Journal of Health Informatics

J Health Inform. 2026;18:e1811
DOI: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1811

ARTIGO COMPLETO
CBIS 2026

Lightweight clinical decision support for premature ventricular contraction detection using unscented Kalman filter

Suporte leve à decisão clínica para detecção de contrações ventriculares prematuras usando filtro de Kalman unscented

Lucas Kenzo Araki[1], Renata Coelho Borges[2] ✉, Victor Baptista Frencl[1]

ORCID: 0009-0001-6574-1980 · 0000-0003-3259-5646 · 0000-0002-8134-2381

[1] Academic Department of Electrotechnics (DAELT-CT), Federal University of Technology – Paraná (UTFPR), Curitiba (PR), Brazil.

[2] Academic Department of Electrotechnics (DAELT-CT), Federal University of Technology – Paraná (UTFPR), Curitiba (PR), Brazil; Graduate Program in Biomedical Engineering (PPGEB-CT), Federal University of Technology – Paraná (UTFPR), Curitiba (PR), Brazil.

ARTICLE INFORMATION

TYPE
Artigo completo · CBIS 2026

IDENTIFIER
e1811

DOI
10.59681/2175-4411.v18.iEspecial.2026.1811

RECEIVED
17 May 2026

ACCEPTED
20 July 2026

REVIEW PROCESS
Double Blind Review

HANDLING EDITOR
Juliano de Souza Gaspar

ACCESS
Open, CC BY 4.0

CORRESPONDING AUTHOR
Renata Coelho Borges
renatacoelho@utfpr.edu.br

LICENSE

This is an open access article published under the Creative Commons CC BY 4.0 license, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited.

CC BY 4.0

ABSTRACT

Objectives: This study proposes a lightweight and interpretable pre-classification framework for automated detection of Premature Ventricular Contractions (PVC) in ambulatory Electrocardiogram (ECG) monitoring. **Methods:** The approach leverages an Unscented Kalman Filter (UKF) combined with a cylindrical dynamic model, called Cylindrical UKF (CUKF) for ECG. Deviations from normal sinus rhythm are quantified via the innovation residual, measured using the Normalized Mean Square Error (NMSE), serving as a digital biomarker. The model was evaluated using the MIT-BIH Arrhythmia Database, with adaptive thresholds determined through 10-fold cross-validation. **Results:** The proposed method achieved a Macro F1-score of 0.8924 ± 0.0160, with average sensitivity of 85.83% ± 3.44% and specificity above 92%. **Conclusion:** This framework offers an efficient alternative to data-intensive approaches, requiring low computational resources and linear time complexity. Its design makes it suitable for real-time deployment in wearable devices and as a pre-screening module within digital health systems, supporting clinical decision-making in large-scale ECG monitoring.

Keywords: Decision Support Systems, Clinical; Digital Health; Arrhythmias, Cardiac.

RESUMO

Objetivos: Este estudo propõe uma estrutura leve e interpretável de pré-classificação para a detecção automatizada de Contrações Ventriculares Prematuras (PVC) no monitoramento ambulatorial por Eletrocardiograma (ECG). **Métodos:** A abordagem utiliza um Filtro de Kalman Unscented (UKF) combinado com um modelo dinâmico cilíndrico, denominado UKF Cilíndrico (CUKF) para ECG. Os desvios em relação ao ritmo sinusal normal são quantificados pelo resíduo de inovação, medido pelo Erro Quadrático Médio Normalizado (NMSE), que atua como biomarcador digital. O modelo foi avaliado com a base MIT-BIH Arrhythmia Database, com limiares adaptativos determinados por validação cruzada com 10 partições. **Resultados:** O método proposto alcançou um F1-score macro de 0,8924 ± 0,0160, com sensibilidade média de 85,83% ± 3,44% e especificidade superior a 92%. **Conclusão:** Esta estrutura oferece uma alternativa eficiente às abordagens intensivas em dados, exigindo poucos recursos computacionais e complexidade de tempo linear. Seu desenho a torna adequada para implantação em tempo real em dispositivos vestíveis e como módulo de pré-triagem em sistemas de saúde digital, apoiando a tomada de decisão clínica no monitoramento de ECG em larga escala.

Descritores: Sistemas de Apoio a Decisões Clínicas; Saúde Digital; Arritmias Cardíacas.

CBIS 2026 XXI CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE

Paper presented at the 21st Brazilian Congress of Health Informatics (CBIS 2026), held from September 23 to 25, 2026, in Brasília (DF), Brazil, and published in the CBIS 2026 section of the Journal of Health Informatics.

HOW TO CITE THIS ARTICLE: Araki LK, Borges RC, Frencl VB. Lightweight clinical decision support for premature ventricular contraction detection using unscented Kalman filter. J Health Inform 2026;18:e1811. doi: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1811

1 INTRODUCTION

Cardiovascular Diseases (CVD) remain the leading cause of mortality worldwide, driving the need for scalable digital health solutions [1,2]. Long-term ambulatory monitoring via Holter recordings is essential for detecting intermittent arrhythmias, yet the massive volume of data, approximately 100,000 cycles per 24 hours, makes manual analysis impractical. This challenge highlights the necessity for automated Clinical Decision Support Systems (CDSS) to efficiently identify abnormal cardiac events in large-scale monitoring.

Among the arrhythmias frequently detected in ambulatory ECG, Premature Ventricular Contractions (PVC) are of particular clinical concern because a high burden of these ectopic beats significantly increases the risk of cardiomyopathies and other adverse outcomes [3,4]. Given the need to accurately quantify this frequency, automated pre-classification systems become essential to streamline clinical workflows, reducing physician fatigue by reliably flagging suspicious segments for targeted review.

While automated arrhythmia detection has been largely dominated by deep learning approaches [5,6], they present limitations for widespread ambulatory deployment: computational resources, which restrict their implementation in low-power wearable devices and embedded monitoring systems [7].

To address these limitations, this work explores a stochastic filtering-based pre-classification strategy. The central idea is that the innovation residual, defined as the difference between the predicted and the measured ECG, can indicate abnormal cardiac activity. Under normal sinus rhythm, the ECG follows a quasi-periodic pattern that can be accurately tracked by a Kalman Filter (KF) [8,9]. In contrast, irregular events introduce deviations that increase the innovation residual.

Based on this principle, we propose the use of a nonlinear KF, called Unscented Kalman Filter (UKF), combined with a cylindrical dynamical model, called Cylindrical Unscented Kalman Filter (CUKF), to perform both signal tracking and abnormal beat detection. The model parameters were tuned for PVC detection. Although the innovation residual is sensitive to deviations from normal rhythm, its performance for other arrhythmias was not evaluated and may require further tuning. The method is designed as a lightweight and interpretable pre-screening module for automated ECG analysis, enabling efficient detection of abnormal segments in wearable and clinical applications.

The main contributions of this work are: (i) a cylindrical ECG model integrated into a nonlinear stochastic filtering framework; (ii) a lightweight PVC pre-classification framework based on CUKF innovation residuals; (iii) the use of the Normalized Mean Square Error (NMSE) as an interpretable digital biomarker; (iv) validation on the MIT-BIH dataset demonstrating competitive performance with low computational cost.

2 ECG MATHEMATICAL MODELING

Accurate cardiac signal tracking using a stochastic filter requires a mathematical model describing the expected dynamics of a normal ECG (sinus rhythm). The synthetic modeling of the ECG traditionally relies on a system of differential equations describing a nonlinear oscillator in Cartesian coordinates $x \in \mathbb{R}$, $y \in \mathbb{R}$, and $z \in \mathbb{R}$. In this classical model proposed by McSharry *et al.* [10], movement in a two-dimensional plane generates a limit cycle, while the variable z modulates physiological amplitude to form the P, Q, R, S, and T waves of the cardiac cycle through the sum of Gaussian functions.

However, the Cartesian representation possesses intrinsic limitations when subjected to recursive stochastic filtering processes, as it couples angular dynamics (e.g., the phase of the heartbeat) with its geometric amplitude. In order to avoid chronic tracking errors, a dynamic model based on a cylindrical representation was adopted in Sameni *et al.* [11]. The transformation to cylindrical coordinates $r \in \mathbb{R}^+$, $\theta \in [0; 2\pi]$, and $z \in \mathbb{R}$ allows a more compact representation of system dynamics, focusing strictly on phase θ and amplitude z states, ensuring a superior morphological stability. Also, Sameni *et al.* [11] state that the dynamic equation related to the r variable could be removed because it is redundant for the stochastic filter estimation. Therefore, the dynamics can be represented by Equation (a).

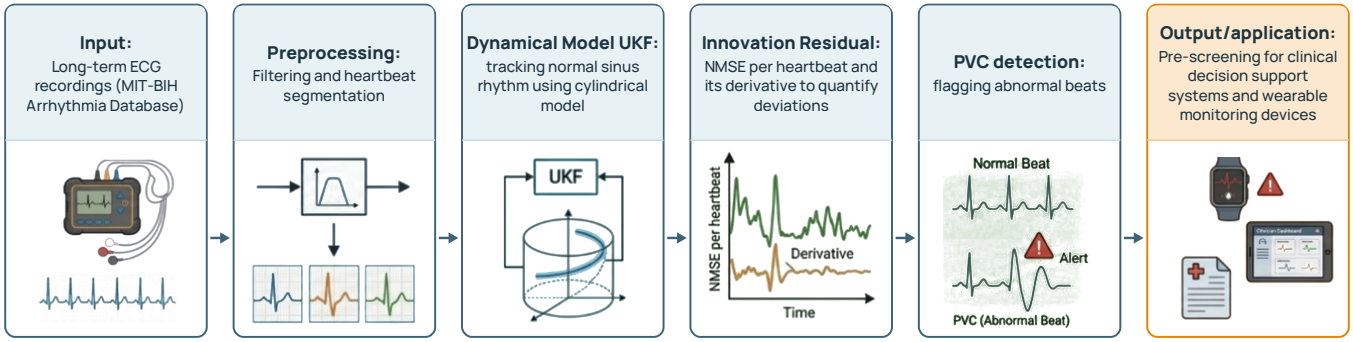
$$\begin{aligned} \dot{\theta}(t) &= \omega, \\ \dot{z}(t) &= -\sum_{i \in \{P,Q,R,S,T\}} \frac{\alpha_i \omega}{b_i^2} \Delta\theta_i(t) \exp\left(-\frac{\Delta\theta_i^2(t)}{2b_i^2}\right) + (z_0 - z(t)), \end{aligned} \tag{a}$$

where $\omega \in \mathbb{R}$ is the angular velocity related to the instantaneous heart rate, $b_i \in \mathbb{R}^+$ denotes the width of each wave, $\Delta\theta_i(t) \in [0; 2\pi]$ is the angular deviation of the trajectory relative to the centers of the Gaussian waves, $\alpha_i \in \mathbb{R}^+$ is the peak amplitude coefficients of the Gaussian functions that model the P, Q, R, S and T components of the ECG. Also, the term $z_0 - z(t)$ refers to the baseline fluctuation [12].

2.1 Mathematical Model Discretization

Most stochastic filters are formulated based on discrete-time mathematical models. The cylindrical dynamic model in Equation (a) is continuous-time based and clearly nonlinear. Furthermore, it is necessary to discretize the mathematical model. This process is performed based on the Euler method, which gives an approximation of the first-order time

Figure 1. Detailed flowchart illustrating the processing pipeline for the PVC pre-classifier.



Source: Prepared by the authors (2026).

Table 1. Simulation parameters for the ECG dynamical model and UKF tuning.

	Parameter	Value
ECG Model Parameters	Peak amplitude (a_i) [mV]	{0.25, -0.15, 1.8, -0.2, 0.5}
	Angular position (θ_i) [rad]	{-π/3, -π/12, 0, π/12, π/2}
	Wave width (b_i) [rad]	{0.25, 0.1, 0.1, 0.1, 0.4}
UKF Tuning Parameters	Sampling period (T)	1/360 s
	Spread of sigma points (α)	10^{-3}
	Prior distribution knowledge (β)	2
	Secondary scaling parameter (p)	0
	Process noise covariance (Q)	diag $(1\times10^{-3}, 1\times10^{-4})$
	Measurement noise covariance (R)	diag $((\omega T)^2/12, 1\times10^{-2})$

Source: Prepared by the authors (2026).

derivative of a continuous-time function.

Applying the Euler method to the cylindrical dynamic model and considering the inclusion of additive stochastic noises to each equation, the equivalent discrete-time mathematical model is obtained as:

$$\theta(k+1) = [\theta(k) + \omega T] \bmod (2\pi) + \nu(k),$$
$$z(k+1) = z(k) - \sum_{i\in\{P,Q,R,S,T\}} T\frac{\alpha_i\omega}{b_i^2}\Delta\theta_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) + \eta(k). \quad \text{(b)}$$

where $\nu(k)$ and $\eta(k)$ are the process stochastic noises related to $\theta(k)$ and $z(k)$, respectively; the baseline fluctuation term is no longer present because it was "absorbed" by $\eta(k)$, modeling physiological imperfections, systemic disturbances from the acquisition sensor and the baseline fluctuation.

Considering Equation (b), we can write this model in a discrete-time state space representation, which is necessary for the stochastic filter formulation. The state equation can be written as:

$$\mathbf{x}(k+1) = \begin{bmatrix} [\theta(k)+\omega T] \bmod (2\pi) \\ z(k) - \sum_{i\in\{P,Q,R,S,T\}} T\frac{\alpha_i\omega}{b_i^2}\Delta\theta_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) \end{bmatrix} + \mathbf{w}(k), \quad \text{(c)}$$

where $\mathbf{x}(k) = [\theta(k)\ z(k)]^\intercal$ is the state vector, $\mathbf{w}(k) = [\nu(k)\ \eta(k)]^\intercal$ is the Gaussian process noise vector with zero mean and covariance matrix $Q(k)$, i.e. $\mathbf{w}(k) \sim \mathcal{N}(0, Q(k))$. The measurement equation can be written as:

$$\mathbf{y}(k) = \begin{bmatrix} \theta(k) \\ z(k) \end{bmatrix} + \mathbf{v}(k), \quad \text{(d)}$$

where $\mathbf{y}(k) = [\phi(k)\ s(k)]^\intercal$ is the measurement vector, $\phi(k)$ is the phase measurement, $s(k)$ is the ECG amplitude measurement, $\mathbf{v}(k) = [\nu(k)\ \eta(k)]^\intercal$ is the Gaussian measurement noise vector with zero mean and covariance matrix $R(k)$, i.e. $\mathbf{v}(k) \sim \mathcal{N}(0, R(k))$.

3 UNSCENTED KALMAN FILTER

Given the strong nonlinearity of the exponential summations composing the cylindrical dynamic model, Equation (b), the application of a linear stochastic filter, such as the KF [13], becomes mathematically unfeasible. In order to deal with those nonlinearities, the Extended Kalman Filter (EKF) [14] is commonly used, promoting a local linearization of the nonlinear model via First-Order Taylor Series (FOTS) expansion. However, this type of linearization frequently fails to capture the ultra-fast transitions of the QRS complex, and depending on the order of the system, obtaining analytical Jacobian matrices consumes substantial processing power.

To resolve the conflict between numerical precision and computational cost, the UKF is introduced [15]. The UKF summarily avoids the use of FOTS expansions and Jacobian matrices by applying the Unscented

Transform (UT) [15,16]. The UT deterministically selects a minimal set of statistical points, called sigma-points, which captures the true mean and true covariance of the current state, instead of approximated values of mean and covariance. Like other stochastic filters, the UKF has two fundamental steps: the prediction step and the update step.

3.1 UKF Prediction Step

Consider a state $\mathbf{x}(k) \in \mathbb{R}^n$ and its estimate $\hat{\mathbf{x}}(k|k) \in \mathbb{R}^n$ based on measurements $Z^k = \{\mathbf{z}(0), ..., \mathbf{z}(k)\}$. Let $\mathbf{f}$: $\mathbb{R}^n \to \mathbb{R}^n$ be nonlinear with $\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k))$. First, in the UKF Prediction Step, a set of $2n+1$ sigma-points $\mathcal{X}(k|k)$ is created, given the updated state vector and covariance matrix, both evaluated at time instant k.

$$\begin{aligned}\hat{\mathcal{X}}(k|k)^{(i)} &= \hat{\mathbf{x}}(k|k) + \left(\sqrt{(n+\lambda)P(k|k)}\right)_i^\top, \ i = 1, \ldots, n,\\ \hat{\mathcal{X}}(k|k)^{(n+i)} &= \hat{\mathbf{x}}(k|k) - \left(\sqrt{(n+\lambda)P(k|k)}\right)_i^\top, \ i = 1, \ldots, n,\end{aligned} \tag{e}$$

where $P(k|k)$ is the estimation error covariance matrix, λ is a parameter that reduces the approximation errors related to high order terms, and $(\cdot)^\top_i$ indicates the i-th column of the matrix. Next, $\mathbf{f}$ is applied to each sigma-point, resulting in a new set of predicted sigma-points, given by:

$$\hat{\mathcal{X}}(k+1|k)^{(i)} = \mathbf{f}(\hat{\mathcal{X}}(k)^{(i)}). \tag{f}$$

After that, the predicted state is obtained as follows:

$$\hat{\mathbf{x}}(k+1|k) = \sum_{i=1}^{2n} W_m^{(i)}\, \hat{\mathcal{X}}(k+1|k)^{(i)}, \tag{g}$$

where $W_m{}^{(i)}$ is the weight related to the i-th sigma-point; these weights are given by:

$$W_m^{(0)} = \lambda/(n+\lambda) \text{ and } W_m^{(i)} = 1/(2(n+\lambda)), \ i = 1, \ldots, 2n. \tag{h}$$

These weights in Equation (h) depend on a set of scaling parameters, as seen in [15]: λ is obtained as $\lambda = \alpha^2(n+p) - n$, where α defines the spread of the sigma-points around the estimation and it has a suggested value of 1×10^{-3}, p is a secondary scalar parameter, usually equal to zero. The last part of the UKF Prediction Step is obtaining the prediction error covariance matrix, given by:

$$P(k+1|k) = Q(k) + \sum_{i=1}^{2n} \{W_c^{(i)}[\hat{\mathcal{X}}(k+1|k)^{(i)} - \hat{\mathbf{x}}(k+1|k)][\hat{\mathcal{X}}(k+1|k)^{(i)} - \hat{\mathbf{x}}(k+1|k)]^\top\}, \tag{i}$$

where $W_c{}^{(i)}$ is the weight related to the i-th sigma-point, related to the covariance matrix, given by:

$$W_c^{(0)} = \lambda/(n+\lambda) + (1 - \alpha^2 + \beta) \text{ and } W_c^{(i)} = 1/(2(n+\lambda)), \ i = 1, \ldots, 2n \tag{j}$$

where β is the prior knowledge of the distribution; for Gaussian distributions, β is commonly set as 2.

3.2 UKF Update Step

In the UKF Update Step, new sigma-points are calculated, as follows, based on the prediction state $\hat{\mathbf{x}}(k+1|k)$ and the prediction error covariance matrix $P(k+1|k)$, as in (g) and (i), respectively:

$$\begin{aligned}\hat{\mathcal{X}}(k+1|k)^{(i)} &= \hat{\mathbf{x}}(k+1|k) + \left(\sqrt{nP(k+1|k)}\right)_i^\top, \ i = 1, \ldots, n,\\ \hat{\mathcal{X}}(k+1|k)^{(n+i)} &= \hat{\mathbf{x}}(k+1|k) - \left(\sqrt{nP(k+1|k)}\right)_i^\top, \ i = 1, \ldots, n.\end{aligned} \tag{k}$$

Next, the new sigma-points in Equation (k) are transformed from the function $\mathbf{h}$ to the predicted measurement sigma-points $\mathcal{Z}(k+1|k)^{(i)}$. These points are combined in order to obtain the predicted measurement $\hat{\mathbf{z}}(k+1|k)$, as follows:

$$\begin{aligned}\hat{\mathcal{Z}}(k+1|k)^{(i)} &= \mathbf{h}(\hat{\mathcal{X}}(k+1|k)^{(i)}),\\ \hat{\mathbf{z}}(k+1|k) &= \sum_{i=1}^{2n} W_m^{(i)}\, \hat{\mathcal{Z}}(k+1|k)^{(i)}.\end{aligned} \tag{l}$$

The predicted measurement error covariance and cross-covariance matrices are given by, respectively:

$$\begin{aligned}P_z(k) &= R(k+1) + \sum_{i=1}^{2n} W_c^{(i)}[\hat{\mathcal{Z}}(k+1|k)^{(i)} - \hat{\mathbf{z}}(k+1|k)][\hat{\mathcal{Z}}(k+1|k)^{(i)} - \hat{\mathbf{z}}(k+1|k)]^\top\\ P_{xz}(k) &= \sum_{i=1}^{2n} W_c^{(i)}[\hat{\mathcal{X}}(k+1|k)^{(i)} - \hat{\mathbf{x}}(k+1|k)][\hat{\mathcal{Z}}(k+1|k)^{(i)} - \hat{\mathbf{z}}(k+1|k)]^\top.\end{aligned} \tag{m}$$

The matrices in (m) are fundamental for obtaining the UKF gain, as follows:

$$K(k+1) = P_{xz}(k)\, P_z^{-1}(k). \tag{n}$$

Finally, the updated state vector and the estimation error covariance matrix can be obtained, as follows:

$$\begin{aligned}\hat{\mathbf{x}}(k+1|k+1) &= \hat{\mathbf{x}}(k+1|k) + K(k+1)[\mathbf{z}(k+1) - \hat{\mathbf{z}}(k+1|k)],\\ P(k+1|k+1) &= P(k+1|k) - K(k+1)P_z K(k+1)^\top.\end{aligned} \tag{o}$$

4 METHODS

To develop the proposed lightweight pre-screening module, a systematic methodology was conducted, encompassing data selection, computational implementation, and performance evaluation, as depicted in the experimental pipeline (Figure 1). Based on the mathematical foundation, the computational implementation was focused on the CUKF. The ECG signals acquired from public-domain datasets provided by PhysioNet, specifically the MIT-BIH Arrhythmia Database [17], were preprocessed and windowed. The CUKF was computationally implemented in the MATLAB software (used under an institutional license provided by the Federal University of Technology – Paraná, UTFPR) and parameterized to act as a healthy sinus rhythm tracker. The automated pre-classification of abnormal cardiac events was established based on the stochastic filter's innovation residual. The NMSE was employed as the core classification criterion. A heartbeat was flagged as abnormal (i.e., a PVC) whenever its corresponding NMSE exceeded a predefined adaptive threshold, indicating a substantial morphological deviation from the expected quasi-periodic dynamic of a normal sinus beat.

4.1 Dataset and Pre-processing

The proposed methodology was validated using the MIT-BIH Arrhythmia Database [17,18], which provides gold-standard annotated ECG recordings. This database contains comprehensive, long-term ambulatory records of patients exhibiting various cardiac arrhythmias, providing a robust foundation for testing the detection algorithm. The validation signals, sampled at 360 Hz, were initially processed using band-pass filtering to reduce baseline wander and high-frequency noise. The study focused on PVC, utilizing 79 signal segments of 20 seconds each, extracted from Lead II across nine distinct patients. This selection totaled 1,299 normal beats and 649 PVC. Although the number of analyzed beats was substantial, the limited number of patients may restrict the generalizability of the results.

4.2 CUKF Parameterization

The computational implementation was based on the proposed CUKF framework. The algorithm was parameterized to function strictly as a normal sinus rhythm tracker, ensuring that any substantial morphological deviation would be reflected in the innovation residual. Specifically, the covariance matrices $P(k|k)$, $Q(k)$, and $R(k)$ were initialized based on heart rate estimates derived from the local average RR interval, allowing the CUKF to adapt its tracking window to the patient's current physiological state.

The simulation parameters for both the cylindrical dynamic model and the UKF are summarized in Table 1. The ECG model parameters, peak amplitude (α_i), angular position (θ_i), and wave width (b_i), were configured to define the standard morphology of the P, Q, R, S, and T waves. Regarding the UKF tuning, the scaling constants (α, β, p) were selected to maintain an optimal sigma-point spread for Gaussian distributions. Furthermore, the process noise (Q) and measurement noise (R) covariance matrices were empirically adjusted to effectively track healthy sinus rhythm, with the sampling period (T) set to 1/360 seconds to maintain consistency with the MIT-BIH Arrhythmia Database specifications.

4.3 Beat Delimitation and Residual Calculation

To evaluate individual beats, the continuous ECG signal was segmented utilizing the estimated phase state ϕ, interpolated within the interval $[-\pi, \pi]$, where $\phi = 0$ strictly aligns with the R-peak. A complete cardiac cycle was thus delimited as the set of samples satisfying the interval $[-\pi, \pi]$. Incomplete beats at the end of the 20-second extracted windows were excluded from the analysis.

The automated detection was based on the CUKF's innovation residual. For each delimited cardiac beat, the mathematical divergence between the observed measurement and the CUKF prediction was quantified using:

$$\mathrm{NMSE} = \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}, \qquad \text{(p)}$$

where $y_i \in \mathbb{R}$ represents the real observed measurement, $\hat{y}_i \in \mathbb{R}$ is the prediction generated by the CUKF, and $\bar{y} \in \mathbb{R}$ is the mean of the observed signal. The index $N \in \mathbb{N}^*$ denotes the number of samples within the specific delimited beat. This normalization ensures that the metric is invariant to the signal's amplitude and baseline fluctuations, resulting in a sturdy digital biomarker for morphological deviations.

4.4 Cross-Validation and Performance Metrics

To assess the diagnostic reliability of the proposed pre-screening strategy, the system's performance was systematically analyzed. Evaluation metrics also included the confusion matrix, Sensitivity, and the Macro F1-score. The automated classification relies on an adaptive decision threshold *Tsh*, wherein beats yielding an NMSE greater than *Tsh* are flagged as abnormal. To ensure robust generalization, a K-Fold Cross-Validation methodology (K=10) was applied. Given the intrinsic class imbalance present in Holter recordings, the diagnostic performance of the pre-screening module was evaluated using Sensitivity, Specificity, and the Macro F1-score [19]. During threshold optimization, *Tsh* was selected to maximize the Macro F1-score, providing a balance between sensitivity and precision. This criterion was adopted because, although the pre-screening stage benefits from high sensitivity, excessive false positives may increase the computational burden of subsequent analysis.

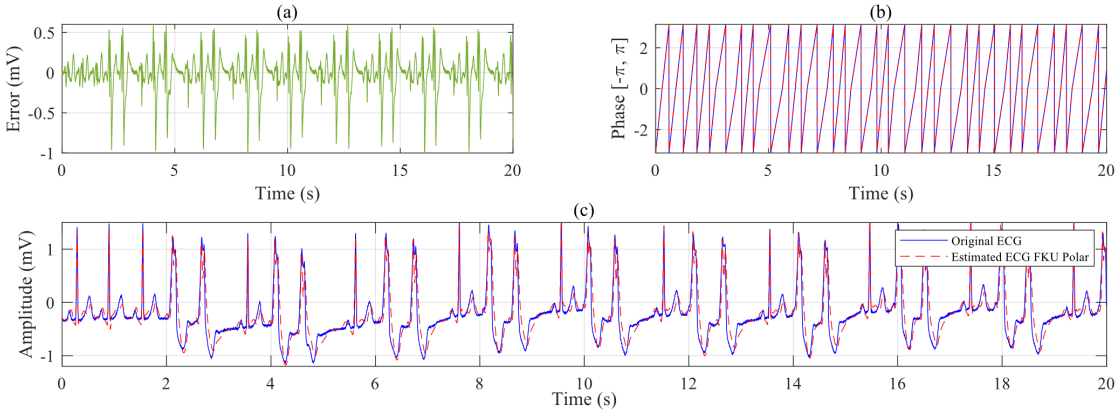
4.5 Ethical Considerations

This study utilized the MIT-BIH Arrhythmia Database, a publicly available and fully anonymized dataset. In accordance with the Brazilian National Health Council (CNS) Resolution No. 510/2016, this research is exempt from review and approval by a Research Ethics Committee (CEPSH).

5 RESULTS

The proposed approach, based on the CUKF framework, demonstrated robust performance in tracking cardiac dynamics and detecting PVC. Under normal sinus rhythm, the innovation residual, representing the instantaneous divergence between the model's prediction and the measured signal, remains bounded and close to zero, as illustrated in Figure 2(a). This stability indicates that the cylindrical dynamic model successfully captures the underlying quasi-periodic morphology of the ECG, establishing a reliable

Figure 2. CUKF curves: (a) Instantaneous error between measured and estimated signals; (b) Comparison between phase measurement and phase estimation; and (c) Comparison between measured and estimated signals.



Source: Prepared by the authors (2026).

baseline for anomaly detection. This low error profile is fundamentally driven by the cylindrical modeling, which explicitly models phase as a state variable, enabling stable synchronization between the estimated and measured phase throughout the cardiac cycle (Figure 2(b)). In contrast, PVC events, characterized by the absence of a P-wave and widened QRS complexes, produced a clear mismatch between the predicted and measured signals. Consequently, the estimated signal closely follows the measured physiological waveform during normal sinus rhythm, resulting in low NMSE values under baseline conditions. However, during PVC events, morphological deviations between the estimated and measured signals produce transient increases in NMSE. These localized NMSE spikes, highlighted in the zoomed visualizations (Figure 2(c)), capture the abnormal waveform dynamics and serve as digital biomarkers for the pre-screening process.

The NMSE-based analysis (Figure 3) shows that such mismatches generate pronounced peaks in the error signal. In particular, the per-beat NMSE and its derivative (Figures 3(a) and 3(b), respectively) provide a clear separation between normal and arrhythmic beats, enabling effective discrimination.

In terms of quantitative metrics extracted from the MIT-BIH dataset analysis, the proposed pre-screening module achieved an excellent balance between sensitivity and precision. The CUKF framework yielded a Sensitivity of 85.83% ± 3.44% and a Macro F1-score of 0.8924 ± 0.0160. Notably, the model's performance is driven by high precision and sturdy rejection of normal morphological variations, generating only 95 false positives across the evaluated segments. In Figure 4, the confusion matrix analysis shows a 92.7% correct classification rate for healthy sinus rhythm class (Specificity). These results demonstrate that the innovation residual of the CUKF is highly reliable for minimizing false alarms, offering competitive performance while maintaining low computational complexity.

6 DISCUSSION

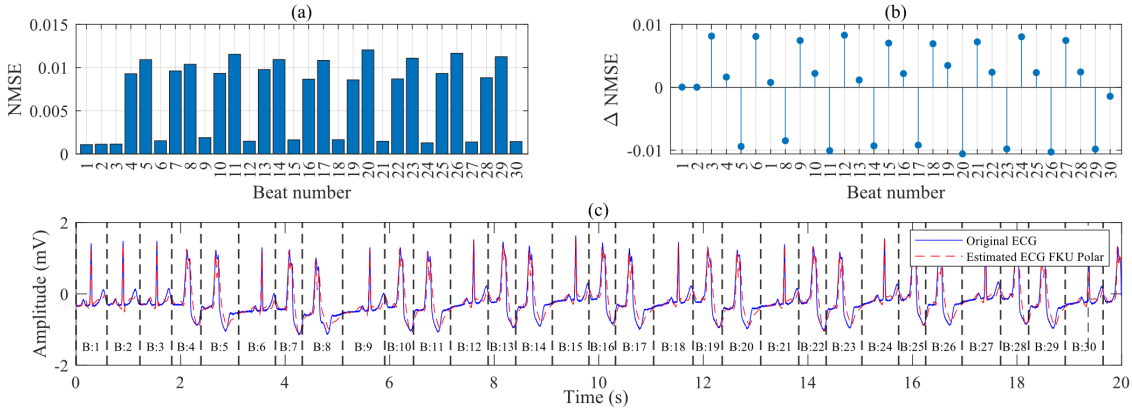
The results show that innovation-residual-based pre-classification offers a transparent alternative for PVC detection in ambulatory ECG monitoring. While recent deep learning architectures report comparable sensitivities on the MIT-BIH dataset [20,21], their complex nature limits physiological interpretability. Achieving an average sensitivity of 85.83% ± 3.44% and a Macro F1-score of 0.8924 ± 0.0160, the proposed framework demonstrates competitive diagnostic performance while providing a metric that inherently reflects morphological deviations. This represents a shift from opaque classifiers toward an interpretable tracking model, with the potential to enhance clinician trust.

6.1 Complexity Analysis and Clinical Integration

Rather than competing directly with highly parameterized machine learning models, the proposed CUKF innovation detector is specifically designed to operate as a pre-screening module within a cascaded ECG analysis pipeline. Deep learning techniques, such as Convolutional Neural Networks (CNN) and Long Short-Term Memory networks (LSTM), and traditional machine learning models, such as Support Vector Machines (SVM) and Random Forests (RFs), achieve excellent accuracy [22-25], but require substantial memory, extensive training datasets, and heavy processing, rendering them inefficient for continuous battery-powered nodes.

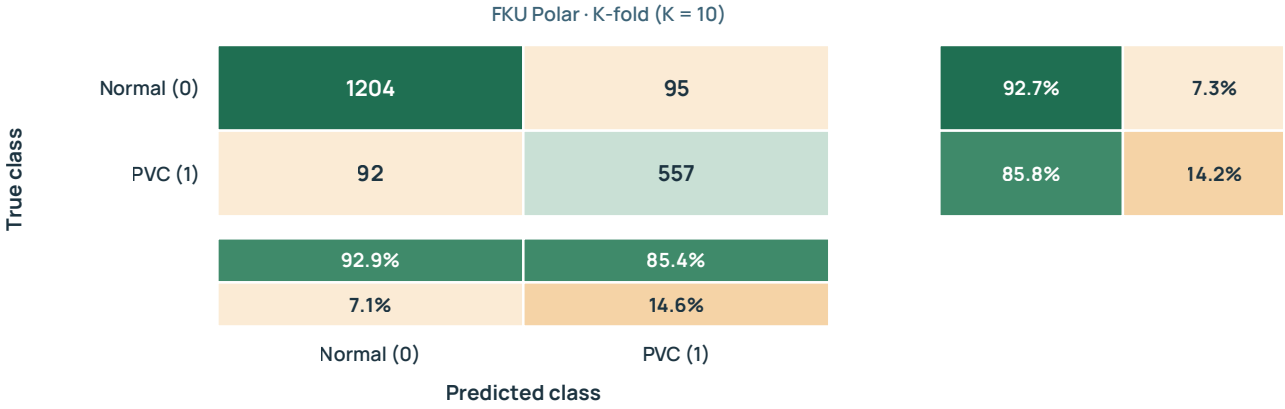
In contrast, the proposed framework is fully unsupervised and requires very few arithmetic operations per cardiac cycle [10,11]. Its low complexity is supported by Big-O analysis. In the standard UKF, the main computational cost arises from Cholesky decomposition and matrix inversion, with complexity $O(n^3+m^3)$ where n denotes the dimension of the state vector and m the dimension of the measurement vector. However, the adopted cylindrical dynamic model reduces the problem to $n = m = 2$, making these operations effectively constant. As a result, the time

Figure 3. NMSE-based analysis for PVC detection: (a) NMSE per heartbeat; (b) NMSE derivative; and (c) Measured versus estimated signals with heartbeat segmentation.



Source: Prepared by the authors (2026).

Figure 4. Confusion matrix analysis of the proposed framework on the MIT-BIH dataset.



Source: Prepared by the authors (2026).

complexity scales linearly with the signal length, $O(N)$, where N is the number of samples, while the memory requirement remains constant, $O(1)$, since only fixed 2 × 2 covariance matrices are stored at each time step.

Based on our experimental evaluation, only approximately 33.5% of the total analyzed heartbeats were flagged as candidate anomalies. In a cascaded architecture, instead of continuously streaming the full ECG signal, the node would only transmit these flagged beats to a secondary device (e.g., a smartphone or cloud server) for final classification. This strategy reduces the telemetry data volume by roughly 66.5%, drastically conserving battery life by restricting both radio activation and downstream complex inferences to instances where they are morphologically justified.

7 CONCLUSION

This study presents a lightweight and interpretable pre-classification framework for the automated detection of PVC in ambulatory ECG monitoring. By leveraging the innovation residual from a UKF coupled with a cylindrical dynamical model, called CUKF, the method provides transparent detection of abnormal cardiac beats with an ultra-low Big-O computational profile. The framework effectively acts as a diagnostic gatekeeper, offering a practical alternative to continuously running data-intensive deep learning algorithms. This approach contributes to scalable digital health systems. In the context of the Brazilian public health system (SUS), integrating algorithms into low-cost wearable devices can democratize access to long-term cardiac monitoring, optimizing telecardiology resources and improving equity in cardiovascular care.

While the framework performed robustly for PVC, its generalizability to other complex arrhythmias, such as Atrial Fibrillation or varying degrees of Atrioventricular Block, remains to be evaluated. Such conditions may require the underlying dynamical model to account for aperiodic P-wave occurrences and irregular R-R intervals. Future work will explore adaptive model tuning for broader anomaly detection and the physical deployment of this framework in microcontrollers to empirically validate the theoretical power savings in real-time clinical support environments.

Author contributions. The authors contributed equally to all stages of the article: conception and design of the study; data collection, analysis, and interpretation; drafting and critical revision of the manuscript; and approval of the final version.

Acknowledgments. The authors would like to acknowledge the Federal University of Technology — Paraná (UTFPR), Fundação Araucária and Superintendência de Ciência, Tecnologia e Ensino Superior (SETI) for their support.

Funding. The authors did not declare any source of funding.

Data and materials availability. The authors did not provide a data and materials availability statement.

Conflicts of interest. The authors did not declare any conflicts of interest.

Use of artificial intelligence. The authors declare that generative AI was used only for language improvement and text revision. No AI tool was used for data generation, mathematical modeling, or scientific conclusions. The authors remain responsible for the final content of the manuscript.

REFERENCES

1. Chong B, Jayabaskaran J, Jauhari SM, Chan SP, Goh R, Kueh MTW, et al. Global burden of cardiovascular diseases: projections from 2025 to 2050. Eur J Prev Cardiol. 2025;32(11):1001-15. doi: https://doi.org/10.1093/eurjpc/zwae281
2. Souza Luiz Y, Alvarenga VM, de Miranda Lopes M, Gomes Silva L. Epidemiological Analysis of Emergency Hospitalizations in Brazil Due to Cardiovascular Diseases From 2013 to 2023. International Journal of Cardiovascular Sciences. 2025;38:e20240200. doi: https://doi.org/10.36660/ijcs.20240200
3. Slaven S, Garg L, Sabzwari R, Barrett C, Tumolo A, Cerbin L, et al. Burden of Premature Ventricular Complexes and Risk of Cardiomyopathy: A Cross-Sectional Study. JACC Clin Electrophysiol. 2025;11(5):894-903. doi: https://doi.org/10.1016/j.jacep.2025.01.004
4. Tsiachris D, Botis M, Doundoulakis I, Bartsioka LI, Tsioufis P, Kordalis A, et al. Electrocardiographic Characteristics, Identification, and Management of Frequent Premature Ventricular Contractions. Diagnostics (Basel). 2023;13(19):3094. doi: https://doi.org/10.3390/diagnostics13193094
5. Liu C, Liu J, Wang D. Energy-Efficient Health Monitoring in Smart Devices: Deep Learning-Based Technique for Wearable ECG Systems. IEEE Trans Consum Electron. 2025;71(2):4736-43. doi: https://doi.org/10.1109/tce.2025.3569456
6. Kodipalli A, Rao T, Ushasree A, Sujatha C. Deep learning for electrocardiogram-based arrhythmia detection and classification: Architectures, challenges, and clinical translation. In: Deep Learning for Cardiac Signal Analysis in Robotic Applications. [place unknown]: Elsevier; 2026. p. 247-66. doi: https://doi.org/10.1016/b978-0-443-45242-0.00004-9
7. Hussain A, Hussain T, Attar RW, Alhomoud A, Alnfiai MM, Alsagri R. Energy-efficient synchronization for body sensor network in the metaverse: an optimized connectivity approach. J Wireless Com Network. 2025;2025(1):7. doi: https://doi.org/10.1186/s13638-025-02433-4
8. Huang X, Zhang F, Wang R, Lin X, Liu H, Fan H. KalmanAE: Deep Embedding Optimized Kalman Filter for Time Series Anomaly Detection. IEEE Trans Instrum Meas. 2023;72:1-11. doi: https://doi.org/10.1109/tim.2023.3329098
9. Hesar HD. A Novel AR-Based Kalman Filtering Framework for ECG Enhancement. In: 2025 32nd National and 10th International Iranian Conference on Biomedical Engineering (ICBME). [place unknown]: IEEE; 2025. p. 1286-92. doi: https://doi.org/10.1109/icbme68496.2025.11392248
10. McSharry PE, Clifford GD, Tarassenko L, Smith LA. A dynamical model for generating synthetic electrocardiogram signals. IEEE Trans Biomed Eng. 2003;50(3):289-94. doi: https://doi.org/10.1109/TBME.2003.808805
11. Sameni R, Shamsollahi MB, Jutten C, Clifford GD. A nonlinear Bayesian filtering framework for ECG denoising. IEEE Trans Biomed Eng. 2007;54(12):2172-85. doi: https://doi.org/10.1109/tbme.2007.897817
12. Truong DA, Hoang KT, Nguyen HH, Nguyen MT, Han HD, Pham-Nguyen L. Baseline Wander Removal in Biomedical Signal Acquisition System for Heart Rate Monitoring. In: 2025 2nd International Conference on Health Science and Technology (ICHST). [place unknown]: IEEE; 2025. p. 1-6. doi: https://doi.org/10.1109/ichst66555.2025.11428790
13. Kalman R. A New Approach to Linear Filtering and Prediction Problems. Journal of Basic Engineering. 1960;82(1):35-45. doi: https://doi.org/10.1115/1.3662552
14. Bar-Shalom Y, Li X, Kirubarajan T. Estimation with Applications to Tracking and Navigation. [place unknown]: Wiley; 2002. doi: https://doi.org/10.1002/0471221279
15. Wan EA, Van Der Merwe R. The unscented Kalman filter for nonlinear estimation. In: Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373). [place unknown]: IEEE; 2000. p. 153-8. doi: https://doi.org/10.1109/asspcc.2000.882463
16. Julier SJ, Uhlmann JK. Unscented Filtering and Nonlinear Estimation. Proc. IEEE. 2004;92(3):401-22. doi: https://doi.org/10.1109/jproc.2003.823141
17. Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. IEEE Eng Med Biol Mag. 2001;20(3):45-50. doi: https://doi.org/10.1109/51.932724
18. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101(23):E215-20. doi: https://doi.org/10.1161/01.cir.101.23.e215
19. He H, Garcia EA. Learning from Imbalanced Data. IEEE Trans Knowl Data Eng. 2009;21(9):1263-84. doi: https://doi.org/10.1109/tkde.2008.239
20. Cai Z, Wang T, Shen Y, Xing Y, Yan R, Li J, et al. Robust PVC Identification by Fusing Expert System and Deep Learning. Biosensors (Basel). 2022;12(4):185. doi: https://doi.org/10.3390/bios12040185

21. Jin L, Zhu S. Premature ventricular contraction detection from partially contaminated ECG signals using local pattern aggregation-based deep learning model. Biomedical Signal Processing and Control. 2026;112:108596. doi: https://doi.org/10.1016/j.bspc.2025.108596

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

23. Kiranyaz S, Ince T, Gabbouj M. Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. IEEE Trans Biomed Eng. 2016;63(3):664-75. doi: https://doi.org/10.1109/TBME.2015.2468589

24. Atwa AEM, Atlam ES, Ahmed A, Atwa MA, Abdelrahim EM, Siam AI. Interpretable Deep Learning Models for Arrhythmia Classification Based on ECG Signals Using PTB-X Dataset. Diagnostics (Basel). 2025;15(15):1950. doi: https://doi.org/10.3390/diagnostics15151950

25. Jaya Prakash A, Nasreddine Belkacem A, Elfadel IM, Jelinek HF, Atef M. Advances in machine and deep learning for ECG beat classification: a systematic review. Front Digit Health. 2025;7:1649923. doi: https://doi.org/10.3389/fdgth.2025.1649923

EDITORIAL NOTES

The Journal of Health Informatics is a scientific journal maintained by the Brazilian Society of Health Informatics (Sociedade Brasileira de Informática em Saúde), a non-profit scientific association dedicated to the development of Health Informatics and Digital Health in Brazil. The editorial process is conducted with scientific independence, free from institutional, commercial, or sponsor interference, and follows international best practices in publication ethics.

This article was double-blind peer reviewed by specialists in the field. The editorial team used artificial intelligence tools, under human supervision, in the initial screening, in checking the text, references and numbers and, when needed, in redrawing figures, charts and tables, which the authors approved in the proof. The tools and safeguards are described in the journal's Policy on the use of artificial intelligence. Editorial decisions were made by people.

SBIS Sociedade Brasileira de Informática em Saúde