Heliyon 12 (2026) e44737

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Capturing non-Gaussianity of RR interval distributions based on Gaussian scale mixture representation

Naoki Hagiyama [a,*], Akira Furui [a,*], Harutoyo Hirano [b], Toshio Tsuji [a,*]

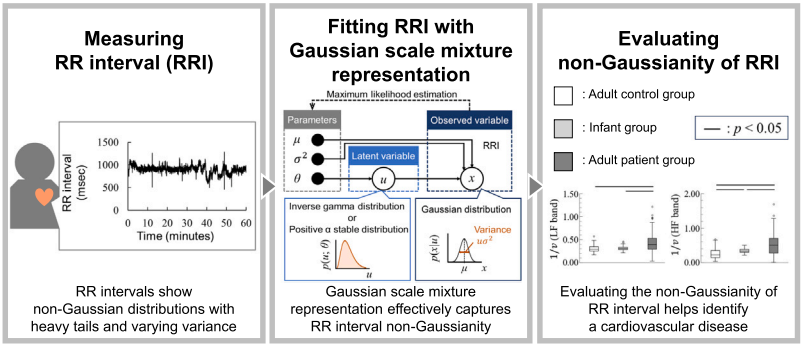
[a] The Graduate School of Advanced Science and Engineering, Hiroshima University, 1-4-1 Kagamiyama, Higashihiroshima, 739-8527, Hiroshima, Japan

[b] Department of Medical Equipment Engineering, Clinical and Educational Collaboration Unit, School of Medical Sciences, Fujita Health University, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, 470-1192, Aichi, Japan

HIGHLIGHTS

- Gaussian scale mixture modeling can capture non-Gaussianity in RR intervals.
- RR intervals show non-Gaussian distributions with heavy tails and varying variance.
- Non-Gaussianity in RR intervals differs between age groups and health conditions.
- Evaluating non-Gaussianity in RR intervals helps identify a cardiovascular disease.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Gaussian scale mixture
Generalized Gaussian distribution
Non-Gaussianity
RR interval
Student's *t*-distribution

ABSTRACT

The time interval between two consecutive R-peaks in an electrocardiography signal is called the RR interval (RRI), and its non-Gaussianity is a valuable marker for assessing cardiovascular disease risk. While RRI non-Gaussianity can be quantified by fitting probability distributions, the selection of an appropriate distribution remains challenging. Although Gaussian scale mixture representation has proven effective for analyzing non-Gaussianity in various biological signals, its application to RRIs has not been explored. This paper examines the application of Gaussian scale mixture representation for quantifying these non-Gaussian characteristics in RRI data. Using open RRI databases from participants of different ages and with various cardiovascular diseases, we demonstrated that this representation, particularly using the Student's *t*-distribution, provides better characterization of RRI distributions compared to conventional Gaussian and stable distributions. The non-Gaussianity indices derived from this approach, when combined with conventional heart rate variability measures, improved the classification of cardiovascular diseases such as congestive

* Corresponding authors.
Email addresses: naokihagiyama@hiroshima-u.ac.jp (N. Hagiyama), akirafurui@hiroshima-u.ac.jp (A. Furui), tsuji-c@bsys.hiroshima-u.ac.jp (T. Tsuji).

https://doi.org/10.1016/j.heliyon.2026.e44737
Received 21 November 2024; Received in revised form 30 January 2026; Accepted 9 March 2026
Available online 21 March 2026

heart failure and hypertension, achieving an average area under the receiver operating characteristic curve of 0.835 across multiple classifiers. These findings establish Gaussian scale mixture representation as an effective framework for quantifying RRI non-Gaussianity, offering potential improvements in cardiovascular risk assessment.

1. Introduction

The autonomic nervous system involuntarily controls vital functions such as circulation, respiration, and body temperature. Electrical signals generated in the sinus node and regulated by the cardiac nerves are transmitted to the atrioventricular node and Purkinje fibers, contracting the myocardium of the atria and ventricles. This electrical activity, controlled by the autonomic nervous system, can be measured using electrocardiography (ECG) [1]. The time interval between two consecutive R-peaks in the ECG signal is called the RR interval (RRI) [2].

Heart rate variability (HRV) is a method to quantify and characterize RRI fluctuations [3]. Traditional HRV analysis employs both time-domain and frequency-domain methods to assess autonomic nervous system activity. Time-domain indices characterize the magnitude of variability, exemplified by overall RRI and the fluctuations between consecutive RRIs. Frequency-domain indices focus on variations within specific frequency bands to characterize the influences of various parts of the autonomic nervous system [4–6]. However, these conventional measures focus primarily on the magnitude of variations, while the statistical properties of the fluctuations themselves may contain additional information.

The amplitude distribution of RRI often deviates from Gaussian distribution, exhibiting heavier tails and more peaked centers [7–10]. These non-Gaussian characteristics reflect the statistical properties of RRI fluctuations that are not fully captured by conventional HRV measures. While time-domain and frequency-domain analyses characterize different aspects of cardiac regulation, such as beat-to-beat variability and specific physiological rhythms, the shape of the amplitude distribution may provide insights into how these regulatory mechanisms manifest in RRI variations. This perspective has motivated investigations into the non-Gaussian nature of RRI distributions.

Various approaches have been developed to analyze these non-Gaussian characteristics, including studies of differences between consecutive RRIs [7], the logarithms of the differences between consecutive RRIs [8], and the detrended RRIs [9,10]. Through these investigations, non-Gaussianity of the amplitude distribution of RRI has emerged as a promising approach for cardiovascular risk assessment [9]. Notably, increased non-Gaussianity in RRI data has been associated with higher mortality risk in patients with congestive heart failure [9,11] and acute myocardial infarction [12,13]. While these findings highlight the clinical significance of RRI non-Gaussianity, its accurate quantification remains challenging. Stable distributions (Lévy stable distributions) have been empirically applied to model such characteristics [7,8,14]. Markov models have also been utilized in HRV analysis to capture dynamic transitions in autonomic regulation [15]. However, identifying suitable probability distributions for accurately representing RRI non-Gaussianity remains an important methodological challenge.

To investigate RRI non-Gaussianity, we explore the Gaussian scale mixture representation as a potential framework for modeling heavy-tailed distributions. This approach represents the RRI distribution as a combination of Gaussian distributions with varying scales [16,17], considering the observation that RRI variability shows temporal changes. The framework suggests possible relationships between variance fluctuations and heavy-tailed distributions, which may be relevant to RRI analysis given the known multiple time scales of autonomic control mechanisms. This mathematical framework offers a quantitative approach to analyze how variance changes relate to the underlying RRI dynamics, potentially providing new insights into cardiovascular regulation. While this approach has been applied to analyze non-Gaussian patterns in other biological signals, particularly in electromyography and electroencephalography [18–20], the potential of this approach for RRI data analysis has not been investigated.

Given this background, we propose a method to evaluate RRI non-Gaussianity based on a Gaussian scale mixture representation. The proposed RRI analysis evaluates RRI amplitude information by modeling the RRI signal through a Gaussian distribution whose variance is treated as a random variable determined by a latent variable. This study focuses on the shape of the amplitude distribution of specific frequency bands in RRI and evaluates non-Gaussianity using a Gaussian scale mixture representation that assumes symmetry.

We conducted comprehensive evaluations of this approach using three complementary analyses: (1) evaluation of the goodness of fit using open source RRI databases, (2) examination of robustness against artificial ectopic beats, and (3) validation of clinical utility through cardiovascular disease detection. Through these analyses, we examined the potential of the Gaussian scale mixture representation as a framework for understanding RRI non-Gaussianity and its clinical applications.

2. Materials and methods

2.1. Gaussian scale mixture representation for RRI modeling

Fig. 1 illustrates the structure of our approach using a graphical model, where white nodes denote random variables and black nodes indicate estimated parameters. In this framework, we model the RRI signal $x \in \mathbb{R}^+$ as a random variable following a Gaussian distribution with mean $\mu \in \mathbb{R}^+$. The variance of the RRI signal is characterized as a random variable, determined by both the scale parameter $\sigma^2 \in \mathbb{R}^+$ of the Gaussian distribution and the latent variable $u \in \mathbb{R}^+$ with parameter θ. We estimate the parameter θ of the variance distribution using maximum likelihood estimation based on the measured RRI signal. By appropriately estimating the probability distribution of the latent variable u, our model can explicitly capture the randomness of the variance.

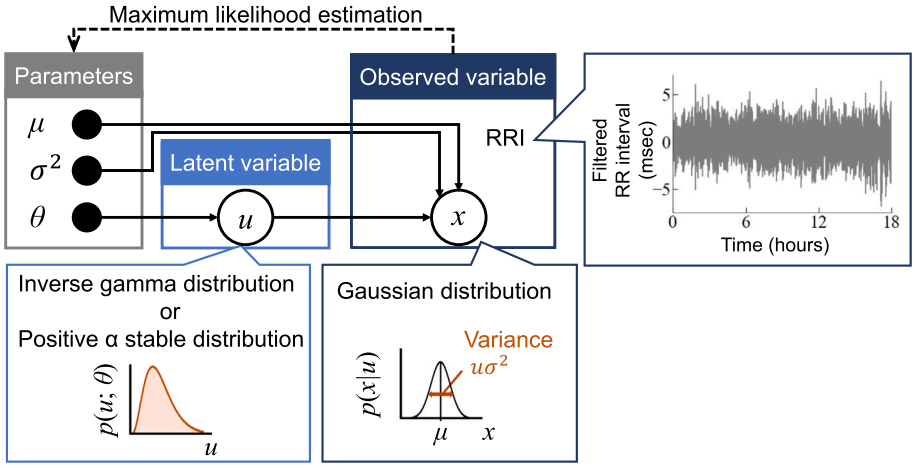


Fig. 1. Schematic overview of the Gaussian scale mixture representation for RRI modeling.

The RRI signal x can be formally expressed using a Gaussian scale mixture representation as follows:

$$p(x) = \int \mathcal{N}(x \mid \mu, u\sigma^2)p(u)du, \tag{1}$$

where $\mathcal{N}(x \mid m, s)$ is a Gaussian distribution with a mean of m and variance of s. In Eq. (1), the variance of the RRI distribution is modulated by the product of latent variable u and scale parameter σ^2. This modulation, driven by the randomness of u, introduces variation in the Gaussian distribution's variance. Consequently, marginal distribution $p(x)$ over u is a non-Gaussian distribution with a heavy tail. As a result, the marginal distribution $p(x)$ obtained by integrating over u yields a non-Gaussian distribution characterized by heavy tails.

We focus on two specific cases of Gaussian scale mixture representations with closed-form solutions: Student's t-distribution and the generalized Gaussian distribution.

Student's t-distribution: Setting the inverse gamma distribution

$$p(u;\nu) = \mathrm{IG}\left(u \,\middle|\, \frac{\nu}{2}, \frac{\nu}{2}\right) \tag{2}$$

as the distribution of the latent variable u yields a Student's t-distribution [21] with ν degrees of freedom. Here, $\mathrm{IG}(a, b)$ represents the probability density function of the inverse gamma distribution with shape parameter a and scale parameter b. The probability density function is given in Eq. (3):

$$\mathrm{IG}(u; a, b) = \frac{b^a}{\Gamma(a)} u^{-a-1} \exp\left(-\frac{b}{u}\right), \tag{3}$$

where $\Gamma(\cdot)$ denotes the gamma function. Substituting Eq. (2) into Eq. (1) and marginalizing u, the scaled mixture expression is attributed to the Student's t-distribution as follows:

$$p(x; \mu, \sigma_t, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi\sigma_t^2}\,\Gamma\left(\frac{\nu}{2}\right)} \left[1 + \frac{(x-\mu)^2}{\nu\sigma_t^2}\right]^{-\frac{\nu+1}{2}}, \tag{4}$$

where σ_t is the scale parameter.

In this study, the inverse of the degrees of freedom, $1/\nu$, is computed to handle the parameters more intuitively. As Student's t-distribution described in Eq. (4), approaches a Gaussian distribution, the inverse of the degrees of freedom $1/\nu$ approaches 0. Therefore, $1/\nu$ can be used to evaluate the non-Gaussian nature of RRI data.

Generalized Gaussian distribution: Setting the distribution

$$p(u;\alpha) = \frac{\alpha\sqrt{2\pi}}{4\Gamma(1/\alpha)} u^{-\frac{3}{2}} S^+_{\alpha/2}\left(\frac{1}{2}u^{-1}\right) \tag{5}$$

as the distribution of latent variable u yields a generalized Gaussian distribution [17] with shape parameter α. Here, $S^+_{\alpha/2}(\cdot)$ is a positive (one-sided) α stable distribution with stability parameter $\alpha/2$. The probability density function of a positive α stable distribution cannot be expressed in closed form [22]. Substituting Eq. (5) into Eq. (1) and marginalizing it for u, the Gaussian scale mixture representation

is attributed to a generalized Gaussian distribution [23,24] as follows:

$$p(x;\mu,\sigma_{GGD},\alpha)=\frac{\alpha}{2\sigma_{GGD}\Gamma(1/\alpha)}\exp\left(-\frac{|x-\mu|^{\alpha}}{\sigma_{GGD}}\right), \tag{6}$$

where σ_{GGD} is the scale parameter. The generalized Gaussian distribution described in Eq. (6) is consistent with a Gaussian distribution when stability parameter $\alpha = 2$ and with a Laplace distribution when stability parameter $\alpha = 1$. Therefore, α can be used to evaluate the non-Gaussianity of RRI data. This study also computed the inverse of α, that is, $1/\alpha$, to handle the parameters more intuitively.

2.2. Database

To evaluate our method across diverse subject characteristics, we utilized six PhysioNet databases (labeled as Databases A-F):

- Database A: *Normal Sinus Rhythm RR Interval Database* [25].
- Database B: *MIT-BIH Normal Sinus Rhythm Database* [25].
- Database C: *RR interval time series from healthy subjects* [25–27].
- Database D: *Congestive Heart Failure RR Interval Database* [25].
- Database E: *BIDMC Congestive Heart Failure Database* [25,28].
- Database F: *Smart Health for Assessing the Risk of Events via the ECG Database* [25,29].

Detailed characteristics of these databases are summarized in Table 1. From each recording, we extracted RRI data for 18 h, beginning 5 min after the start of recording. To ensure data quality, we excluded recordings that contained either consecutive missing intervals exceeding 5 min or had a total length of less than 18 h due to measurement errors or missing values. The complete data screening process is illustrated in Fig. 2.

Based on the cardiovascular disease and age of the subjects, they were classified into three groups: an adult control group with no cardiovascular disease (> 20 years old); an infant group with no cardiovascular disease (< 1 year old); and an adult patient group with cardiovascular disease (> 20 years old). As a result, there were 67 data points in the adult control group, 65 in the infant group, and 146 in the adult patient group.

Table 1
Demographic and clinical characteristics of study participants.

Database	A	B	C	D	E	F
Number of data	54	18	147	29	15	139
Sex (male/female/unknown)	30/24/0	5/13/0	72/67/8	8/2/21	11/4/0	90/49/0
Age (years)	28.5–76	26–50	0.083–55	34–79	22–71	72–77
ECG sampling rate (Hz)	128	128	128	128	250	128
Cardiovascular disease	None	None	None	Congestive heart failure	Congestive heart failure	Hypertension

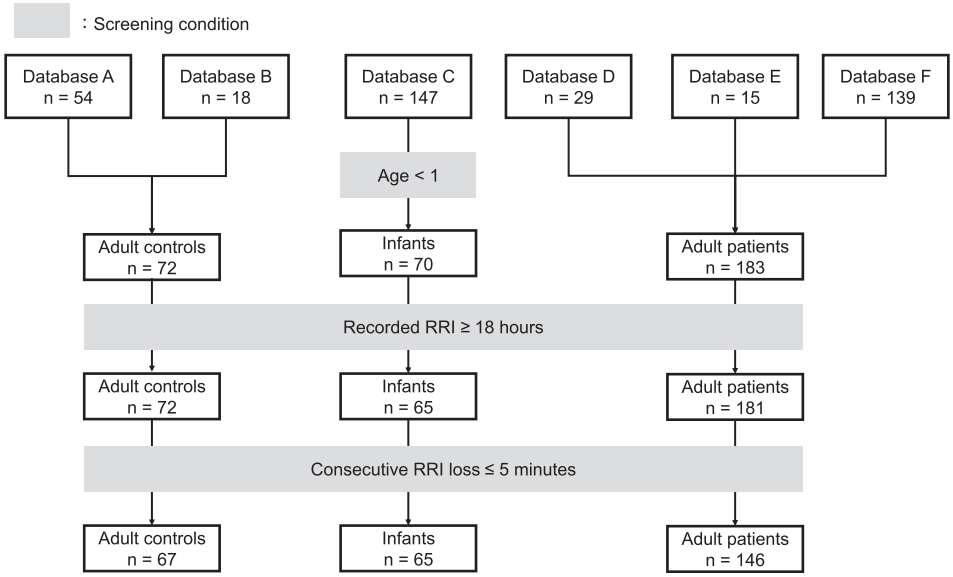


Fig. 2. Flow diagram of the data screening process.

2.3. Preprocessing

Based on previous findings that RRI non-Gaussianity characteristics vary across frequency bands [9], we analyzed the RRI signal separately in LF and HF bands. The frequency bands were determined based on previous studies, with LF set to 0.04–0.15 Hz and HF set to 0.15–0.40 Hz [3]. This frequency-specific analysis is particularly important as the behavior of the latent variable in the Gaussian scale mixture representation may differ between frequency bands.

The preprocessing of N instances of RRI data $\mathbf{r} = \{r_i\}_{i=1}^{N}$ consisted of several steps. First, to exclude RRIs not associated with normal sinus rhythm, we removed outliers based on the interquartile range (IQR): RRIs outside the range $[\mathrm{Q1} - 1.5\mathrm{IQR}, \mathrm{Q3} + 1.5\mathrm{IQR}]$ were excluded. We then applied a linear resampling process at 4 Hz to the outlier removed RRI data, followed by a fifth-order digital Butterworth filter to separate the LF and HF components. The cutoff frequencies were set to 0.04–0.15 Hz for the LF band and 0.15–0.40 Hz for the HF band, and we applied the filter in both forward and reverse directions to eliminate phase distortion (zero-phase filtering).

The filtered RRI data were standardized to zero mean and unit variance, denoted as sRRI. We distinguish these standardized signals as "sRRI (LF band)," and "sRRI (HF band)" for their respective frequency bands. Using maximum likelihood estimation, we fitted these sRRI data to both Student's t-distribution and generalized Gaussian distribution. The non-Gaussianity of each frequency band was characterized by two parameters: $1/\nu$ and σ_t for the Student's t-distribution, and $1/\alpha$ and σ_{GGD} for the generalized Gaussian distribution.

2.4. Experimental procedures

2.4.1. Goodness-of-fit of Gaussian scale mixture representation for RRI data

We evaluated the goodness-of-fit of both Student's t-distribution and generalized Gaussian distribution by fitting them to the sRRI data (both LF and HF bands) from all groups. The RRI length (w_{RRI}) was set to 18 h to capture the characteristics of long-term RRI distributions. For comparison, we also fitted Gaussian, Laplace, Cauchy, and Lévy stable distributions as baseline models.

The goodness-of-fit was quantified using the Euclidean distance (L^2 norm) between the fitted probability distribution and the empirical distribution of the sRRI data. Given a fitted probability distribution $p(x)$ and an empirical sRRI distribution $q(x)$, the L^2 norm was calculated as:

$$d^2(p(z_i), q(z_i)) = \sqrt{\sum_{i=1}^{100} (p(z_i) - q(z_i))^2}, \tag{7}$$

where z_i is defined as follows:

$$z_i = x_{\min} + \frac{n(x_{\max} - x_{\min})}{100}. \tag{8}$$

Here, $x_{\max}$ is the maximum value of x and $x_{\min}$ is the minimum value of x. We compared the goodness-of-fit among different probability distributions using Wilcoxon signed-rank test of the L^2 norms calculated according to Eqs. (7) and (8).

To examine how the length of the RRI data affects the model's performance, we analyzed the relationship between w_{RRI} and both the L^2 norm and non-Gaussianity indices. We progressively reduced w_{RRI} from 18 h to 5 min. For w_{RRI} less than 9 h, we extracted multiple non-overlapping segments from the original 18-h dataset (e.g., two 9-h segments for $w_{RRI} = 9$ h).

2.4.2. Robustness of RRI non-Gaussianity indices

To assess the robustness of our non-Gaussianity indices, we examined their sensitivity to artificial ectopic beats. We quantified the changes in these indices by comparing their values before and after introducing artificial ectopic beats into the RRI data.

The artificial ectopic beats were inserted at a rate of $z\%$ using the following procedure. We first generated an indicator array $\mathbf{s} = \{s_i\}_{i=1}^{N}$ of length N (matching the RRI signal length) with all elements initially set to 0. We then randomly selected $\lfloor Nz/200 \rfloor$ elements, ensuring that no adjacent samples were chosen and assigned 1 or -1 with equal probability.

The modified RRI signal $\hat{\mathbf{r}} = \{\hat{r}_i\}_{i=1}^{N}$ was computed as:

$$\hat{r}_i = \begin{cases} r_i & s_i = 0 \\ 1.25 r_i & s_i = 1 \\ 0.75 r_i & s_i = -1. \end{cases} \tag{9}$$

To ensure that artificially altered RRI values according to Eq. (9) do not propagate to subsequent intervals, modifications were redistributed to the following RRI. Specifically, when the RRI at time i was increased by 1.25 times or reduced to 0.75 times its original value, the corresponding adjustment was subtracted from the subsequent RRI at $i + 1$. This modification better aligns with physiological mechanisms and prevents unrealistic distortions in the RRI sequence.

We evaluated the robustness by comparing the non-Gaussianity indices calculated from both the original (r_i) and modified ($\hat{r}_i$) RRI signals. The ectopic beats were designed to create 25% variations in RRI samples, with insertion rates of 0.1%, 0.5%, and 1.0%. For comparison, we also calculated conventional frequency indices (*LF* and *HF*) [3]. These frequency indices were computed by applying a Hanning window to the resampled RRIs and calculating the power spectral density using Welch's method [30]. The Wilcoxon signed-rank test was used to compare the absolute percentage changes in each index following ectopic beat insertion.

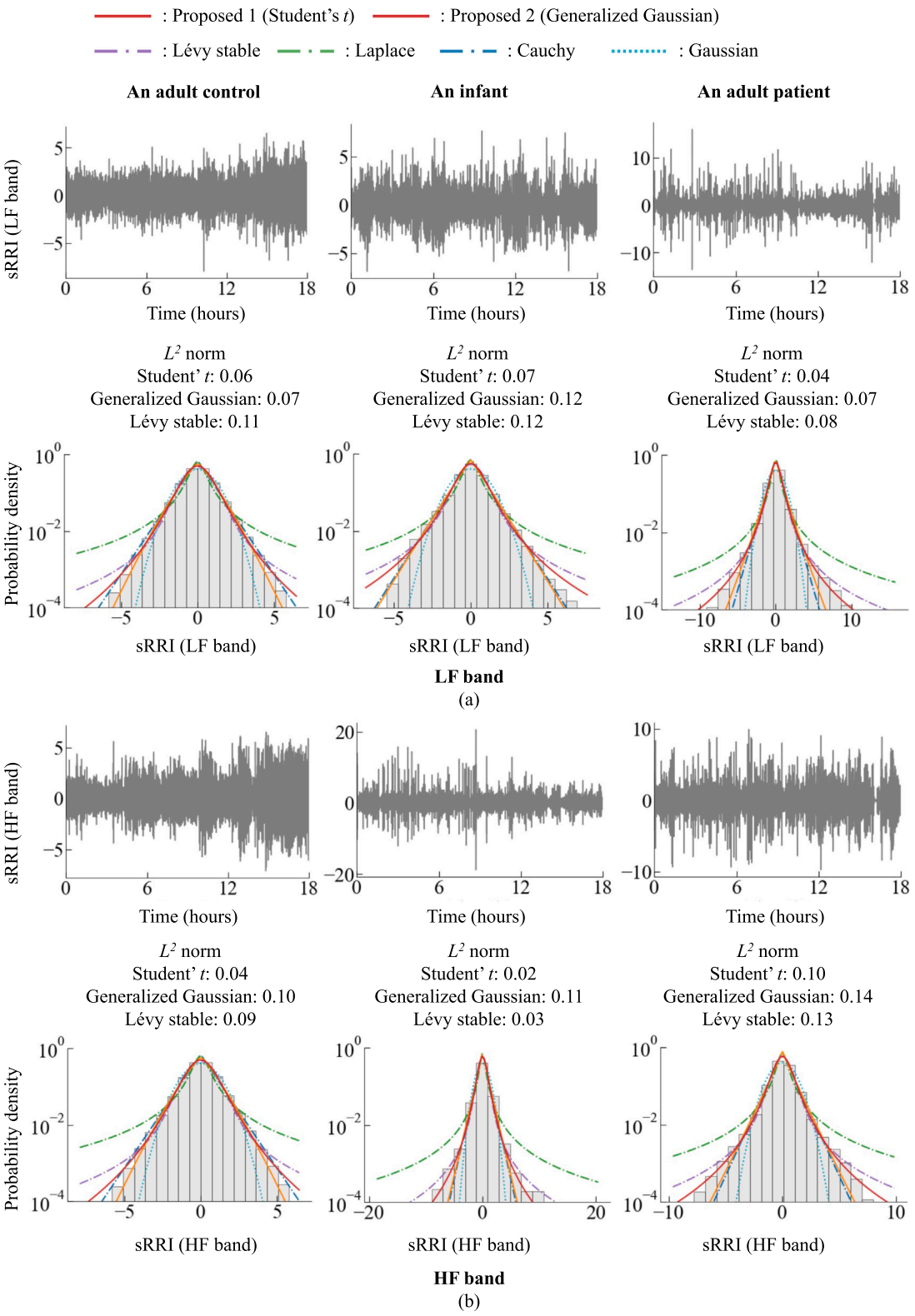


Fig. 3. Representative examples of the sRRI waveform, sRRI histogram, and established distribution. (a) LF band. (b) HF band. Each row shows results for an adult control (left), infant (middle), and adult patient (right). The L^2 norm of the fitted probability distribution is shown in the upper left corner of each histogram.

2.4.3. Ability of non-Gaussianity indices to identify cardiovascular disease

We analyzed the diagnostic potential of non-Gaussianity indices by setting w_{RRI} to 18 h and examining their ability to distinguish between different subject groups. The analysis consisted of two parts: a comparison among three groups (adult control, infant, and adult patient groups) and a binary classification between adult control and patient groups.

For the three-group comparison, we used the Brunner-Munzel test [31], a rank-based method suitable for non-normally distributed data. We then focused on the binary classification between adult control and patient groups, evaluating three different feature sets:

- **Conventional-set** includes the three conventional indices [3] as the *average of the RRIs*, *the standard deviation of the RRIs*, and *LF/HF*.
- ***t*-set** consists of the non-Gaussianity indices obtained by fitting with Student's t-distribution, i.e., $\{1/\nu, \sigma_t\}$ for both LF and HF bands.
- **GGD-set** consists of the non-Gaussianity indices obtained by fitting with a generalized Gaussian distribution, i.e., $\{1/\alpha, \sigma_{GGD}\}$ for both LF and HF bands.

To investigate the relationships between the input features, we calculated the correlation coefficients between the non-Gaussianity indices (t-set) and conventional HRV indices (Conventional-set).

The classification analysis employed four different models: linear discriminant analysis (LDA), logistic regression (LR), support vector machine (SVM), and random forest (RF). Using stratified 5-fold cross-validation to maintain class distributions, we split the dataset into training and test sets in a 4:1 ratio. Both the training and test data were standardized using the mean and standard deviation computed from the training data. To optimize model hyperparameters, the training data was further divided in a 3:1 ratio into training and validation sets, again using stratified 4-fold cross-validation.

Classification performance was evaluated using the area under the receiver operating characteristic curve (AUC) [32]. The hyperparameters were optimized through a random search over 500 iterations to maximize the validation set AUC. To ensure robust evaluation, we repeated this process 10 times with different random seeds and reported the average AUC across these iterations for each model.

2.5. Statistical analysis and software

For multiple group comparisons, we adjusted the significance level using the Holm method, with the base significance level set at 5%. Statistical analyses were performed using *SciPy* [33]. We used *SciPy* to fit the Student's t-distribution, generalized Gaussian distribution, Laplace distribution, Cauchy distribution, and Gaussian distribution, while Lévy stable distributions were fitted using *Pylevy* [34]. For analyses with w_{RRI} less than 18 h, we included only those cases where *Pylevy*'s parameter optimization successfully converged for all distributions. The classification analyses were implemented using *scikit-learn* [35].

3. Results

3.1. Goodness-of-fit of a Gaussian scale mixture representation for RRI data

Fig. 3(a) and (b) present representative examples of sRRI waveforms, their corresponding histograms, and fitted distributions in the LF and HF bands, respectively. Notable sudden changes in variance were observed in both LF and HF bands for adult patients and in the LF band for infants, suggesting enhanced non-Gaussian characteristics in these distributions.

Fig. 4(a) and (b) compare the L^2 norm across different distributions in the LF and HF bands of 18-h sRRI (w_{RRI} = 18 h), respectively. Among all tested distributions, the Student's t-distribution exhibited the lowest L^2 norm, indicating the best fit, while the Gaussian distribution showed the highest L^2 norm, suggesting the poorest fit. The L^2 norm of the generalized Gaussian distribution was significantly lower than that of the Lévy stable, Laplace, Cauchy, and Gaussian distributions for the LF band.

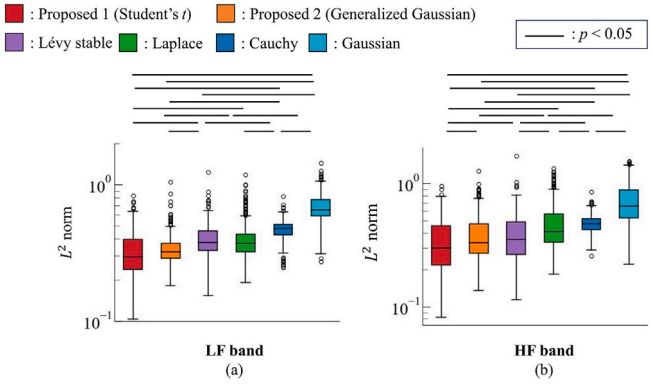


Fig. 4. Comparison of L^2 norms across probability distributions for 18-h RRI recordings. (a) LF band. (b) HF band.

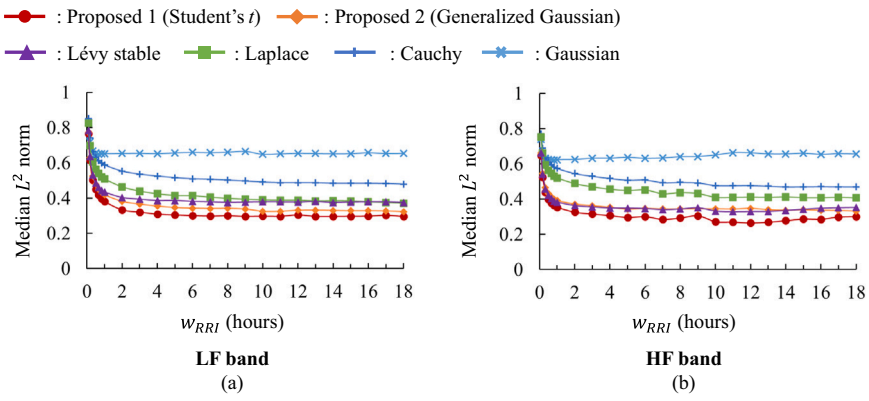


Fig. 5. Effect of window length (w_{RRI}) on L^2 norms for different probability distributions. (a) LF band. (b) HF band.

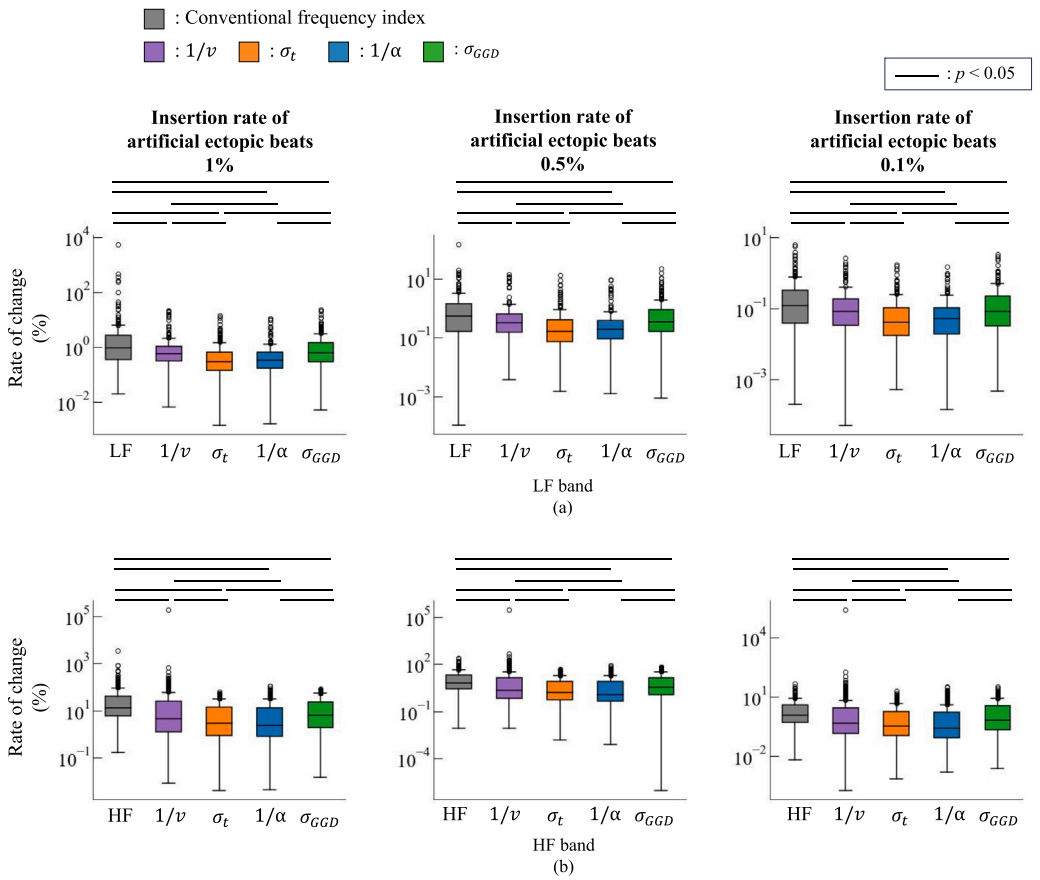


Fig. 6. Effects of artificial ectopic beat insertion on the absolute rate of change in specific parameters. (a) LF band. (b) HF band.

Fig. 5(a) and (b) show the relationship between the median L^2 norm and w_{RRI} for each probability distribution in the LF and HF bands of sRRI, respectively. The median L^2 norm of the Student's t-distribution was the lowest, regardless of the value of w_{RRI}.

3.2. Robustness of RRI non-Gaussianity indices

Fig. 6(a) and (b) show the absolute rate of change due to artificial ectopic beats in the LF and HF bands of sRRI, respectively. The non-Gaussianity indices calculated by the proposed method had a significantly smaller absolute rate of change than the conventional frequency index for the same frequency band. This trend persisted even as the insertion rate of artificial ectopic beats increased from 0.1% to 1%.

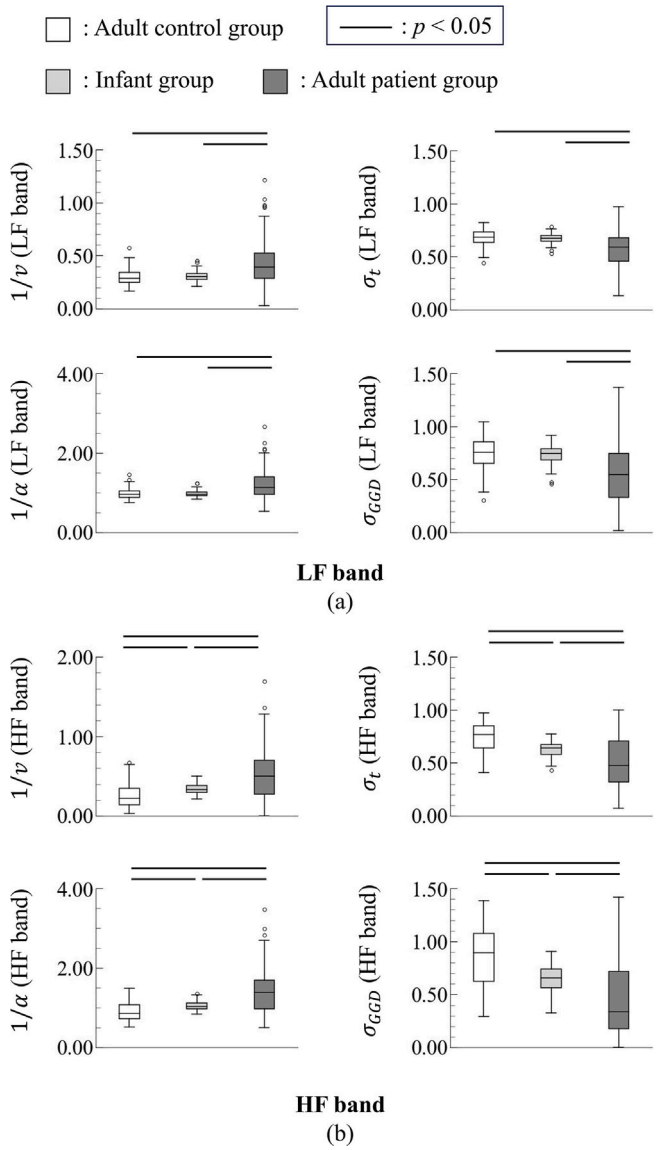


Fig. 7. Comparison of non-Gaussianity indices among the adult control group, infant group, and adult patient group. (a) LF band. (b) HF band.

3.3. Ability of non-Gaussianity indices to identify cardiovascular disease

Fig. 7(a) and (b) show the non-Gaussianity indices in the LF and HF bands, respectively, among the adult control group, infant group, and adult patient group. Statistical analysis revealed significant differences between the adult patient group and the other groups across all parameters, indicating distinctive non-Gaussian characteristics in patients with cardiovascular diseases. Specifically, the distribution of both the adult patient group and the infant group exhibited greater non-Gaussianity ($1/\nu$, $1/\alpha$) and narrower width (σ_t, σ_{GGD}) in both the LF and HF bands compared to that of the adult control group. While the adult patient group showed significant differences from adult controls in both the LF and HF bands, the infant group showed significant differences from adult controls only in the HF band.

The correlation analysis revealed significant relationships between our non-Gaussianity indices and conventional HRV measures (Table 2). Specifically, $1/\nu$ showed moderate positive correlations with RRI standard deviation in both frequency bands (LF: $r = 0.474$, HF: $r = 0.498$; $p < 0.05$) and negative correlations with LF/HF ratio (LF: $r = -0.285$, HF: $r = -0.372$; $p < 0.05$). These relationships suggest that increased non-Gaussianity is associated with both higher overall variability and reduced sympathovagal balance.

Table 3 reports the AUC of various classifiers in distinguishing an adult control group from an adult patient group based on different input features. The addition of our non-Gaussianity indices to conventional indices consistently improved the AUC across all classifiers, demonstrating the complementary value of non-Gaussianity measures in distinguishing between control and patient groups.

Table 2
The correlation coefficients between the non-Gaussianity indices of Conventional-set and t-set.

Conventional HRV features	$1/\nu$ (LF band)	$1/\nu$ (HF band)	σ_t (LF band)	σ_t (HF band)
The average of the RRIs	−0.090	−0.128	0.056	0.068
The standard deviation of the RRIs	0.474*	0.498*	−0.440*	−0.448*
LF/HF	−0.285*	−0.372*	0.298*	0.372*

* $p < 0.05$

Table 3
Classification performance (AUC) for cardiovascular disease detection using different feature sets and classifiers. Standard deviations are in parentheses; **bold** and underlined values indicate the highest and second-highest AUC, respectively.

Input feature	Classifier				Average
	LDA	LR	SVM	RF	
Conventional-set (baseline)	0.717 (0.007)	0.715 (0.018)	0.845 (0.006)	0.853 (0.017)	0.782
t-set	0.664 (0.019)	0.633 (0.029)	0.628 (0.023)	0.625 (0.017)	0.637
GGD-set	0.676 (0.020)	0.628 (0.024)	0.615 (0.036)	0.621 (0.012)	0.635
Conventional & t-set	0.791 (0.004)	0.774 (0.038)	**0.869 (0.027)**	**0.868 (0.014)**	0.825
Conventional & GGD-set	0.789 (0.008)	0.764 (0.028)	0.855 (0.013)	0.866 (0.015)	0.818
Conventional & t & GGD-set	**0.797 (0.010)**	**0.821 (0.017)**	0.860 (0.020)	0.862 (0.025)	**0.835**

LDA = linear discriminant analysis; LR = logistic regression; SVM = support vector machine; RF = random forest.

Since classification tasks can be sensitive to class imbalance, we ensured methodological robustness by implementing stratified cross-validation and AUC metrics. The consistent performance across multiple classifiers suggests that our approach effectively mitigates potential biases caused by the uneven distribution of class sizes.

4. Discussion

This study demonstrates that the Gaussian scale mixture representation provides an effective framework for characterizing RRI non-Gaussianity, with several key findings. First, we showed that RRI distributions are consistently non-Gaussian across different subject groups. Second, we found that the Student's t-distribution, a specific case of Gaussian scale mixture representation, provides the best fit for RRI data. Finally, we demonstrated that non-Gaussianity indices derived from this framework can enhance the detection of cardiovascular diseases.

The Student's t-distribution, generalized Gaussian distribution, and Lévy stable distribution showed superior goodness-of-fit compared to other tested distributions (Figs. 4 and 5). These distributions share an important mathematical property: they include simpler distributions as special cases—the Student's t-distribution encompasses both Gaussian and Cauchy distributions, the generalized Gaussian distribution includes Laplace and Gaussian distributions, and the Lévy stable distribution incorporates both Gaussian and Cauchy distributions. This flexibility in distribution shape likely contributes to their better fitting performance. The consistently poor fit of the Gaussian distribution across both LF and HF bands provides strong evidence for the non-Gaussian nature of RRI signals.

Among these distributions, the Gaussian scale mixture representation—specifically Student's t-distribution and generalized Gaussian distribution—demonstrated a better fit to the RRI data than the Lévy stable distribution. The Student's t-distribution, in particular, achieved the lowest L^2 norm across all conditions (w_{RRI} and frequency band). This superior performance can be attributed to its ability to represent heavy-tailed distributions with infinite kurtosis when the degrees of freedom ν is less than four. While the generalized Gaussian distribution can model both heavier and lighter tails than the Gaussian distribution (depending on whether α is less than or greater than two), the absence of lighter-tailed distributions in our RRI data likely explains the better performance of the Student's t-distribution.

The results indicate that the proposed method's fitting accuracy decreases significantly when the length of the RRI data is less than one hour (Fig. 5). Therefore, to effectively apply the method, it is recommended that RRI data samples be at least one hour long to achieve a stable assessment of non-Gaussianity. However, increasing the analysis window size beyond this threshold is not necessarily desirable, as it could alter the time scale of the captured physiological phenomena. The appropriate window length should be carefully considered based on the specific application.

The proposed non-Gaussianity indices demonstrated greater robustness to artificial ectopic beats compared to conventional frequency indices (*LF* and *HF*) (Fig. 6). This difference may be attributed to the distinct nature of these measures: frequency indices inherently depend on the temporal sequence of the data, while our non-Gaussianity indices characterize the overall distribution of values given a sufficient number of RRI samples. This property makes our approach particularly suitable for long-term RRI analysis, where artifacts and ectopic beats are common.

The analysis revealed distinctive non-Gaussian characteristics in both frequency bands across different groups. Particularly notable was the pronounced non-Gaussianity in the LF band for patients with cardiovascular conditions (congestive heart failure or hypertension) and in the HF band for both patient and infant groups (Fig. 7). The non-Gaussianity indices showed significant correlations with both the standard deviation of RRIs and LF/HF ratio (Table 2). However, the moderate strength of these correlations, combined with the improved classification performance when combining conventional and non-Gaussianity indices (Table 3), indicates that

our approach provides complementary information beyond traditional HRV measures. These results suggest that non-Gaussianity indices capture distinct statistical properties of RRI variations that are not fully characterized by conventional time- and frequency-domain analyses. While these statistical patterns appear to provide complementary information about RRI dynamics, their precise physiological interpretation requires further investigation.

5. Conclusion and limitations

This study examined the application of Gaussian scale mixture representation for evaluating RRI non-Gaussianity. The Gaussian scale mixture representation, particularly using the Student's t-distribution, effectively characterizes RRI distributions and outperforms conventional models in goodness-of-fit. Our approach enables robust quantification of non-Gaussian characteristics through well-defined distributional parameters and demonstrates potential utility in cardiovascular disease detection, especially when combined with conventional HRV indices. Future studies should focus on investigating intra-individual variations in RRI non-Gaussianity under controlled measurement conditions. Such investigations would help validate the clinical applicability of this approach and potentially lead to improved cardiovascular risk assessment tools.

Several limitations of this study warrant consideration. First, the proposed method was applied to RRIs from participants with heart failure and hypertension, and future studies should explore its generalizability across a wider range of cardiovascular conditions. Additionally, analyzing RRIs measured in environments with noise and motion artifacts may lead to non-Gaussian assessments that deviate from actual physiological changes. RRI measurements from wearable devices are particularly susceptible to such issues [36]. While we employed zero-phase filtering to eliminate phase distortion, the frequency band separation process can still introduce amplitude attenuation, particularly at signal edges, which may affect the evaluation of non-Gaussianity.

Second, this study achieved an AUC value of 0.835 for classifying cardiovascular patients, which indicates moderate accuracy [37, 38]. While this represents an improvement over conventional methods, the analysis was conducted using existing databases rather than prospectively collected clinical data. Future studies should focus on collecting more controlled data and conducting prospective validation to confirm the clinical utility of these non-Gaussian indices.

Third, we evaluated RRI non-Gaussianity using a Gaussian scale mixture representation that assumes symmetry in the amplitude distribution. However, several studies have reported that the distribution of RRI is asymmetrical [39,40]. Moreover, while our model captures cardiac activity phenomenologically, it does not explicitly incorporate underlying physiological mechanisms. Future work should focus on developing models that can account for both asymmetrical distributions and physiological mechanisms while maintaining the analytical advantages of the scale mixture approach.

CRediT authorship contribution statement

Naoki Hagiyama: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Akira Furui:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization. **Harutoyo Hirano:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Toshio Tsuji:** Writing – review & editing, Supervision, Resources, Project administration, Methodology.

Funding statement

This work was supported in part by JSPS KAKENHI Grant Number JP22H00197 and JP23K17321.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Toshio Tsuji reports that financial support was provided by JSPS KAKENHI. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data for this article can be found online at doi:10.1016/j.heliyon.2026.e44737.

Data availability statement

The authors used publicly available databases. All databases can be found here: https://physionet.org. The data and code used in the experiment to evaluate the ability of non-Gaussianity indices to identify cardiovascular diseases are provided in the Appendix.

References

[1] R. Gordan, J.K. Gwathmey, L.-H. Xie, Autonomic and endocrine control of cardiovascular function, World J. Cardiol. 7 (2015) 204–214, https://doi.org/10.4330/wjc.v7.i4.204

[2] J. Bolea, P. Laguna, J.M. Remartínez, E. Rovira, A. Navarro, R. Bailón, Methodological framework for estimating the correlation dimension in HRV signals, Comput. Math. Methods Med. 2014 (2014) 129248, https://doi.org/10.1155/2014/129248

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

[4] S. Akselrod, D. Gordon, F.A. Ubel, D.C. Shannon, A.C. Berger, R.J. Cohen, Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control, Science 213 (1981) 220–222, https://doi.org/10.1126/science.6166045

[5] B. Pomeranz, R.J. Macaulay, M.A. Caudill, I. Kutz, D. Adam, D. Gordon, K.M. Kilborn, A.C. Barger, D.C. Shannon, R.J. Cohen, et al., Assessment of autonomic function in humans by heart rate spectral analysis, Am. J. Physiol.-Heart Circ. Physiol. 248 (1985) H151–H153, https://doi.org/10.1152/ajpheart.1985.248.1.H151

[6] G.G. Berntson, J.T. Cacioppo, K.S. Quigley, Respiratory sinus arrhythmia: autonomic origins, physiological mechanisms, and psychophysiological implications, Psychophysiology 30 (1993) 183–196, https://doi.org/10.1111/j.1469-8986.1993.tb01731.x

[7] C.-K. Peng, J. Mietus, J.M. Hausdorff, S. Havlin, H.E. Stanley, A.L. Goldberger, Long-range anticorrelations and non-gaussian behavior of the heartbeat, Phys. Rev. Lett. 70 (1993) 1343–1346, https://doi.org/10.1103/physrevlett.70.1343

[8] B.L. Lan, M. Toda, Fluctuations of healthy and unhealthy heartbeat intervals, Europhys. Lett. 102 (2013) 18002, https://doi.org/10.1209/0295-5075/102/18002

[9] K. Kiyono, J. Hayano, E. Watanabe, Z.R. Struzik, Y. Yamamoto, Non-Gaussian heart rate as an independent predictor of mortality in patients with chronic heart failure, Heart Rhythm 5 (2008) 261–268, https://doi.org/10.1016/j.hrthm.2007.10.030

[10] J. Hayano, E. Yuda, Assessment of autonomic function by long-term heart rate variability: beyond the classical framework of lf and hf measurements, J. Physiol. Anthropol. 40 (2021) 21, https://doi.org/10.1186/s40101-021-00272-y

[11] S. Gouveia, M. Scotto, On modelling RR tails in heart rate variability studies: an extreme value analysis, in: 2015 Computing in Cardiology Conference (CinC), 2015, pp. 777–780, https://doi.org/10.1109/CIC.2015.7411026

[12] J. Hayano, K. Kiyono, Z.R. Struzik, Y. Yamamoto, E. Watanabe, P.K. Stein, L.L. Watkins, J.A. Blumenthal, R.M. Carney, Increased non-Gaussianity of heart rate variability predicts cardiac mortality after an acute myocardial infarction, Front. Physiol. 2 (2011), https://doi.org/10.3389/fphys.2011.00065

[13] J. Hayano, N. Ueda, M. Kisohara, E. Yuda, R.M. Carney, J.A. Blumenthal, Survival predictors of heart rate variability after myocardial infarction with and without low left ventricular ejection fraction, Front. Neurosci. 15 (2021), https://doi.org/10.3389/fnins.2021.610955

[14] B.J. West, M. Turalska, Hypothetical control of heart rate variability, Front. Physiol. 10 (2019), https://doi.org/10.3389/fphys.2019.01078

[15] T.T. Zin, P. Tin, Markov chain modelling for heart rate variability analysis: bridging artificial intelligence and physiological data, in: 2023 IEEE 13th International Conference on Consumer Electronics - Berlin (ICCE-Berlin), 2023, pp. 163–166, https://doi.org/10.1109/ICCE-Berlin58801.2023.10375625

[16] D.F. Andrews, C.L. Mallows, Scale mixtures of normal distributions, J. R. Stat. Soc. Ser. B 36 (2018) 99–102, https://doi.org/10.1111/j.2517-6161.1974.tb00989.x

[17] M. West, On scale mixtures of normal distributions, Biometrika 74 (1987) 646–648, https://doi.org/10.1093/biomet/74.3.646

[18] A. Furui, H. Hayashi, T. Tsuji, A scale mixture-based stochastic model of surface EMG signals with variable variances, IEEE Trans. Biomed. Eng. 66 (2019) 2780–2788, https://doi.org/10.1109/TBME.2019.2895683

[19] A. Furui, R. Onishi, A. Takeuchi, T. Akiyama, T. Tsuji, Non-gaussianity detection of EEG signals based on a multivariate scale mixture model for diagnosis of epileptic seizures, IEEE Trans. Biomed. Eng. 68 (2021) 515–525, https://doi.org/10.1109/TBME.2020.3006246

[20] S. Fukuda, A. Furui, M.G. Machizawa, T. Tsuji, Stochastic fluctuation in EEG evaluated via scale mixture model for decoding emotional valence, in: 2024 IEEE/SICE International Symposium on System Integration (SII), IEEE, 2024, pp. 567–572, https://doi.org/10.1109/SII58957.2024.10417430

[21] Student, The probable error of a mean, Biometrika 6 (1908) 1–25, https://doi.org/10.2307/2331554

[22] S.H.T. Lihn, A theory of asset return and volatility under stable law and stable lambda distribution (November 15, 2017), 2017, Quantitative Finance, Forthcoming, Available at SSRN: https://ssrn.com/abstract=3046732.

[23] T. Gneiting, Normal scale mixtures and dual probability densities, Am. J. Physiol.-Heart Circ. Physiol. 59 (1997) 375–384, https://doi.org/10.1080/00949659708811867

[24] J.A. Palmer, K. Kreutz-Delgado, B.D. Rao, S. Makeig, Modeling and estimation of dependent subspaces with non-radially symmetric and skewed densities, in: Independent Component Analysis and Signal Separation: 7th International Conference, ICA 2007, London, UK, September 9–12, 2007. Proceedings 7, Springer, 2007, pp. 97–104, https://doi.org/10.1007/978-3-540-74494-8_13

[25] R.L. Goldsmith, J.T. Bigger Jr, R.C. Steinman, J.L. Fleiss, Comparison of 24-hour parasympathetic activity in endurance-trained and untrained young men, J. Am. Coll. Cardiol. 20 (1992) 552–558, https://doi.org/10.1016/0735-1097(92)90007-a

[26] I.M. Irurzun, L. Garavaglia, M.M. Defeo, T. Mailland, RR interval time series from healthy subjects (version 1.0.0), PhysioNet (2021), https://doi.org/10.13026/51yd-d219

[27] L. Garavaglia, D. Gulich, M.M. Defeo, J. Thomas Mailland, I.M. Irurzun, The effect of age on the heart rate variability of healthy subjects, PLoS ONE 16 (2021) e0255894, https://doi.org/10.1371/journal.pone.0255894

[28] D.S. Baim, W.S. Colucci, E.S. Monrad, H.S. Smith, R.F. Wright, A. Lanoue, D.F. Gauthier, B.J. Ransil, W. Grossman, E. Braunwald, Survival of patients with severe congestive heart failure treated with oral milrinone, J. Am. Coll. Cardiol. 7 (1986) 661–670, https://doi.org/10.1016/s0735-1097(86)80478-8

[29] P. Melillo, R. Izzo, A. Orrico, P. Scala, M. Attanasio, M. Mirra, N. De Luca, L. Pecchia, Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis, PLoS ONE 10 (2015) e0118504, https://doi.org/10.1371/journal.pone.0118504

[30] P. Welch, The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, IEEE Trans. Audio Electroacoust. 15 (1967) 70–73, https://doi.org/10.1109/TAU.1967.1161901

[31] E. Brunner, U. Munzel, The nonparametric Behrens-Fisher problem: asymptotic theory and a small-sample approximation, Biom. J. 42 (2000) 17–25, https://doi.org/10.1002/(SICI)1521-4036(200001)42:1<17::AID-BIMJ17>3.0.CO;2-U

[32] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recognit. 30 (1997) 1145–1159, https://doi.org/10.1016/S0031-3203(96)00142-2

[33] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 contributors, SciPy 1.0: fundamental algorithms for scientific computing in Python, Nature Methods 17 (2020) 261–272, https://doi.org/10.1038/s41592-019-0686-2

[34] J.M. Miotto, Module levy [software], 2020 https://pylevy.readthedocs.io/en/latest/levy.html (Accessed 17 July 2023).

[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[36] K. Eguchi, R. Aoki, S. Shimauchi, K. Yoshida, T. Yamada, R-R interval outlier processing for heart rate variability analysis using wearable ECG devices, Adv. Biomed. Eng. 7 (2018) 28–38, https://doi.org/10.14326/abe.7.28

[37] J.A. Swets, Measuring the accuracy of diagnostic systems, Science 240 (1988) 1285–1293, https://doi.org/10.1126/science.3287615

[38] J.E. Fischer, L.M. Bachmann, R. Jaeschke, A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis, Intens. Care Med. 29 (2003) 1043–1051, https://doi.org/10.1007/s00134-003-1761-8

[39] P. Guzik, J. Piskorski, T. Krauze, A. Wykretowicz, H. Wysocki, Heart rate asymmetry by Poincaré plots of RR intervals, Biomed. Eng. Biomed. Tech. 51 (2006) 272–275, https://doi.org/10.1515/BMT.2006.054

[40] A. Rohila, A. Sharma, Asymmetric spread of heart rate variability, Biomed. Signal Process. Control. 60 (2020) 101985, https://doi.org/10.1016/j.bspc.2020.101985