BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 31 No. 3. 2026

UDC 004.891:616-7
DOI: 10.62660/bcstu/3.2026.173

A method for constructing a feature space based on the phase portrait of a biosignal in analytical software subsystems

Vadym Matsiievskyi
Postgraduate Student, Senior Lecturer
Zhytomyr Polytechnic State University
10005, 103 Chudnivska Str., Zhytomyr, Ukraine
https://orcid.org/0000-0003-3947-1363
Tetiana Nikitchuk*
PhD in Technical Sciences, Associate Professor
Zhytomyr Polytechnic State University
10005, 103 Chudnivska Str., Zhytomyr, Ukraine
https://orcid.org/0000-0002-9068-931X

Abstract. The development of software for biomedical monitoring platforms built on internet of medical things architectures requires computational criteria for biosignal analysis that are robust to the heterogeneity of sensor hardware and to random data losses in wireless communication channels, because the conventional geometric characteristics of the phase portrait are highly sensitive to these factors, which complicates the aggregation of data from different devices in distributed systems. The aim of the study is to present a method for constructing the feature space of the predictive software components of a biomedical monitoring platform on the basis of geometric invariants of the biosignal phase portrait that are robust to variations in the sensor sampling rate and to data packet loss. The phase space was reconstructed from the discrete signal using the differential method and the time-delay method based on Takens' embedding theorem, while feature extraction employed three newly developed computational criteria: the dynamic phase asymmetry index, the normalised fractal density of the trajectory and the phase coefficient of shape variation. The properties of the criteria were verified on a synthetic sample of two hundred pulse wave segments at three sampling rates and three levels of artificial packet loss, while their contribution to the classification quality of critical physiological states was assessed on a multi-source clinical sample of one hundred and thirty patients by means of recurrent neural networks, support vector machines and random forests, with statistical significance tested by the Wilcoxon signed-rank test. The dynamic phase asymmetry index proved to be asymptotically invariant to the sampling rate: the coefficient of variation of its values was 2.89% against 63.28% for the integral attractor area. The normalised fractal density of the trajectory significantly outperformed the box-counting method in robustness to packet loss from ten per cent of lost samples onwards. The superiority of the combined phase feature vector over time-domain heart rate variability indices was not confirmed on the synthetic sample, which indicates the need for clinical validation of the diagnostic value of the proposed criteria. The developed criteria have a computational complexity suitable for stream processing on the edge nodes of distributed monitoring platforms and ensure the robustness of the feature vector to sensor hardware heterogeneity, which makes it possible to aggregate data from different devices without a systematic bias in the predictions of predictive models

Keywords: phase portrait; IoMT; feature extraction; dynamic phase asymmetry index; normalised fractal density; hardware invariance; biomedical signals

Article's History: Received: 07.05.2026; Revised: 27.08.2026; Accepted: 21.09.2026; Published: 07.10.2026

Suggested Citation:
Matsiievskyi, V., & Nikitchuk, T. (2026). A method for constructing a feature space based on the phase portrait of a biosignal in analytical software subsystems. *Bulletin of Cherkasy State Technological University*, 31(3), 173-185. doi: 10.62660/bcstu/3.2026.173.

*Corresponding author (ntm@ztu.edu.ua)

Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/)

INTRODUCTION

A. Rancea *et al.* (2024), in a systematic review of edge computing in healthcare, highlight its advantages for near-real-time decision-making and efficient communication in internet of medical things (IoMT) environments, while also identifying open challenges in data privacy, security, and computational offloading at the edge – challenges that directly motivate the shift described above from standalone information systems to complex, high-load software systems built on IoMT architectures and big-data processing technologies. R.T. Sutton *et al.* (2020), in a state-of-the-art review of clinical decision support systems, identify the design of highly reliable software capable of real-time analysis of heterogeneous data as a key applied task in this domain, while also cataloguing common pitfalls and potential harms that must be addressed for such systems to be safely adopted in clinical practice.

The effectiveness of such systems depends critically on the soundness of architectural decisions, on the methods used to aggregate multimodal data, and on the seamless integration of intelligent agents (machine-learning models) into the software environment. The application of conventional methods often results in the loss of important diagnostic features. Integrating entropy- and fractal-dimension-based algorithms, by contrast, makes it possible to detect latent dynamic changes that precede the clinical manifestation of disease – as demonstrated by A. Rohila & A. Sharma (2019), who computed phase entropy from the distribution of points on the phase plane as an informative measure of heart-rate-variability complexity – to construct robust feature vectors for the subsequent training of neural networks, and to adapt the software logic to the individual complexity of a given patient's physiological processes.

There is thus a contradiction between the high heterogeneity and noisiness of biomedical data in distributed IoMT networks and the requirements of high accuracy, invariance and speed placed on software analytical models, which necessitates the development of new software-algorithmic solutions. Building an adaptive data-transformation pipeline, formalising the computation of nonlinear phase criteria, and integrating them with predictive neural-network microservices are fundamental prerequisites for constructing highly reliable software platforms for predictive monitoring. The hardware foundation of such a platform is provided by wearable photoplethysmographic sensors; P.H. Charlton *et al.* (2023), in a review documenting substantial variability in signal quality across devices from different manufacturers, systematise the capabilities and limitations of such sensors. This confirms the relevance of the chosen topic and defines the direction of the present study.

The aim of this study was to develop a method for constructing the feature space of the predictive software components of a biomedical monitoring platform, based on geometric invariants of the biosignal phase portrait, for improving the accuracy of predictive diagnostics and supporting clinical decision-making. The following tasks are addressed within the article: to propose a method of feature space construction that ensures the invariance of the feature vector to fluctuations in sensor sampling rate and to random packet loss in communication channels, and to substantiate its computational complexity and suitability for stream processing; to present improvements to the algorithmic support for constructing and analysing biosignal phase portraits in real time, specifically in terms of the automatic selection of the computational pathway according to the morphological type of the input signal and the entropy-based assessment of phase-plane filling; to experimentally confirm the invariance properties of the developed criteria and to assess their contribution to the classification quality of critical physiological states.

The scientific novelty of the study was as follows: a method for constructing the feature space of the predictive software components of a biomedical monitoring platform, based on geometric invariants of the biosignal phase portrait, was developed for the first time – including, in particular, the dynamic phase asymmetry index (IDA), defined as the ratio of the integral euclidean distances of the trajectory from the centroid in the regions of positive and negative velocity – together with an improved method for constructing and analysing biosignal phase portraits in real time, distinguished by a software algorithm for the automatic selection of the computational pathway according to the morphological type of the signal.

LITERATURE REVIEW

The choice of methods for processing biomedical signals is a critical stage in software architecture design, as it determines the requirements for computational power and database architecture. Conventional analytical approaches often prove insufficient for processing signals of biological origin owing to their high non-stationarity and stochasticity. A. Subasi & S.M. Qaisar (2023), in a review of signal-acquisition and feature-extraction techniques for biomedical signals, note that conventional linear methods of biosignal analysis – time-domain and frequency-domain analysis – remain the basic toolkit of biomedical monitoring software systems owing to their ease of implementation and low computational cost. However, they rely on assumptions of linearity and stationarity that hold only approximately for complex physiological systems: integral heart rate variability indices reflect the balance of autonomic regulation, yet the nonlinear structure of inter-system interactions remains beyond their scope (Rohila & Sharma, 2019; Olivieri *et al.*, 2024).

Time-domain analysis operates on signal characteristics in the original time domain – the mean value,

variance, root-mean-square (RMS) value, and autocorrelation (Subasi & Qaisar, 2023). In electrocardiogram (ECG) analysis, time-domain parameters are used to assess QRS-complex morphology, detect arrhythmias, and monitor interval durations; R. Hemmati *et al.* (2025), studying the influence of body mass index on electrocardiographic parameters in healthy and cardiovascular patients, demonstrate that these parameters are themselves sensitive to the patient's anthropometric characteristics. In heart rate variability (HRV) analysis, time-domain indices, in particular SDNN, serve as markers of autonomic nervous system activity (Olivieri *et al.*, 2024). The advantages of this group of methods are ease of implementation, low computational complexity, and a direct physical interpretation of the values, which makes it suitable for execution on wearable devices with limited computational resources (Nelson *et al.*, 2020).

At the same time, biosignals are non-stationary, and a substantial proportion of physiological processes are nonlinear; consequently, time-domain indices cannot fully characterise the complex nonlinear interactions and chaotic dynamics that are characteristic of biosignals (Rohila & Sharma, 2019; Subasi & Qaisar, 2023). A further limitation in distributed monitoring systems is the dependence of the time-domain index values themselves on registration conditions: a review of wearable photoplethysmography capabilities documents substantial variability in signal quality across devices from different manufacturers (Charlton *et al.*, 2023), while S. Singh *et al.* (2024), in a systematic review and meta-analysis of the impact of skin pigmentation on pulse oximetry and wearable pulse-rate accuracy, demonstrate a systematic bias in the estimates that is not attributable to the patient's physiological state.

Frequency-domain analysis transforms the signal from the time domain to the frequency domain, making it possible to study its spectral composition; the primary tool is the Fourier transform (FFT), which allows the identification of dominant frequencies, harmonics, and spectral bands for diagnostic purposes (Subasi & Qaisar, 2023). It is effective for stationary signals with pronounced periodic components, but it loses temporal information: the spectrum characterises the frequency composition of the signal over the entire observation period without indicating precisely when particular components appear, which is critical for non-stationary biosignals. Nonlinear phenomena are, in this case, interpreted as linear sums of frequency components without accounting for their nonlinear interaction, as a result of which the fractal properties and chaotic dynamics of the signal remain unrepresented (Rohila & Sharma, 2019).

To overcome these limitations, time-frequency transforms are used in biosignal processing practice, allowing the signal to be simultaneously localised in time and frequency and enabling changes in its frequency composition to be tracked. For instance, S. Nagai *et al.* (2017) applied the stationary wavelet transform to remove motion artefacts from wearable ECG recordings, while Y. Ye *et al.* (2016) combined nonlinear adaptive filtering with signal decomposition to eliminate motion artefacts in photoplethysmography. These methods effectively solve the problem of signal cleaning, but do not, by themselves, form a compact feature vector suitable for training predictive models (Subasi & Qaisar, 2023).

Nonlinear methods – fractal analysis, entropy measures, and the phase-plane method – are tools for detecting latent dynamic changes in complex systems that cannot be described by linear models. Fractal analysis quantitatively assesses the complexity and self-similarity of data structures and reveals patterns at different scales, while fractal dimension characterises the degree to which the system's trajectory fills the space (Grassberger & Procaccia, 1983; Hentschel & Procaccia, 1983). Entropy measures assess the uncertainty and degree of chaoticity of a dynamic system: in particular, phase entropy, computed from the distribution of points on the phase plane, has proved to be an informative measure of heart rate variability complexity (Rohila & Sharma, 2019). Geometric characteristics of the phase portrait, primarily Poincaré-plot parameters, are widely used in cardiology, but M. Brennan *et al.* (2001), directly investigating whether existing measures of Poincaré-plot geometry reflect nonlinear features of heart rate variability, show that they capture the nonlinear properties of the rhythm only partially.

Reconstruction of the phase space from a single scalar biosignal recording relies on the embedding theorem (Takens, 1981), with embedding parameters determined using the mutual information function (Fraser & Swinney, 1986) and the false-nearest-neighbours method (Kennel *et al.*, 1992). J. de Pedro-Carracedo *et al.* (2020), in a case study applying phase-space reconstruction to a photoplethysmographic signal, confirm the suitability of this apparatus for such signals experimentally**,** while the morphological features of PPG-waveform derivatives, which determine the choice of computational pathway, have been systematised in a review on the detection of PPG-waveform fiducial points (Suboh *et al.*, 2022). To evaluate the strengths and limitations of existing approaches, a comparative analysis of biomedical signal processing method groups across key engineering criteria is presented in Table 1.

Table 1. Comparative analysis of biomedical signal processing method groups by key engineering criteria

Method category	Algorithm examples	Advantages	Limitations	Complexity
Linear (time-domain analysis)	statistical parameters (SDNN, RMSSD)	minimal resource consumption, ease of real-time implementation	loss of information about the signal's internal dynamics	O(n)

Continued Table 1.

Method category	Algorithm examples	Advantages	Limitations	Complexity
Linear (frequency-domain analysis)	fast Fourier transform (FFT), autoregression	high accuracy for stationary periodic processes	unsuitability for signals with a time-varying spectrum	O(n log n)
Time-frequency (hybrid)	wavelet transform (DWT/CWT), S-transform	simultaneous localisation of the signal in time and frequency	difficulty in selecting the optimal mother wavelet	O(n log n)
Nonlinear (complexity)	approximate entropy, DFA analysis	detection of latent pathologies through regularity analysis	high sensitivity to sample length and noise	$O(n^2)$
Phase space	Poincaré plots, attractors	visual clarity and the possibility of visual diagnostics	high complexity of automatic parameter interpretation	O(n), average

Source: compiled by the authors based on M. Brennan *et al*. (2001), Y. Ye *et al*. (2016), S. Nagai *et al*. (2017), A. Rohila & A. Sharma (2019), B.W. Nelson *et al*. (2020), and A. Subasi & S.M. Qaisar (2023)

A separate group is formed by software-engineering factors that affect the quality of the feature vector in distributed platforms. Distributional shift between data sources can disrupt the performance of machine-learning models (Dockès *et al.*, 2021). Empirical Bayes harmonisation methods (Johnson *et al.*, 2007) are applied to eliminate it. Their modifications are also applied for this purpose (Zhang *et al.*, 2020; Chen *et al.*, 2022). Sample losses in wireless channels are systematic rather than purely random in nature (Braem *et al.*, 2024). The processing pipeline is therefore supplemented with data-recovery methods – multiple imputation by chained equations (van Buuren & Groothuis-Oudshoorn, 2011) and non-parametric imputation (Stekhoven & Bühlmann, 2012). Generative approaches are also included (Yoon *et al.*, 2018). Simplified schemes such as last-observation-carried-forward lead to biased estimates (Lachin, 2015).

Offloading analytics to edge nodes is an established architectural response to latency and privacy requirements in medical applications (Shi *et al.*, 2016), and secure architectures for remote patient monitoring form a distinct direction within IoMT-platform engineering (Bhattacharya *et al.*, 2025). Clinical decision-support systems combine rules with machine-learning models (Sutton *et al.*, 2020; Kiernan *et al.*, 2023). In practice, however, an excessive rate of false alerts remains a key barrier to their adoption (Chromik *et al.*, 2022). Interpretability requirements also limit the use of opaque models in tasks with a high cost of error (Rudin, 2019; Amann *et al.*, 2020).

Thus, within the development of intelligent software, particular attention should be paid to methods of nonlinear dynamics – entropy-based and fractal analysis – since biomedical signals such as heart rate variability possess the property of self-similarity and a high structural complexity, and the application of conventional methods often results in the loss of important diagnostic features (Rohila & Sharma, 2019). At the same time, the works reviewed do not formulate invariance requirements for the geometric features of the phase portrait with respect to sensor hardware specifications or to incompleteness of the data stream, even though it is precisely these factors that determine the suitability of the feature vector for aggregation in distributed IoMT platforms. This constitutes an unresolved part of the overall problem and justifies the further consideration of the method for constructing a feature space based on the biosignal phase portrait, presented in the next section.

MATERIALS AND METHODS

The developed method was validated on two samples. The first (sample S) is synthetic, generated by a software pulse-wave simulator; it was used for controlled verification of the criteria's invariance properties with respect to hardware factors (sampling rate and packet loss), as it allows these factors to be isolated from the variability of real physiological states. Each segment of sample S has a duration of 10 s; the heart rate and the systolic-phase fraction within a segment were assigned randomly within the ranges of 55-95 bpm and 0.28-0.45, respectively, and the parameter ranges of the two classes (normal/pathological) were deliberately overlapped to avoid trivial linear separability of the classification task. The sample size is 200 segments per experiment; this size provides a statistical power of at least 0.8 for detecting a difference in the coefficients of variation of the compared criteria at a significance level of 0.05, and is consistent with the sample sizes used in comparable studies of the sensitivity of phase-analysis features to hardware factors.

The second sample (sample 1) is a clinical sample of 130 patients from three sources – the MIT-BIH Arrhythmia Database (Moody & Mark, 2001), the MIMIC-III critical care database (Johnson *et al.*, 2016), and the PPG-DaLiA wearable dataset (Reiss *et al.*, 2019) – used to assess the contribution of the phase criteria to the classification quality of critical physiological states. The total volume is 12,241 analysed windows, with 42.3% of windows corresponding to a critical state. The split into training, validation, and test subsets was performed by patient (rather than by window) in a 60/20/20 ratio, to avoid information leakage between subsets. The size of this sample (130 patients, 12,241 windows) was determined by the availability of retrospective data collected between 1975 and 1979 (MIT-BIH), between 2001 and 2012 (MIMIC-III), and in 2019 (PPG-DaLiA) at the time of the study; it exceeds the minimum size required for stable training of the classification models used, given the adopted split ratio.

The input data for both samples is the corrected vector of biosignal amplitude values.

The input data specified for both samples were obtained at the output of the ComBat-MICE data pre-harmonisation service, which combines empirical Bayes correction of inter-source effects (Johnson *et al.*, 2007) with multiple imputation of missing values (van Buuren & Groothuis-Oudshoorn, 2011):

$$Y_{corrected} = [y(t_1), y(t_2), \ldots, y(t_N)], \quad (1)$$

where $Y_{corrected}$ – the vector of corrected biosignal amplitude values, which is fed into the phase-space reconstruction algorithm; $y(t_1)$ – the signal amplitude value at time t_i, i = 1. ..., N, obtained after data harmonisation by the ComBat-MICE service; N – the length of the analysis segment in samples, determined by the observation-window duration and the sensor's sampling rate.

Reconstruction of the phase space. The mathematical foundation of the method is the concept of mapping the state of a dynamic system in a two-dimensional phase space, developed in the works of T.M. Nikitchuk *et al.* (2023). For software implementation in IoMT architectures, where biosignals are processed as discrete real-time streams, the continuous mathematical apparatus requires discrete formalisation and optimisation for computational complexity. In medical software engineering, two software-oriented approaches to phase-space reconstruction from a discrete vector (1) are applied: the differential method and the time-delay method based on the embedding theorem (Takens, 1981); the choice of approach is automated according to the morphological type of the input signal.

The differential method («signal-velocity» coordinates) forms the phase plane by computing the velocity vector of the process (the first derivative) $\dot{Y}$. To minimise computational latency at the edge level and to prevent amplification of high-frequency digital noise, an optimised numerical differentiation algorithm based on a five-point central difference is applied:

$$y_k = \frac{-y_{k+2}+8y_{k+1}-8y_{k-1}+y_{k-2}}{12\Delta t}, \quad (2)$$

where Δt – the sampling period, determined by the hardware specification of the sensor.

The software window forms the phase trajectory as a set of points $Pk(xk,vk)$, where the coordinate axes are normalised by the software service using a *min-max* scheme:

$$x_k = \frac{y_k - \min(Y)}{\max(Y) - \min(Y)}, v_k = \frac{\dot{y}_k - \min(\dot{Y})}{max(\dot{Y}) - min(\dot{Y})}. \quad (3)$$

The differential method is optimal for processing photoplethysmographic (PPG) signals, since the morphology of the PPG waveform contains smooth, quasi-harmonic transitions, and its first derivative clearly reflects the anacrotic and dicrotic phases of the pulse cycle (Suboh *et al.*, 2022).

For signals with a pronounced impulsive character and a high gradient of change (in particular ECG, where the R-wave forms a virtually vertical amplitude jump), computing the derivative leads to mathematical singularity and algorithmic instability. In this case, the phase space is reconstructed using the Takens time-delay method, forming coordinates based on the time shift τ:

$$P_k = [y(t_k), y(t_k + \tau)], \quad (4)$$

The optimal delay step τ is determined automatically by one of two software services: using the mutual information function (Fraser & Swinney, 1986), where τ is selected at the point of the first local minimum (which guarantees maximum statistical independence of the coordinates), or using the autocorrelation function (ACF), where τ is determined as the time over which the signal's autocorrelation decays to $\frac{1}{e}$ or 0.1. The sufficiency of the embedding dimension is verified using the false-nearest-neighbours method (Kennel *et al.*, 1992).

Feature-extraction criteria. In the classical methodology of phase analysis, the principal geometric criteria are the integral attractor area S_{pulse} and the fractal dimension computed by the box-counting method (Grassberger & Procaccia, 1983). Under distributed IoMT-system conditions, these metrics exhibit high sensitivity to sampling rate and to random packet loss, which creates technical biases when aggregating data from different devices. To ensure the invariance of features to hardware specification, three computational criteria have been developed: the dynamic phase asymmetry index, the normalised fractal density of the trajectory, and the phase coefficient of shape variation.

Dynamic phase asymmetry index (IDA). The criterion is designed to assess the rate of change of vascular tone between the systolic (ejection) and diastolic (outflow) phases. It is computed as the ratio of the integral Euclidean distances of the trajectory in symmetric segments of the normalised phase plane:

$$IDA = \frac{\left(\sum_{\{k \in Q_1\}} \sqrt{(x_k - x_c)^2 + (v_k - v_c)^2}\right)}{\left(\sum_{\{k \in Q_2\}} \sqrt{(x_k - x_c)^2 + (v_k - v_c)^2}\right)}, \quad (5)$$

where $Q_1 = \{k : v_k \geq 0\}$ – the subset of phase-trajectory points with non-negative velocity; $Q_2 = \{k : v_k < 0\}$ – the subset of points with negative velocity; (x_c, v_c) – the geometric centre of mass (centroid) of the phase portrait, computed by the software module:

$$x_c = \frac{1}{N}\sum_{k=1}^{N} x_k, v_c = \frac{1}{N}\sum_{k=1}^{N} v_k. \quad (6)$$

The criterion is defined only under the condition $Q2 \neq \emptyset$. If the analysis segment is short or the signal section is monotonic, so that the denominator of formula (5) equals zero, the IDA value is undefined: the

software module returns a sentinel value, logs the event in the logging service, and skips the current window without passing it to the predictive service. The admissible range of the criterion is IDA ∈ (0; +∞); a value of IDA = 1 corresponds to complete symmetry of the trajectory relative to the centroid between the ejection and outflow phases, an increase in IDA indicates a lengthening of the systolic branch, and a decrease indicates predominance of the diastolic branch.

This parameter enables the software to detect latent disturbances in cardiovascular regulation (for example, vascular-wall stiffness in elderly patients) based on the geometric deformation of the limit cycle. Normalised fractal density of the trajectory (NFD**).** Unlike the standard box-counting method, which only records the fact that the trajectory crosses a grid cell, the developed algorithm computes the entropy density of the state-point distribution in space. The phase-window matrix of dimension $M \times M$ is populated with the probability values of signal samples falling into the corresponding sector A_{ij}:

$$NFD = -\frac{1}{2\ln M}\sum_{i=1}^{M}\sum_{j=1}^{M} p\left(A_{ij}\right)\ln p\left(A_{ij}\right),$$
$$\text{де } p\left(A_{ij}\right) = \frac{n_{ij}}{N}, \qquad (7)$$

where n_{ij} – the number of trajectory points located inside sector (i, j); N – the total length of the analysis segment; where M – the linear dimension of the square phase grid M × M.

The limits of the scale have the following interpretation: NFD → 0 corresponds to degeneration of the trajectory (all state points localised in a single sector, the attractor collapses); NFD → 1 corresponds to uniform filling of the phase grid (the trajectory loses its deterministic structure, characteristic of a stochastic regime or strong noise).

By construction, NFD is a normalised Shannon entropy of the phase-grid fill distribution and is therefore related to the information dimension D_1. The difference is that D_1 is defined as the limit at $\varepsilon \to 0$ and is an unbounded quantity, whereas NFD is computed at a fixed M, is guaranteed to lie within [0; 1], and remains correct under random sample dropout, since it is normalised by the actual number of recorded points N. This qualifies NFD as a refinement of the known method of entropy-based estimation of phase-plane filling that ensures feature invariance to data gaps and hardware fluctuations.

Phase coefficient of shape variation (FCV). To quantitatively assess the microstructural instability of the biosignal, a criterion has been developed that measures the variance of the radius vectors of trajectory points relative to the centroid over L consecutive pulse cycles:

$$FCV = \frac{1}{L}\sum_{m=1}^{L}\left(\frac{\sigma_R^{(m)}}{\bar{R}^{(m)}}\right). \qquad (8)$$

where $\bar{R}^{(m)}$ and $\sigma_R^{(m)}$ – the mean value and standard deviation of the distances from the points of the m-th phase portrait to its geometric centre.

The criterion makes it possible to detect microarrhythmias and non-stationary disturbances directly during stream processing. To integrate the method into the platform's overall software architecture, the feature-extraction process has been formalised as an algorithm optimised for execution at the edge-gateway level or on a cloud computing service. Generalised characteristics of the developed and baseline criteria are presented in Table 2.

Table 2. Computational complexity and invariance properties of the feature-extraction criteria

Criterion	Time complexity	Memory	Invariance to f_s	Robustness to sample loss
IDA (5)	O(N)	O(1)	yes (ratio of sums)	high (both sums decrease proportionally)
NFD (7)	$O(N + M^2)$	$O(M^2)$	partial	high (normalised by the actual N)
FCV (8)	O(N)	O(L)	partial	medium
S_pulse (baseline)	O(N)	O(1)	no (linear scaling)	low
Box-counting (baseline)	$O(N + M^2)$	$O(M^2)$	no	low

Source: compiled by the authors based on analytical derivations (see Eqs. (5)-(8))

The invariance of IDA to sampling rate can be justified analytically as follows. Let a signal segment of duration T be digitised at a rate fs, corresponding to $N = T \cdot fs$ samples. Both sums in formula (5) are integral (Riemann) sums approximating curvilinear integrals along the corresponding branches of the phase trajectory. When fs increases by a factor of λ, the number of samples in each of the subsets Q_1 and increases by approximately the same factor λ, while the distribution of point-to-centroid distances is preserved, since the coordinates were previously normalised using the min–max scheme (3). The numerator and denominator therefore scale identically, and their ratio remains constant:

$$IDA(\lambda fs) \approx IDA(fs).$$

The integral attractor area S_{pulse}, by contrast, is an additive quantity and scales linearly with increasing sample density, which accounts for its sensitivity to sensor hardware specification. An empirical verification of this claim at fs = 125/250/500 Hz.

Regarding the behaviour of the criteria under random sample loss: when the stream is thinned by a

fraction ρ, both sums in formula (5) decrease proportionally, so the IDA estimate remains unbiased as long as the losses are uniformly distributed with respect to the cycle phases. For NFD, normalising the probabilities by the actual number of recorded points N ensures the correctness of the entropy estimate under an incomplete stream; the increase in estimate variance at is compensated by prior data recovery using the multiple imputation (MICE) module.

The method was tested on three hypotheses using synthetic sample S and multi-source sample 1 (with the exception of H3, which was tested on sample 1). For all statistical comparisons, the Wilcoxon signed-rank test for paired samples was applied at a significance level of $\alpha = 0.05$; the confidence intervals of the gain were estimated by bootstrapping over the differences in indicators across paired runs.

Hypothesis H1 – sensitivity of IDA to sampling rate. Segments generated at a rate of 500 Hz were selected from sample S. Each segment was subjected to decimation to 250 Hz and 125 Hz with anti-aliasing filtering, producing three versions of the same physiological process that differed exclusively in the hardware sampling-rate parameter. For each version, the IDA criterion (5)–(6), the normalised fractal density NFD (7), and the baseline criteria were computed – the integral attractor area S_{pulse} and the fractal dimension by the box-counting method; the phase-grid dimension M was fixed identically across all versions. As the measure of the criterion's sensitivity to sampling rate, the coefficient of variation of the criterion's values across the set of frequencies CV (k)), f_s {125; 250; 500} Hz, and the maximum relative deviation from the value obtained at the highest frequency, δ_{max} (k). The hypothesis was accepted if the median CV(IDA) across the set of segments was statistically significantly lower than the median CV(S_{pulse}), and at the same time δ_{max}(IDA) did not exceed 5%.

Hypothesis H2 – robustness of NFD to packet loss. For each segment of samples S and 1, reference values of NFD and of the fractal dimension by the box-counting method were computed on the full stream. Artificial thinning of the stream was then performed, uniformly removing a fraction $\rho \in \{5; 10; 20\}$% of samples, simulating packet loss in a wireless communication channel, as well as a block-loss model (removal of contiguous intervals). For each thinning level, the same criteria were computed and the relative deviation from the reference value was determined. The experiment was carried out in two configurations: without prior data recovery and with prior data recovery using the multiple imputation (MICE) module. The hypothesis was accepted if, for all three levels ρ of ρ, the mean relative deviation of the NFD estimate was statistically significantly lower than the corresponding deviation of the box-counting estimate.

Hypothesis H3 – contribution of the phase criteria to classification quality. On sample 1, four groups of feature sets were compared: baseline (time-domain heart rate variability indices and SpO2 level), frequency-domain (spectral variability indices and SpO2), phase (the IDA, NFD, FCV criteria), and combined:

$$V = [HRV, SpO2, IDA, NFD]. \quad (9)$$

The study has limitations related to data volume and source: the invariance tests (H1, H2) were performed on a synthetic sample, while the diagnostic-value test (H3) was performed on a sample of a size typical for clinical studies, without independent external verification. The quantitative estimates obtained should therefore be regarded as characteristics of the criteria themselves under controlled conditions, rather than as a definitive assessment of their clinical significance.

Each feature set was fed into three classifiers (LSTM, SVM with an RBF kernel, Random Forest) for binary classification of the critical physiological state; the sensitivity of the result to the data-harmonisation method (ComBat, Z-normalisation, no harmonisation) was additionally examined. Within the combined set (9), an ablation analysis was performed F_1 – the sequential removal of each feature followed by model retraining. Classification quality was assessed using the Accuracy, F_1-measure, and AUC metrics over five independent runs. The hypothesis was accepted if the F_1-measure of the combined set (9) statistically significantly exceeded the -measure of the baseline set according to the Wilcoxon test at a significance level of 0.05, and the lower bound of the confidence interval of the gain was positive.

This study used retrospective, fully de-identified data from publicly available datasets (MIT-BIH Arrhythmia Database, MIMIC-III, and PPG-DaLiA), each of which was collected in accordance with the ethical principles of the Declaration of Helsinki (World Medical Association, 2013) and approved by the institutional review board of the respective originating institution. As only de-identified, publicly released data were used, no additional institutional ethics approval was required for the present secondary analysis.

RESULTS

Table 3 presents the mean values of the criteria across the sample $n = 200$ segments, duration 10 s, M = 32) and δ_{max} the medians of the CV and indicators.

Table 3. Sensitivity of the feature-extraction criteria to sampling rate

Criterion	fs = 125 Hz	fs = 250 Hz	fs = 500 Hz	CV, %	δ_{max}, %
IDA (5)	1.229	1.244	1.158	2.89	6.90
NFD (7)	0.622	0.610	0.651	3.28	7.34
FCV (8)	0.385	0.430	0.293	17.41	54.09

Continued Table 3.

Criterion	fs = 125 Hz	fs = 250 Hz	fs = 500 Hz	CV, %	δ_{max}, %
S_{pulse} (baseline)	4.015	2.892	8.224	63.28	77.34
Fractal dimension by the box-counting method	1.112	1.175	1.128	2.32	4.91

Source: compiled by the authors

The superiority of IDA over the integral attractor area S_pulse is confirmed with a wide margin: the coefficient of variation of IDA is 2.89% versus 63.28% for the S_{pulse} (21.9 times lower), and the difference is statistically significant according to the one-sided Wilcoxon test ($\rho = 2.1 \cdot 10^{-34}$). At the same time, the median of the maximum relative deviation δ_{max} (IDA) is 6.90% and exceeds the established threshold of 5%, so under the second acceptance-criterion condition, hypothesis H1 is not fully confirmed, since IDA should be classified as an asymptotically invariant criterion rather than a strictly invariant one.

Two additional refining results were also obtained. First, the box-counting fractal dimension proved to be no less robust to changes in sampling rate than IDA (CV 2.32% versus 2.89%, the difference being non-significant, $\rho = 0{,}86$); this does not contradict the main result, which concerns the integral attractor area, but it calls for a more precise formulation of the advantage of IDA – it lies in the combination of invariance with O(N) computational complexity and sensitivity to phase asymmetry, rather than in superiority over all known criteria on any single robustness indicator taken in isolation. Second, the phase coefficient of variation FCV shows no invariance to sampling rate (CV 17.41%, δ_{max} 54.09%), since it depends on the per-cycle segmentation procedure, which is sensitive to sample density; FCV is therefore applicable only as an indicator of hardware degradation within a fixed sensor configuration.

Hypothesis H2 is confirmed under substantial losses and is not confirmed under small ones. Under the uniform loss model, the superiority of NFD over the box-counting method is statistically significant at $\rho = 10\%$ (6.12% versus 7.81%, $\rho = 5.6 \cdot 10^{-19}$) and $\rho = 20\%$ (11.06% versus 13.68%, $\rho = 1.9 \cdot 10^{-27}$), whereas at $\rho = 5\%$ the difference is non-significant (2.81% versus 2.97%, $\rho = 0.70$). Under the block loss model, the picture is similar: a significant advantage at $\rho = 10\%$ ($\rho = 0.033$) and $\rho = 20\%$ ($\rho = 5.3 \cdot 10^{-25}$), while at $\rho = 5\%$ the NFD estimate is even inferior to the baseline (9.13% versus 6.43%). Since the acceptance criterion required a significant advantage at all three levels, hypothesis H2 should be considered only partially confirmed – for $\rho \geqslant 10\%$.

Table 4. Robustness of the criteria to packet loss in the data stream, δ (%)

Criterion	Loss model	ρ = 5 %	ρ = 10 %	ρ = 20 %	ρ = 20% with MICE
NFD (7)	uniform	2.81	6.12	11.06	6.89
NFD (7)	block	9.13	9.69	9.34	–
Box-Counting	uniform	2.97	7.81	13.68	–
Box-Counting	block	6.43	10.18	13.40	–
IDA (5)	uniform	3.55	4.32	5.38	–
IDA (5)	block	7.79	8.19	9.94	–

Source: compiled by the authors

Values are given as the mean relative deviation of the estimate from the reference value computed on the full stream ($n = 200$ segments of sample S, $f_s = 125$ Hz, $M = 32$). Separately, the compensation of the increased estimate variance by the multiple imputation module was confirmed: at $\rho = 20\%$ the NFD deviation decreases from 11.06% to 6.89%, i.e. by 37.7%. At the same time, at $\rho = 5\%$ and $\rho = 10\%$ imputation does not improve, but slightly worsens the estimate (3.90% versus 2.81% and versus 6.12% respectively), since it introduces its own recovery error, which for small gaps exceeds the error being eliminated. H3 is verified in Tables 5 and 6. The tables show mean values over five independent runs, with the standard deviation across runs given after the F_1-measure.

Table 5. Comparison of feature sets by classification quality of critical states (sample 1, $n = 12{,}241$ windows, 130 patients)

Feature set	Predictor	Accuracy	F_1	AUC
Time-domain HRV indices + SpO2	LSTM	0.858	0.840 ± 0.039	0.917
Frequency-domain HRV indices + SpO2	LSTM	0.796	0.771 ± 0.036	0.881
Phase criteria (IDA, NFD, FCV)	LSTM	0.572	0.441 ± 0.237	0.654
Combined set (9)	LSTM	0.804	0.764 ± 0.118	0.900
Combined set (9)	SVM (RBF)	0.769	0.719 ± 0.086	0.848
Combined set (9)	Random Forest	0.760	0.713 ± 0.083	0.842
Combined set (9), Z-score instead of ComBat	LSTM	0.807	0.789 ± 0.057	0.877
Combined set (9), without harmonisation	LSTM	0.742	0.684 ± 0.115	0.857

Source: compiled by the authors

Table 6. Results of the ablation analysis of the combined feature vector (9)

Vector configuration	F_1	Change in F_1 relative to the full vector	Confidence interval of the change	p-value
Full vector (9)	0.764	-	-	-
Without IDA	0.809	+0.044	[−0.012; +0.142]	0.813
Without NFD	0.761	−0.004	[−0.103; +0.096]	1.000
Without HRV	0.719	−0.045	[−0.136; +0.072]	0.438
Without SpO2	0.638	−0.127	[−0.210; −0.074]	0.063

Source: compiled by the authors

Confidence intervals of the change were computed by bootstrapping over the F_1 differences on paired runs; significance was assessed using the Wilcoxon test for paired samples. Hypothesis H3 was not confirmed on sample. F_1-measure of the combined set (9) is 0.764, whereas the set of time-domain heart rate variability indices together with the saturation level gives 0.840; the difference is not significant in favour of the combined set ($\rho = 0.91$ according to the one-sided Wilcoxon test). The ablation analysis is consistent with this result: removing IDA from the vector does not decrease but increases the F_1-measure by 0.044, removing NFD changes it within the margin of error (−0.004), whereas removing the saturation level decreases F_1 by 0.127. The phase criteria, taken alone, yield $F_1 = 0.441$ at AUC = 0.654, i.e. in this setting they carry substantially less information about the class than the rhythm variability indices.

DISCUSSION

The empirical mechanism behind the advantage of IDA over S_{pulse} differs from what would be expected from the simple assumption of the latter's additivity. If S_{pulse} scaled linearly with sample density, then across the transition 125 → 250 → 500 Hz its value would be expected to roughly double and quadruple; instead, the mean values are 4.015, 2.892 and 8.224 respectively, i.e. the dependence is neither monotonic nor proportional. A control computation on the unnormalised phase plane gives ratios of 1.07 and 0.93 instead of the expected 2 and 4, with a coefficient of variation of only 5.1% – the phase-portrait area as such is practically independent of sample density.

The reason for the discrepancy lies elsewhere: both criteria are computed on a plane normalised according to formula (3), and the scale coefficient of this normalisation is determined by the range of the derivative, which is itself unstable (its coefficient of variation across the set of frequencies is 47.4%, since the extreme values of the derivative depend on the passband, which is fixed for a given sampling rate). The IDA criterion is a ratio of two sums computed on the same plane, so the common scale factor cancels out in it; the criterion S_{pulse} is an absolute quantity and fully inherits the instability of the normalisation. Thus, the correct explanation of IDA's invariance is as follows: it is ensured not by this quantity failing to be additive, but by its being a ratio in which the normalisation scale factor cancels out. This explanation extends to any source of change in the phase-portrait scale, not only to changes in sampling rate.

This pattern of sampling-rate sensitivity is consistent with findings for related geometric descriptors of quasi-periodic biosignals. E. Mejía-Mejía *et al.* (2022), simulating PPG signals with known pulse-rate-variability content, found that the Poincaré-plot ellipse area S and the short-term descriptor SD1 showed statistically significant deviations from the gold standard at sampling rates below 128 Hz, whereas the long-term descriptor SD2 remained statistically indistinguishable from the gold standard even at 32 Hz. Unlike these area- and dispersion-based descriptors, IDA avoids this sensitivity specifically because it is a ratio rather than an absolute quantity – a mechanism not previously identified in the literature on geometric phase-space descriptors.

A control check on the simulator artefact rules out an alternative explanation of the result. The pulse-wave model implemented in the simulator has a break in the first derivative at the boundary between adjacent cycles, and decimation with anti-aliasing filtering interacts with this break. To confirm that the result is not a consequence of this particular feature of the generator, the experiment was repeated on signals pre-smoothed with a low-pass filter with a cutoff frequency of 12 Hz, which removes the break. The conclusion does not change: the coefficient of variation of IDA is 2.75% versus 2.89% in the main run, and for S_{pulse} it is 61.23% versus 63.28%. The relationship between the criteria is preserved, so it is determined by the properties of the criteria themselves rather than by a feature of the generator.

The practical significance of this result is that, when S_{pulse} is used as a feature, a predictive model trained on data from one sensor produces systematically biased predictions on data from a sensor with a different sampling rate, whereas with IDA the bias does not exceed a few percentage points – a property that is critical for heterogeneous IoMT deployments, where sensors from different manufacturers operate at different sampling rates.

The non-significance of the NFD advantage at small loss fractions ($\rho = 5\%$) has a natural explanation: at this level of loss, both estimates lie within the range of their own statistical error, and the advantage of normalising entropy by the actual number of recorded points has not yet manifested itself. The practical conclusion for the software architecture is to dispatch the recovery strategy according to the length of the gap: multiple imputation should be enabled only once the loss

fraction exceeds a threshold (approximately $\rho \geqslant 10\%$), rather than applying it unconditionally, since for small gaps it introduces its own error that exceeds the one being eliminated.

The conditional dispatch between LOCF and multivariate imputation adopted here is broadly consistent with recent benchmarking of imputation strategies for critical-care time series. M. Poette *et al.* (2026), benchmarking multiple imputation strategies on ICU monitoring data from MIMIC-IV under real-world-inspired missingness scenarios, found that methods lacking explicit temporal modelling, including MICE and MissForest, performed less favourably than simpler alternatives, with linear interpolation remaining a robust baseline for short gaps in physiological time series. Unlike this benchmark, which evaluates each imputation strategy uniformly across a dataset, the present work dispatches between LOCF and MICE within a single signal according to gap length, combining the two strategies rather than treating them as mutually exclusive alternatives.

The comparison of loss models also outlines the boundary of applicability of the unbiasedness of the IDA estimate: under uniform thinning, the IDA deviation at $\rho = 20\%$ is 5.38%, whereas under the block model it is 9.94% – nearly twice as high – which is consistent with the condition of a uniform distribution of gaps relative to the cycle phases, under which the theoretical justification was obtained.

This result contrasts with reports in the wider HRV literature, where combining nonlinear descriptors with conventional time-domain features has generally been found to improve classification performance. A.S. Eltrass *et al.* (2022), combining deep-learning features with linear and nonlinear HRV and ECG measures for multi-class ECG classification, reported that this combination improved diagnostic performance compared with time-domain features alone. The diagnostic contribution of the developed phase criteria was assessed on a synthetic sample; confirming their added value for classification therefore requires further verification on a larger sample of clinically labelled data. Thus, the result of testing hypothesis H3 requires careful interpretation in the context of the synthetic nature of sample *S*: the simulator's morphological model is not a clinically validated model of pathological states, and the distribution of information between rhythm and pulse-wave shape in it is set by construction rather than established clinically. The experiment conducted can therefore reliably answer questions about the formal properties of the criteria – invariance, robustness, computational complexity – but not questions about their clinical informativeness. Two practical conclusions follow from this. First, a gain in classification accuracy from adding geometric phase criteria does not arise automatically from the mere fact of computing them, so any claim of such a gain requires confirmation on clinically labelled data. Second, it is methodologically important to control the overlap of class distributions in the synthetic generator: in earlier trial setups, where the class parameter ranges did not overlap, the task turned out to be linearly separable by the saturation level alone, and any feature set yielded $F_1 = 1.000$, which illustrates the risk of a spurious "confirmation" of a hypothesis on synthetic data when the generator's parameters are chosen carelessly.

CONCLUSIONS

A method has been developed for constructing the feature space of the predictive software components of a biomedical monitoring platform, based on geometric invariants of the biosignal phase portrait, which combines the differential method and the Takens time-delay method for phase-space reconstruction with three computational criteria – the dynamic phase asymmetry index (IDA), the normalised fractal density of the trajectory (NFD), and the phase coefficient of shape variation (FCV).

Experimental verification on a synthetic sample (200 segments) and a multi-source clinical sample (130 patients, 12,241 windows) yielded the following results. Hypothesis H1 was partially confirmed: the advantage of IDA over the integral attractor area is 2.89% versus 63.28% in terms of the coefficient of variation ($\rho = 2.1 \cdot 10^{-34}$), but the maximum deviation of 6.90% exceeds the established tolerance of 5%, so the criterion is classified as asymptotically, rather than strictly, invariant to sampling rate. Hypothesis H2 was partially confirmed: the advantage of NFD over the box-counting method is statistically significant only at a lost-packet fraction of $\rho \geqslant 10\%$. Hypothesis H3 was not confirmed on the synthetic basis: the combined feature vector [HRV, SpO2, IDA, NFD] does not outperform the set of time-domain heart rate variability indices in terms of the F_1-measure, which requires confirmation of the diagnostic value of the phase criteria on clinically labelled data.

The developed method provides computationally efficient (O(N) for IDA and FCV, $O(N + M^2)$ for NFD) criteria suitable for stream processing on the edge nodes of IoMT platforms, whose robustness to hardware heterogeneity has been empirically confirmed within refined boundaries of applicability. A promising direction for further research is to repeat the described experimental verification methodology on clinically labelled recordings from real sources, which is a necessary condition for a definitive conclusion regarding the independent diagnostic value of the proposed criteria within the predictive component's feature vector.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V.I. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20, article number 310. doi: 10.1186/s12911-020-01332-6.

[2] Bhattacharya, P., Mukherjee, A., Bhushan, B., Gupta, S.K., Gadekallu, T.R., & Zhu, Z. (2025). A secured remote patient monitoring framework for IoMT ecosystems. *Scientific Reports*, 15, article number 22882. doi: 10.1038/s41598-025-04774-y.

[3] Braem, C.I.R., Yavuz, U.S., Hermens, H.J., & Veltink, P.H. (2024). Missing data statistics provide causal insights into data loss in diabetes health monitoring by wearable sensors. *Sensors*, 24(5), article number 1526. doi: 10.3390/s24051526.

[4] Brennan, M., Palaniswami, M., & Kamen, P. (2001). Do existing measures of Poincaré plot geometry reflect nonlinear features of heart rate variability? *IEEE Transactions on Biomedical Engineering*, 48(11), 1342-1347. doi: 10.1109/10.959330.

[5] Charlton, P.H., *et al.* (2023). The 2023 wearable photoplethysmography roadmap. *Physiological Measurement*, 44, article number 111001. doi: 10.1088/1361-6579/acead2.

[6] Chen, A.A., Beer, J.C., Tustison, N.J., Cook, P.A., Shinohara, R.T., & Shou, H. (2022). Mitigating site effects in covariance for machine learning in neuroimaging data. *Human Brain Mapping*, 43(4), 1179-1195. doi: 10.1002/hbm.25688.

[7] Chromik, J., Klopfenstein, S.A.I., Pfitzner, B., Sinno, Z.-C., Arnrich, B., Balzer, F., & Poncette, A.-S. (2022). Computational approaches to alleviate alarm fatigue in intensive care medicine: A systematic literature review. *Frontiers in Digital Health*, 4, article number 843747. doi: 10.3389/fdgth.2022.843747.

[8] de Pedro-Carracedo, J., Fuentes-Jimenez, D., Ugena, A.M., & Gonzalez-Marcos, A.P. (2020). Phase space reconstruction from a biological time series: A photoplethysmographic signal case study. *Applied Sciences*, 10(4), article number 1430. doi: 10.3390/app10041430.

[9] Dockès, J., Varoquaux, G., & Poline, J.-B. (2021). Preventing dataset shift from breaking machine-learning biomarkers. *GigaScience*, 10(9), article number giab055. doi: 10.1093/gigascience/giab055.

[10] Eltrass, A.S., Tayel, M.B., & Ammar, A.I. (2022). Automated ECG multi-class classification system based on combining deep learning features with HRV and ECG measures. *Neural Computing and Applications*, 34(11), 8755-8775. doi: 10.1007/s00521-022-06889-z.

[11] Fraser, A.M., & Swinney, H.L. (1986). Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33(2), 1134-1140. doi: 10.1103/PhysRevA.33.1134.

[12] Grassberger, P., & Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena*, 9(1-2), 189-208. doi: 10.1016/0167-2789(83)90298-1.

[13] Hemmati, R., Parvenous, M., Bahrami Zanjanbar, D., Kia, S.H., Soleimani, A., & Bakhshaei, S. (2025). Body mass index and its influence on electrocardiographic parameters in healthy and cardiovascular patients. *Obesity Science & Practice*, 11(6), article number e70107. doi: 10.1002/osp4.70107.

[14] Hentschel, H.G.E., & Procaccia, I. (1983). The infinite number of generalized dimensions of fractals and strange attractors. *Physica D: Nonlinear Phenomena*, 8(3), 435-444. doi: 10.1016/0167-2789(83)90235-X.

[15] Johnson, A.E.W., Pollard, T.J., Shen, L., Lehman, L.H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L.A., & Mark, R.G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3, article number 160035. doi: 10.1038/sdata.2016.35.

[16] Johnson, W.E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8(1), 118-127. doi: 10.1093/biostatistics/kxj037.

[17] Kennel, M.B., Brown, R., & Abarbanel, H.D.I. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, 45, 3403-3411. doi: 10.1103/PhysRevA.45.3403.

[18] Kierner, S., Kucharski, J., & Kierner, Z. (2023). Taxonomy of hybrid architectures involving rule-based reasoning and machine learning in clinical decision systems: A scoping review. *Journal of Biomedical Informatics*, 144, article number 104428. doi: 10.1016/j.jbi.2023.104428.

[19] Lachin, J.M. (2015). Fallacies of last observation carried forward analyses. *Clinical Trials*, 13(2), 161-168. doi: 10.1177/1740774515602688.

[20] Mejía-Mejía, E., May, J.M., & Kyriacou, P.A. (2022). Effects of using different algorithms and fiducial points for the detection of interbeat intervals, and different sampling rates on the assessment of pulse rate variability from photoplethysmography. *Computer Methods and Programs in Biomedicine*, 218(C), article number 106724. doi: 10.1016/j.cmpb.2022.106724.

[21] Moody, G.B., & Mark, R.G. (2001). The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3), 45-50. doi: 10.1109/51.932724.

[22] Nagai, S., Anzai, D., & Wang, J. (2017). Motion artefact removals for wearable ECG using stationary wavelet transform. *Healthcare Technology Letters*, 4(4), 138-141. doi: 10.1049/htl.2016.0100.
[23] Nelson, B.W., Low, C.A., Jacobson, N., Arean, P., Torous, J., & Allen, N.B. (2020). Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. *npj Digital Medicine*, 3(1), article number 90. doi: 10.1038/s41746-020-0297-4.
[24] Nikitchuk, T.M., Andreiev, O.V., Korenivska, O.L., & Medvediev, M.G. (2023). Model of an automated biotechnical system for analyzing pulseograms as a kind of edge devices. *Journal of Edge Computing*, 2(1), 64-83. doi: 10.55056/jec.627.
[25] Olivieri, F., Biscetti, L., Pimpini, L., Pelliccioni, G., Sabbatinelli, J., & Giunta, S. (2024). Heart rate variability and autonomic nervous system imbalance: Potential biomarkers and detectable hallmarks of aging and inflammaging. *Ageing Research Reviews*, 101, article number 102521. doi: 10.1016/j.arr.2024.102521.
[26] Poette, M., Mouysset, S., Ruiz, D., Pey, V., Alliot, J.-M., & Minville, V. (2026). Benchmarking imputation strategies for missing time-series data in critical care using real-world-inspired scenarios. *Scientific Reports*, 16, article number 8116. doi: 10.1038/s41598-026-39035-z.
[27] Rancea, A., Anghel, I., & Cioara, T. (2024). Edge computing in healthcare: Innovations, opportunities, and challenges. *Future Internet*, 16(9), article number 329. doi: 10.3390/fi16090329.
[28] Reiss, A., Indlekofer, I., Schmidt, P., & Van Laerhoven, K. (2019). Deep PPG: Large-scale heart rate estimation with convolutional neural networks. *Sensors*, 19(14), article number 3079. doi: 10.3390/s19143079.
[29] Rohila, A., & Sharma, A. (2019). Phase entropy: A new complexity measure for heart rate variability. *Physiological Measurement*, 40(10), article number 105006. doi: 10.1088/1361-6579/ab499e.
[30] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1, 206-215. doi: 10.1038/s42256-019-0048-x.
[31] Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5), 637-646. doi: 10.1109/JIOT.2016.2579198.
[32] Singh, S., Bennett, M.R., Chen, C., Shin, S., Ghanbari, H., & Nelson, B.W. (2024). Impact of skin pigmentation on pulse oximetry blood oxygenation and wearable pulse rate accuracy: Systematic review and meta-analysis. *Journal of Medical Internet Research*, 26, article number e62769. doi: 10.2196/62769.
[33] Stekhoven, D.J., & Bühlmann, P. (2012). MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112-118. doi: 10.1093/bioinformatics/btr597.
[34] Subasi, A., & Qaisar, S.M. (2023). Signal acquisition preprocessing and feature extraction techniques for biomedical signals. In S.M. Qaisar, H. Nisar & A. Subasi (Eds.), *Advances in non-invasive biomedical signal sensing and processing with machine learning* (pp. 25-52). Cham, Switzerland: Springer. doi: 10.1007/978-3-031-23239-8_2.
[35] Suboh, M.Z., Jaafar, R., Nayan, N.A., Harun, N.H., & Mohamad, M.S.F. (2022). Analysis on four derivative waveforms of photoplethysmogram (PPG) for fiducial point detection. *Frontiers in Public Health*, 10, article number 920946. doi: 10.3389/fpubh.2022.920946.
[36] Sutton, R.T., Pincock, D., Baumgart, D.C., Sadowski, D.C., Fedorak, R.N., & Kroeker, K.I. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. *NPJ Digital Medicine*, 3, article number 17. doi: 10.1038/s41746-020-0221-y.
[37] Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand & L.-S. Young (Eds.), *Dynamical systems and turbulence, Warwick 1980* (Lecture Notes in Mathematics, Vol. 898, pp. 366-381). Berlin, Germany: Springer. doi: 10.1007/BFb0091924.
[38] van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1-67. doi: 10.18637/jss.v045.i03.
[39] World Medical Association. (2013). *WMA Declaration of Helsinki – Ethical principles for medical research involving human subjects*. Retrieved from https://www.wma.net/policies-post/wma-declaration-of-helsinki/doh-oct2013/.
[40] Ye, Y., Cheng, Y., He, W., Hou, M., & Zhang, Z. (2016). Combining nonlinear adaptive filtering and signal decomposition for motion artifact removal in wearable photoplethysmography. *IEEE Sensors Journal*, 16(19), 7133-7141. doi: 10.1109/JSEN.2016.2597265.
[41] Yoon, J., Jordon, J., & van der Schaar, M. (2018). GAIN: Missing data imputation using generative adversarial nets. In J. Dy & A. Krause (Eds.), *Proceedings of the 35th International Conference on Machine Learning* (PMLR Vol. 80, pp. 5689-5698). Retrieved from https://proceedings.mlr.press/v80/yoon18a.html?.
[42] Zhang, Y., Parmigiani, G., & Johnson, W.E. (2020). ComBat-seq: Batch effect adjustment for RNA-seq count data. *NAR Genomics and Bioinformatics*, 2(3), article number lqaa078. doi: 10.1093/nargab/lqaa078.

Метод формування простору ознак на основі фазового портрета біосигналу в аналітичних підсистемах програмного забезпечення

Вадим Мацієвський
Аспірант, старший викладач
Державний університет «Житомирська політехніка»
10005, вул. Чуднівська, 103, м. Житомир, Україна
https://orcid.org/0000-0003-3947-1363
Тетяна Нікітчук
Кандидат технічних наук, доцент
Державний університет «Житомирська політехніка»
10005, вул. Чуднівська, 103, м. Житомир, Україна
https://orcid.org/0000-0002-9068-931X

Анотація. Розвиток програмного забезпечення платформ біомедичного моніторингу на основі архітектур інтернету медичних речей вимагає обчислювальних критеріїв аналізу біосигналів, стійких до неоднорідності апаратного забезпечення сенсорів і випадкових втрат даних у каналах бездротового зв'язку, оскільки традиційні геометричні характеристики фазового портрета демонструють високу чутливість до цих факторів, що ускладнює агрегацію даних із різних пристроїв у розподілених системах. Мета дослідження – представити метод формування простору ознак предиктивних програмних компонентів платформи біомедичного моніторингу на основі геометричних інваріантів фазового портрета біосигналу, стійких до коливань частоти дискретизації сенсорів та втрат пакетів даних. Для реконструкції фазового простору за дискретним сигналом застосовано диференціальний метод та метод часового запізнення за теоремою Такенса, а для вилучення ознак – три розроблені обчислювальні критерії: індекс динамічної фазової асиметрії, нормалізовану фрактальну щільність траєкторії та фазовий коефіцієнт варіації форми. Властивості критеріїв перевірено на синтетичній вибірці із двохсот сегментів пульсової хвилі за трьома частотами дискретизації та трьома рівнями штучних втрат пакетів, а їхній внесок у якість класифікації критичних фізіологічних станів – на багатоджерельній клінічній вибірці зі ста тридцяти пацієнтів методами рекурентних нейронних мереж, опорних векторів та random forest з перевіркою статистичної значущості за критерієм Вілкоксона. Індекс динамічної фазової асиметрії виявився асимптотично інваріантним до частоти дискретизації: коефіцієнт варіації його значень становив 2,89 % проти 63,28 % для інтегральної площі атрактора. Нормалізована фрактальна щільність траєкторії статистично значуще перевершила метод підрахунку клітинок за стійкістю до втрат пакетів, починаючи з десяти відсотків втрачених відліків. Перевагу комбінованого вектора фазових ознак над часовими показниками варіабельності серцевого ритму на синтетичній вибірці не підтверджено, що вказує на потребу клінічної валідації діагностичної цінності запропонованих критеріїв. Розроблені критерії мають обчислювальну складність, придатну для потокової обробки на периферійних вузлах розподілених платформ моніторингу, і забезпечують стійкість вектора ознак до апаратної гетерогенності сенсорів, що дозволяє агрегувати дані різних пристроїв без систематичного зміщення прогнозів предиктивних моделей

Ключові слова: фазовий портрет; IoMT; вилучення ознак; індекс динамічної фазової асиметрії; нормалізована фрактальна щільність; апаратна інваріантність; біомедичні сигнали