Mathematical modeling

Математическое моделирование

UDC 51.7
https://doi.org/10.32362/2500-316X-2025-13-1-144-156
EDN WPRKRW

RESEARCH ARTICLE

Multivariate discriminant analysis of the electrocardiogram

Polina A. Sakharova [@], **Vyacheslav A. Balandin**

MIREA – Russian Technological University, Moscow, 119454 Russia
[@] *Corresponding author, e-mail: polinka6777@gmail.com*

Abstract

Objectives. The article presents a study of heart rate variability using multivariate discriminant analysis. Representing an effective statistical method of classification, discriminant analysis can be used to divide objects into groups based on differences in the parameters characterizing these objects. The effectiveness of multivariate discriminant analysis, which is actively used in medicine to diagnose cardiovascular pathologies, is due to the wide range of analyzed parameters: statistical, spectral, and autocorrelation. The aim of the work is to identify the parameters of variational pulsometry, which provide the best distinction between healthy patients and patients with arrhythmia, by means of discriminant analysis.

Methods. The durations of cardiac intervals of patients aged 63–72 years, which had been placed in the open database of biomedical signals PhysioNet.org, were used as initial data. When selecting the arguments of the discriminant function, priority was given to parameters that were weakly correlated with each other, had a normal distribution, and differed between healthy and ill patients. The statistical significance of differences between the parameters of the two groups was tested using Student's *t*-test and Mann–Whitney *U* test.

Results. Two discriminant functions were obtained: the first depended on three time-domain parameters, while the second included one spectral and one autocorrelation parameter in addition to time-domain parameters. In both cases, the average values of the discriminant function for healthy and sick patients were calculated. The statistical significance of differences in the average values of the discriminant function in the two groups was investigated using Student's *t*-test.

Conclusions. The values of the first discriminant function are shown to differ insignificantly between healthy and sick patients, while the inclusion of autocorrelation and spectral parameters in the number of arguments of the discriminant function provides pronounced and statistically significant differences between patients of the two groups. Thus, the high significance of spectral and autocorrelation parameters in arrhythmia diagnosis was demonstrated.

Keywords: heart rate variability, variational pulsometry, spectral analysis, autocorrelation analysis, RR-intervals, multivariate discriminant analysis, discriminant function

• **Submitted:** 26.03.2024 • **Revised:** 09.06.2024 • **Accepted:** 10.12.2024

For citation: Sakharova P.A., Balandin V.A. Multivariate discriminant analysis of the electrocardiogram. *Russian Technological Journal.* 2025;13(1):144–156. https://doi.org/10.32362/2500-316X-2025-13-1-144-156, https://elibrary.ru/WPRKRW

Financial disclosure: The authors have no financial or proprietary interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Многофакторный дискриминантный анализ электрокардиограммы

П.А. Сахарова [@], **В.А. Баландин**

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] *Автор для переписки, e-mail: polinka6777@gmail.com*

Резюме

Цели. Статья посвящена исследованию вариабельности сердечного ритма с помощью многофакторного дискриминантного анализа. Дискриминантный анализ является эффективным статистическим методом классификации, позволяющим разбивать объекты на группы исходя из различий между характеризующими эти объекты параметрами. Эффективность многофакторного дискриминантного анализа, который активно используется в медицине для диагностики сердечно-сосудистых патологий, обусловлена широким набором анализируемых параметров: статистических, спектральных и автокорреляционных. Цель работы – выявление методом дискриминантного анализа параметров вариационной пульсометрии, которые обеспечивают наилучшее различение между здоровыми пациентами и пациентами с аритмией.

Методы. В качестве исходных данных использовались длительности кардиоинтервалов пациентов возраста 63–72 лет, размещенные в открытой базе биомедицинских сигналов PhysioNet.org. При выборе аргументов дискриминантной функции преимущество отдавалось слабо коррелирующим между собой параметрам, имеющим нормальное распределение и различающимся у здоровых и больных пациентов. Статистическая значимость различий между параметрами двух групп проверялась с помощью *t*-критерия Стьюдента и *U*-критерия Манна – Уитни.

Результаты. Получены две дискриминантные функции: первая зависела от трех временных параметров; вторая, помимо временных, включала один спектральный и один автокорреляционный. В обоих случаях были рассчитаны средние значения дискриминантной функции для здоровых и больных пациентов. Статистическая значимость различий средних значений дискриминантной функции в двух группах исследовалась с помощью *t*-критерия Стьюдента.

Выводы. Показано, что значения первой дискриминантной функции незначительно различаются у здоровых и больных пациентов, в то время как включение автокорреляционного и спектрального параметров в число аргументов дискриминантной функции обеспечивает выраженные и статистически значимые различия между пациентами двух групп. Тем самым продемонстрирована высокая значимость спектральных и автокорреляционных параметров в диагностике аритмии.

Ключевые слова: вариабельность сердечного ритма, вариационная пульсометрия, спектральный анализ, автокорреляционный анализ, RR-интервалы, многофакторный дискриминантный анализ, дискриминантная функция

• **Поступила:** 26.03.2024 • **Доработана:** 09.06.2024 • **Принята к опубликованию:** 10.12.2024

Для цитирования: Сахарова П.А., Баландин В.А. Многофакторный дискриминантный анализ электрокардиограммы. *Russian Technological Journal.* 2025;13(1):144–156. https://doi.org/10.32362/2500-316X-2025-13-1-144-156, https://elibrary.ru/WPRKRW

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Currently, cardiovascular diseases (CVDs) are the main cause of death worldwide [1]. Among them, a special place is occupied by diseases associated with heart rhythm disorders characterized by various arrhythmias [2]. According to statistics, about one third of people with CVDs suffer from arrhythmias [3]. In view of high mortality rates from CVDs, timely diagnosis of cardiovascular system disorders becomes especially important.

Along with traditional amplitude-time analysis, an important approach to diagnosing CVD involves the study of heart rate variability (HRV) [4]. Various methods are used for quantitative determination of HRV indices, among which the following can be emphasized [5]:

1. Methods based on statistical transformations. These include temporal analysis of HRV, whose numerical characteristics are average value of cardiac interval duration (RR normal-to-normal interval, RRNN), standard deviation (SD) of normal-to-normal intervals (SDNN) of the cardiac cycle (standard deviation of normal-to-normal intervals, SDNN), percentage of consecutive normal-to-normal intervals that differ by more than 50 ms (pNN50), square root of the mean sum of squares of the successive differences (root mean square of the successive differences, RMSSD);
2. Geometric methods, including scatterographic and histographic analysis, whose objects are mode (Mo)—the value of cardiointerval duration that occurs most frequently in the sample, mode amplitude (AMo)—the share of cardiointervals (in %) that fall within the modal interval, coefficient of variation (in %) $CV = \frac{SD}{RRNN}$, variation range (delta RR interval, dRR)—the difference between the maximum and minimum values of cardiointervals, as well as a set of indices, of which the Baevsky tension index (TI) of the regulatory systems is the most widespread $TI = \frac{AMo}{2Mo \cdot dRR}$;
3. Wave structure analysis methods, comprising:

- spectral method analyzing the power of the spectrum of RR-intervals in the region of very low frequencies (VLF) 0.004–0.03 Hz, low frequencies (LF) 0.04–0.15 Hz, high frequencies (HF) 0.15–0.40 Hz, normalized values of the last two parameters (LF_{norm}, HF_{norm}), centralization index CI = (HF + LF)/VLF, and vagosympathetic interaction index LF/HF;
- autocorrelation method that calculates the shift number at which the autocorrelation function first becomes negative (C0) and the value of the autocorrelation function at the first shift (C1).

The description of the parameters and their diagnostic significance are given in detail in [6].

These parameters are calculated from rhythmograms, which are obtained from the electrocardiograms (ECGs) as time differences between the occurrence of consecutive R beats. Figure 1 shows how the intervals between the occurrence of R beats on the ECG are plotted on the rhythmogram on the ordinate axis and numbered on the abscissa axis.

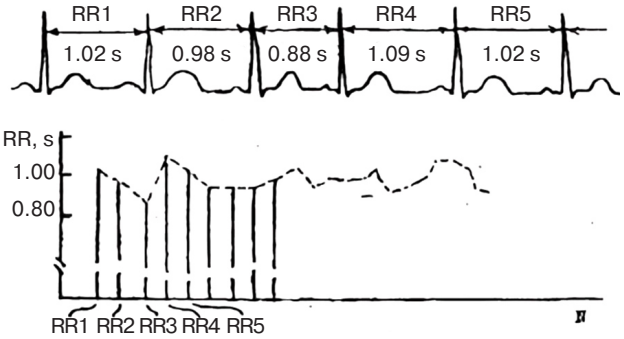

Fig. 1. Obtaining a rhythmogram from the original ECG[1]

The obtained values of the indicators are analyzed by a number of statistical methods to establish significant differences between healthy and sick patients. Among such methods, multivariate discriminant analysis (MDA), which consists in constructing a linear combination of the most informative features that would best ensure the difference between groups, assumes an important place. The effectiveness of MDA application in cardiology was demonstrated, in particular, in [7], which resulted in a model for the distribution of children and adolescents into groups with different cardiovascular system conditions, providing an accuracy of 98.1%.

In [8], it is shown that the significance of MDA based only on statistical parameters of variation pulsometry significantly depends on the type of cardiac pathology. In particular, in arrhythmia, the difference between the discriminant function (DF) values of healthy and sick patients is ~25%, which is significantly lower than that of chronic heart failure, for instance.

Thus, the present study set out to identify the most informative features among the above-mentioned features that would significantly increase the differences of DF in the presence or absence of arrhythmia.

MATERIALS AND METHODS

The initial materials for the work were fragments of ECG examination results, namely, RR interval durations. These records were obtained from the PhysioBank open

[1] Methodical development of practical training on pathophysiology for 3rd year students of general medicine and pediatrics departments. https://patfizo.narod.ru/read/heartprakt.htm (in Russ.). Accessed June 05, 2024.

Table 1. HRV indices of a healthy patient at different ECG recording durations

Parameter \ Duration	1 min	2 min	5 min	10 min	15 min	20 min	30 min	Norm [8]
Heart rate (HR), bpm	82	82	79	78	79	80	81	60–90
RRNN, ms	736	734	760	773	764	751	744	660–937
dRR, ms	125	172	250	250	289	312	312	310–450
SDNN, ms	26	32	46	42	49	52	54	40–80
Mo, ms	750	750	750	789	789	750	750	870–930
AMo, %	59	48	45	45	40	37	35	32–38
CV, %	3.5	4.4	6	5.5	6.4	6.9	7.2	3–12
pNN50, %	7.2	6.7	6.1	5.4	4.7	4.6	5.6	1–9
RMSSD, ms	27.7	26.1	26.5	25.6	24.7	24.7	25.7	20–50
TI, c.u.	315	186	120	115	87	78	75	80–150

database of biomedical signals made available via the PhysioNet portal[2]. Data on healthy patients were taken from the Normal Sinus Rhythm RR Interval Database. Records of sick patients diagnosed with arrhythmias were taken from the MIT-BIH Arrhythmia Database, which contains ECG records of various cardiac rhythm disturbances. It should be noted that in this database there is no differentiation of sick patients according to the type of arrhythmia. Therefore, in this study, the selection of patients was randomized according to the specific type of arrhythmia.

In the present work, the records of 10 healthy patients and 10 patients diagnosed with arrhythmia were selected from the given databases for the purposes of analysis. The age range of the patients is 63–72 years.

To estimate the effective duration of ECG recording, statistical HRV parameters of a healthy patient were preliminarily calculated using rhythmograms of 1, 2, 5, 10, 15, 20, and 30 min duration. The values of the most common temporal and histographic parameters are given in Table 1.

The Norm column of the Table 1 shows the ranges of values of HRV parameters that, according to the authors [5, 9], meet the clinical norm[3].

Within the framework of the task of this work, it is inappropriate to consider recordings of 1-and 2-min duration for which the values of some indices (SDNN, AMo, pNN50, dRR) and TI index differ significantly from the average values.

The advisability of using 5-min recording segments as base samples when analyzing data regardless of the duration of registration is noted in [10]. The consideration of two or three such consecutive segments confirms the conditions of physiological status stability. In the case of rhythm disturbances (arrhythmia), however, it is better to consider recordings having a duration of at least 10 min.

As can be seen from Table 1, SDNN, TI parameters stabilize only at a recording duration of 15 min. Thus, in the present work, recordings of 15 min duration were used to calculate statistical parameters. The choice of this duration is also argued by the fact that, as shown in [8], with 5-min recordings used for rapid diagnosis of CVD, the differences between healthy patients and patients with arrhythmia are not so significant.

It is recommended to estimate spectral and autocorrelation parameters during short ECG recordings (lasting about 5 min) due to their rather rapid stabilization [11]; moreover, at this interval, their changes can be considered as stationary processes.

RESULTS AND DISCUSSION

Table 2 shows HRV parameters for healthy patients; Table 3—for patients with arrhythmia. Confidence intervals were calculated according to the standard methodology for small samples using Student's *t*-test for 95% confidence probability.

[2] The Research Resource for Complex Physiologic Signals. https://physionet.org. Accessed March 24, 2024.

[3] Analysis of heart rate variability. http://protein.bio.msu.ru/~akula/varCI/VarCI.htm (in Russ.). Accessed March 24, 2024.

Table 2. Values of HRV parameters of healthy patients

Parameter \ Patient	1	2	3	4	5	6	7	8	9	10	Average	Norm [8]
HR, bpm	75	82	75	89	108	76	87	77	79	77	82.5 ± 7.3	60–90
RRNN, ms	795	728	805	674	553	789	689	784	762	779	735.8 ± 56.1	660–937
dRR, ms	282	305	398	343	227	375	368	391	422	329	344 ± 43	310–450
SDNN, ms	49	61	41	80	54	76	78	64	73	64	64 ± 9	40–80
Mo, ms	828	727	797	742	594	773	633	797	711	812	741.4 ± 55.5	870–930
AMo, %	41	34	54	27	45	25	41	42	26	40	37.5 ± 6.7	32–38
CV, %	6.1	8.4	5.1	11.8	10	9.6	11.6	8.1	9.6	8.2	8.9 ± 1.5	3–12
pNN50, %	3.9	1.6	2.1	0.7	0.2	13.6	3.1	3.2	8.1	0.5	3.7 ± 3.0	1–9
RMSSD, ms	24	19	20	15	10	41	21	24	31	16	22.1 ± 6.3	20–50
TI, c.u.	88	77	85	53	167	44	88	67	43	74	78.6 ± 25.3	80–150
VLF, %	51	61	74.3	69	53.5	49	72.5	60.4	49.8	40.4	58.1 ± 8.0	15–30
LF, %	35.8	34.5	19.8	23.8	37.9	32	23.1	30.6	34.1	54.6	32.6 ± 7.0	15–40
HF, %	13.2	4.6	5.9	7.2	8.6	19	4.4	9	16.1	5	9.3 ± 3.7	15–25
LF/HF	2.7	7.6	3.3	3.3	4.4	1.6	5.3	3.4	2.1	11	4.5 ± 2.1	1.5–2
LFnorm	73.1	88.3	77	76.9	81.6	62.2	84	77.2	67.9	91.7	78.0 ± 6.4	41.2–60
HFnorm	26.9	11.7	23	23.1	18.4	37.8	16	22.8	32.1	8.3	22.0 ± 6.4	40–58.8
C0	16	31	81	46	44	37	51	28	52	25	41.1 ± 13.1	–
C1	0.72	0.94	0.86	0.93	0.77	0.51	0.87	0.88	0.7	0.89	0.81 ± 0.10	–

Table 3. Values of HRV parameters of sick patients

Parameter \ Patient	1	2	3	4	5	6	7	8	9	10	Average	Norm [8]
HR, bpm	76	74	71	83	66	79	82	74	88	73	76.6 ± 4.6	60–90
RRNN, ms	789	808	841	720	905	759	735	811	680	820	786.8 ± 46.7	660–937
dRR, ms	500	403	342	405	281	711	673	775	997	575	566 ± 160	310–450
SDNN, ms	46	39	46	32	56	83	79	71	147	78	67.7 ± 23.8	40–80
Mo, ms	792	825	869	731	903	761	712	842	581	836	785.2 ± 66.9	870–930
AMo, %	50	50	47	58	32	69	39	37	17	43	44.2 ± 10.2	32–38
CV, %	5.8	4.8	5.4	4.4	6.2	11	10.8	8.8	21.6	9.6	8.8 ± 3.7	3–12
pNN50, %	7.1	26.2	13.2	7.2	26.4	15.7	33.9	14.7	80.4	34.5	25.9 ± 15.4	1–9
RMSSD, ms	54	55	56	45	47	141	126	101	258	110	99.3 ± 47.2	20–50
TI, c.u.	63	75	80	98	63	64	41	28	15	44	57 ± 18	80–150
VLF, %	4.8	4.5	2.1	3.7	9.7	4.1	19.7	14.2	12.3	11	8.6 ± 4.1	15–30
LF, %	6.5	9.6	25.3	3.6	8	8.6	12.3	10.2	15	17.4	11.7 ± 4.5	15–40
HF, %	88.7	85.9	72.6	92.7	82.2	87.3	68	75.6	72.7	71.7	79.7 ± 6.2	15–25
LF/HF	0.07	0.11	0.35	0.04	0.1	0.1	0.18	0.14	0.21	0.24	0.15 ± 0.07	1.5–2
LFnorm	6.8	10	25.8	3.7	8.9	9	15.3	11.9	17.1	19.6	12.8 ± 4.7	41.2–60
HFnorm	93.2	90	74.2	96.3	91.1	91	84.7	88.1	82.9	80.4	87.2 ± 4.7	40–58.8
C0	1	1	2	2	4	1	1	1	1	1	1.5 ± 0.7	–
C1	−0.04	−0.09	0.28	0.05	0.34	−0.5	−0.4	−0.3	−0.22	−0.06	−0.09 ± 0.20	–

The diagram in Fig. 2, which shows the *p*-values of Student's *t*-test (in cases of normal distribution of the trait in both groups) and Mann–Whitney *U* test (in cases where at least in one sample the distribution of the trait is different from normal), demonstrates the differences between the indicators in the groups more clearly. The dashed line corresponds to the selected confidence level of $p = 0.05$.

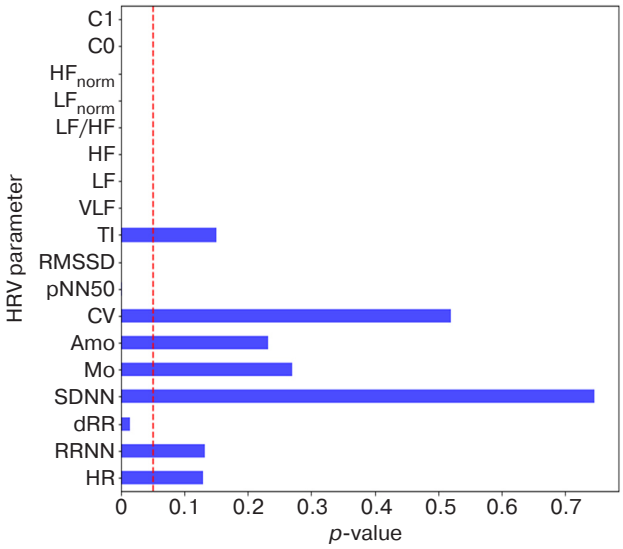

Fig. 2. Diagram of *p*-values for HRV parameters

From an analysis of the parameters of Tables 2 and 3 and the *p*-values of the diagram in Fig. 2, we can conclude that spectral and autocorrelation parameters have the greatest differentiating power for this pathology. Consequently, the inclusion of these parameters in the DF will increase the differentiation between sick and healthy patients. To confirm this assumption, two DFs are compared:

- dependent only on statistical parameters of HRV (stage 1);
- dependent on the complex of HRV parameters: statistical, autocorrelation and spectral (stage 2).

The decision as to which function most adequately separates patients into two clusters will be made based on the results of Student's *t*-test assessing the significance of differences in average values in the two samples.

Stage 1. Discriminant function of statistical indices

1.1. Parameter selection for discriminant analysis

Parameters of an effective DF that carry different information and do not repeat each other should meet the following requirements:

- it is desirable that they have a normal distribution;
- parameters should not be significantly correlated among themselves;
- parameters should be well differentiated in representatives of two groups (differences of average values in the group of healthy and the group of patients should be statistically significant).

An analysis of the data presented in Tables 2 and 3 shows that the values of dRR, RMSSD and pNN50 are very different in healthy and sick patients: the values of dRR are on average 65% higher in the sick group than in the healthy group; RMSSD is 349% higher in the sick group, while pNN50 is 600% higher in the sick group, respectively. Due to RMSSD and pNN50 being highly correlated [12], i.e., interchangeable, RMSSD was used for further analysis.

The AMo parameter usually has a good prognostic significance: exceeding the value of 50% by this parameter is considered as the presence of cardiovascular disease [5]. Therefore, despite the fact that the mode amplitude in patients is on average only 18% higher than in healthy individuals, the decision was taken to include the mode amplitude in the set of DF arguments.

All statistical tests were performed at a significance level of $\alpha = 0.05$.

The hypothesis of normality of parameter distribution was tested using the Shapiro–Wilk test suitable for small samples. The results are summarized in Table 4. The Shapiro–Wilk test statistic is denoted by the letter *W*. The null hypothesis H_0 represents the assumption that a given distribution does not contradict the normal distribution.

It should be noted that, due to irregular rhythmic disturbances in sick patients, the distribution of RR intervals does not tend to normal; thus, the law of distribution of HRV indices may also differ from normal.

Since $W > W_{crit}$, the null hypothesis is not rejected; therefore, there is no reason to believe that the distribution of parameters AMo, dRR, and RMSSD in the group of healthy patients differs from normal.

When applying Student's *t*-test to identify the significance of differences between the average values for each parameter in the group of patients and healthy individuals, it is necessary for the distribution in both samples to correspond to the normal distribution. Table 4 shows that the distribution of the RMSSD parameter in the group of patients with arrhythmia differs from normal. In this case, the *t*-test was replaced by a nonparametric analog comprising a calculation of the Mann–Whitney *U* test, which is less sensitive to distribution deviations from normal and allows us to compare the expression of the parameter in two samples.

The results of Student's *t*-test and the Mann–Whitney *U* test are shown in Table 5. The null hypothesis H_0 in case of the *t*-test rests in the assumption that the differences between the average values in the groups of healthy and sick people are insignificant (or the distribution of the trait in the two groups is the same in the case of the *U* test).

Table 4. Results of testing the hypothesis of normality of distribution of HRV parameters

Group	HRV parameter	Critical value W_{crit}	Calculated value W	Accepted hypothesis
Healthy	AMo	0.842	0.9224	H_0
	RMSSD	0.842	0.9194	H_0
	dRR	0.842	0.9576	H_0
	LF/HF	0.842	0.835	H_0
	C0	0.842	0.9249	H_0
Sick	AMo	0.842	0.9850	H_0
	RMSSD	0.842	0.7976	H_1
	dRR	0.842	0.9533	H_0
	LF/HF	0.842	0.9277	H_0
	C0	0.842	0.6033	H_1

Table 5. Results of Student's *t*-test and Mann–Whitney *U* test

HRV parameters with normal distribution	Critical value t_{crit}	Calculated value t	Accepted hypothesis
AMo	2.101	1.237	H_0
LF/HF	2.101	4.757	H_1
dRR	2.101	3.035	H_1
HRV parameters, the distribution of which differs from the normal distribution	Critical value U_{crit}	Calculated value U	Accepted hypothesis
RMSSD	23	0	H_1
C0	23	0	H_1

As shown in Table 5 the RMSSD and dRR parameters differ markedly between sick and healthy patients. This is because the *t*-statistics value for dRR exceeds the tabulated value, while the *U* value, which is interpreted differently, does not exceed the tabulated value for RMSSD; however, the mode amplitude in this case showed poor discriminatory power. This is most likely due to insufficient sample size, as well as the fact that extrasystoles in patients with arrhythmias strongly affect indices such as TI and AMo. As a result, their values differ little from normal. Nevertheless, in the present work, this indicator was left in place due to its typically high prognostic significance and weak correlation with other HRV parameters. The effectiveness of DF can be additionally evaluated by the value of the coefficient in front of the AMo parameter: it should not be greater than that of the most informative features.

The correlation coefficients were calculated using Pearson's rule and are contained in Table 6 (3 first rows and 3 first columns).

Table 6. Correlation between HRV statistical parameters

Parameter / Parameter	AMo	RMSSD	dRR	LF/HF	C0
AMo	1	−0.52	−0.22	0.22	0.28
RMSSD	−0.52	1	0.58	−0.53	−0.06
dRR	−0.22	0.58	1	−0.30	0.44
LF/HF	0.22	−0.53	−0.30	1	−0.29
C0	0.28	−0.06	0.44	−0.29	1

As Table 6 shows, the correlation between the parameters is weak to moderate (according to the Cheddock scale).

1.2. Standardization and normalization

In the literature on classification methods, standardization [13, 14] or minimax normalization [15, 16] is often used to eliminate differences between parameter units. Otherwise,

DF weights can be misleading regarding the importance of parameters.

Figure 3 shows a characteristic view of a 15-min rhythmogram section with a duration of 1 min, as well as values of HRV parameters calculated from the full (15-min) rhythmogram. Due to large differences in the values of parameters, normalization is necessary to obtain adequate results.

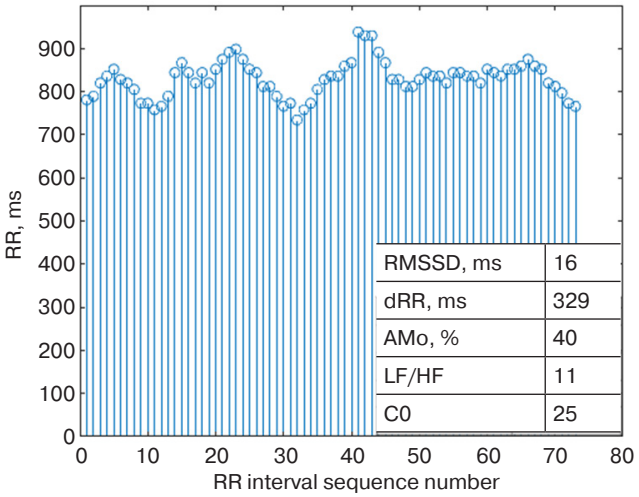

Fig. 3. Rhythmogram section of a healthy patient and HRV parameters

There is no consensus on which method of converting values to a single scale should be preferred. In order to study the differences between the two methods and to identify the most effective one for this study, it was decided to construct two DFs, the parameters of which:

1) have been preliminarily standardized;
2) have been preliminarily normalized.

Standardization was performed using the formula $\frac{x - x_{av}}{\sigma}$, where σ is the SDNN; x_{av} is the average value of the parameter.

Normalization was carried out according to the formula $\frac{x - x_{min}}{x_{max} - x_{min}}$, where x_{min} is the minimum value of the trait; x_{max} is the maximum value of the trait.

1.3. Discriminant analysis

Multivariate discriminant analysis was performed according to the methodology detailed in [17].

The raw data are presented in matrix form. A vector of average values is formed for each trait and each class. Then centered and covariance matrices are calculated. Based on the covariance matrices for two classes (healthy and sick), a general covariance matrix and its inverse matrix are calculated. In order to find the vector of DF coefficients, it is necessary to multiply the matrix inverse to the general covariance matrix by the difference of the two centered matrices.

In the case of standardization, the DF has the following form:

$$DF_s = 1.27 \cdot dRR - 1.73 \cdot AMo - 4.28 \cdot RMSSD. \quad (1)$$

In the case of normalization DF has the following form:

$$DF_n = 4.85 \cdot dRR - 7.217 \cdot AMo - 17.397 \cdot RMSSD. \quad (2)$$

Table 7 shows the results of checking whether the distribution of DF values conforms to the normal law.

According to the Shapiro–Wilk test, the distribution of values of the two DFs in both groups does not contradict the normal distribution. Consequently, to assess the significance of the differences between the averages, we apply Student's *t*-test.

In the case of classification by both DF_s and DF_n (1.44 and 1.94, respectively), the calculated values of Student's *t*-test were less than the critical value of t_{crit} equal to 2.26. Therefore, there is no reason to reject the hypothesis that the average values of DFs in healthy and sick people do not differ.

Figure 4 shows the point distribution of DF values for healthy and sick patients in the case of standardization (a) and normalization (b). The average values of DF are shown with a solid line, while the confidence intervals of DF values are shown with dashed lines. Here, a conclusion about the low efficiency of discrimination becomes obvious, since the confidence intervals of PF values for both groups overlap, and it is not possible to reliably separate the two clusters.

Table 7. Results of testing the hypothesis of normality of distribution of DF values

Group	DF	Critical value W_{crit}	Calculated value W	Accepted hypothesis
Healthy	DF_s	0.842	0.9170	H_0
	DF_n	0.842	0.9157	H_0
Sick	DF_s	0.842	0.8573	H_0
	DF_n	0.842	0.8745	H_0

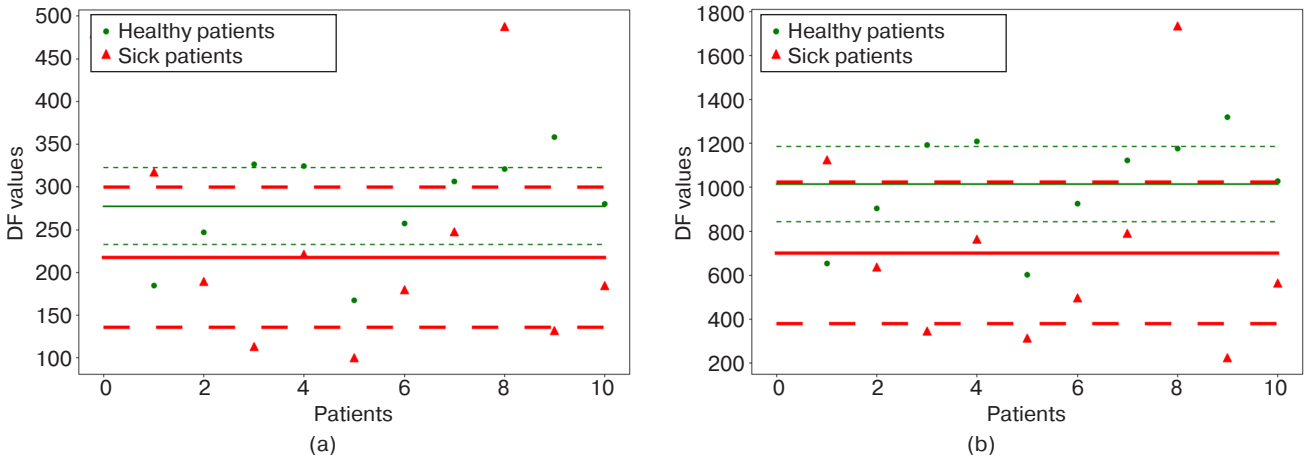

Fig. 4. DF values in case of standardization (a) and normalization (b) (3 parameters) (p = 0.95)

The low efficiency of DF_s and DF_n functions confirms the assumption that statistical parameters are insufficient for reliable arrhythmia detection. This conclusion is confirmed by the results of [8], in which the values of DFs depending on five statistical parameters differed by no more than 25% in healthy patients and patients with arrhythmia.

It is worth noting that both scaling methods assigned the same rank to the traits in terms of their level of contribution to the DF: RMSSD, AMo, dRR (in descending order of importance).

Stage 2. Discriminant function of time, spectral and autocorrelation indices

2.1. Selection of parameters for discriminant analysis

To the previously selected statistical parameters, we added two parameters among the most different in sick and healthy patients: C0 and LF/HF.

From Table 4 it can be seen that the value of W-statistics in the group of sick patients in the case of C0 does not exceed the critical tabular value; since the distribution of this parameter contradicts the normal distribution, while LF/HF statistics correspond to the tabular value within the margin of error, the distribution of this parameter can be considered as not contradicting the normal distribution.

The significance of the differences between the parameters in the two groups was tested using Student's t-test (LF/HF) and the Mann–Whitney U test (C0). Test results are given in Table 5. In both cases, the alternative hypothesis of a statistically significant difference in the values of the parameters in the two groups was confirmed.

The new parameters are weakly or moderately correlated with those introduced earlier. The correlation coefficients between all parameters are shown in Table 6.

2.2. Standardization and normalization of the new parameters

Standardization and normalization of LF/HF and C0 were performed using the calculation formulas given in Section 1.2.

2.3. Discriminant analysis

For standardized parameters, the DF is as follows:

$$DF_s = 0.804 \cdot dRR + 2.732 \cdot AMo + 2.381 \times \\ \times RMSSD - 8.242 \cdot C0 - 5.194 \cdot LF/HF. \quad (3)$$

For normalized parameters DF takes the form:

$$DF_n = 3.357 \cdot dRR + 11.418 \cdot AMo + 9.623 \times \\ \times RMSSD - 27.568 \cdot C0 - 19.162 \cdot LF/HF. \quad (4)$$

Since the hypothesis of normal distribution of DF values was confirmed (the calculated values of W-statistics are given in Table 8), we apply Student's t-test in order to determine the effectiveness of discrimination.

The Student's t-test confirmed the statistical significance of the differences between the averages in the two samples: the value of t-statistics in the case of discrimination by DF_s is equal to 6.67, by DF_n is equal to 6.27, i.e., in both cases it exceeds the critical value of 2.26.

Figure 5 shows the point distribution of DF values for healthy and sick patients in the case of standardization (a) and normalization (b). Average DF values are shown by a solid line, confidence intervals of DF values are shown by a solid range for healthy and sick patients.

It can be seen that the confidence intervals do not overlap, allowing the groups of healthy patients and patients with arrhythmia to be reliably distinguished.

Table 8. Results of testing the hypothesis of normality of distribution of DF values

Group	DF	Critical value W_{crit}	Calculated value W	Accepted hypothesis
Healthy	DF_s	0.842	0.9688	H_0
	DF_n	0.842	0.9682	H_0
Sick	DF_s	0.842	0.9393	H_0
	DF_n	0.842	0.9395	H_0

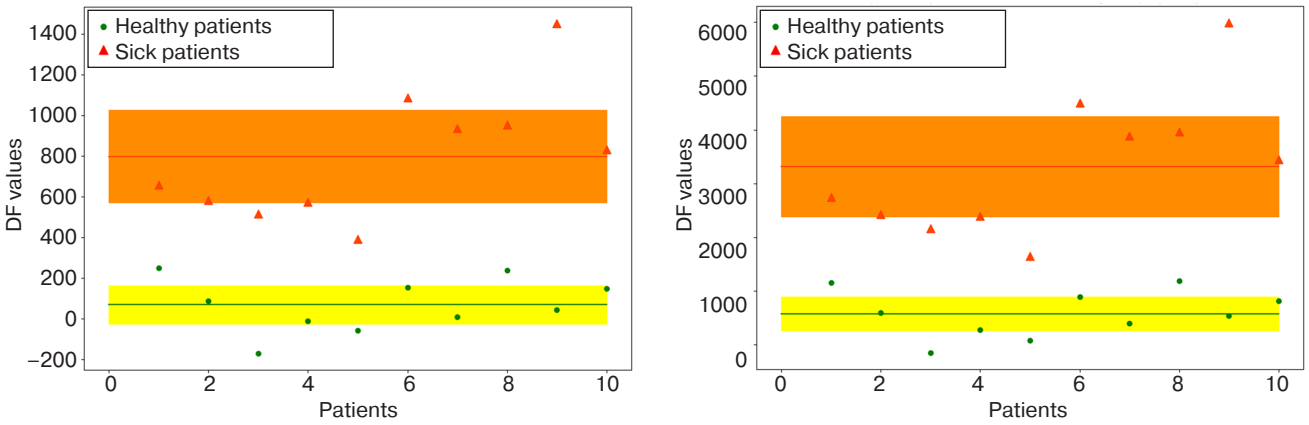

Fig. 5. DF values in case of standardization (a) and normalization (b) (5 parameters)

Both functions arranged the parameters in the same order in decreasing order of their significance: C0, LF/HF, AMo, RMSSD, dRR. Here it should also be noted that the scatter of DF values relative to the average value is identical for both types of scaling, indicating the consistency of the results obtained by standardization and normalization methods.

CONCLUSIONS

The values of variation pulsometry indices for 10 healthy patients and 10 patients with arrhythmia were calculated in this work.

When selecting informative features for MDA, priority was given to the features that have high prognostic significance and those that meet the following criteria: normal distribution, weak correlation, and significantly differing between sick and healthy patients. Two groups of features were formed: a group of statistical parameters (RMSSD, dRR, AMo) and a more representative group that additionally included one spectral and one autocorrelation parameter (RMSSD, dRR, AMo, LF/HF, C0).

DFs were constructed for both groups using different approaches to scaling the values of informative features. Standardization and normalization are shown to lead to the same results: coincidence of the distinguishing ability of the functions and the same estimation of parameter contributions to the DF.

Student's *t*-test clearly demonstrated high classification ability of DF in the case of adding LF/HF and C0 parameters to statistical parameters; DF values for sick and healthy patients overlap neither in the case of standardization nor that of normalization. Thus, it has been shown that MDA can be used to effectively identify differences between patients with arrhythmias and healthy patients. However, it should be noted that the study was conducted on small samples (10 healthy and 10 sick patients); for this reason, these preliminary conclusions are subject to refinement on larger samples.

Authors' contributions

P.A. Sakharova—processing initial data, calculation of parameters, and analysis of the results.

V.A. Balandin—problem statement, development of the research program, analysis of the results.

REFERENCES

1. Ermoshkin V.I. The proposed mechanism of arrhythmia human heart. *Obrazovatel'nyi vestnik "Soznanie" = Educational Bulletin "Consciousness"*. 2013;15(6):4–15 (in Russ.).
2. Getman S.I. Characteristics of Disturbances of Heart Rhythm and Conduction Among Patients Attending Consultation of a Cardiologist at Ambulatory Stage. *Kardiologiya.* 2018;58(6):20–28 (in Russ.). https://doi.org/10.18087/cardio.2018.6.10130
3. Arzykulov Zh.A., Omarov A.A., Kituev B.B., Tursunova F.A., Eshtai A.A., Pavlova N.G. Innovations in the treatment of cardiac arrhythmia. *Vestnik khirurgii Kazakhstana = Bulletin of Surgery in Kazakhstan*. 2012;4(32):4–5 (in Russ.).
4. Ziep B.M., Taratukhin E.O. Heart rate variability assessment and its potential. *Rossiiskii kardiologicheskii zhurnal = Russian Journal of Cardiology*. 2011;6:69–75 (in Russ.).
5. Babunts I.V., Miridzhanyan E.M., Mashaekh Yu.A. *Azbuka analiza variabel'nosti serdechnogo ritma* (*The ABC of Heart Rate Variability Analysis*). Stavropol: Print-Master; 2002. 112 p. (in Russ.).
6. Baevskii R.M., Ivanov G.G., Gavrilushkin A.P., et al. Analysis of heart rate variability using various electrocardiographic systems (Part 1). *Vestnik aritmologii = Journal of Arrhythmology.* 2002;24:65–86 (in Russ.).
7. Bykh A.I., Vysotskaya E.V., Porvan A.P., et al. Using of discriminant analysis for diagnosis chronic cardiac insufficiency on teenagers. *Vestnik Natsional'nogo tekhnicheskogo universiteta Khar'kovskii politekhnicheskii institut. Seriya: Informatika i modelirovanie = Bulletin of the National Technical University Kharkov Polytechnic Institute. Series: Informatics and Modeling*. 2010;31:16–22 (in Russ.).
8. Silkina U.I., Balandin V.A. Discriminant analysis of variational pulsometry parameters. *Rossiiskii tekhnologicheskii zhurnal.* 2020;8(3):81–91 (in Russ.). https://doi.org/10.32362/2500-316X-2020-8-3-81-91
9. Novikov A.A., Smolensky A.V., Mikhailova A.V. Approaches to assessing heart rate variability (literature review). *Vestnik novykh meditsinskikh tekhnologii. Elektronnoe izdanie = J. New Medical Technologies.* 2023;17(3):85–94 (in Russ.). https://doi.org/10.24412/2075-4094-2023-3-3-3
10. Baevsky R.M., Ivanov G.G. Cardiac Rhythm variability: theoretical aspects and opportunities of clinical application. *Ul'trazvukovaya i funktsional'naya diagnostika = Ultrasonnd & Functional Diagnostics*. 2001;3:108–127 (in Russ.).
11. Alejnikova T.V. Heart rate variability (literature review). *Problemy zdorov'ya i ekologii = Health and Ecology Issues*. 2012;1:17–23 (in Russ.). https://doi.org/10.51523/2708-6011.2012-9-1-3
12. Besedina S.A., Balandin V.A. Correlation of heart rhythm variability parameters. *Nauchnyi dialog: Molodoi uchenyi = Scientific Dialog: Young Scientist*. 2017. P. 33–37 (in Russ.). https://doi.org/10.18411/spc-22-11-2017-11
13. Kulikov A.L., Bezdushnii D.I., Osokin V.Yu. Application of linear discriminant analysis for classification of emergency grid emergency states. *Vestnik Ivanovskogo gosudarstvennogo energeticheskogo universiteta = Vestnik of Ivanovo State Power Engineering University*. 2020;5:38–47 (in Russ.). https://doi.org/10.17588/2072-2672.2020.5.038-047
14. Sukiasyan A.G., Markina V.S., Mitrofanov D.P., Shabalina U.M. Clustering departments of integrated groups of enterprises according to level of risk based on methods of multivariate statistical analysis. *Fundamental'nye issledovaniya = Fundamental Research*. 2019;5:115–125 (in Russ.).
15. Budko O.N., Senko E.V. Classification functions for express diagnostics of the degree of disease with arterial hypertension. In: *BIG DATA Advanced Analytics: Collection of Materials of the Fourth International Scientific and Practical Conference*. 2018. P. 360–365 (in Russ.). Available from URL: https://elib.grsu.by/doc/48339
16. Pavlenkov M.N., Smirnova N.A. Development of technology for the assessment of stability of an enterprise of the chemical complex. *Statistika i Ekonomika = Statistics and Economics*. 2013;3:66–69 (in Russ.).
17. Sizykh D.S., Sizykh N.V. Features of teaching methods of discriminant analysis for training specialists in the direction of "Business Informatics". In: Isaev Y.N., Pavlov I.V. (Eds). *Obrazovanie: opyt i perspektivy razvitiya* (*Education: Experience and Development Prospects*). Cheboksary: Sreda; 2019. P. 107–133 (in Russ.). https://doi.org/10.31483/r-33128

СПИСОК ЛИТЕРАТУРЫ

1. Ермошкин В.И. Предполагаемый механизм возникновения аритмии сердца человека. *Образовательный вестник «Сознание»*. 2013;15(6):4–15.
2. Гетман С.И. Распространенность нарушений ритма сердца и проводимости среди обратившихся за медицинской помощью к кардиологу на амбулаторном этапе. *Кардиология*. 2018;58(6):20–28. https://doi.org/10.18087/cardio.2018.6.10130
3. Арзыкулов Ж.А., Омаров А.А., Китуев Б.Б., Турсунова Ф.А., Ештай А.А., Павлова Н.Г. Инновации в лечении аритмии сердца. *Вестник хирургии Казахстана*. 2012;4(32):4–5.
4. Зиеп Б.М., Таратухин Е.О. Возможности методики вариабельности сердечного ритма. *Российский кардиологический журнал*. 2011;6:69–75.
5. Бабунц И.В., Мириджанян Э.М., Машаех Ю.А. *Азбука анализа вариабельности сердечного ритма*. Ставрополь: Принт-мастер; 2002. 112 р.
6. Баевский Р.М., Иванов Г.Г., Гаврилушкин А.П. и др. Анализ вариабельности сердечного ритма при использовании различных электрокардиографических систем (часть 1). *Вестник аритмологии*. 2002;24:65–86.
7. Бых А.И., Высоцкая Е.В., Порван А.П., Рак Л.И. и др. Использование дискриминантного анализа для диагностики хронической сердечной недостаточности у подростков. *Вестник Национального технического университета Харьковский политехнический институт. Серия: Информатика и моделирование.* 2010;31:16–22.

8. Силкина У.И., Баландин В.А. Дискриминантный анализ параметров вариационной пульсометрии. *Российский технологический журнал*. 2020;8(3):81–91. https://doi.org/10.32362/2500-316X-2020-8-3-81-91
9. Новиков А.А., Смоленский А.В., Михайлова А.В. Подходы к оценке показателей вариабельности сердечного ритма (обзор литературы). *Вестник новых медицинских технологий.* 2023;17(3):85–94. https://doi.org/10.24412/2075-4094-2023-3-3-3
10. Баевский Р.М., Иванов Г.Г. Вариабельность сердечного ритма: теоретические аспекты и возможности клинического применения. *Ультразвуковая и функциональная диагностика*. 2001;3:108–127.
11. Алейникова Т.В. Вариабельность сердечного ритма (обзор литературы). *Проблемы здоровья и экологии*. 2012;1:17–23. https://doi.org/10.51523/2708-6011.2012-9-1-3
12. Беседина С.А., Баландин В.А. Корреляция показателей вариабельности сердечного ритма. *Научный диалог: Молодой ученый*. 2017. С. 33–37. https://doi.org/10.18411/spc-22-11-2017-11
13. Куликов А.Л., Бездушный Д.И., Осокин В.Ю. Применение линейного дискриминантного анализа для классификации аварийных режимов электрической сети. *Вестник Ивановского государственного энергетического университета*. 2020;5:38–47. https://doi.org/10.17588/2072-2672.2020.5.038-047
14. Сукиасян А.Г., Маркина В.С., Митрофанов Д.П., Шабалина У.М. Кластеризация подразделений интегрированной группы предприятий по уровню риска на основе методов многомерного статистического анализа. *Фундаментальные исследования*. 2019;5:115–125.
15. Будько О.Н., Сенько Е.В. Классификационные функции для экспресс-диагностики степени заболевания артериальной гипертензией. В сб.: *BIG DATA и анализ высокого уровня: материалы Четвертой Международной научно-практической конференции.* 2018. С. 360–365. URL: https://elib.grsu.by/doc/48339
16. Павленков М.Н., Смирнова Н.А. Разработка технологии оценки устойчивости предприятия химического комплекса. *Статистика и Экономика.* 2013;3:66–69.
17. Сизых Д.С., Сизых Н.В. Особенности преподавания методов дискриминантного анализа для подготовки специалистов по направлению «Бизнес-информатика». В кн.: *Образование: опыт и перспективы развития*; под ред. Ю.Н. Исаева, И.В. Павлова. Чебоксары: ИД «Среда»; 2019. С. 107–133. https://doi.org/10.31483/r-33128

About the authors

Polina A. Sakharova, Bachelor, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: polinka6777@gmail.com. https://orcid.org/0009-0008-1323-9308

Vyacheslav A. Balandin, Cand. Sci. (Phys.-Math.), Assistant Professor, Department of Biocybernetics Systems and Technologies, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: admiral49@mail.ru. Scopus Author ID 7003691025, RSCI SPIN-code 1288-9918

Об авторах

Сахарова Полина Александровна, бакалавр, Институт искусственного интеллекта, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: polinka6777@gmail.com. https://orcid.org/0009-0008-1323-9308

Баландин Вячеслав Алексеевич, к.ф.-м.н., доцент, кафедра биокибернетических систем и технологий, Институт искусственного интеллекта, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: admiral49@mail.ru. Scopus Author ID 7003691025, SPIN-код РИНЦ 1288-9918

Translated from Russian into English by Lyudmila O. Bychkova
Edited for English language and spelling by Thomas A. Beavitt

MIREA – Russian Technological University.
78, Vernadskogo pr., Moscow, 119454 Russian Federation.
Publication date January 30, 2025.
Not for sale.

МИРЭА – Российский технологический университет.
119454, РФ, г. Москва, пр-т Вернадского, д. 78.
Дата опубликования 30.01.2025 г.
Не для продажи.