scientific reports

OPEN

Computational method of the cardiovascular diseases classification based on a generalized nonlinear canonical decomposition of random sequences

Igor Atamanyuk [1,2✉], Yuriy Kondratenko [3], Valerii Havrysh [2] & Yuriy Volosyuk [2]

Decision support systems can seriously help medical doctors in the diagnosis of different diseases, especially in complicated cases. This article is devoted to recognizing and diagnosing heart disease based on automatic computer processing of the electrocardiograms (ECG) of patients. In the general case, the change of the ECG parameters can be presented as a random sequence of the signals under processing. Developing new computational methods for such signal processing is an important research problem in creating efficient medical decision support systems. Authors consider the possibility of increasing the diagnostic accuracy of cardiovascular diseases by implementing of the new proposed computational method of information processing. This method is based on the generalized nonlinear canonical decomposition of a random sequence of the change of cardiogram parameters. The use of a nonlinear canonical model makes it possible to significantly simplify the maximum likelihood criterion for classifying diseases. This simplification is provided by the transition from a multi-dimensional distribution density of cardiogram parameters to a product of one-dimensional distribution densities of independent random coefficients of a nonlinear canonical decomposition. The absence of any restrictions on the class of random sequences under study makes it possible to achieve maximum accuracy in diagnosing cardiovascular diseases. Functional diagrams for implementing the proposed method reflecting the features of its application are presented. The quantitative parameters of the core of the computational diagnostic procedure can be determined in advance based on the preliminary statistical data of the ECGs for different heart diseases. That is why the developed method is quite simple in terms of computation (computing complexity, accuracy, computing time, etc.) and can be implemented in medical computer decision systems for monitoring cardiovascular diseases and for their diagnosis in real time. The results of the numerical experiment confirm the high accuracy of the developed method for classifying cardiovascular diseases.

Medical statistics show[1–3] that currently the main cause of death in the world (more than 30%) is diseases of the cardiovascular system (including among people of working age). Therefore, timely high-precision diagnostics of heart diseases, prevention and treatment at an early stage of the development of the disease acquire exceptional relevance. To improve the accuracy of diagnosing the state of the heart in recent decades, computer systems for automatic analysis[4–6] of electrocardiographic data obtained during the processing of an electrocardiographic signal have been widely used. Automatic analysis of electrocardiograms is a rather complex theoretical problem. First of all, this is due to the physiological origin of the signal[7–9], which is the reason for its indeterminacy, diversity, variability, unpredictability, non-stationarity and susceptibility to numerous types of interference.

[1]Warsaw University of Life Science, Nowoursynowska Str. 166, 02-787 Warsaw, Poland. [2]Mykolayiv National Agrarian University, Georgi Gongadze Str. 9, Mykolaiv 54020, Ukraine. [3]Petro Mohyla Black Sea National University, 68th Desantnykiv Str. 10, Mykolaiv 54003, Ukraine. ✉email: ihor_atamaniuk@sggw.edu.pl

At present, the analysis process, as a rule, is a study of isopotential and other maps generated from the data obtained using the software supplied with the signal recording apparatus.

In medical practice, the conclusions of cardiologists about patients' diagnoses have, as usual, qualitative or verbal character and are not always confirmed by enough number of quantitative data. In special or difficult situations with disease recognition, diagnosis errors by young or insufficiently experienced medical doctors are possible, and the real diagnostic process may be significantly extended until a final correct decision about the truth diagnosis.

Decision support systems can seriously help medical doctors in the decision-making processes about the diagnosis of different heart diseases, especially in complicated cases. The most perspective approach is based on the recognizing and diagnosing heart disease using automatic computer processing of the electrocardiograms (ECG) of patients.

In the general case, the change of the ECG parameters can be presented as a random sequence of the signals under processing. Developing new computational methods for such signal processing is an important research problem in creating efficient medical decision support systems.

In this article, the authors consider the possibility of increasing the diagnostic accuracy of cardiovascular diseases by implementing the computational method, which is based on the generalized nonlinear canonical decomposition of a random sequence of the change of cardiogram parameters. The absence of any restrictions on the class of random sequences under study makes it possible to achieve maximum accuracy in diagnosing cardiovascular diseases. The main advantage is that quantitative parameters of the computational diagnostic procedure can be determined in advance based on the preliminary statistical data of the ECGs for different heart diseases. That is why the developed method is quite simple in terms of computation (computing complexity, accuracy, computing time, etc.) and can be implemented in medical computer decision systems for monitoring cardiovascular diseases and for their diagnosis in real time.

Thus, the development of efficient mathematical models and computation methods for identifying the high-accuracy individual characteristics of an electrocardiogram (with subsequent classification), as well as the creation of an automated computer diagnostic support system, is an urgent and important task in "medicine–computer science" multidisciplinary research.

The rest of the article covers multiple aspects related to the topic discussion. "Background and analysis of the related works" section consists of the analysis of the related works in the field of ECG processing. In "Problem statement" section authors formulate the problem statement. "Solution" section deals with the development of the computation method and corresponding mathematical model based on the generalized nonlinear canonical decomposition of a random sequence of the change of cardiogram parameters. "Results of the numerical experiment" section represents the modeling and simulation results for different existing methods and comparative results with the proposed computational method. The paper ends with a conclusion in "Conclusion" section.

Background and analysis of the related works

Diagnostics of electrocardiograms consist of three successive stages: (a) preliminary processing, (b) feature extraction (normalization), and (c) classification. Let us analyze all these stages consequently.

Fist stage—Preprocessing reduces signal measurement noise by smoothing the electrocardiogram signal, reducing drift suppression and baseline deviation. The most common existing methods used to reduce signal noise are (a) second order low pass and (b) high pass Butterworth filters[10], (c) Daubechies wavelet[11] and (d) orthogonal wavelet-filter[12]. Besides, for baseline adjustment, such techniques as median filter, linear phase high pass filter, mean median filter and others are used also.

Second stage—Feature extraction is an interactive process that includes a series of automatic data transformation procedures. In cases with a large number of measurements-features that describe the characteristics of the input signal, the correlation and factor analysis of data can be used to reduce the dimension of the problem. According to the extraction and analysis methods, the features can be divided into the following categories:

- Temporary features[13] (these features are described in the time domain, representing amplitude, slope and heart rate);
- Spectral features[14,15] (features are defined in the frequency domain, account for spectral concentration, normalized spectral moments);
- Time-frequency/wavelet features[16,17] (features extracted from the results of the wavelet transform applied to the electrocardiogram signal);
- Signs of the complexity of geometric distortions[18] (these signs include various calculations related to the complexity of the considered segment of the electrocardiogram).

The stage of feature extraction ends with the optimization of their number, which allows reducing the set of redundant functions, reducing computational costs and improving the overall performance of the system. This step uses the following three main categories of feature selection methods: (a) wrapper methods (recursive feature elimination[14]; direct feature selection[19]; genetic algorithms[20]; (b) filter methods (correlation)[21]; Chi-Square[15]; analysis of variance (ANOVA)[22]; ReliefF[23]; (c) built-in methods[24].

To date, a large number of approaches have been developed to solve the problem of diagnosing cardiovascular diseases using various mathematical methods.

Third stage—The classification process can be carried out in several iterations, depending on the chosen recognition scheme. In some cases, the results obtained at this stage require a revision of the entire processing scheme as a whole. The most common classification methods are: discriminant function[25]; cluster analysis[26];
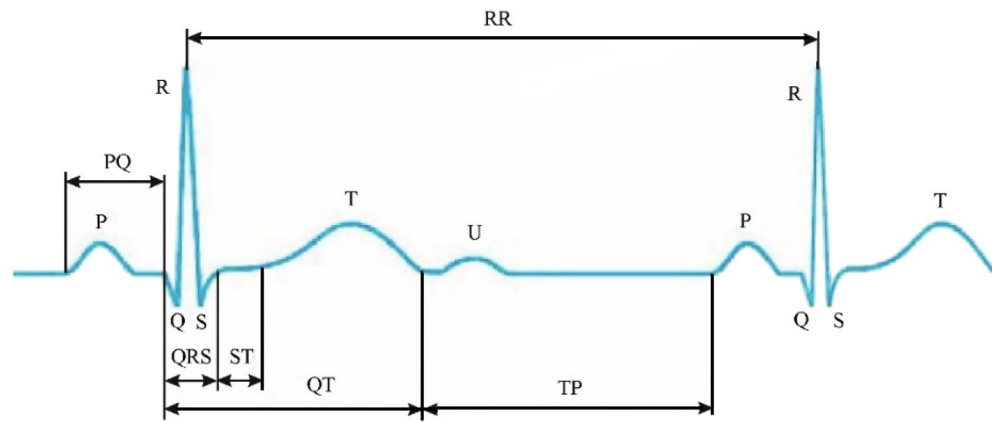
Figure 1. Cardiogram parameters.

artificial neural network[27]; Naive Bayes classifier[28]; support vector machine[29]; k-Nearest Neighbors (k-NNs)[30]; Decision Trees (DT)[31].

Several different approaches for ECG analysis are based on a chaos theory[32], a combination of statistical, geometric, and nonlinear heart rate variability features[32], a semantic web ontology and heart failure expert system[33], signal averaging method, multivariate analysis[34], RPCA—recursive principal component analysis[35], SPSA—simultaneous perturbation stochastic approximation method[36], ABT—Amplitude Based Technique FDBT—First Derivative Based Technique, SDBT—Second Derivative Based Technique[37], Hilbert transform[38] and others.

At the same time, each of the above-mentioned methods has its drawbacks and limitations. That is why the need to develop new effective methods of medical diagnostics has not lost its relevance.

Thus, the change in the values of the electrocardiogram has a stochastic character; therefore, for the diagnosis of cardiovascular diseases, it is necessary to use methods for recognizing random functions and random sequences.

The main method for recognition of the realizations of random sequences is the Bayes decision rule, according to which a decision about the belonging of the realization to a certain class, for which the posterior probability is maximum, is made. The method is theoretically accurate, however, as well as many of its modifications (the Neyman-Pearson, Wald criterion, etc.[39]) is applied in conditions when the stochastic properties of classes of random sequences are fully known. If the prior probabilities of the classes of random sequences are not known, then equal values are assigned to them and the decision rule is modified into the maximum likelihood criterion. The criterion is especially important when solving problems of recognition, in which unlikely events cannot be excluded from consideration (diagnostics of emergency technical systems, medical diagnostics, etc.). However, for the maximum likelihood method, as well as for the Bayes rule, the problem of approximating the multivariate distribution density for a random sequence with a large number of sampling points remains unresolved.

The use of the canonical expansion of Pugachev[40] makes it possible to pass the decision rule from a multivariate distribution density to the product of one-dimensional distribution densities of uncorrelated random coefficients. However, this approach is valid only for Gaussian random sequences.

The aim of our study is to eliminate the abovementioned disadvantage and develop a diagnostic method that takes into account non-linear stochastic features of changes in cardiogram parameters.

Problem statement

The accumulated volume of statistical data on various cardiovascular diseases makes it possible to determine with high accuracy the characteristics $E\left[C^{\xi_g}\left(i - r_{g-1}\right)\cdots C^{\xi_2}(i - r_1)C^{\xi_1}(i)\right], \sum_{j=1}^{g}\xi_j \leq N, r_j = \overline{1, i-1}, i = \overline{1, I}$ ($E[]$-expected value) of random sequences $C(i/k)$, $i = \overline{1, I}, k = \overline{1, K}$ describing the changes in the values of information signs of an electrocardiogram (K includes ($K - 1$) different diseases and one normal state of a person, I—the number of information signs of an electrocardiogram). For example, $C(i/1)$, $i = \overline{1, I}$—Marfan syndrome; $C(i/2)$, $i = \overline{1, I}$ pulmonary embolism; $C(i/3)$, $i = \overline{1, I}$—heart attack; $C(i/4)$, $i = \overline{1, I}$—cardiomyopathy; $C(i/5)$, $i = \overline{1, I}$—pericardial disease; $C(i/6)$, $i = \overline{1, I}$—rheumatic heart disease; $C(i/7)$, $i = \overline{1, I}$—stroke; $C(i/8)$, $i = \overline{1, I}$—normal state of a person; in this case $K = 8$.

As a result of electrocardiography, a certain sequence of values $c(i)$, $i = \overline{1, I}$ can be obtained. It is necessary to determine to which class $k^* \in \{1, .., K\}$ this realization $c(i)$, $i = \overline{1, I}$ belongs.

Solution

Taking into account the specific features of the problem of diagnosing cardiovascular diseases, to analyze the cardiogram, as a rule[41,42], a random sequence $\{C\} = \{C(1), C(2), ..., C(I)\}$, $I = 14$ with fourteen elements is used. Each element corresponds to one most informative parameter of the electrocardiogram (Fig. 1), in particular: $C(1)$ is the height of the tooth P; $C(2)$—the width of the tooth P; $C(3)$—the height of the tooth Q; $C(4)$—the interval PQ; $C(5)$—the height of the first tooth R; $C(6)$—the interval QRS; $C(7)$—the height of the tooth S; $C(8)$

—the interval RR; $C(9)$—the height of the tooth T; $C(10)$—the interval QT; $C(11)$—the interval ST; $C(12)$—the interval TP; $C(13)$—the width of the tooth U; $C(14)$—the height of the tooth U.

A universal method for recognizing a random sequence is the maximum likelihood criterion[43,44], according to which the decision on whether the realization $\vec{c} = \{c(1), ..., c(14)\}$ belongs to the class $k^* \in \{1, .., K\}$ is made when the following condition is met[45,46]:

$$k^* = \arg\max_k \{f_I(\vec{c}/k)\}, \ I = 14, \tag{1}$$

where $f_I(\vec{c}/k)$, $k = \overline{1, K}$, $I = 14$ is the conditional distribution density of the features $\vec{c}$, provided that the realization belongs to this class.

Thus, to solve the recognition problem, it is necessary to obtain the estimate of unknown densities $f_I(\vec{c}/k)$, $k = \overline{1, K}$, which, in turn, taking into account rather large dimension ($I = 14$) of the function, is a complex and time-consuming procedure. In the case of using the simplifying assumption ($E[C^\nu(j)C^\mu(i)] \neq 0$, $E[C^{\xi_g}(i - p_{g-1})...C^{\xi_2}(i - p_1)C^{\xi_1}(i)] = 0$) about the presence in a random sequence of only stochastic relations between two arbitrary parameters, the problem is greatly simplified by means of transition from the sequence $C(i)$, $i = \overline{1, 14}$ to the analysis of a set of independent random coefficients $P_i^{(N)}$, $i = \overline{1, 14}$ of the canonical expansion[47,48]:

$$C(i) = E[C(i)] + \sum_{\nu=1}^{i-1} \sum_{j=1}^{N} P_\nu^{(j)} T_{\lambda\nu}^{(j)}(i) + P_i^{(1)}, \quad i = \overline{1, 14}; \tag{2}$$

$$P_i^{(\lambda)} = C^\lambda(i) - \sum_{\nu=1}^{i-1} \sum_{j=1}^{N} P_\nu^{(j)} T_{\lambda\nu}^{(j)}(i) - \sum_{j=1}^{\lambda-1} P_i^{(j)} T_{\lambda i}^{(j)}(i), \quad i = \overline{1, 14}. \tag{3}$$

The coordinate functions $T_{h\nu}^{(\lambda)}(i)$ are determined from the relation[49,50]:

$$T_{h\nu}^{(\lambda)}(i) = \frac{1}{D_\lambda(\nu)} \left(E\left[C^\lambda(\nu)C^h(i)\right] - E\left[C^\lambda(\nu)\right] E\left[C^h(i)\right] - \sum_{\rho=1}^{\nu-1} \sum_{j=1}^{N} D_j(\rho) \right.$$
$$\left. \times \ T_{\lambda\mu}^{(j)}(\nu) T_{h\rho}^{(j)}(i) - \sum_{j=1}^{\lambda-1} D_j(\nu) T_{\lambda\nu}^{(j)}(\nu) T_{h\nu}^{(j)}(i) \right), \quad \nu = \overline{1, i}, \ i = \overline{1, 14}. \tag{4}$$

where $D_\lambda(i)$—the variances of the random coefficients $P_i^{(\lambda)}$

$$D_\lambda(i) = E\left[C^{2\lambda}(i)\right] - E^2\left[C^\lambda(i)\right] - \sum_{\rho=1}^{i-1} \sum_{j=1}^{N} D_j(\rho)$$
$$\times \ \left\{T_{\lambda\rho}^{(j)}(i)\right\}^2 - \sum_{j=1}^{\lambda-1} D_j(i) \left\{T_{\lambda i}^{(j)}(i)\right\}^2, \quad i = \overline{1, 14}. \tag{5}$$

In this case, the substitution of $\vec{c}$ by the vector $\vec{p}$ considering $f_I(\vec{p}/k) = \prod_{i=1}^{I} f_1\left(p_i^{(N)}/k\right)$, $k = \overline{1, K}$, $I = 14$ allows us to write the decision rule in the following form

$$k^* = \arg\max_j \prod_{i=1}^{I} f_1\left(p_i^{(N)}/k\right), \quad k = \overline{1, K}, I = 14. \tag{6}$$

The problem of recognition, therefore, is reduced to a successive approximation of twelve one-dimensional distribution densities.

The decision rule (1) is significantly simplified, however, the transition from the vector $\vec{c}$ to the vector $\vec{p}$ is possible provided that the random sequence $C(i)$, $i = \overline{1, 14}$ has only stochastic relations $E[C^\nu(j)C^\mu(i)]$.

To eliminate all existing probabilistic relationships $E\left[C^{\xi_g}(i - r_{g-1})...C^{\xi_2}(i - r_1)C^{\xi_1}(i)\right]$, we introduce into consideration the array of random variables:

$$C = \begin{bmatrix} C(i) \\ C^2(i) \\ \cdots \\ C^{N-1}(i) \\ C^{\xi_2^{(2)}}\left(i - r_1^{(2)}\right) C^{\xi_1^{(2)}}(i) \\ \cdots \\ C^{\xi_g^{(g)}}\left(i - r_{g-1}^{(g)}\right) \cdots C^{\xi_2^{(g)}}\left(i - r_1^{(g)}\right) C^{\xi_1^{(g)}}(i) \end{bmatrix}. \tag{7}$$

The parameters of the matrix C are determined by the expressions

$$i = \overline{1,I}; \; \xi_\mu^{(g)} = \overline{1, \xi_\mu'^{(g)}}, \; \mu = \overline{1,l}; \; r_j^{(g)} = \overline{r_{j-1}^{(g)} + 1, r_j'^{(g)}}, \; j = \overline{1, l-1}, r_0^{(g)} = 0; \; l = \overline{1, N-1};$$

$$r_j'^{(g)} = \begin{cases} 0, & \text{for } \left(j \neq \overline{1, g-1}\right) \vee (g = 1); \\ i - g + j, & \text{for } j = \overline{1, g-1}, \; g > 1; \end{cases}$$

$$\xi_\mu'^{(g)} = N - 1 - g + \mu - \sum_{j=1}^{\mu-1} \xi_j'^{(g)}, \mu = \overline{1, g}.$$

A priori information $E\left[C^{\xi_g}\left(i - r_{g-1}\right) \cdots C^{\xi_2}(i - r_1) C^{\xi_1}(i)\right]$ about the sequence $C(i)$, $i = \overline{1, 14}$ can be obtained by determining the cross-correlation of the elements the array C. Consider this array as a vector random sequence $\vec{C}$, each component of which corresponds to a row of the array C. Applying the vector linear canonical decomposition to $\vec{C}$ gives the following expression for the first component of (7).

$$
\begin{aligned}
C(i) =\; & E[C(i)] + \sum_{\nu=1}^{i-1} \sum_{\xi_1^{(1)}=1}^{N-1} P_{\xi_1^{(1)}}(\nu) \omega\left(\nu, \xi_1^{(1)}/i; 1\right) + P_1(i) \\
& + \sum_{\nu=1}^{i-1} \sum_{g=2}^{M(\nu)} \sum_{r_1^{(g)}=1}^{r_1'^{(g)}} \cdots \sum_{r_{g-1}^{(g)}=r_{g-2}^{(g)}+1}^{r_{g-1}'^{(g)}} \sum_{\xi_1^{(g)}=1}^{\xi_1'^{(g)}} \cdots \sum_{\xi_g^{(g)}=1}^{\xi_g'^{(g)}} P_{r_1^{(g)} \cdots r_{g-1}^{(g)}; \xi_1^{(g)} \cdots \xi_g^{(g)}}(\nu) \\
& \times \omega\left(\nu; r_1^{(g)} \cdots r_{g-1}^{(g)}, \xi_1^{(g)} \cdots \xi_g^{(g)}/i; 1\right), \quad i = \overline{1, 14},
\end{aligned}
\tag{8}
$$

where $M(\nu) = \begin{cases} N - 1, & \text{for } \nu \geq N - 1; \\ \nu, & \text{for } \nu < N - 1. \end{cases}$

The coordinate functions $\omega(\nu; \alpha_1/i; b_1 \cdots b_{m-1}; a_1 \cdots a_m)$, $\omega(\nu; \beta_1, \ldots \beta_{n-1}; \alpha_1, \ldots \alpha_n/i; b_1, \ldots b_{m-1}; a_1, \ldots a_m)$ of the canonical decomposition (8) are determined by the relations:

$$
\begin{aligned}
\omega(\nu; \alpha_1/i; b_1 \cdots b_{m-1}; a_1 \cdots a_m) =\; & \frac{1}{D_{\alpha_1}(\nu)} \Big\{ E\left[C^{\alpha_1}(\nu) C^{a_m}(i - b_{m-1}) \cdots C^{a_1}(i)\right] \\
& - E\left[C^{\alpha_1}(\nu)\right] M\left[C^{a_m}(i - b_{m-1}) \cdots C^{a_1}(i)\right] \\
& - \sum_{\lambda=1}^{\nu-1} \sum_{\xi_1^{(1)}=1}^{N-1} D^{\xi_1^{(1)}}(\lambda) \omega\left(\lambda; \xi_1^{(1)}/\nu; \alpha_1\right) \omega\left(\lambda; \xi_1^{(1)}/i; b_1 \cdots b_{m-1}; a_1 \cdots a_m\right) \\
& - \sum_{\xi_1^{(1)}=1}^{\alpha_1-1} D^{\xi_1^{(1)}}(\nu) \omega\left(\nu; \xi_1^{(1)}/\nu; \alpha_1\right) \omega\left(\nu; \xi_1^{(1)}/\nu; b_1 \cdots b_{m-1}; a_1 \cdots a_m\right) \\
& - \sum_{\lambda=1}^{\nu-1} \sum_{g=2}^{M(\lambda)} \sum_{r_1^{(g)}=1}^{r_1'^{(g)}} \cdots \sum_{r_{g-1}^{(g)}=r_{g-2}^{(g)}+1}^{r_{g-1}'^{(g)}} \sum_{\xi_1^{(g)}}^{\xi_1'^{(g)}} \cdots \sum_{\xi_g^{(g)}}^{\xi_g'^{(g)}} D^{r_1^{(g)} \cdots r_{g-1}^{(g)}; \xi_1^{(g)} \cdots \xi_g^{(g)}}(\lambda) \\
& \times \omega\left(\lambda; r_1^{(g)} \cdots r_{g-1}^{(g)}; \xi_1^{(g)} \cdots \xi_g^{(g)}/\nu; \alpha_1\right) \\
& \times \omega\left(\lambda; r_1^{(g)} \cdots r_{g-1}^{(g)}; \xi_1^{(g)} \cdots \xi_g^{(g)}/\nu; b_1 \cdots b_{m-1}; a_1 \cdots a_m\right) \Big\}, \quad \nu = \overline{1, 14}.
\end{aligned}
\tag{9}
$$

$$\omega\left(v; \beta_1, \ldots \beta_{n-1}; \alpha_1, \ldots \alpha_n/i; b_1, \ldots b_{m-1}; a_1, \ldots a_m\right) = \frac{1}{D_{\beta_1, \ldots \beta_{n-1}; \alpha_1, \ldots \alpha_n}(v)}$$

$$\left\{ E\left[C^{\alpha_n}(v - \beta_{n-1}) \cdots C^{\alpha_1}(v) \times C^{a_m}(i - b_{m-1}) \cdots C^{a_1}(i)\right] - M\left[C^{\alpha_n}(v - \beta_{n-1}) \cdots C^{\alpha_1}(v)\right]\right.$$

$$\times E\left[C^{a_m}(i - b_{m-1}) \cdots C^{a_1}(i)\right] - E\left[C^{\alpha_n}(v - \beta_{n-1}) \cdots C^{\alpha_1}(v)\right] E\left[C^{a_m}(i - b_{m-1}) \cdots C^{a_1}(i)\right]$$

$$- \sum_{\lambda=1}^{v-1} \sum_{\xi_1^{(1)}=1}^{N-1} D^{\xi_1^{(1)}}(\lambda) \omega\left(\lambda; \xi_1^{(1)}/v; \beta_1, \ldots \beta_{n-1}; \alpha_1, \ldots \alpha_n\right) \omega\left(\lambda; \xi_1^{(1)}/i; b_1 \cdots b_{m-1}; a_1 \cdots a_m\right)$$

$$- \sum_{\lambda=1}^{v-1} \sum_{g=2}^{M(\lambda)} \sum_{r_1^{(l)}=1}^{p_1^{'(g)}} \cdots \sum_{r_{g-1}^{(g)}=r_{g-2}^{(g)}+1}^{p_{g-1}^{'(g)}} \sum_{\xi_1^{(g)}=1}^{\xi_1^{'(g)}} \cdots \sum_{\xi_g^{(g)}=1}^{\xi_g^{'(g)}} D^{r_1^{(g)} \ldots r_{g-1}^{(g)}; \xi_1^{(g)} \ldots \xi_g^{(g)}}(\lambda)$$

$$\times \omega\left(\lambda; r_1^{(g)} \cdots r_{g-1}^{(g)}; \xi_1^{(g)} \cdots \xi_g^{(g)}/v; \beta_1, \ldots \beta_{n-1}; \alpha_1, \ldots \alpha_n\right)$$

$$\times \omega\left(\lambda; r_1^{(g)} \cdots r_{g-1}^{(g)}; \xi_1^{(g)} \cdots \xi_g^{(g)}/i; b_1, \ldots b_{m-1}; a_1, \ldots a_m\right) \tag{10}$$

$$- \sum_{l=2}^{n-1} \sum_{p_1^{(g)}=1}^{p_1^{'(g)}} \cdots \sum_{p_{g-1}^{(g)}=p_{g-2}^{(g)}+1}^{p_{g-1}^{'(g)}} \sum_{\xi_1^{(g)}=1}^{\xi_1^{'(g)}} \cdots \sum_{\xi_g^{(g)}=1}^{\xi_g^{'(g)}} D^{p_1^{(g)} \ldots p_{g-1}^{(g)}; \xi_1^{(g)} \ldots \xi_g^{(g)}}(v)$$

$$\times \omega\left(v; p_1^{(g)} \cdots p_{g-1}^{(g)}; \xi_1^{(g)} \cdots \xi_g^{(g)}/v; \beta_1, \ldots \beta_{n-1}; \alpha_1, \ldots \alpha_n\right)$$

$$\times \omega\left(v; b_1, \ldots b_{m-1}; a_1, \ldots a_m/i; p_1^{(g)} \cdots p_{g-1}^{(g)}; \xi_1^{(g)} \cdots \xi_g^{(g)}\right)$$

$$- \sum_{r_1^{(n)}=1}^{r_1^{*(n)}} \cdots \sum_{r_{l-1}^{(n)}=r_{l-2}^{(n)}+1}^{r_{n-1}^{*(n)}} \sum_{\xi_1^{(n)}=1}^{\xi_1^{*(n)}} \cdots \sum_{\xi_n^{(n)}=1}^{\xi_n^{*(n)}} D^{r_1^{(n)} \ldots r_{n-1}^{(n)}; \xi_1^{(n)} \ldots \xi_n^{(n)}}(v)$$

$$\times \omega\left(v; r_1^{(n)} \cdots r_{n-1}^{(n)}; \xi_1^{(n)} \cdots \xi_n^{(n)}/v; \beta_1, \ldots \beta_{n-1}; \alpha_1, \ldots \alpha_n\right)$$

$$\left. \times \omega\left(v; r_1^{(n)} \cdots r_{n-1}^{(n)}; \xi_1^{(n)} \cdots \xi_n^{(n)}/i; b_1, \ldots b_{m-1}; a_1, \ldots a_m\right)\right\}, \quad v = \overline{1, 14};$$

where $D^{\alpha_1}(v)$, $D^{\beta_1, \ldots \beta_{n-1}; \alpha_1, \ldots \alpha_n}(v)$—variances of random coefficients $P_{\alpha_1}(v)$, $P_{\beta_1, \ldots \beta_{n-1}; \alpha_1, \ldots \alpha_n}(v)$.

The values $b_1, \ldots, b_{m-1}$; $a_1, \ldots, a_m$ change in the intervals $b_\mu \in \left[b_{\mu-1}^{(m)}; b_\mu^{'(m)}\right]$ and $a_\mu \in \left[1; a_\mu^{'(m)}\right]$, $m = \overline{1 - N}$ respectively. The right boundaries of the intervals are determined from the formulas

$$b_\mu^{'(m)} = \begin{cases} 0, & \left(\mu \neq \overline{1, m-1}\right) \vee (m = 1); \\ i - m + \mu, & \mu = \overline{1, m-1}, m > 1; \end{cases}$$

$$a_\mu^{'(m)} = N - 1 - m + \mu - \sum_{j=1}^{\mu-1} a_j^{(m)}, \mu = \overline{1, m}.$$

The block diagram of the algorithm for calculating the parameters of the canonical expansion (8) is shown in Fig. 2.

Expression (8) is a non-linear canonical decomposition of the investigated sequence with full consideration of the stochastic properties $E\left[C^{\xi_g}(i - r_{g-1}) \ldots C^{\xi_2}(i - r_1) C^{\xi_1}(i)\right]$. Therefore, the random coefficients $P_{N-1}(1)$, $P_{1;1,N-2}(2)$, $P_{1,2;1,1,N-3}(3)$, $\ldots$, $P_{14-N+1,14-N+2,\ldots,11;1,1\ldots1}(14)$, calculated at the last iteration, are independent random variables and the decision rule takes the form:

$$k^* = \arg\max_k \left\{ f_1\left(p_{N-1}(1)\right) f_1\left(p_{1;1,N-2}(2)\right) \ldots f_1\left(p_{14-N+1,14-N+2,\ldots,13;1,1\ldots1}(14)\right)/k, k = \overline{1, K}\right\} \tag{11}$$

The absence of assumptions about the form of the distribution density of random variables $P_{N-1}(1)$, $P_{1;1,N-2}(2)$, $P_{1,2;1,1,N-3}(3)$, $\ldots$, $P_{14-N+1,14-N+2,\ldots,13;1,1\ldots1}(14)$ leads to the need to use nonparametric methods to describe them. The simplest and most efficient approach under these conditions is the use of nonparametric Parzen-type estimates[51].

The computational method for diagnosing cardiovascular diseases consists in the realization of the following stages:

(1) Estimation of moment functions $E\left[C^{\xi_g}(i - r_{g-1}) \ldots C^{\xi_2}(i - r_1) C^{\xi_1}(i)\right]$, $\sum_{j=1}^{g} \xi_j \leq N$, $r_j = \overline{1, i-1}$, $i = \overline{1, 14}$

based on statistical information $c_l(i)$, $i = \overline{1, 14}$, $l = \overline{1, L}$;

(2) Formation of canonical decompositions (8) for various classes (diseases) of random sequences;

(3) Synthesis of one-dimensional distribution densities of independent random coefficients $P_{N-1}(1)$, $P_{1;1,N-2}(2)$, $P_{1,2;1,1,N-3}(3)$, $\ldots$, $P_{14-N+1,14-N+2,\ldots,13;1,1\ldots1}(14)$;
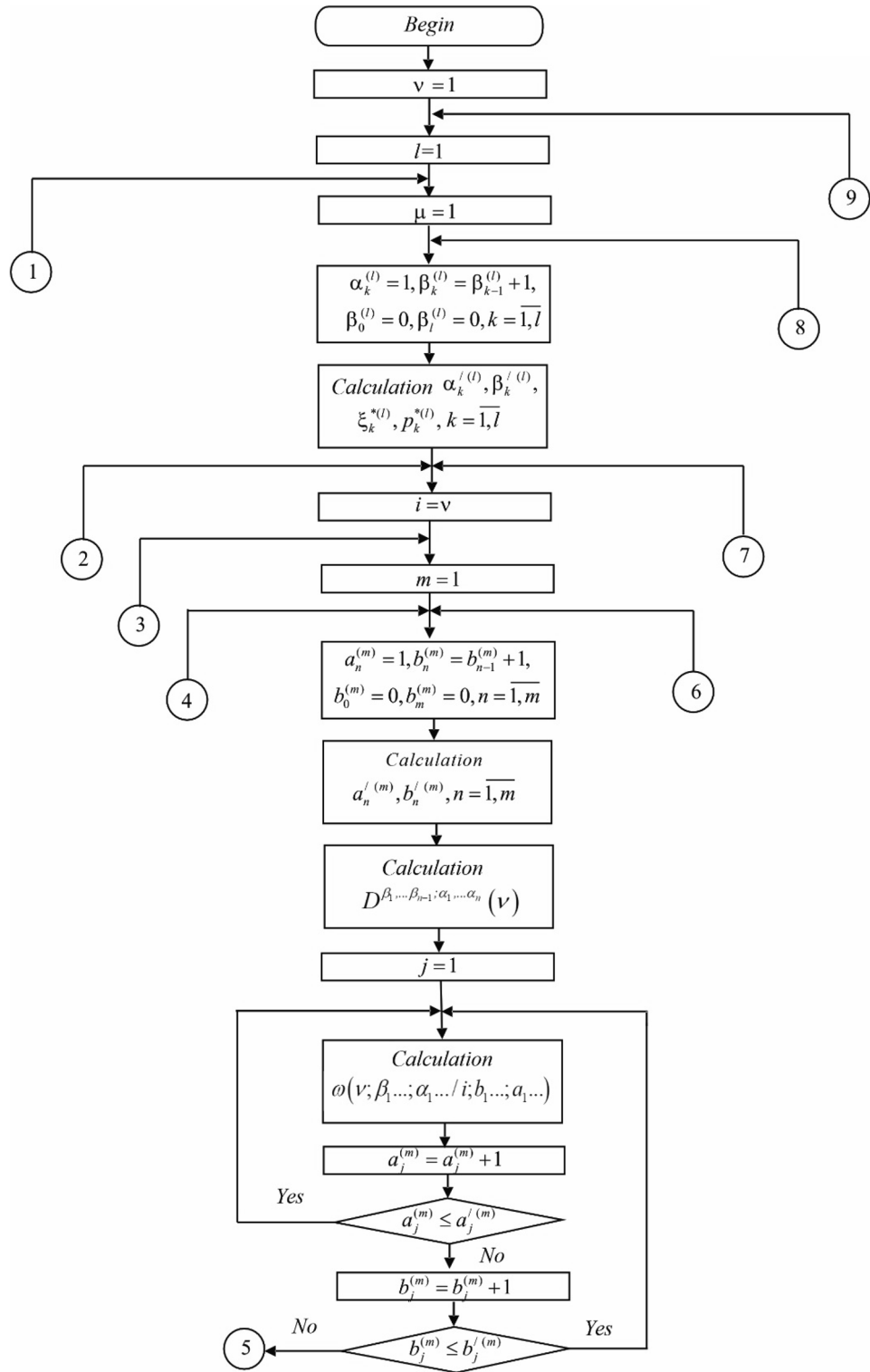
Figure 2. Block diagram of the algorithm for calculating the parameters of the canonical expansion (8).

(4) Calculation of the values $p_{N-1}(1), p_{1;1,N-2}(2), p_{1,2;1,1,N-3}(3), \ldots, p_{14-N+1,14-N+2,\ldots,13;1,1\ldots1}(14)$ for some cardiogram $\vec{c} = \{c(1), \ldots, c(14)\}$;

(5) Determination of the belonging of the cardiogram $\vec{c} = \{c(1), \ldots, c(14)\}$ to a certain class based on the decision rule (11).

The diagram of the functioning of the system of diagnostics of cardiovascular diseases based on the developed method is shown in Fig. 3.
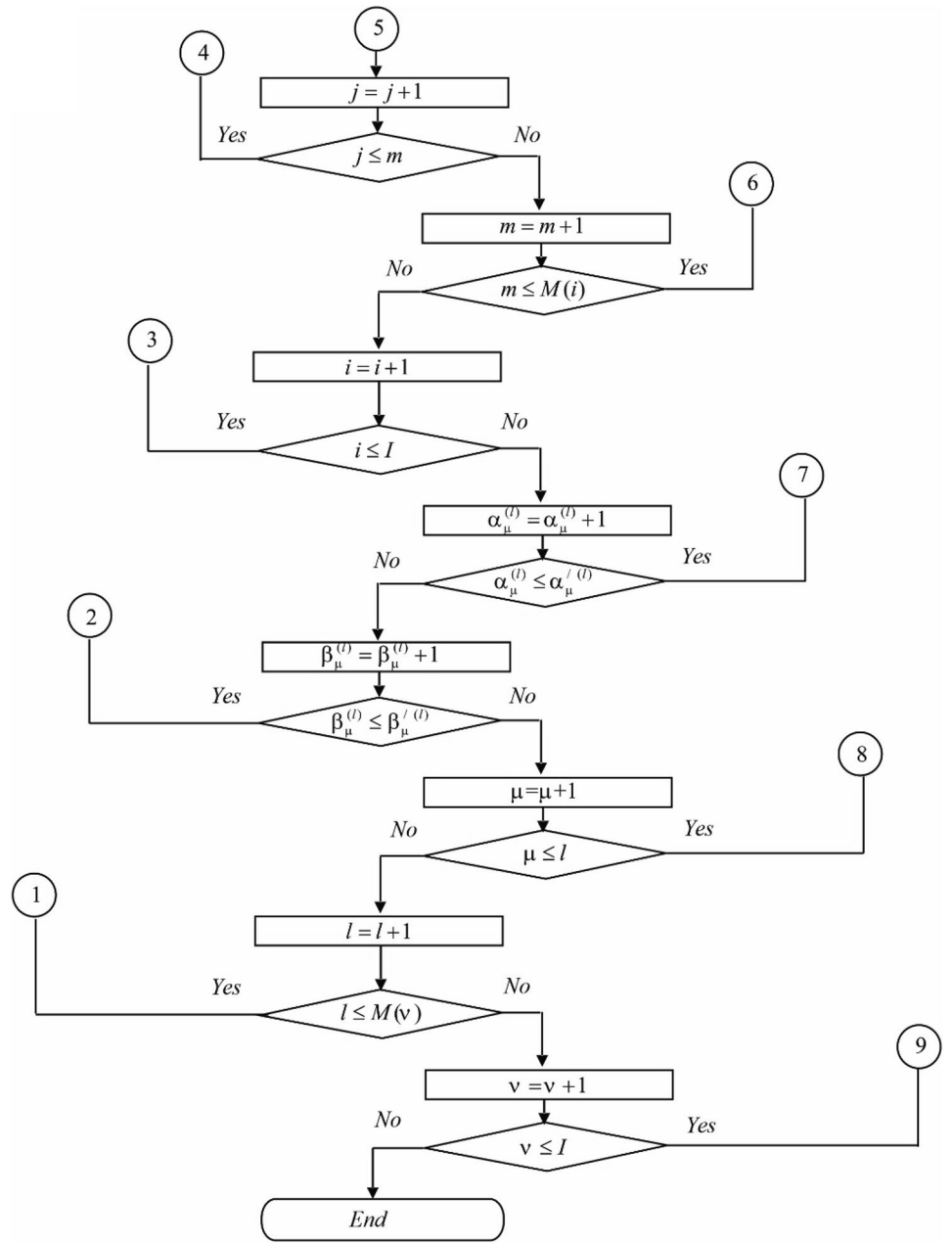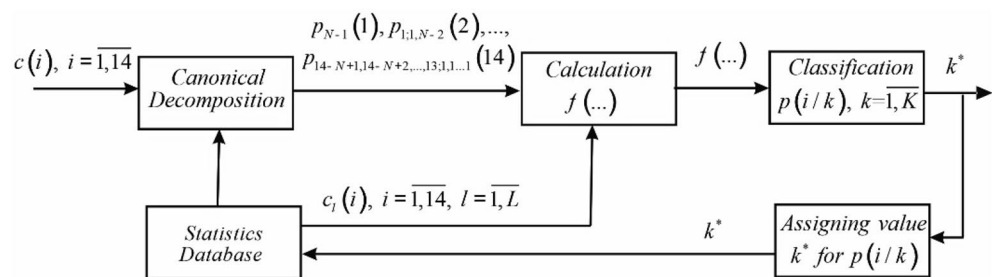
Figure 2. (continued)

Figure 3. Diagram of the functioning of the system of diagnostics of cardiovascular diseases.

Results of the numerical experiment

The proposed method was tested on the basis of statistical data of nine cardiovascular diseases: (a) mild neurocirculatory dystonia, $\{C(i)/1\}$, $i = \overline{1, 14}$; (b) neurocirculatory dystonia of moderate degree, $\{C(i)/2\}$, $i = \overline{1, 14}$; (c) severe neurocirculatory dystonia, $\{C(i)/3\}$, $i = \overline{1, 14}$; (d) stenocardia of the first functional class, $\{C(i)/4\}$, $i = \overline{1, 14}$;

(e) hypertrophy of myocardium, $\{C(i)/5\}$, $i = \overline{1, 14}$; (f) severe arrhythmia, $\{C(i)/6\}$, $i = \overline{1, 14}$; (g) aortic stenosis, $\{C(i)/7\}$, $i = \overline{1, 14}$; (h) stenocardia of the second functional class, $\{C(i)/8\}$, $i = \overline{1, 14}$; (i) stenocardia of the third functional class, $\{C(i)/9\}$, $i = \overline{1, 14}$.

For the numerical experiment two hundred different cardiograms for each disease $\{C(i)/k\}$, $i = \overline{1, 14}$, $k = \overline{1, K}$, $K = 9$ were used from the Physikalisch-Technische Bundesanstalt (PTB) dataset, which is a publicly available database. This database is compiled by the National Metrology Institute of Germany. It contains combinations of digitized ECGs of both normal and abnormal subjects' recordings, which are provided for research[52].

Testing the statistical hypothesis about the independence of random coefficients $P_{N-1}(1)$, $P_{1;1,N-2}(2)$, $P_{1,2;1,1,N-3}(3)$, …, $P_{14-N+1,14-N+2,…,13;1,1.....1}(14)$ based on the criterion χ^2 [53] and the Bloom criterion[54], showed the truth of the hypothesis about the independence of coefficients at $N = 4$ for all six sequences with a probability not less than $P_D = 0,98$ (for a linear decomposition there is a statistically significant relationship between random coefficients; for $N = 3$, there are not enough grounds for accepting the hypothesis about independence of random coefficients for $i = 3$, $i = 6$, $i = 7$, $i = 8$). Thus, the canonical expansion (8) for $N = 4$ (random coefficients are a array of independent random variables) with the corresponding sets of coordinate functions are adequate models[40] of the studied random sequences $\{C(i)/k\}$, $i = \overline{1, 14}$, $k = \overline{1, 9}$. The following methods were used for recognition: (a) linear criterion (rule (6) for $E\big[C(j)C(i)\big] \neq 0$, $E\big[C^\nu(j)C^\mu(i)\big] = 0$, $\nu + \mu > 2$); (b) polynomial criterion (6) [55]; (c) fuzzy logic method [56–58]; (d) neural network [59,60] based on the Daubechies wavelet function of the fourth order and the Levenberg–Marquardt algorithm for learning; (e) generalized non-linear criterion (11). Let us consider mentioned criteria in details:

(a) *Linear criterion*

$$k^* = \arg \max_k \prod_{i=1}^{14} f_1\big(p_i/k\big), \quad k = \overline{1, 6}, \tag{12}$$

where p_i, $i = \overline{1, I}$, $I = 14$ are random values of uncorrelated random coefficients P_i, $i = \overline{1, 14}$:

$$P_i = C(i) - E[C(i)] - \sum_{\nu=1}^{i-1} P_\nu T_\nu(i), \quad i = \overline{1, 14}.$$

Coordinate functions

$$T_\nu(i) = \frac{1}{D(\nu)} \left(E[C(\nu)C(i)] - E[C(\nu)]E[C(i)] - \sum_{\rho=1}^{\nu-1} D(\rho) \right.$$
$$\left. \times\; T_\rho(\nu)T_\rho(i) \right), \quad \nu = \overline{1, i}, \quad i = \overline{1, 14}.$$

Variances of random coefficients

$$D_\lambda(i) = E\Big[C^{2\lambda}(i)\Big] - E^2\Big[C^\lambda(i)\Big] - \sum_{\rho=1}^{\lambda-1} D(\rho)\big\{T_\rho(i)\big\}^2, \quad i = \overline{1, 14}.$$

(b) *Polynomial criterion*

$$k^* = \arg \max_j \prod_{i=1}^{14} f_1\Big(p_i^{(3)}/k\Big), \quad k = \overline{1, 6}. \tag{13}$$

Random coefficients

$$P_i^{(\lambda)} = C^\lambda(i) - \sum_{\nu=1}^{i-1}\sum_{j=1}^{3} P_\nu^{(j)} T_{\lambda\nu}^{(j)}(i) - \sum_{j=1}^{\lambda-1} P_i^{(j)} T_{\lambda i}^{(j)}(i), \quad i = \overline{1, 14}.$$

Coordinate functions

$$T_{hv}^{(\lambda)}(i) = \frac{1}{D_\lambda(v)} \left(E\left[C^\lambda(v)C^h(i)\right] - E\left[C^\lambda(v)\right]E\left[C^h(i)\right] - \sum_{\rho=1}^{v-1}\sum_{j=1}^{3} D_j(\rho) \right.$$

$$\left. \times \; T_{\lambda\mu}^{(j)}(v)T_{h\rho}^{(j)}(i) - \sum_{j=1}^{\lambda-1} D_j(v)T_{\lambda v}^{(j)}(v)T_{hv}^{(j)}(i) \right), \quad v = \overline{1,i}, \quad i = \overline{1,14}.$$

Variances of random coefficients

$$D_\lambda(i) = E\left[C^{2\lambda}(i)\right] - E^2\left[C^\lambda(i)\right] - \sum_{\rho=1}^{i-1}\sum_{j=1}^{3} D_j(\rho)$$

$$\times \; \left\{T_{\lambda\rho}^{(j)}(i)\right\}^2 - \sum_{j=1}^{\lambda-1} D_j(i)\left\{T_{\lambda i}^{(j)}(i)\right\}^2, \quad i = \overline{1,14}.$$

(c) *Fuzzy logic method for medical diagnostics*
    This method is based on the implementation of fuzzy system with hierarchical structure of Rule Base [56].

    Input parameters: p_1—increase of double product per one kilogram of the body weight of the sick; p_2—increase of double product per one kilogram of physical exertion; p_3—coefficient of phosphorilation; p_4—age of the sick; p_5—double product of pulse on arterial tension; p_6—adenosinetriphosphoric acid; p_7—adenosine diphosphoric acid; p_8—adenylic acid; p_9—coefficient of the ratio of lactic and pyruvic acid content; p_{10}—maximal consumption of oxygen per one kilogram of the body weight of the sick; p_{11}—increase of double product in the response for submaximal physical exertion; p_{12}—tolerance to physical activity.
    Expressions for the determination of the diagnosis are of the form:

$$\begin{aligned}
c &= f_c\left(p_1, p^{(1)}, p^{(2)}\right), \\
p^{(1)} &= f_{p^{(1)}}\left(p_2, p_3, p_4, p_5, p_{10}, p_{11}\right), \\
p^{(2)} &= f_{p^{(2)}}\left(p_6, p_7, p_8, p_9, p_{12}\right)
\end{aligned}$$

(14)

where values c: c_1—mild neurocirculatory dystonia, $\{C(i)/1\}$, $i = \overline{1,14}$; c_2—neurocirculatory dystonia of moderate degree, $\{C(i)/2\}$, $i = \overline{1,14}$; c_3- severe neurocirculatory dystonia, $\{C(i)/3\}$, $i = \overline{1,14}$; c_4—stenocardia of the first functional class, $\{C(i)/4\}$, $i = \overline{1,14}$; c_5- hypertrophy of myocardium, $\{C(i)/5\}$, $i = \overline{1,14}$; c_6—severe arrhythmia, $\{C(i)/6\}$, $i = \overline{1,14}$; c_7- aortic stenosis, $\{C(i)/7\}$, $i = \overline{1,14}$; c_8—stenocardia of the second functional class, $\{C(i)/8\}$, $i = \overline{1,14}$; c_9—stenocardia of the third functional class, $\{C(i)/9\}$, $i = \overline{1,14}$. Tables 1, 2 and 3 are the information base for the formation of a system of fuzzy logic equations that connect the membership functions of the diagnosis and input variables.
    Tables 1, 2, 3 use the abbreviation for fuzzy terms: L—low, BA—below the average, A—average, AA—above average, H—high.
    For example, if $p_1 = A$, $p^{(1)} = H$, $p^{(2)} = AA$, according to the last line of Table 1, the diagnosed disease is stenocardia of the third functional class.
    The value H for parameter $p^{(1)}$ is accepted in three cases (rule base Table 2):

(a) $p_2 = L, p_3 = L, p_4 = L, p_5 = AA, p_{10} = L, p_{11} = L;$
(b) $p_2 = BA, p_3 = L, p_4 = BA, p_5 = H, p_{10} = L, p_{11} = BA;$
(c) $p_2 = L, p_3 = BA, p_4 = BA, p_5 = AA, p_{10} = L, p_{11} = L.$

    Parameter $p^{(2)}$ is equal to AA if one of the conditions is met (rule base Table 3):

(a) $p_6 = BA, p_7 = A, p_8 = BA, p_9 = A, p_{12} = A;$
(b) $p_6 = AA, p_7 = BA, p_8 = A, p_9 = BA, p_{12} = BA;$
(c) $p_6 = L, p_7 = A, p_8 = A, p_9 = BA, p_{12} = A.$

(iv) *Neural network*
    Daubechies wavelet function of the fourth order and the Levenberg–Marquardt algorithm for learning were used [59].
    Expressions for the determination of approximation coefficients and detailing of discrete wavelet transform are presented in the form:

p_1	$p^{(1)}$	$p^{(2)}$	c
L	L	L	c_1
L	BA	BA	c_1
BA	BA	L	c_1
BA	BA	BA	c_2
A	BA	BA	c_2
BA	BA	A	c_2
A	BA	A	c_3
AA	AA	BA	c_3
AA	A	A	c_3
AA	A	AA	c_4
A	AB	AA	c_4
BA	AA	AA	c_4
BA	A	AA	c_5
BA	H	AA	c_5
BA	A	A	c_5
BA	H	A	c_6
BA	A	H	c_6
BA	H	H	c_6
A	AA	AA	c_7
AA	AA	AA	c_7
H	A	A	c_7
A	H	A	c_8
AA	AA	H	c_8
H	AA	AA	c_8
H	H	H	c_9
AA	H	AA	c_9
A	H	AA	c_9

Table 1. Knowledge base according to equation $c = f_c\left(p_1, p^{(1)}, p^{(2)}\right)$.

p_2	p_3	p_4	p_5	p_{10}	p_{11}	$p^{(1)}$
H	H	H	L	H	H	L
H	AA	H	BA	H	H	L
AA	H	AA	L	H	H	L
AA	AA	H	BA	H	AA	BA
H	H	AA	A	H	H	BA
AA	AA	H	BA	AA	AA	BA
A	A	A	A	A	A	A
AA	AA	A	BA	AA	A	A
A	AA	AA	A	AA	AA	A
BA	A	BA	AA	BA	BA	AA
BA	BA	A	A	L	BA	AA
A	BA	BA	AA	BA	A	AA
L	L	L	AA	L	L	H
BA	L	BA	H	L	BA	H
L	BA	BA	AA	L	L	H

Table 2. Knowledge base according to equation $p^{(1)} = f_{p^{(1)}}\left(p_2, p_3, p_4, p_5, p_{10}, p_{11}\right)$.

$$W_\gamma\left(j_0, k\right) = \frac{1}{M} \sum_x F(x)\gamma_{j_0,k}(x),$$

$$W_\phi\left(j, k\right) = \frac{1}{M} \sum_x F(x)\phi_{j,k}(x),$$

p_6	p_7	p_8	p_9	p_{12}	$p^{(2)}$
H	H	H	H	H	L
AA	H	AA	AA	AA	L
H	AA	H	A	AA	L
AA	AA	A	A	AA	BA
A	AA	A	AA	H	BA
A	H	AA	AA	AA	BA
A	A	A	AA	AA	A
AA	AA	A	A	A	A
AA	A	AA	AA	A	A
BA	A	BA	A	A	AA
AA	BA	A	BA	BA	AA
L	A	A	BA	A	AA
L	L	L	L	BA	H
BA	L	BA	L	L	H
L	BA	BA	L	BA	H

Table 3. Knowledge base according to equation $p^{(2)} = f_{p^{(1)}}(p_6, p_7, p_8, p_9, p_{12})$.

Heading level	{C(i)/1}(%)	{C(i)/2}(%)	{C(i)/3}(%)	{C(i)/4}(%)	{C(i)/5}(%)
Linear criterion (12)	79	83	89	77	82
Polynomial criterion (13)	92	93	96	91	95
Fuzzy logic method (14)	87	89	93	87	92
Neural network	86	87	91	89	94
Non-linear criterion (11)	99	99	100	99	100

Table 4. The results of the diagnostics of cardiovascular diseases $\{C(i)/k\}$, $k = \overline{1,5}$ (% of correct diagnoses).

Heading level	{C(i)/6}(%)	{C(i)/7}(%)	{C(i)/8}(%)	{C(i)/9}(%)
Linear criterion (12)	85	86	81	84
Polynomial criterion (13)	95	96	93	96
Fuzzy logic method (14)	93	94	88	93
Neural network	93	92	89	94
Non-linear criterion (11)	100	100	100	100

Table 5. The results of the diagnostics of cardiovascular diseases $\{C(i)/k\}$, $k = \overline{6,9}$ (% of correct diagnoses).

where $\gamma_{j,k}(x)$, $\phi_{j,k}(x)$ is a family of basic functions.

Output signal of each of separate neuron of output layer was forming as

$$y(k) = \frac{1}{M} F \left(\sum_{i=0}^{K} w_{ki} F \left(\sum_{j=0}^{N} w_{ij} x \right) \right)$$

Continuous sigmoid bipolar function $F(x) = th(x)$ was used as activation function of each separate neuron. Tables 4, 5 shows the comparative diagnostic results.

The data in Tables 4, 5 indicate the low efficiency of the linear criterion (12) (the minimum amount of a priori information is used: $E\left[C(j)C(i)\right]$. The use of additional stochastic relations $(E\left[C^\nu(j)C^\mu(i)\right])$ in the criterion (13) makes it possible to achieve an increase in the accuracy of solving the problem of diagnosing cardiovascular diseases compared to (12). The maximum accuracy of diagnostics is achieved by applying the proposed decision rule (11) by maximizing the use of the stochastic properties $(E\left[C^{\xi_l}(i - r_{l-1}) \cdots C^{\xi_2}(i - r_1)C^{\xi_1}(i)\right])$ of the studied random sequences. The existing set of wavelet functions and the lack of a rigorous mathematical apparatus for analyzing fuzzy equations significantly limit the quality of decision-making about a cardiovascular disease based on the fuzzy logic method and neural network.

Conclusion

A computational method for computer systems for automated diagnosis of cardiovascular diseases based on a generalized nonlinear canonical decomposition of a random sequence of change of cardiograms has been obtained. The use of the canonical model made it possible to form the decision rule for the maximum distribution density in the form of a product of one-dimensional distribution densities of random coefficients. The canonical decomposition does not impose any significant restrictions (linearity, stationarity, Markov property, monotony, ergodicity, etc.) on the class of random sequences under study, which makes it possible to maximally take into account the stochastic characteristics of sequences related to various cardiovascular diseases.

Taking into account the recurrent regularity of calculations, the diagnostic method is quite simple in terms of computation and allows using an arbitrary number of input parameters. A significant advantage of the method is the ability to use characteristics not directly related to the cardiogram (age of the patient, blood pressure, etc.).

During the operation of the diagnostic system based on the proposed computational method, new diseases unknown to medicine can be identified in the case of a significant difference in the values of the likelihood function for the investigated cardiogram and the classified cardiograms of known diseases.

The results of the numerical experiment indicate a high reliability of the diagnostics of cardiovascular diseases based on the proposed method.

Data availability

The datasets generated and analysed during the current study are available in the Physikalisch-Technische Bundesanstalt (PTB) repository, https://physionet.org/physiobank/database/.

References

1. Yun, S., Oh, K. The Korea national health and nutrition examination survey data linked cause of death data. *J. Epidemiol. Health* e2022021 (2022).
2. Mazumder, O. *et al.* Synthetic PPG signal generation to improve coronary artery disease classification: Study with physical model of cardiovascular system. *J. Biomed. Health Inf.* **26**(5), 2136–2146 (2022).
3. Bair E. *et al.* The use of paclitaxel-coated devices in the treatment of peripheral Arterial disease is not associated with increased mortality or amputations. *J. Ann. Vasc. Surgery* (79), 387 (2022).
4. Zhao, Y., Zeng, Q., Li, J. & Jiang, X. Digital subtraction angiography image features under the deep learning algorithm in cardiovascular interventional treatment and nursing for vascular restenosis. *J. Comput. Math. Methods Med.* (2022).
5. Sinha, D., Sharma, A. & Sharma, S. Automated detection of coronary artery disease comparing arterial fat accumulation using CNN. *J. Electron. Imaging* **31**(5), 051405 (2022).
6. Yu, C., Che, Y., Sun, G., Zhao, X. & Liu, B. Research on diagnosis architecture of cardiovascular diseases based on multimedical images. *J. Comput. Math. Methods Med.* (2022).
7. Darmawahyuni, A. *et al.* Deep learning-based electrocardiogram rhythm and beat features for heart abnormality classification. *J. Comput. Sci.* **8**, e825 (2022).
8. Zang, X., Li, B., Zhao, L., Yan, D. & Yang, L. End-to-end depression recognition based on a one-dimensional convolution neural network model using two-lead ECG signal. *J. Med. Biol. Eng.* 1–9 (2022).
9. Fuentes-Aguilar, R. Q., Pérez-Espinosa, H., & Filigrana-de-la-Cruz, M. A. Biosignals analysis (heart, phonatory system, and muscles). In *Biosignal Processing and Classification Using Computational Learning and Intelligence* 7–26 (2022).
10. Tripathy, R. K. *et al.* Detection of life threatening ventricular arrhythmia using digital taylor fourier transform. *J. Front. Physiol.* https://doi.org/10.3389/fphys.2018.00722 (2018).
11. Oh, S. L. *et al.* Shockable versus nonshockable life-threatening ventricular arrhythmias using dwt and nonlinear features of ECG signals. *J. Mech. Med. Biol.* **17**(07), 1740004. https://doi.org/10.1142/S0219519417400048 (2017).
12. Sharma, M., Tan, R.-S. & Acharya, U. R. Detection of shockable ventricular arrhythmia using optimal orthogonal wavelet filters. *J. Neural Comput. Appl.* **32**(20), 15869–15884 (2020).
13. Arafat, M. A., Chowdhury, A. W. & Hasan, M. K. A simple time domain algorithm for the detection of ventricular fibrillation in electrocardiogram. *J. Signal Image Video Process.* **5**(1), 1–10. https://doi.org/10.1007/s11760-009-0136-1 (2011).
14. Granitto, P. M., Furlanello, C., Biasioli, F. & Gasperi, F. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *J. Chemom. Intell. Lab. Syst.* **83**(2), 83–90. https://doi.org/10.1016/j.chemolab.2006.01.007 (2006).
15. Pławiak, P. Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals. *J. Swarm Evol. Comput.* **39**, 192–208. https://doi.org/10.1016/j.swevo.2017.10.002 (2018).
16. Zhdanov, A. E. *et al.* OculusGraphy: Literature review on electrophysiological research methods in ophthalmology and electroretinograms processing using wavelet transform. In *2020 International Conference on e-Health and Bioengineering (EHB), Iasi, Romania*, 1–6 (IEEE, 2020). https://doi.org/10.1109/EHB50910.2020.9280221
17. Li, Y., Bisera, J., Weil, M. H. & Tang, W. An algorithm used for ventricular fibrillation detection without interrupting chest compression. *IEEE Trans. Biomed. Eng.* **59**(1), 78–86. https://doi.org/10.1109/TBME.2011.2118755 (2012).
18. Sinha, N. & Das, A. Automatic diagnosis of cardiac arrhythmias based on three stage feature fusion and classification model using DWT. *J. Biomed. Signal Process. Control* **62**, 102066. https://doi.org/10.1016/j.bspc.2020.102066 (2020).
19. Ververidis, D. & Kotropoulos, C. Fast and accurate sequential floating forward feature selection with the Bayes classifier applied to speech emotion recognition. *J. Signal Process.* **88**(12), 2956–2970. https://doi.org/10.1016/j.sigpro.2008.07.001 (2008).
20. Bhuvaneswari Amma, N. G. Cardiovascular disease prediction system using genetic algorithm and neural network. In *Proceedings of 2012 International Conference on Computing, Communication and Applications* 1–5 (2012). https://doi.org/10.1109/ICCCA.2012.6179185
21. Sireesha, M. Classification model for prediction of heart disease using correlation coefficient technique. *Int. J. Adv. Trends Comput. Sci. Eng.* **9**(2), 2116–2123. https://doi.org/10.30534/ijatcse/2020/185922020 (2020).
22. Lehmann, E. L. & Romano, J. P. *Testing Statistical Hypotheses* 3rd edn. (Springer, 2008).
23. Fisher, R. A. Statistical methods for research workers. In *Breakthroughs in Statistics* (eds Kotz, S. & Johnson, N. L.) 66–70 (Springer, 1992). https://doi.org/10.1007/978-1-4612-4380-9_6.
24. Kira, K. & Rendell, L. A. The feature selection problem: Traditional methods and a new algorithm. In *AAAI-92 Proceedings* 129–134 (1992). Available at: https://aaai.org/Papers/AAAI/1992/ AAAI92–020.pdf (Accessed November 29, 2020).
25. Firoozabadi, R., Gregg, R. E., Babaeizadeh, S. & Laciar, E. Identification of exercise-induced ischemia using QRS slopes. *J. Electrocardiol.* **49**, 55–59 (2016).

26. Shen, C.-P. *et al.* Detection of cardiac arrhythmia in electrocardiograms using adaptive feature extraction and modified support vector machines. *Expert Syst. Appl.* **39**, 7845–7852 (2012).
27. Ronzhina, M. *et al.* Spectral and higher-order statistics analysis of ECG: application to study of ischemia in rabbit isolated hearts. In *2012 Computing in Cardiology Conference (CinC)* Vol 36, 645–648 (2012).
28. Sayadi, O., Shamsollahi, M. B., Clifford, G. D., Verleysen, M. & Bencherif, M. A. Robust detection of premature ventricular contractions using a wave-based bayesian framework. *IEEE Trans. Biomed. Eng.* **57**, 353–362 (2010).
29. Tseng, Y.-L., Lin, K.-S., Jaw, F.-S. & Laciar, E. Comparison of support-vector machine and sparse representation using a modified rule-based method for automated myocardial ischemia detection. *J. Comput. Math. Methods Med.* **2016**, 1–8 (2016).
30. Doquire, G., de Lannoy, G., François, D., Verleysen, M. & Bencherif, M. A. Feature selection for interpatient supervised heart beat classification. *J. Physiol. Meas.* **31**, 903–920 (2010).
31. Shebanin, V. *et al.* Application of fuzzy predicates and quantifiers by matrix presentation in informational resources modeling. In *Proceedings of XII International Conference "MEMSTECH 2016", Lviv-Poljana,* 146–149 (2016). https://doi.org/10.1109/MEMSTECH.2016.7507536
32. Jovic, A. & Bogunovic, N. Electrocardiogram analysis using a combination of statistical, geometric, and nonlinear heart rate variability features. *J. Artif. Intell. Med.* **51**(3), 175–186 (2011).
33. Prcela, M., Gamberger, D. & Jovic, A. Semantic web ontology utilization for heart failure expert system design. *J. Stud. Health Technol. Inf.* **136**, 851–856 (2008).
34. Biel, L., Pettersson, O., Philipson, L. & Wide, P. ECG analysis: A new approach in human identification. *J. Instrum. Meas.* **50**(3), 808–812 (2001).
35. Pawar, T., Anantakrishnan, N. S., Chaudhuri, S. & Duttagupta, S. P. Impact analysis of body movement in ambulatory ECG. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 5453–5456 (2007).
36. Gerencsér, L., Kozmann, G., Vágó, Z. & Haraszti, K. The use of the SPSA method in ECG analysis. *J. IEEE Trans. Biomed. Eng.* **49**(10), 1094–1101 (2002).
37. Fang, Q., Sufi, F. & Cosic, I. *A Mobile Device Based ECG Analysis System* (NTECH Open Access Publisher, 2008).
38. Benitez, D., Gaydecki, P. A., Zaidi, A. & Fitzpatrick, A. P. The use of the Hilbert transform in ECG signal analysis. *J. Comp. Biol. Med.* **31**(5), 399–406 (2001).
39. Box, G. E. P. & Jenkins, G. M. *Time–series analysis, Forecasting and Control* (Holden-Day, 1970).
40. Kudritsky, V. D. *Filtering, Extrapolation and Recognition Realizations of Random Functions* 176 (FADA Ltd, 2001).
41. Rosenblum, M. G., Pikovsky, A. S. & Kurths, J. Synchronization approach to analysis of biological systems. *Random Fluctuat. World Celebr. Two Decad. Fluctuat. Noise Lett.* 335–344 (2022).
42. Peng, H., Wu, X., Wen, Y., Lin, J. & Guan, W. Myocardial infarction and stroke risks in multiple sclerosis patients: A two-sample Mendelian randomization study. *J. Mult. Scler. Relat. Disord.* **58**, 103501 (2022).
43. Chabchoub, S., Mansouri, S. & Ben Salah, R. Signal processing techniques applied to impedance cardiography ICG signals–a review. *J. Med. Eng. Technol.* https://doi.org/10.1080/03091902.2022.2026508 (2022).
44. Bickel, D. Coherent checking and updating of Bayesian models without specifying the model space: Adecision-theoretic semantics for possibility theory. *Int. J. Approx. Reason.* **142**, 81–93 (2022).
45. Shebanin, V., *et al.* Canonical mathematical model and information technology for cardio-vascular diseases diagnostics. In: *Proceedings 14th International Conference CADSM 2017* 438–440 (2017). https://doi.org/10.1109/CADSM.2017.7916170
46. Kondratenko, Y., *et al.* University curricula modification based on advancements in information and communication technologies. In *Proceedings volume of ICTERI 2016, Kyiv on the 21st–24th of June, 2016* 184–199 (2016).
47. Atamanyuk, I. P. Algorithm of extrapolation of a nonlinear random process on the basis of its canonical decomposition. *J. Kibern. Sist. Anal.* **2**, 131–138 (2005).
48. Atamanyuk, I., Kondratenko, Y., Shebanin, V. & Mirgorod, V. Method of polynomial predictive control of fail-safe operation of technical systems. In *Proceedings XIIIth International Conference CADSM 2015, Polyana-Svalyava, Ukraine* 248–251 (2015). https://doi.org/10.1109/CADSM.2015.7230848
49. Atamanyuk, I. P. Optimal polynomial extrapolation of realization of a random process with a filtration of measurement errors. *J. Automat. Inf. Sci.* **41**(8), 38–48. https://doi.org/10.1615/JAutomatInfScien.v41.i8.40 (2009).
50. Atamanyuk, I., Kondratenko, Y. & Sirenko, N. Management system for agricultural enterprise on the basis of its economic state forecasting, complex systems: Solutions and challenges in economics, management and engineering. In *Studies in systems, decision and control* Vol. 125 (eds Christian, B.-V. *et al.*) 453–470 (Springer, 2018). https://doi.org/10.1007/978-3-319-69989-9_27
51. Parzen, E. An estimation of a probability density function and mode. *Ann. Math. Stat.* **33**(3), 1065–1076 (1962).
52. https://physionet.org/physiobank/database/
53. Yongcheng, Q. & Yingchao, Z. Empirical likelihood method for complete independence test on high dimensional data. arXiv preprint arXiv:2201.08492 (2022).
54. Li, H., Zhang, H. & Jiang, H. Combining power of different methods to detect associations in large data sets. *J. Brief. Bioinform.* **23**(1), bbab 488 (2022).
55. Atamanyuk I. & Kondratenko Y. Calculation method for a computer's diagnostics of cardiovascular diseases based on canonical decompositions of random sequences. ICT in education, research and industrial applications: Integration, harmonization and knowledge transfer. In *Proceedings of the 11 International Conference ICTERI-2015*, (eds S. Batsakis, et al.), CEUR-WS, Vol. 1356, 108–120 (2015).
56. Rotshtein, A. P. *Intellectual Technologies of Identification: Fuzzy Logic, Genetic Algorithms, Neuron Networks* 320 (UNIVERSUM-Vinnitsa, Vinnitsa, 1999).
57. Zadeh, L. A. Fuzzy Sets. *J Inf. Control* **8**, 338–353 (1965).
58. Zadeh, L. A. The role of fuzzy logic in modeling, identification and control. *J. Model. Identif. Control* **15**(3), 191–203 (1994).
59. Grigoriev, D. S. & Spitsin, V. G. The application of neural network and discrete wavelet transform for the analysis and classification of electrocardiograms. *J. Bull. Tomsk Polytech. Univ.* **5**, 57–61 (2012).
60. Gupta, V., Saxena, N. K., Kanungo, A. Gupta, A. & Kumar, P. A review of different ECG classification/detection techniques for improved medical applications. *Int. J. Syst. Assur. Eng. Manag.* 1–15 (2022).

Author contributions

I.A., Y.K. wrote the main manuscript textV.H., Y.V. performed numerical experiments.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to I.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.