Investigation of Models for Prognosis of Critical Values of Non-Invasive Electrophysiological Parameters of Pregnant Women with Abnormalities of Heart Rate

Maxim Mitrokhin, Andrey Kuzmin, Nikita Dyatlov, Fagim Rakhmatullov, Natalia Mitrokhina, Alan Alimuradov
Alexander Tychkov
Penza State University
Penza, Russia
mmax83@mail.ru, flickerlight@inbox.ru, sakedas@gmail.com, fagim.rakhmatullov@mail.ru, {natena1, alansapfir}@yandex.ru, tychkov-a@mail.ru

***Abstract*—The article presents the research materials on the development of models and methods for prognosis of critical values in the number and duration of atrial fibrillation for trimesters in pregnant women. The choice of cardiac activity characteristics, most influencing the predicted values, obtained in the framework of non-invasive electrophysiological studies, is substantiated. A new two-step method of selection of predictors for creation of models of electrophysiological parameters of heart prediction is proposed. It allows to identify the most valuable dependencies of target parameters. Based on the selected characteristics, a research is conducted to build several types of prediction models, and to determine the indicators of their effectiveness**

I. Introduction

Currently, cardiovascular diseases occupy a leading place in the worldwide mortality [1], ischemic heart disease, a stroke, arterial hypertension, heart rhythm disturbances, and other diseases being traditionally distinguished among them.

Heart rhythm disturbances are widespread pathological conditions characterized by changes in frequency, rhythm, and sequence of cardiac contractions. One of the most common arrhythmias is atrial fibrillation (AF), in which there is a chaotic contraction of individual atrial muscle fibers at a frequency of 350-900 beats per minute due to a physiological abnormality of atrioventricular conduction. Such contractions lead to discoordination of atrial and ventricular work, and indiscriminate ventricular contractions [2].

Atrial fibrillation is dangerous primarily for its thromboembolic complications, namely, stroke and systemic embolism of the circulatory system [3]. Therefore, timely detection of arrhythmia will avoid premature death and disability of a person.

In recent years, due to the widespread use of non-invasive methods for the diagnosis of atrial fibrillation (24-hour ECG monitoring [4], transesophageal electrophysiological study of the heart [5]), and the success of their interventional treatment (in particular, radiofrequency ablation [6]), there is a tendency to reduce the negative consequences of this disease. However, for some groups of patients, for example, pregnant women, there remain a number of unresolved issues. In particular, in view of the complexity of carrying out intracardiac electrophysiological studies due to radiation stress, triggering and supporting mechanisms of atrial fibrillation and arrhythmia during pregnancy trimesters remain unstudied [7], [8].

It is known that the organism of a pregnant woman undergoes a serious hormonal (increase in the level of estrogen and progesterone), hemodynamic (increase in the volume of circulating blood, stroke volume, minute volume of heart, heart rate (HR), total peripheral resistance (TPR) of blood vessels), and vegetative alterations (imbalance in the activity of the sympathetic and parasympathetic nervous system). These changes lead to the onset and progression of heart rhythm disturbances [9], [10]. Atrial fibrillation during pregnancy is fraught with intrauterine growth retardation due to chronic hypoxia, thromboembolism, miscarriage, and early Caesarean section [10], [11].

In this regard, it is essential to predict the course of atrial fibrillation in pregnant women for trimesters.

Modern technical equipment of medical institutions allows to collect and accumulate a significant amount of diagnostic information that can be used to assess the patient's condition, and to identify the risk of occurrence and frequency of arrhythmia, and to take measures to prevent it, including prescribing preventive medication.

Thus, it is necessary to develop new methods of information processing that allow a doctor to assess the current status of a pregnant woman by analyzing the received non-invasive electrophysiological indicators and, most importantly, to predict the course of atrial fibrillation at any gestation period, and to start preventive medication in time.

Creation of models for prediction requires a set of researches that includes:

- Collecting ECG records of pregnant women with arrhythmias into a database;

- Analysis of parameters for identifying target predictable parameters that are the most suitable for prediction of clinical course and estimation of current conditions of pregnant women;
- Selection of predictors from the set of monitoring parameters for creation of models for prediction of values of target parameters;
- Creation of models for prediction of target parameters that meet the requirements of accuracy of prediction and easiness of application by the practicing doctor.

II. Prediction of Cardiac Monitoring Parameters

A. The current state

When examining pregnant women, the choice of diagnostic methods is severely limited in view of their possible negative impact on fetal health. Methods for studying heart rhythm disorders during pregnancy should be non-invasive, atraumatic, convenient and inexpensive. These criteria are met by methods of Holter ECG monitoring (HECG) and transesophageal electrophysiological study of the heart (TEEPS), which allow determining critical deviations of electrophysiological parameters of heart activity accompanying the development of dangerous complications for the fetus and pregnant woman.

The method of 24-hour outpatient ECG monitoring was proposed more than half a century ago by Norman Holter [12], but has not lost its relevance nowadays. The main purpose of the HECG is to identify, document and characterize abnormal cardiac activity under the patient's daily activity. Currently, the HECG is used in the diagnosis of paroxysmal heart rhythm disturbances, which are difficult to document with a standard ECG [13].

The spectrum of parameters obtained in the course of the HECG is quite large:

- minimum, maximum and average heart rate (HR);
- heart rate variability;
- episodes of brady- and tachycardia;
- measurement of heart cycle intervals;
- number of single, paired and group supraventricular (SVES) and ventricular extrasystoles (VES);
- supraventricular (SVT) and ventricular tachycardia (VT);
- pause rhythm;
- ischemic episodes of ST segment elevation and depression;
- detection and analysis of atrial fibrillation (AF), etc.

All described parameters are registered during the day, being grouped into day, night and total ones [14].

The TEEPS method is used for the differential diagnosis of some supraventricular tachyarrhythmias, which look the same on ECG and HECG. It is also used to provoke and arrest supraventricular tachycardias, and evaluate the function of the conduction system of the heart [5], [15]. Its effectiveness is similar to an invasive costly intracardiac electrophysiological study using X-rays [16], but, unlike the latter, the conduct of TEEPS is safe for pregnant women and children [17].

The following indicators are evaluated in the course of the TEEPS method:

- standard (SNRT) and corrected sinus node recovery time (cSNRT);
- effective refractory periods of the atrioventricular node (AVN-EPR);
- left atrium (A-ERP);
- fast and slow pathways of the atrioventricular node (ERP beta, ERP alpha);
- Wenckebach point (WP);
- intervals of basic (St1-R1) and extrastimulus (St2-R2) with programmed stimulation;
- upper (UW) and the lower window (LW);
- cycle of tachycardia (RR-tachycardia);
- time of ventriculoatrial (VA);
- atrioventricular (AV) conduction, etc.

In view of such a wide range of parameters being analyzed, it is often difficult for the clinician to interpret endpoints, and identify those that play a key role in the development of atrial fibrillation. In addition, the analysis of these indicators in pregnant women with atrial fibrillation is rare, and the triggering factors of atrial fibrillation in this group of patients are not described.

Thus, it seems important to analyze the indices obtained by the methods of HECG and TEEPS in pregnant women, and to distinguish ones that play the most important role in the development of paroxysms of atrial fibrillation in these patients using the mathematical analysis. Based on the analysis of the key indicators obtained, it is necessary to build a model for prognosis of the course of atrial fibrillation for pregnancy trimesters.

B. The source data of research

Data collection was carried out at the clinical base of Penza State University (Penza, Russia) using the HECG method on the Schiller MT-101 apparatus (Schiller AG, Switzerland), and the TEEPS method using a complex for non-invasive electrophysiological studies of Astrocard Polysystem-EP/L (Meditek Ltd., Russia).

To conduct the study, a database containing information on 90 pregnant women with a paroxysmal form of atrial fibrillation, which were divided into two groups, was created. The first group included 31 pregnant women with asymptomatic paroxysms of atrial fibrillation (group AS31), and the second group included 59 pregnant women with symptomatic paroxysms of atrial fibrillation (group S59).

All studied women have undergone the HECG and the TEEPS at the beginning of pregnancy (source data), in each trimester of pregnancy, and in the postpartum period. Thus, a total of 450 HEGG records, and 450 TEEPS records were analyzed.

In the course of the HECG, the following parameters were assessed:

- minimum, maximum, average HR;
- number of single, paired and group SVES;
- number and duration of AF paroxysms;

- mean ventricular contraction rate (VCR) during paroxysmal AF;
- number of single and paired VES;
- sum of SVES and VES.

During the performance of the TEEPS, the following indicators were evaluated:

- cSNRT;
- AVN-EPR;
- A-ERP;
- WP;
- St1R1, St2R2, St2R2 increment;
- St2R2/St1R1, St2R2/ AVN-EPR.

Thus, 49 indicators were analyzed in total.

C. Determination of dependencies in the source data

It is known that the most vivid characteristics of arrhythmia are the number and duration of atrial fibrillation. They also characterize the severity of arrhythmia. Therefore, these characteristics are selected as predictors.

Fig. 1 shows an average atrial fibrillation number and total duration of atrial fibrillation in groups of symptomatic and asymptomatic patients in the first, second and third trimesters.

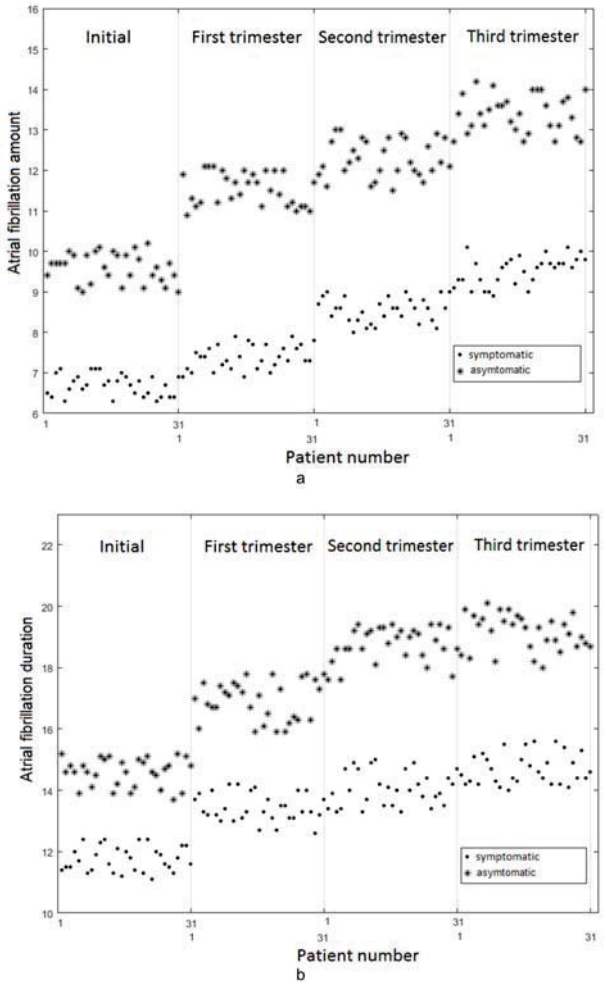

Fig. 1. Average number of atrial fibrillations (a), and their total duration divided by trimesters (b) in symptomatic and asymptomatic patients

The analysis of the values of the predicted features shows that for the group of asymptomatic patients the values of both the number of fibrillations and their total duration significantly exceed the values of the same features in the group of symptomatic patients. Similar effects are observed for most other symptoms. This allows us to conclude that there is a need for separate consideration of two groups of patients.

At the next stage of the analysis, a relationship between the values of the predicted characteristics and other indicators was established. Pair dependencies between the features are strong enough; however, the form of each dependency may be quite different from a linear one.

For example, Fig. 2 shows the scatter plot for the pairwise values of the following characteristics: Atrial fibrillation amount (AF amount), Minimal night heart rate (Min night HR), Total number of supraventricular extrasystoles (Total SVE), Increment of the interval S2-V2 (S2-V2 increment), Effective refractory period of atrioventricular node (AVN-ERP) for symptomatic and asymptomatic patient groups.

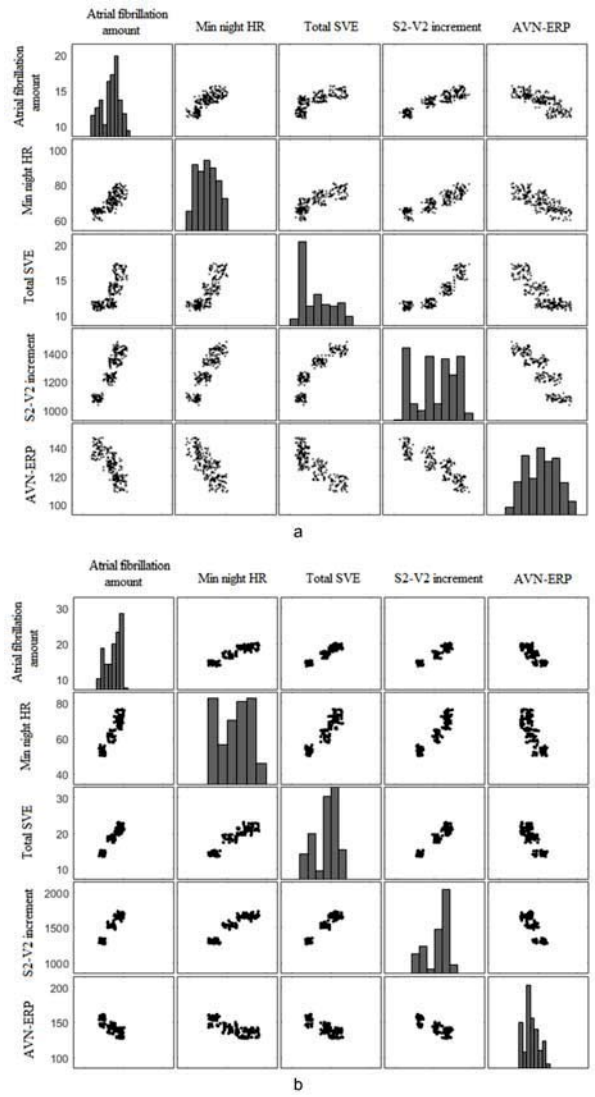

Fig. 2. Scatter plot for the pairwise values of some characteristics in the symptomatic (a) and asymptomatic (b) patient groups during the three trimesters

The type of dependence for one pair of characteristics can differ in groups. For example, for the total number of supraventricular extrasystoles, paired relationships with other symptoms in the asymptomatic patient group can be approximated by a straight line, and they are nonlinear in the symptomatic group.

Before the construction of the prediction model, the selection of characteristics was carried out in each of the patient groups. While selecting the characteristics it is necessary to keep their good interpretability by a physician. Thus the known methods of decreasing the number of dimension of characteristics space such as principal component analysis and independent component analysis are not applicable because the resulting new characteristics are a combination of the initial ones and lose their physiological interpretation. Therefore the new two-steps method is proposed. The first stage is aimed at decreasing the number of parameters space dimension by excluding the characteristics which have high pair correlation from the set. The pairs of features with a coefficient of pair correlation greater than 0.8 were removed from the set. Such features lead to instability of the parameters of the prediction model and therefore, one feature from a pair of such characteristics was left. Thus, the dimensionality of the feature space was reduced to 15 features.

At the second stage, the selection of characteristics was performed by the iterative procedure of the neighborhood component analysis (NCA) [18]. This procedure defines the feature of weights for minimization the average leave one out (LOO) loss over the prior data:

$$L(w) = \frac{1}{N}\sum_{i=1}^{N} l_i + \lambda \sum_{m=1}^{M} w_m^2 \qquad (1)$$

where N is a number of observations in training set $S=\{(\mathbf{x}_i, y_i), i=1,2, ..., N\}$, $\mathbf{x}_i \in \Re^M$ is the feature vector, y_i is response variable, l_i is loss function of nearest neighbor regression, λ is regularization term, that prevents many weights w to have large value.

In NCA feature selection for regression is considered that:

- the randomly picked point $\mathbf{x}_r$ from S is the "reference point" for $\mathbf{x}$, where $\mathbf{x}$ are predictors for predicted value $\hat{y}$;
- the response value $\hat{y}$ at $\mathbf{x}$ is equal to the response value of the reference point $\mathbf{x}_r$.

Then, the probability that some point $\mathbf{x}_j$ from S is the "reference point" for $\mathbf{x}$ is higher, if the closer $\mathbf{x}_j$ to $\mathbf{x}$, and the less the difference $\hat{y} - y_j$ are. Proximity can be estimated by some distance function d.

The LOO procedure attempts to predict the response for $\mathbf{x}_i$ using the data in S^{N-1} (excluding the point $(\mathbf{x}_i, y_i)$). The probability that point $\mathbf{x}_j$ is picked as the reference point for $\mathbf{x}_i$ is

$$p_{ij} = \frac{K(d(\mathbf{x}_i, \mathbf{x}_j))}{\sum_{j=1, j\neq i}^{N-1} K(d(\mathbf{x}_i, \mathbf{x}_j))} \qquad (2)$$

and the average value of loss function l_i of difference between present y_i and predicted $\hat{y}_i$ is

$$l_i(y_i, \hat{y}_i) = \sum_{j=1, j\neq i}^{N-1} p_{ij} l_i(y_i, y_j) \qquad (3)$$

The above mentioned K is kernel function.

In the current research, the mean squared error (MSE) was selected as the loss function $l_i(y_i, y_j)$, defined as

$$l_i(y_i, y_j) = \frac{1}{N}\sum_{i=1}^{N} (y_i - y_j)^2 \qquad (4)$$

The regularization term λ is defined by cross-validation procedure (Fig. 3).

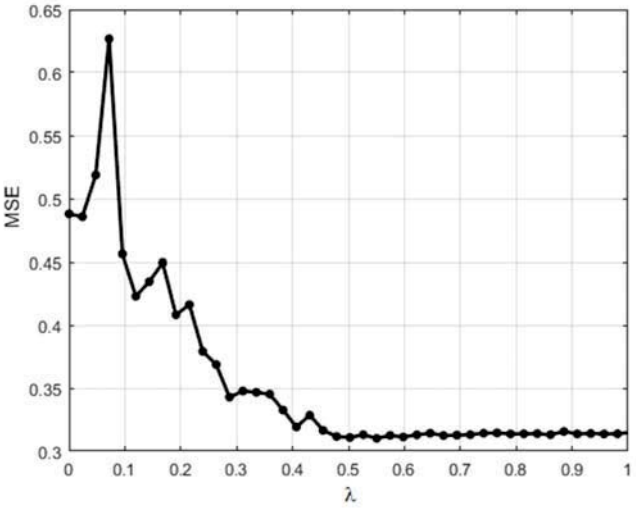

Fig. 3. Graph of MSE dependence on the regularization term λ defined by cross-validation procedure

The value of the regularization term, ensuring a minimum MSE λ=0.551 was used in feature selection by NCA. As a result of selection, feature sets most significant for predicting the number and total duration of atrial fibrillation in two groups of patients were formed (Table I).

TABLE I. SETS OF PREDICTOR VARIABLES

	Atrial fibrillation number	**Atrial fibrillation duration**
Symptomatic	Maximum night HR, Total number of SVE, VE, Total number of ventricular extrasystoles, AVN-ERP, A-ERP, S1-V1 min interval	Average night HR, Total number of SVE, VE, AVN-ERP, A-ERP, S1-V1 min interval
Asymptomatic	Average night HR, Total number of SVE and VE, cSNRT, A-ERP, S1-V1 min interval	Average night HR, Total number of SVE and VE, cSNRT, AVN-ERP, A-ERP, S1-V1 min interval

Abbreviation in Table I: HR – heart rate, SVE – supraventricular extrasystoles, VE – ventricular extrasystoles, AVN-ERP – Effective refractory period of atrioventricular node, A-ERP – atrial effective refractory period, S1-V1 min interval – minimal interval between basic stimulus S1 and ventricular response V1 are obtained, cSNRT – corrected sinus node recovery time.

D. Prediction models

One of the most difficult stages in the construction of prediction models based on a priori data is the choice of a model of optimal complexity.

This choice is often assigned on the model's developer, which puts the quality of the resulting decision in dependence on his experience and qualification. To avoid this dependence, one can use the group method of data handling (GMDH) [19].

The GMDH method is a family of algorithms directed on the choice of models of optimal complexity in terms of the parameters number in a given class of models.

The method involves the construction of real-valued features using $\mathbf{x}=(x^1,x^2,\ldots, x^M)$ based on a priori data retrieval S of the optimal dependence of the output variable $y=f(\mathbf{z})$, where $\mathbf{z}$ is the secondary feature set calculated as possible combinations of the initial ones. For example, for the two features, a secondary set $\mathbf{z}=(x^1,x^2, (x^1)^2, x^1\cdot x^2, (x^2)^2, \ldots, (x^2)^R)$ can be formed.

The GMDH includes the following steps.

1) A class of models (support functions) is selected, in which a sequential search for models of increasing complexity (number of parameters) is performed. The polynomials are frequently used for the model search, in particular, Kolmogorov-Gabor polynomial:

$$y = a_0 + \sum_{i=1}^{M} a_i x^i + \sum_{i=1}^{M}\sum_{j=1}^{M} a_{ij} x^i x^j + \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M} a_{ijk} x^i x^j x^k + \ldots \quad (5)$$

where $A=(a_1, a_2, \ldots, a_{12}, a_{21}, \ldots, a_{ij}, \ldots)$ is the vector of model coefficients.

2) In the selected class, various variants of models are constructed, the quality of which is determined within the framework of the cross-validation procedure. The generation of models of increasing complexity is carried out until the optimum of the external quality criterion of the model is found.

A combinatorial algorithm for model generating is the simplest one. It has an iterative structure. Such an algorithm generates all possible linear combinations of the initial arguments, the complexity of which does not exceed a predetermined F-number of the secondary features in the form of productions of the original ones, that is, $x^1 \to z^1, x^2 \to z^2, (x^1)^2 \to z^3, x^1\cdot x^2 \to z^4, \ldots, (x^M)^R \to z^F$. Models of non-decreasing complexity are constructed for all $\mathbf{z}$ combinations of input arguments $\mathbf{x}$:

$$\begin{aligned} y_1 &= a_{10} + a_{11} z^1, \\ y_2 &= a_{20} + a_{21} z^2, \\ y_3 &= a_{30} + a_{31} z^1 + a_{32} z^2, \\ &\vdots \\ y_k &= a_{k0} + a_{k1} z^1 + a_{k2} z^2 + \ldots + a_F z^F. \end{aligned} \quad (6)$$

The parameters of each model are estimated by the method of least squares in the training sample; the model quality is estimated by an external criterion, for example, the root-mean-square error in the control sample.

The GMDH application allows formalizing the choice procedure for the order of the model complexity.

For example, in Fig. 4 the MSE graph of the count of atrial fibrillation estimated for prediction models in the symptomatic patient group according to the formed set of features is given.

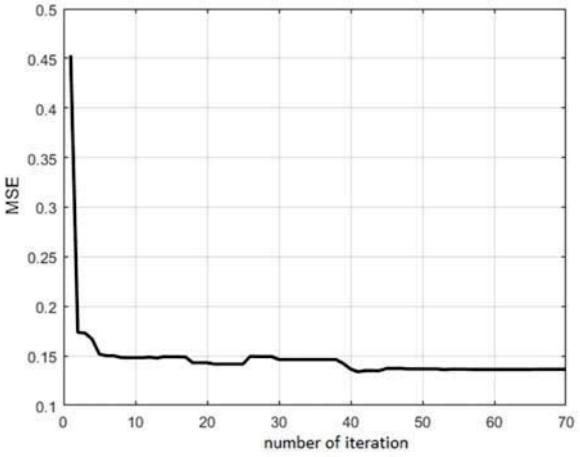

Fig. 4. Graph of MSE dependence on the control sample in selecting the model using the GMDH

The model obtained at the 41st iteration of the method has the lowest MSE score, but at the same time the high complexity (number of parameters). The graph shows that the acceptable accuracy of the prediction is provided by the models after fifth iteration of the GMDH. In practice, using a model with fewer numbers of parameters is more justified with a small loss of accuracy.

A prediction example of the count of atrial fibrillation for two patients of the symptomatic group, not included in the training sample, is shown in Fig. 5.

The estimation of the accuracy of prognostic models on the fifth iteration GMDH based on the data that are not included in the training set shows that MSE of prediction of the count of atrial fibrillation and the total duration of atrial fibrillation at the border of the third trimester are not more than 0.2 and 0.38 respectively. Border of prognosis confidence interval based on the confidence probability 0.95 for the count of atrial fibrillation is ±0.278, and for the total duration of atrial fibrillation it is ±0.441.

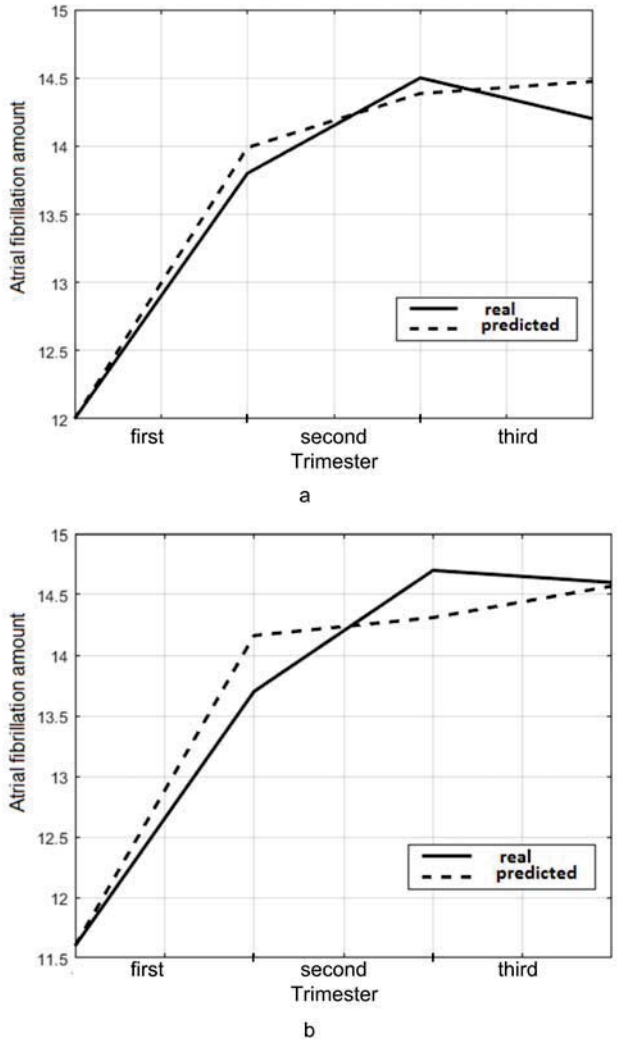

Fig. 5. A prediction example of the count of atrial fibrillation for two patients of the symptomatic group not included in the training sample: (a) Patient 1 and (b) Patient 2.

III. Conclusion

Modern methods for data analysis provide a wide range of tools for identifying complex dependencies of a large number of factors. Thus, in the conducted research, a correlation of some indicators of the course of arrhythmias in pregnant women obtained with the help of HECG and TEEPS was established. The developed models for predicting the number and duration of atrial fibrillation for trimesters allow not only assessing the current status of a pregnant woman, but also timely to predict diseases and to plan the therapeutic measures. It is extremely important to construct a prediction model of arrhythmias course during pregnancy in initially non-pregnant women, as well as in patients during the short term of pregnancy, which allows carrying out the stratification of risk based on the calculation of the number and duration of paroxysms of atrial fibrillation at different stages of pregnancy.

In future, it is planned to work out the question of hormonal indices' influence on cardiac activity, and "personalization" of prediction models to consider the individual characteristics of the disease course in each patient.

Acknowledgment

The authors are grateful to the Russian Science Foundation for the financial support of the project No. 17-71-20029.

References

[1] World Health Organization, *"Mortality Database: Cause of Death Query Online"*. Web: http://apps.who.int/healthinfo/statistics/mortality/causeofdeath_quer.

[2] R.C.W. Christopher, "Atrial Fibrillation: The Most Common Arrhythmia", *Texas Heart Institute Journal,* vol. 27(3), 2000, pp. 257–267.

[3] H. Naess, U.W. Andreassen, L. Thomassen et al, "A score for paroxysmal atrial fibrillation in acute ischemic stroke". *International Journal of Stroke*, 2017, in press.

[4] A. Galli, F. Ambrosini, F. Lombardi, "Holter Monitoring and Loop Recorders: From Research to Clinical Practice", *Arrhythmia and Electrophysiology Review*, vol. 5(2), 2016, pp. 136-143.

[5] B. Brembilla-Perrot, D. Beurrier, P. Houriez et al. "Utility of transesophageal atrial pacing in the diagnostic evaluation of patients with unexplained syncope associated or not with palpitations". *International Journal in Cardiology*, vol. 96(3), 2004, pp. 347-53.

[6] N. Mujović, M. Marinković, R. Lenarczyk et al, "Catheter Ablation of Atrial Fibrillation: An Overview for Clinicians", *Advances in Therapy*, 2017, in press.

[7] V. Katsi, G. Georgiopoulos, M. Marketou et al, "Atrial fibrillation in pregnancy: a growing challenge", *Current Medical Research and Opinion*, vol. 33(8), 2017, pp. 1497-1504.

[8] L. Eckardt, K.G. Häusler, U. Ravens, "ESC guidelines on atrial fibrillation 2016 : Summary of the most relevant recommendations and modifications", *Herz*, vol. 41(8), 2016, pp. 677-683.

[9] F.K. Rakhmatullov, S.V. Klimova, A.M. Kuryaeva, N.E. Dyatlov, E.G. Zinovieva, L.F. Burmistrova, "Influence of Pregnancy on Frequency of Occurrence of Extrasystoles and Paroxisms of Reciprocal Atrioventricular Junctional Tachycardia", *Bulletin of High Educational Institutions. Volga Region. Medical Sciences,* vol. 2(34), 2015, pp. 103-112. (In Russian)

[10] J.H. McAnulty, "Arrhythmias in pregnancy", *Cardiology Clinics*, vol. 30(3), 2012, pp. 425-434.

[11] S.H. Chang, C.F. Kuo, I.J. Chou et al, "Outcomes Associated With Paroxysmal Supraventricular Tachycardia During Pregnancy", *Circulation,* vol. 7, 2017, pp. 616-618.

[12] N.J. Holter, "New method for heart studies", *Science,* vol. 134(3486), 1961, pp. 1214-1220.

[13] J.S. Steinberg., N. Varma, I. Cygankiewicz et al. "2017 ISHNE-HRS expert consensus statement on ambulatory ECG and external cardiacmonitoring/telemetry", *Heart Rhythm*, vol. 14(7), 2017, pp. 55-96.

[14] F.K. Rakhmatullov, S.V. Klimova, A.M. Kuryaeva, N.E. Dyatlov, E.G. Zinovieva, L.F. Burmistrova, "Values of Conductive System of the Heart of Women with Asymptomatic Paroxysms", *Bulletin of High Educational Institutions. Volga Region. Medical Sciences,* vol. 1(33), 2015, pp. 87-87. (In Russian)

[15] H.H. Haaland, T.H. Morstøl, J. Vegsundvåg et al, "Diagnostic transoesophageal atrial stimulation", *Journal of the Norwegian Medical Association*, vol. 123(18), 2003, pp. 2577-2579.

[16] A. Akin, S. Özer, T. Karagöz et al, "Sensitivity of transesophageal electrophysiologic study in children with supraventricular tachycardia on electrocardiography", *Pacing and Clinical Electrophysiology*, vol. 37(8), 2014, pp. 1002-1008.

[17] J.C.R. Lee, G. Wetzel, K. Shannon, "Maternal arrhythmia management during pregnancy in patients with structural heart disease", *Progress in Pediatric Cardiology*, vol. 19(1), 2004, pp. 71-82.

[18] J. Goldberger, S. Roweis, G. Hinton, R. Salakhutdino, "Neighbourhood Components Analysis", *Advances in Neural Information Processing Systems*, vol. 17, 2004, pp. 513–520.

[19] A.G. Ivakhnenko, "Polynomial Theory of Complex Systems", *IEEE Transactions on Systems, Man and Cybernetics,* vol. SMC-1, no.4, 1971, pp. 364–378.