascular health.[9] As a non-invasive indicator of autonomic control, HRV quantifies the constant interplay between the sympathetic and parasympathetic nervous system's inputs to the heart. From a clinical perspective, a high level of HRV suggests that the cardiovascular system is resilient and able to cope with stress.[10] Furthermore, understanding the stratification of ANS regulatory patterns may provide valuable insights into the individual cardiovascular health profiles and inform targeted prevention strategies.

The integration of artificial intelligence (AI) with a focus on machine learning offers significant potential to advance the analysis of HRV and deepen our understanding of the regulatory mechanisms of the ANS. This data-driven approach is crucial for shifting the focus from population-level analyses to more precise and personalized evaluations of autonomic health.[4,5,11-14] which underpins the rationale for this research. The present study aims to explore ANS regulation a critical aspect of cardiovascular health through the application of AI-based methods. By stratifying individuals into distinct groups based on their HRV patterns, this study proposes a novel framework for categorizing and analyzing variations in autonomic function. Importantly, it investigates heterogeneity within a healthy population, rather than attempting to differentiate between healthy individuals and those with disease. It is hypothesized that, even among asymptomatic young men, there are distinct patterns of autonomic regulation that are physiologically meaningful and may represent different trajectories of cardiovascular health long before clinical disease manifests.

This stratification can be achieved through an integrated analytical framework combining principal component analysis (PCA) and K-means clustering. PCA, an unsupervised dimensionality reduction technique, is used during data pre-processing to maximize variance preservation within a reduced two-dimensional space, thereby facilitating improved visualization and subsequent clustering.[15,16] The K-means algorithm is then applied to these reduced dimensions to group individuals according to their position in the new analytical space. As an iterative partitioning method, K-means require a pre-specified number of clusters ($k$) and operate by minimizing intra-cluster distance while maximizing inter-cluster distance to form distinct, coherent groups.[17,18] This study aims to develop and validate a method that integrates PCA and K-means clustering to identify distinct patterns of autonomic regulation in healthy men using HRV data.

## 2. State of the art

### 2.1. HRV

HRV refers to the variation in the time intervals between successive heartbeats.[19] As a well-established indicator of ANS activity, HRV offers a non-invasive means of assessing cardiac autonomic regulation.[20] Historically, accurately evaluating autonomic function has presented a significant challenge for cardiologists. A breakthrough occurred in 1981 when Akselrod *et al.*[21] demonstrated that specific components of HRV correspond directly to parasympathetic and sympathetic nervous system activity. Building on this, a 1987 study by Kleiger *et al.*[22] established the clinical prognostic value of HRV, showing that it serves as a predictor of mortality risk following myocardial infarction. In recent years, HRV analysis has been widely recognized as a reliable, non-invasive method for evaluating ANS modulation at the heart's sinus node.[23]

Time-domain analysis of HRV quantifies the variability in RRI using direct statistical measurements. Key parameters include the mean normal-to-normal (NN) mean R-R interval (MRR), standard deviation of NN intervals (SDNN), root mean square of successive RRI differences (RMSSD), and the proportion of adjacent NN intervals differing by more than 50 ms (pNN50).[24] The

reliability of these metrics is highly dependent on the integrity of the RRI data, as artifacts such as ectopic beats can significantly distort the results. Consequently, the application of filtering techniques is a crucial preprocessing step. A commonly used approach involves excluding any RRI that deviates by more than 20% from the preceding normal interval, thereby ensuring data accuracy.[25]

Spectral analysis of HRV quantifies the power distribution of different frequency components within the sinus rhythm. The two primary methods employed are the non-parametric Fourier transform which decomposes the signal into constituent sinusoids[26] – and the parametric autoregressive model, which estimates the spectrum using a predictive model of the RRI.[27] Despite their methodological differences, both methods provide broadly comparable assessments of HRV spectra, and neither demonstrates a clear advantage over the other.

Spectral HRV analysis typically distinguishes two main frequency bands: the high-frequency (HF) band (0.15 – 0.40 Hz), which reflects parasympathetic modulation associated with respiratory sinus arrhythmia, and the low-frequency (LF) band (0.04 – 0.15 Hz), which represents a combination of sympathetic and parasympathetic influences on baroreflex regulation.[28] The LF/HF ratio is often calculated to estimate sympathovagal balance, with the HF band serving as an index of parasympathetic tone and the LF band representing integrated autonomic output.[29]

Similar to the time-domain analysis, the accuracy of HRV spectral analysis is highly dependent on data quality and requires careful handling of arrhythmias. A common approach is to exclude RRIs immediately before and after ectopic beats and replace them with interpolated values based on adjacent, true RRIs. However, the exclusion of more than two RRIs surrounding an ectopic beat is generally avoided due to the risk of compromising signal continuity.[30]

Following the Task Force report,[24] various non-linear metrics have emerged to analyze the complex, multi-causal, and potentially chaotic nature of HRV. These metrics apply techniques such as Lyapunov and Hurst exponents, coarse-grained spectral analysis, detrended fluctuation analysis, and entropy measures to capture the interplay of humoral, hemodynamic, and electrophysiological factors influencing HRV. While their precise physiological interpretations remain under investigation, these methods have shown promising potential in differentiating the effects of conditions such as stress and diabetes on HRV.[31,32]

Novel metrics based on instantaneous heart rate acceleration and deceleration have also been introduced. In 2006, Bauer *et al*.[33] developed phase-rectified signal averaging (PRSA), a statistical technique designed to analyze quasi-periodic signals in non-stationary or noisy data. By applying PRSA to NN intervals, they proposed deceleration capacity (DC) and acceleration capacity metrics derived from the coherent averaging of RRIs that exhibit increases or decreases, respectively. These indices aim to assess sympathetic modulation of sinoatrial node acceleration and deceleration, independent of other physiological factors.

The DC index, in particular, has gained significant attention due to its promising clinical implications. Notably, studies have demonstrated its superior predictive power for mortality following acute myocardial infarction compared to the widely used left ventricular ejection fraction.[34] Furthermore, a strong relationship has been observed between DC and the risk of sudden cardiac death in individuals with Chagas disease.[35] Studies have also reported a significant correlation between DC and high levels of physical conditioning,[36,37] suggesting its potential use as a valuable marker of cardiovascular health and fitness.

### 2.2. PCA

PCA is a dimensionality reduction technique that transforms a set of correlated variables into a smaller number of uncorrelated linear combinations, known as principal components (PCs). These components are ordered to capture as much of the total variance in the original dataset as possible.[38]

The first PC captures the largest proportion of the total variance. The second PC explains the maximum remaining variance, with the constraint that it is uncorrelated with the first. This process continues sequentially, with each subsequent component capturing a decreasing proportion of the remaining variance and remaining uncorrelated with all previously derived components.[39]

Although PCA can theoretically continue until all variance is accounted for, it is typically stopped after extracting a smaller number of PCs that collectively explain a significant proportion of the total variance.[38] The eigenvalue associated with each PC represents the amount of variance it explains higher eigenvalues indicate greater explanatory power.[40]

### 2.3. Cluster analysis

Cluster analysis encompasses a range of statistical techniques used to group an initially unclassified set of cases, subjects, or objects into relatively homogeneous groups, or clusters, based on observed characteristics.[41] The primary goal is to identify underlying group structures without prior knowledge of group membership. Also

referred to as classification analysis or numerical taxonomy, cluster analysis differs fundamentally from discriminant analysis, which requires predefined groups. In contrast, cluster analysis aims to identify previously unknown groupings inherent in the data.[42]

The cluster analysis process typically involves a series of key steps: defining the research problem; selecting an appropriate distance or similarity measure; choosing a clustering algorithm; determining the optimal number of clusters; interpreting the characteristics of each identified cluster; and evaluating the validity of the resulting cluster solution.[43] Careful selection of variables is essential and should be guided by research hypotheses, prior studies, and the researcher's informed judgment. Similarly, the choice of distance or similarity measure is critical; for instance, Euclidean distance is frequently used.[44]

Clustering methods are broadly classified as hierarchical, non-hierarchical, or two-stage.[43] Hierarchical approaches build a nested structure of clusters either agglomeratively (bottom-up) or divisively (top-down) and do not require a predefined number of clusters. The results of these methods are visualized using dendrograms, in which branch lengths indicate inter-cluster distances. In contrast, non-hierarchical methods such as K-means require the number of clusters to be defined in advance. The choice of method depends on the distance measure used, and the resulting clusters must be interpretable and relevant to the research objectives.

## 3. Materials and methods

### 3.1. Study population

This cross-sectional study was conducted in Macapá, Brazil, and involved 80 healthy, young male participants (22.0 ± 2.8 years). Participants were recruited based on a low-risk profile for CVD.[45] Exclusion criteria included smoking, a history of cardiopulmonary disease, or the current use of any medication. All participants provided verbal informed consent before enrollment. The study protocol was approved by the Human Research Ethics Committee of the Federal University of Amapá (CAAE: 50150121.1.0000.0003) and conducted in accordance with the principles of the Declaration of Helsinki[46] and Resolution 510/2016 of the National Health Council.

### 3.2. Anthropometric assessment

Before enrollment, all participants received a detailed explanation of the study protocol, including measurement procedures and estimated duration. Participants were instructed to wear appropriate attire (e.g., light clothing, no shoes) and to avoid carrying objects. Height (cm) and weight (kg) were measured using a calibrated mechanical scale (Filizola, Brazil). All anthropometric measurements were performed by a trained assessor to minimize inter-rater variability and ensure data accuracy.

### 3.3. Experimental procedure and data acquisition

The tests were conducted in a quiet room maintained at a temperature of 22°C. Participants were instructed to refrain from strenuous physical activity for 24 h and to avoid consuming alcohol, caffeine, or large meals for at least 3 h before their session. Upon arrival at the laboratory, participants rested quietly in a supine position for 10 min while breathing spontaneously. RRIs were recorded throughout this period using a Polar V800 heart rate monitor (Polar, Finland) with a sampling rate of 1,000 Hz. The monitor was positioned over the xiphoid process of the sternum. The first 5 min of data were discarded to allow for signal stabilization, and the subsequent 5 min were used for analysis. The tachograms of RRI were transferred via an infrared interface to Polar Precision Performance SW software version 3.0 (Polar, Finland), which automatically corrected the RRI using a moving average filter. The data were then saved as ".txt" files.

### 3.4. HRV analysis

For the time-domain analysis, the following parameters were calculated: MRR, SDNN, RMSSD, and the pNN50. For the frequency-domain analysis, spectral analysis was performed using the Welch periodogram method (256-point segments, 128-point overlap, and a Hanning window). This yielded normalized power for the LF (0.04 – 0.15 Hz) and HF (0.15 – 0.40 Hz) bands, both expressed as percentages. All parameters were computed in accordance with the guidelines established by the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology,[24] and were implemented in MATLAB 2020.b (MathWorks, United States).

### 3.5. Statistical analysis

Descriptive statistics are presented as mean ± standard deviation. The Shapiro–Wilk test was employed to assess the normality of the data distribution.

An 80 × 6 matrix of normalized HRV data derived from the RRI tachograms was used for dimensionality reduction to two dimensions using PCA. PCA, a dimensionality reduction technique, transforms correlated variables into uncorrelated PCs via eigenvalue decomposition of the covariance matrix.[15] K-means clustering an unsupervised learning algorithm – was then applied to the reduced-dimensionality data,[16] making it well-suited to classify individuals into distinct groups based on their ANS regulation. Cluster assignment was based on the Euclidean distance metric (Equation I). Cluster centroids were

iteratively recalculated until convergence (Equation II) by minimizing the sum of squared errors (Equation III).

The following equations were applied in the K-means clustering process:

$$d(p,q) = \sqrt{\sum_{i=1}^{n}\left((q_i - p_i)\right)^2} \quad \text{(I)}$$

$$m = \frac{1}{n_k}\sum_{i\epsilon C_k} x_i \quad \text{(II)}$$

Where:

- $x_i$ is the HRV values
- $n_k$ is the number of points in $C_k$
- $m$ is the centroid of a cluster
- $d$ is the Euclidean distance
- $p$ is a data point
- $q$ is a cluster centroid
- $q_i$ is the i-th attribute of data point q
- $p_i$ is the i-th attribute of data point p.

$$J_k = \sum_{i=1}^{k}\sum_{i\epsilon C_k}(x_i - m_k)^2 \quad \text{(III)}$$

Where:

- $J_k$ is the within-cluster sum of squares (WCSS), which is the objective function for K-means
- $k$ is the number of clusters
- $m_k$ is the centroid of cluster k.

To identify distinct subgroups within the dataset, the elbow method was used to determine the optimal number of clusters ($k$) by evaluating the WCSS, while silhouette analysis measures how well each data point fits within its assigned cluster. Once the optimal k was established, the K-means algorithm partitioned the data by iteratively refining cluster centroids until membership stabilized. The quality of this final partition was visually validated using a silhouette plot, which graphically displays the cohesion and separation of the resulting clusters.[47] To complement this analysis, agglomerative clustering was conducted using Ward's linkage method with a Euclidean distance metric, and the output was visualized as a dendrogram.

After identifying three clusters, a one-way analysis of variance was conducted to assess whether there are significant differences in HRV parameters among the groups. Additionally, *post hoc* Tukey's tests were applied for pairwise comparisons. The magnitude of these differences was evaluated using Cohen's $d$ effect size, and statistical significance was assessed using 95% confidence intervals (CI) for the mean difference. All statistical analyses were performed in MATLAB 2020b (MathWorks, United States) with a significance level set at $\alpha = 0.05$.

## 4. Results

Table 1 presents the anthropometric, physical, and HRV data of the participants. Normality tests indicate that age, body mass, height, and MRR follow a Gaussian distribution ($p \geq 0.05$), and the low standard deviations support the homogeneity of the sample. However, SDNN, RMSSD, pNN50, LF, and HF do not follow a normal distribution ($p < 0.05$). These findings highlight the importance of HRV data normalization in the context of machine learning, particularly when applying techniques such as PCA and K-means.

After dimensionality reduction using PCA, a non-hierarchical K-means clustering algorithm was applied to the first two PCs. The algorithm was initialized randomly

**Table 1. Anthropometric characteristics and heart rate variability parameters of the participants**

| Variables | Mean±standard deviation | *p*-value |
|---|---|---|
| Age (years) | 22.0±2.8 | 0.200 |
| Body mass (kg) | 65.2±6.9 | 0.935 |
| Height (cm) | 171.0±6.5 | 0.745 |
| MRR (ms) | 935.0±132.2 | 0.571 |
| SDNN (ms) | 62.8±30.9 | 0.008 |
| RMSSD (ms) | 72.7±44.6 | 0.001 |
| pNN50 (%) | 36.6±24.5 | 0.007 |
| LF (%) | 49.0±21.8 | 0.015 |
| HF (%) | 51.8±22.2 | 0.031 |

Abbreviations: HF: High-frequency; LF: Low-frequency; MRR: Mean R-R interval; pNN50: The proportion of adjacent normal-to-normal intervals differing by more than 50 ms; RMSSD: The root mean square of successive differences between adjacent intervals; SDNN: The standard deviation of all normal-to-normal intervals.

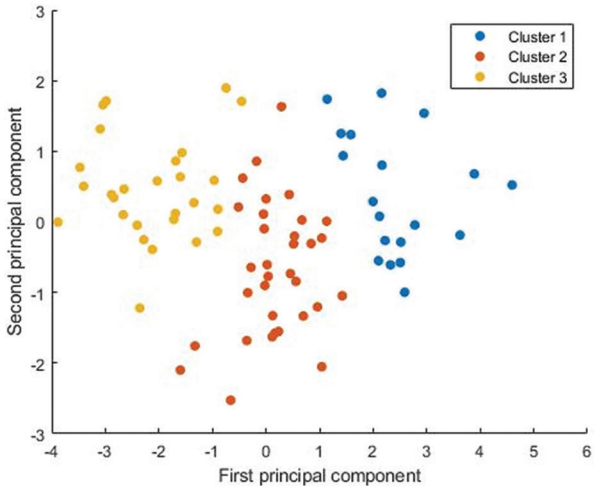


**Figure 1.** K-means clustering based on principal component coefficients derived from normalized heart rate variability data

and run for 10 iterations, resulting in the partitioning of the data into three distinct clusters, as shown in Figure 1.

Figure 2 shows the WCSS as a function of $k$, which is used to determine the optimal number of clusters in the dataset. The graph reveals a distinct "elbow" at $k = 3$, where the WCSS reduction plateaus. This suggests that three clusters represent the most appropriate choice, as adding additional clusters beyond this point results in only marginal improvements in cluster compactness. The elbow method, therefore, supports the selection of $k = 3$ as the optimal number of clusters for subsequent analysis.

As shown in Table 2, cluster 3 ($n = 19$) demonstrates significantly higher HRV parameters compared to cluster 1 ($n = 33$) and cluster 2 ($n = 28$) ($p = 0.001$). *Post hoc* analysis further confirms that cluster 1 exhibits significantly different HRV parameters compared to both cluster 2 and cluster 3 ($p$=0.001).

**Table 2. Comparison of heart rate variability parameters among the identified clusters**

| Variables | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| MRR (ms) | 1034.7±129.6 | 919.2±121.2 | 886.1±113.0 |
| SDNN (ms) | 101.1±24.8 | 59.9±20.4 | 40.1±18.1 |
| RMSSD (ms) | 132.8±33.0 | 68.1±25.7 | 37.5±21.7 |
| pNN50 (%) | 67.2±10.7 | 37.9±16.7 | 14.4±13.4 |
| LF (%) | 34.4±18.0 | 40.6±15.5 | 68.9±15.8 |
| HF (%) | 69.2±15.6 | 59.6±15.4 | 30.8±16.0 |

Note: Data are presented as mean±standard deviation.
Abbreviations: HF: High-frequency; LF: Low-frequency; MRR: Mean R-R interval; pNN50: The proportion of adjacent normal-to-normal intervals differing by more than 50 ms; RMSSD: The root mean square of successive differences between adjacent intervals; SDNN: The standard deviation of all normal-to-normal intervals.

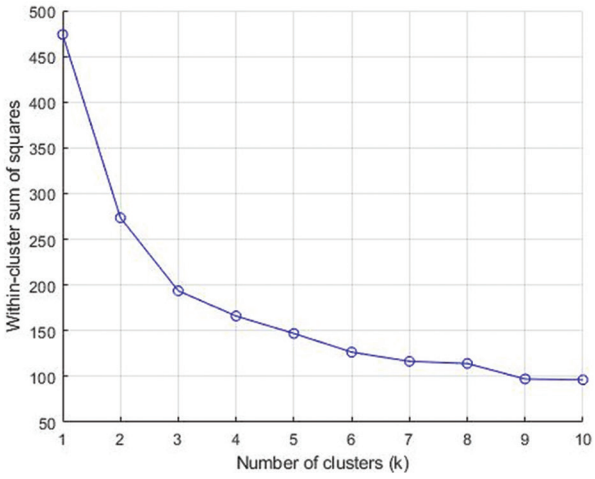


**Figure 2.** Within-cluster sum of squares plotted against different values of $k$ to determine the optimal number of clusters

Cluster 1 demonstrates a significantly greater parasympathetic profile than cluster 2. This is supported by large and statistically significant differences in mean RRI (mean difference = 122.18; 95% CI = 52.67 – 191.70; $d$ = 0.99), SDNN (mean difference = 44.35; 95% CI = 32.40 – 56.31; $d$ = 2.09), RMSSD (mean difference = 69.81; 95% CI = 55.80 – 83.81; $d$ = 2.80), and pNN50 (mean difference = 32.68; 95% CI = 25.30 – 40.06; $d$ = 2.49). HF power – an established marker of vagal tone – is also significantly higher in cluster 1, with a medium-to-large effect size ($d$ = 0.67). In contrast, no significant difference in LF power is observed between these two clusters (95% CI = −14.06 – 2.42).

The differences between cluster 1 and cluster 3 are even more substantial. Cluster 1 exhibits significantly higher values across all indices of overall HRV and parasympathetic activity, with very large effect sizes for mean RRI ($d$ = 1.32), SDNN ($d$ = 3.10), RMSSD ($d$ = 3.77), pNN50 ($d$ = 4.48), and HF ($d$ = 3.28). Furthermore, cluster 1 exhibits significantly lower LF power than cluster 3 (mean difference = −40.18, $d$ = −2.75), indicating reduced sympathetic modulation compared to cluster 3.

Although no statistically significant difference in MRR is observed between clusters 2 and 3 (95% CI = −29.45 – 92.78), their autonomic modulation profiles differ significantly. Cluster 2 exhibits significantly greater parasympathetic activity than cluster 3, as evidenced by substantial differences in SDNN, RMSSD, pNN50, and HF power. In addition, cluster 2 exhibits significantly lower LF power ($d$ = −2.82), indicating increased vagal tone and reduced sympathetic modulation compared to cluster 3.

The silhouette index, a commonly used measure of cluster quality, was calculated to evaluate the PCA-based grouping of normalized HRV data. The silhouette index values indicate effective cluster separation, with low intra-cluster variability and high inter-cluster dissimilarity. The silhouette scores are 0.397 for one cluster, 0.481 for two clusters, and 0.556 for three clusters. The highest silhouette coefficient is observed at $k = 3$, supporting the selection of three clusters as the optimal solution (Figure 3).

The agglomerative hierarchical clustering dendrogram provides an alternative view of the dataset's structure (Figure 4). While the longest vertical linkage representing the primary bifurcation suggests a two-cluster solution, further analysis reveals a distinct substructure within one of the main branches. This hierarchical arrangement indicates that a three-cluster model may more accurately reflect the underlying granularity of the data. Thus, the finer details of the dendrogram further support the three-cluster solution identified by the elbow and silhouette

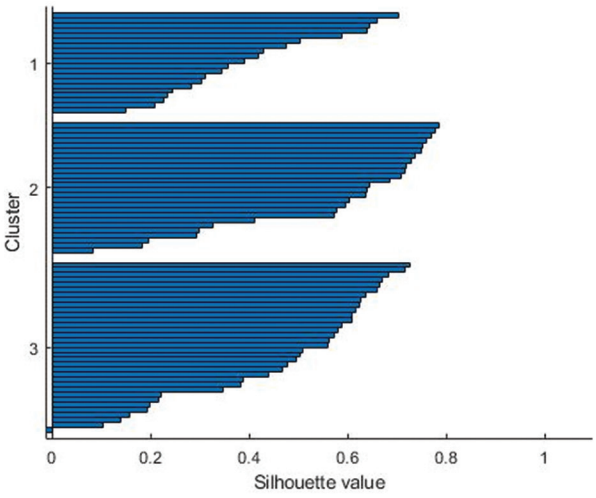


**Figure 3.** Silhouette plot for identifying the optimal number of clusters

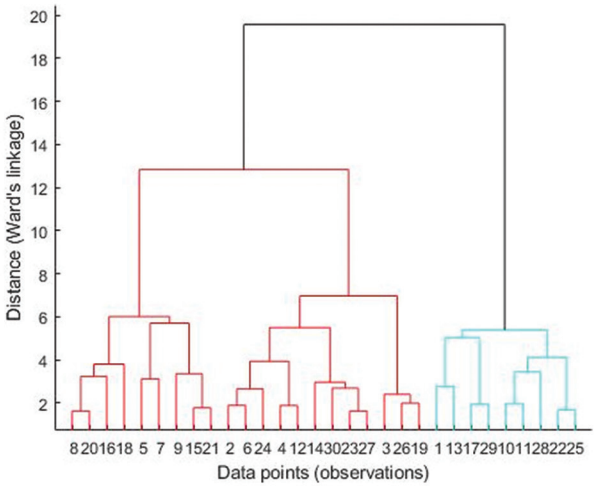


**Figure 4.** Hierarchical cluster analysis of the heart rate variability dataset

methods, highlighting a nested and interpretable structure within the dataset.

## 5. Discussion

The primary aim of this study is to develop and validate a method for identifying distinct patterns of autonomic regulation in healthy men, using a combined PCA and K-means clustering approach applied to HRV data. The findings demonstrate that this approach is both feasible and effective in stratifying cardiovascular autonomic function, revealing significant physiological heterogeneity within a homogeneous and healthy sample.

HRV is an important biomarker for assessing cardiovascular health, serving as a non-invasive indicator of the ANS and its modulation of cardiac function.[48] It reflects the balance between sympathetic and parasympathetic activity, with implications for both physical and mental well-being.[49] Reduced HRV is generally associated with a higher risk of CVD and physiological stress,[3,50,51] whereas elevated HRV is typically indicative of good health.[38,52-54] The advent of wearable technology has made HRV monitoring more practical, enabling early detection of cardiovascular dysfunction and supporting proactive health management strategies.[55] However, to date, no studies have examined the use of combined PCA and K-means clustering techniques to identify individual risk profiles based on HRV data.

Pasquini *et al.*[54] applied PCA to identify five PCs in ANS time series data that collectively explained 75% of the variance during emotional reactivity tasks. K-means clustering revealed five distinct ANS states corresponding to specific emotions, such as awe and sadness.

However, the present study validates a methodological framework integrating PCA and K-means to identify distinct patterns of autonomic regulation in healthy men using HRV data. This approach demonstrates variability in autonomic balance, even within a healthy population. The findings highlight the potential of these techniques to stratify individuals based on cardiovascular autonomic function and suggest that lower HRV profiles may indicate individuals who could benefit from targeted interventions to improve autonomic balance.

A critical component of the analysis involves selecting the optimal number of clusters. Notably, both the elbow method and hierarchical dendrogram support $k = 3$ as a plausible solution, whereas the silhouette index favors a more conservative $k = 2$ solution. Rather than interpreting this discrepancy as a contradiction, it provides valuable insights into the hierarchical structure of the data. The higher silhouette score for $k = 2$ indicates the presence of two primary, well-defined groups. However, selecting $k = 3$, as recommended by the elbow method, provides greater physiological granularity by revealing an "intermediate" group that would otherwise remain undetected in a binary classification. This methodological approach allows for a more detailed and clinically interpretable classification of autonomic regulation.

In addition, these findings offer promising potential for clinical application. The ability to stratify healthy individuals into potential autonomic "risk" profiles provides insights into the development of personalized preventive strategies. A *post hoc* analysis of the three-cluster solution identified distinct, hierarchically ordered autonomic profiles. Cluster 3 (high-vagal-tone profile; $n = 33$) exhibits the highest values across all HRV parameters, indicating enhanced parasympathetic modulation. This profile is typically associated with good cardiovascular health, greater stress resilience, and efficient physical conditioning.

Cluster 2 (intermediate autonomic profile; $n = 28$) represents a moderate autonomic state, with HRV values significantly lower than those of cluster 1 but higher than those of cluster 3. Cluster 1 (low-vagal-tone or sympathetic predominant profile; $n = 19$) exhibits the lowest HRV values of parasympathetic activity and the highest relative LF power. While these individuals are clinically healthy, this profile may indicate a subclinical state of autonomic imbalance or reduced adaptive capacity, potentially indicating elevated long-term cardiovascular risk.

The identification of these three distinct profiles in a young, healthy population represents a central finding of this study. While the literature establishes that reduced HRV is associated with disease, the present findings suggest that a spectrum of autonomic function exists even in the absence of clinical pathology. This observation aligns with previous research, including that by Pasquini *et al*.,[54] who applied similar techniques to identify autonomic states during emotional reactivity. The present study further highlights interindividual variability at rest, suggesting that these autonomic profiles may represent individual physiological traits.

The use of machine learning, particularly K-means clustering, underscores the potential of AI to enhance HRV analysis and deepen understanding of ANS regulation. This data-driven approach enables the identification of distinct autonomic patterns that may be obscured by traditional statistical methods. The combined PCA and K-means clustering method serves as an effective tool for exploring complex physiological data and identifying meaningful subgroups within a population. Furthermore, the application of machine learning techniques to predict individual risk based on HRV profiles holds significant promise for enabling more personalized and effective preventive strategies.[4,5,11-13]

It is crucial to interpret these findings within the context of the study's design. Unlike studies that aim to distinguish between healthy individuals and those with cardiac disease, a supervised classification problem this study addresses an unsupervised discovery question: what distinct autonomic profiles exist within a healthy population? This approach is based on the hypothesis that physiologically meaningful patterns may reflect distinct long-term cardiovascular health trajectories, even among asymptomatic individuals, before the onset of clinical disease. Therefore, the identification of three distinct clusters should not be viewed as a limitation due to the absence of a patient control group, but rather as a validation of the study's primary objective, which explores an under-investigated area of autonomic profiling in healthy populations.

However, several limitations should be acknowledged in this study. A key limitation is the homogeneous nature of the sample, which consisted of 80 healthy young men from Macapá, Brazil. Consequently, the findings should be interpreted as proof of concept within this specific demographic. The study emphasizes that the identified cluster patterns may not be directly generalizable to women, other age groups, ethnic backgrounds, or populations with different health and lifestyle characteristics. Therefore, validating the method is a priority, and future studies are recommended to include larger, more diverse, and multi-center cohorts to determine the broader applicability of these autonomic profiles. Furthermore, the cross-sectional design of the study limits the ability to draw conclusions about the temporal stability of these profiles or their predictive value for future health outcomes.

## 6. Conclusion

This study highlights that the application of combined PCA and K-means clustering to HRV data is a valid and effective approach for identifying distinct patterns of autonomic regulation in healthy men. Three distinct physiological profiles, high vagal tone, intermediate vagal tone, and low vagal tone, were identified, even within a demographically homogeneous cohort. These findings reinforce the importance of HRV as a sensitive biomarker of cardiovascular health and underscore the potential of machine learning techniques to advance personalized preventive strategies. However, future longitudinal studies are needed to examine the stability of these autonomic profiles and their association with long-term health outcomes across more diverse populations.

## Acknowledgments

The author would like to thank the Physical Education Department at the Federal University of Amapá for allowing them to use their laboratory facilities and for helping them to recruit participants.

## Funding

This research was funded by the Amapá Research Support Foundation through its public call 003/2018, specifically within the "Research Program for the Unified Health System (SUS): management in Health-PPSUS." The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Conflict of interest

The author declares no conflicts of interest.

## Author contributions

This is a single-authored article.

## Ethics approval and consent to participate

The study protocol was ethically approved by the Human Research Ethics Committee of the Federal University of Amapá (CAAE: 50150121.1.0000.0003; approval number: 5.121.013) and conducted in accordance with the Declaration of Helsinki. All participants provided verbal informed consent before their inclusion in the study.

## Consent for publication

Verbal informed consent was obtained from all participants before their inclusion in the study. The consent form explicitly stated that anonymized data would be analyzed and subsequently used for publication in scientific journals. All data were fully anonymized before analysis, and the results are presented in an aggregated format to ensure that no individual participant can be identified.

## Availability of data

The dataset generated and analyzed during the current study is not publicly available, since it contains sensitive clinical information that is subject to data protection regulations.

## References

1. Timmis A, Vardas P, Townsend N, *et al*. European society of cardiology: Cardiovascular disease statistics 2021. *Eur Heart J*. 2022;43(8):716-799.

   doi: 10.1093/eurheartj/ehab892

2. Chorin E, Hochstadt A, Schwartz AL, Matz G, Viskin S, Rosso R. Continuous heart rate monitoring for automatic detection of life-threatening arrhythmias with novel biosensing technology. *Front Cardiovasc Med*. 2021;8:707621.

   doi: 10.3389/fcvm.2021.707621

3. Zeid S, Buch G, Velmeden D, *et al*. Heart rate variability: Reference values and role for clinical profile and mortality in individuals with heart failure. *Clin Res Cardiol*. 2024;113(9):1317-1330.

   doi: 10.1007/s00392-023-02248-7

4. Materko W, Fernandes DF, Façanha CCR, *et al*. A machine learning approach to developing an accurate stratification of type 2 diabetes mellitus based on heart rate variability parameters using the K-means clustering technique in elderly women. *Gazz Med Ital Arch Sci Med*. 2024;183:44-50.

   doi: 10.23736/S0393-3660.23.05096-9

5. Materko W, Fernandes DF, Sadala MN, Pureza D, Alberto AAD, Pena FPS. Evaluation on heart rate variability parameters in elderly with type 2 diabetes mellitus using principal component analysis. *Gazz Med Ital Arch Sci Med*. 2022;181:879-884.

   doi: 10.23736/S0393-3660.22.04782-9

6. Gillinov S, Etiwy M, Wang R, *et al*. Variable accuracy of wearable heart rate monitors during aerobic exercise. *Med Sci Sports Exerc*. 2017;49(8):1697-1703.

   doi: 10.1249/MSS.0000000000001284

7. Materko W, Dos Reis Façanha CC, Guedes GC, *et al*. Temporal cross-correlation between Polar® heart rate monitor interface board and ECG to measure RR interval at rest. *Isokinet Exerc Sci*. 2024;32(1):59-64.

   doi: 10.3233/IES-230061

8. Latino F, Tafuri F. Wearable sensors and the evaluation of physiological performance in elite field hockey players. *Sports (Basel)*. 2024;12(5):124.

   doi: 10.3390/sports12050124

9. Trevizani GA, Nasario-Junior O, Benchimol-Barbosa PR, Silva LP, Nadal J. Cardiac autonomic changes in middle-aged women: Identification based on principal component analysis. *Clin Physiol Funct Imaging*. 2016;36(4):269-273.

   doi: 10.1111/cpf.12222

10. Perrone MA, Volterrani M, Manzi V, Barchiesi F, Iellamo F. Heart rate variability modifications in response to different types of exercise training in athletes. *J Sports Med Phys Fitness*. 2021;61(10):1411-1415.

    doi: 10.23736/S0022-4707.21.12480-6

11. Rafieisangari R, Shiri N. A multi-adaptive neuro-fuzzy inference system with variable thresholds for heartbeat classification. *Artif Intell Health*. 2024;1(4):43-60.

    doi: 10.36922/aih.3367

12. Nshimiyimana M, Niyogisubizo J, Ninteretse JD. Heartbeat classification using various machine learning models: A comparative study. *Artif Intell Health*. 2024;1(4):61-72.

    doi: 10.36922/aih.3543

13. Materko W, Miranda SAM, Bezerra THL, De Oliveira Figueira CAM. Heart rate variability in soccer players and the application of unsupervised machine learning. *Explor Cardiol*. 2025;3:101241.

    doi: 10.37349/ec.2025.101241

14. Materko W. Stratification fitness aerobic based on heart rate variability during rest by principal component analysis and K-means clustering. *J Exerc Physiol*. 2018;21(1):91-101.

15. Kaya IE, Pehlivanlı AÇ, Sekizkardeş EG, Ibrikci T. PCA based clustering for brain tumor segmentation of T1w MRI images. *Comput Methods Programs Biomed*. 2017;140:19-28.

    doi: 10.1016/j.cmpb.2016.11.011

16. Jolliffe IT, Cadima J. Principal component analysis: A review

and recent developments. *Philos Trans A Math Phys Eng Sci*. 2016;374(2065):20150202.

doi: 10.1098/rsta.2015.0202

17. Ikotun AM, Absalom E, Abualigah L, Abuhaija B, Jia H. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Inform Sci*. 2023;622:178-210.

doi: 10.1016/j.ins.2022.11.139

18. Sinaga KP, Yang MS. Unsupervised K-Means Clustering Algorithm. *IEEE Access*. 2020;8:80716-80727.

doi: 10.1109/access.2020.2988796

19. Ishaque S, Khan N, Krishnan S. Trends in heart-rate variability signal analysis. *Front Digit Health*. 2021;3:639444.

doi: 10.3389/fdgth.2021.639444

20. Kleiger RE, Stein PK, Bigger JT Jr. Heart rate variability: Measurement and clinical utility. *Ann Noninvasive Electrocardiol*. 2005;10(1):88-101.

doi: 10.1111/j.1542-474X.2005.10101.x

21. Akselrod S, Gordon D, Ubel FA, Shannon DC, Barger AC, Cohen RJ. Power spectrum analysis of heart rate fluctuations: A quantitative probe of beat-to-beat cardiovascular control. *Science*. 1981;213(4504):220-222.

doi: 10.1126/science.6166045

22. Kleiger RE, Miller JP, Bigger JT Jr., Moss AJ, The Multicenter Post-Infarction Research Group. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *Am J Cardiol*. 1987;59(4):256-262.

doi: 10.1016/0002-9149(87)90795-8

23. Sammito S, Thielmann B, Böckelmann I. Update: Factors influencing heart rate variability-a narrative review. *Front Physiol*. 2024;15:1430458.

doi: 10.3389/fphys.2024.1430458

24. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: Standards of measurement, physiological interpretation and clinical use. *Circulation*. 1996;93(5):1043-1065.

doi: 10.1161/01.cir.93.5.1043

25. Clifford GD, Mcsharry PE, Tarassenko L. Characterizing artefact in the normal human 24-hour RR time series to aid identification and artificial replication of circadian variations in human beat to beat heart rate using a simple threshold. *Comput Cardiol*. 2002;29:129-132.

doi: 10.1109/CIC.2002.1166724

26. Chemla D, Young J, Badilini F, *et al*. Comparison of fast Fourier transform and autoregressive spectral analysis for the study of heart rate variability in diabetic patients. *Int J Cardiol*. 2005;104(3):307-313.

doi: 10.1016/j.ijcard.2004.12.018

27. Staffini A, Svensson T, Chung UI, Svensson AK. Heart rate modeling and prediction using autoregressive models and deep learning. *Sensors (Basel)*. 2021;22(1):34.

doi: 10.3390/s22010034

28. Hayano J, Yuda E. Assessment of autonomic function by long-term heart rate variability: Beyond the classical framework of LF and HF measurements. *J Physiol Anthropol*. 2021;40(1):21.

doi: 10.1186/s40101-021-00272-y

29. Jarczok MN, Weimer K, Braun C, *et al*. Heart rate variability in the prediction of mortality: A systematic review and meta-analysis of healthy and patient populations. *Neurosci Biobehav Rev*. 2022;143:104907.

doi: 10.1016/j.neubiorev.2022.104907

30. Choi A, Shin H. Quantitative analysis of the effect of an ectopic beat on the heart rate variability in the resting condition. *Front Physiol*. 2018;9:922.

doi: 10.3389/fphys.2018.00922

31. Immanuel S, Teferra MN, Baumert M, Bidargaddi N. Heart rate variability for evaluating psychological stress changes in healthy adults: A scoping review. *Neuropsychobiology*. 2023;82(4):187-202.

doi: 10.1159/000530376

32. Ardashev A, Loskutov A, Passman R, Zhelyakov E, Rytkin E, Efimov I. Theoretical and practical aspects of the nonlinear dynamics' methods of heart rate variability analyses in tachyarrhythmia patients underwent radiofrequency catheter ablation. *Cardiovasc Eng Technol*. 2025;16(2):190-201.

doi: 10.1007/s13239-024-00766-7

33. Bauer A, Kantelhardt JW, Bunde A, *et al*. Phase-rectified signal averaging detects quasi-periodicities in non-stationary data. *Physica A*. 2006;364:423-434.

doi: 10.1016/j.physa.2005.08.080

34. Bauer A, Kantelhardt JW, Barthel P, *et al*. Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction: Cohort study. *Lancet*. 2006;367(9523):1674-1681.

doi: 10.1016/S0140-6736(06)68735-7

35. Alberto AC, Pedrosa RC, Zarzoso V, Nadal J. Association between circadian Holter ECG changes and sudden cardiac death in patients with Chagas heart disease. *Physiol Meas*. 2020;41(2):025006.

doi: 10.1088/1361-6579/ab6ebc

36. Materko W, Bartels R, Pecanha T, Lima JRP, Carvalho ARS, Nadal J. Maximum oxygen uptake prediction model based on heart rate variability parameters for young healthy adult

males at rest. *Open Access Biostat Bioinform*. 2018;2(1):1-7.

doi: 10.31031/oabb.2018.02.000536

37. Nasario-Junior O, Benchimol-Barbosa PR, Pedrosa RC, Nadal J. Assessment of autonomic function by phase rectification of RR-interval histogram analysis in Chagas disease. *Arq Bras Cardiol*. 2015;104(6):450-455.

doi: 10.5935/abc.20150032

38. Greenacre M, Groenen PJF, Hastie T, *et al*. Principal component analysis. *Nat Rev Methods Primers*. 2022;2(1):100.

doi: 10.1038/s43586-022-00184-w

39. Peres-Neto PR, Jackson DA, Somers KM. How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Comput Stat Data Anal.* 2005;49(4):974-997.

doi: 10.1016/j.csda.2004.06.015

40. Cadima J, Jolliffe IT. Loading and correlations in the interpretation of principle compenents. *J Appl Stat*. 1995;22(2):203-214.

doi: 10.1080/757584614

41. Dalmaijer ES, Nord CL, Astle DE. Statistical power for cluster analysis. *BMC Bioinformatics*. 2022;23(1):205.

doi: 10.1186/s12859-022-04675-1

42. Lawton M, Ben-Shlomo Y, May MT, *et al*. Developing and validating Parkinson's disease subtypes and their motor and cognitive progression. *J Neurol Neurosurg Psychiatry*. 2018;89(12):1279-1287.

doi: 10.1136/jnnp-2018-318337

43. Scitovski R, Sabo K, Martínez-Álvarez F, Ungar Š. *Cluster Analysis and Applications*. Berlin: Springer; 2021.

44. Sneath PHA. A method for testing the distinctness of clusters: A test of the disjunction of two clusters in Euclidean space as measured by their overlap. *Math Geol.* 1977;9(2):123-143.

doi: 10.1007/bf02312508

45. Holtrop J, Bhatt DL, Ray KK, *et al*. Impact of the 2021 European society for cardiology prevention guideline's stepwise approach for cardiovascular risk factor treatment in patients with established atherosclerotic cardiovascular disease. *Eur J Prev Cardiol*. 2024;31(6):754-762.

doi: 10.1093/eurjpc/zwae038

46. World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191-2194.

doi: 10.1001/jama.2013.281053

47. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math.* 1987;20:53-65.

doi: 10.1016/0377-0427(87)90125-7

48. Raju S, Woo H, Fawzy A, *et al.* Decreased cardiac autonomic function is associated with higher exacerbation risk and symptom burden in chronic obstructive pulmonary disease. J *COPD Found*. 2023;10(3):328-334.

doi: 10.15326/jcopdf.2023.0410

49. Sundas A, Contreras I, Navarro-Otano J, Soler J, Beneyto A, Vehi J. Heart rate variability over the decades: A scoping review. *PeerJ*. 2025;13:e19347.

doi: 10.7717/peerj.19347

50. Osborne MT, Shin LM, Mehta NN, Pitman RK, Fayad ZA, Tawakol A. Disentangling the links between psychosocial stress and cardiovascular disease. *Circ Cardiovasc Imaging*. 2020;13(8):e010931.

doi: 10.1161/circimaging.120.010931

51. Vaccarino V, Bremner JD. Stress and cardiovascular disease: An update. *Nat Rev Cardiol*. 2024;21(9):603-616.

doi: 10.1038/s41569-024-01024-y

52. Faust O, Hong W, Loh HW, *et al*. Heart rate variability for medical decision support systems: A review. *Comput Biol Med*. 2022;145:105407.

doi: 10.1016/j.compbiomed.2022.105407

53. Lu L, Zhu T, Morelli D, *et al*. Uncertainties in the analysis of heart rate variability: A systematic review. *IEEE Rev Biomed Eng*. 2024;17:180-196.

doi: 10.1109/rbme.2023.3271595

54. Pasquini L, Noohi F, Veziris CR, *et al.* Dynamic autonomic nervous system states arise during emotions and manifest in basal physiology. *Psychophysiology*. 2023;60(4):e14218.

doi: 10.1111/psyp.14218

55. Turcu AM, Ilie AC, Ştefăniu R, *et al*. The impact of heart rate variability monitoring on preventing severe cardiovascular events. *Diagnostics (Basel)*. 2023;13(14):2382.

doi: 10.3390/diagnostics13142382