AUTOMATION AND CONTROL IN ENGINEERING

Comparative Analysis of Machine Learning Methods for Prediction of Heart Diseases

I. V. Stepanyan[a],*, Ch. A. Alimbayev[b], M. O. Savkin[c], D. Lyu[c], and M. Zidun[c]

[a] *Mechanical Engineering Research Institute of the Russian Academy of Sciences, Moscow, Russia*
[b] *U. Joldasbekov Institute of Mechanics and Engineering, Almaty, Kazakhstan*
[c] *Peoples' Friendship University of Russia, Moscow, Russia*
*e-mail: neurocomp.pro@gmail.com

Abstract—The problem of combating cardiovascular diseases is becoming increasingly important due to the high level of disability and mortality from heart disease. In this paper, a study of methods for predicting heart disease using electrocardiography and machine learning algorithms was conducted. In total, during the study, 75 000 numerical experiments with various machine learning algorithms and their parameters were conducted. Based on the comparative analysis, the models and methods of machine learning were selected that gave the best results. The following methods were applied: logistic regression, k-nearest neighbors algorithm, decision tree, support-vector machine, Bayesian classifier, random forest, and deep neural networks. The selected models were generalized to identify their parameters and effective application.

INTRODUCTION

According to the World Health Organization, mortality from cardiovascular disease (CVD) has long occupied a leading position in the world. However, in Russia and Kazakhstan, people die from diseases of the circulatory system almost twice as often as, for example, in European countries [1, 2]. More than others, it threatens people with chronic cardiac insufficiency, which most often develops as a result of arterial hypertension, coronary heart disease, rheumatic malformations, and anemia of various origin.

Timely medical assistance is of decisive importance for preserving the life and health of the sufferers, reducing disability and mortality. Long-term, often lifelong, drug treatment, as well as its high cost, dictates the need to pay more and more attention to the early primary prevention of these diseases. To reduce the risks of life-threatening arrhythmia, it is necessary to improve the systems for diagnosing and processing electrocardiosignals (ECG). A significant part of existing research recommends machine learning algorithms for predicting heart disease.

MATERIALS AND METHODS

In cardiodiagnostics, the QRS complex is widely used, which displays the process of propagation of excitation through the myocardium of the ventricles of the heart (the so-called "depolarization" of the ventricles). The QRS complex consists of Q, R, and S waves on an ECG (Fig. 1). Intelligent systems based on machine learning methods, compared to traditional methods, are able to predict accurately some heart diseases at an early stage, even with complex input data [3]. To study and achieve effective recognition of ECG data, a convolutional neural network is presented in [4] to perform encoding of a single QRS complex with the addition of entropy features. The study is aimed at determining the combination of signal information and providing the best result for subsequent classification of cardiac signals. The analyzed information included the raw ECG signal, its entropy characteristics, and extracted QRS complexes. The methods used in this study for calculation features based on entropy and R-wave detection had limitations in use due to their high computational complexity.
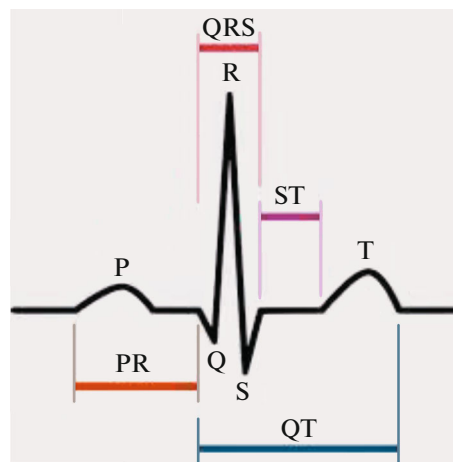
Fig. 1. Scheme of depolarization of the heart ventricles on the ECG (structure of the QRS complex).

In cardiology, ultrasound investigations are used to diagnose heart disease associated with myocardial infarction. Research [5] is aimed to develop methods for segmenting the left ventricle on ultrasound images to check for myocardial movement during a heartbeat. The proposed method uses machine learning methods, such as active contour and convolutional neural networks. A hybrid approach in combination with linear and nonlinear characteristics extracted from ECG and heart rate variability (HRV) has been proposed and described in [6] for multiclass ECG classification based on a deep neural network. The use of this method improves the efficiency of ECG diagnostics by combining optimized deep learning functions with efficient aggregation of ECG functions and HRV indicators based on dynamic chaos theory. Although the proposed approach has been shown to be a promising tool for ECG classification, it should be further developed by examining large ECG datasets with many more patients to identify different classes of heart disease, including heart rhythm classes and myocardial infarction classes.

Paper [7] presents a method for predicting heart diseases based on support vector machines (SVM) supplemented with fuzzy fusion at the decision level. Research [8] combines cloud computing with machine learning by combining statistical datasets and applying fusion methods to ensure the accuracy and consistency of predictions. Real-time patient indicators were retrieved from the cloud repository and processed using a fuzzy model. The model used an artificial neural network, decision tree, and Bayesian method. In [9], to determine algorithms for classifying coronary heart disease (CHD), polygenic risk scales, logistic regression, the naive Bayes model, random forests, support vector machines, and gradient boosting were compared. The resulting models were tested on an independent data set. As a result, it was found that polygenic risk scales turned out to be the most effective algorithm for classifying CHD.

An extended convolutional neural network with deep learning support was developed in [10] to predict cardiovascular disease and determine the level of health risk. The test results showed comparison with approaches such as the deep neural network, the recurrent neural network, and the neural network ensemble method. The system was implemented on the platform of the Internet of Medical Things and a medical decision support system. Paper [11] presents the HealthCloud system for monitoring the health status of patients using machine learning and cloud computing. This study was aimed to integrate information from various sources needed to describe heart disease in detail with an accurate prognosis. The presence of heart diseases was determined using support vector machines, k-nearest neighbors, neural networks, and logistic regression. In [12], machine learning methods were used to construct predictive models of tachyarrhythmia after acute myocardial infarction. However, the authors noted that the machine learning approach needs further validation and optimization before clinical application.

The great importance and intensive development of machine learning methods for the study of heart disease is evidenced by the increased number of publications in this area [13−22]. At the same time, the number of publications on this topic is growing, which is associated not only with the relevance and high social significance of the detection of cardiovascular diseases, but also with the development of machine learning methods themselves. Analysis of the literature showed that, in the presence of a well-developed mathematical apparatus and significant amounts of clinical data, there is a scientific problem of determining the optimal parameters of machine learning methods and their application to improve the quality of a

Table 1. Description of data attributes

No. of attribute	Name of attributes	Values of attributes
I	Age	Numerical value from 29 to 78, age
II	Gender	0, male; 1, female.
III	Heart pain	1, typical angina pectoris
		2, atypical angina pectoris
		3, nonanginal pain
		4, asymptomatic
IV	Blood pressure at rest	Numerical value
V	Serum cholesterol	Numerical value
VI	Fasting blood pressure	Numerical value
VII	ECG at rest	0, norm
		1, having ST-T
		2, hypertrophy
VIII	Maximum heart rate (HR)	Numerical value
IX	Exercise-induced angina	1, yes
		0, no
X	Exercise-induced ST segment depression compared to rest (ST is a part of the ECG curve that corresponds to the period of the cardiac cycle)	Numerical value
XI	Slope of the peak ST segment under load	1, rise
		2, flat
		3, rise
XII	Number of large vessels stained on fluoroscopy	Numerical value, 0−3
XIII	Thalassemia	3, normal
		6, fixed defect
		7, reversible defect
XIV	Presence of heart diseases	1, yes
		0, no

multicriteria analysis of the state of the human cardiovascular system, primary diagnosis, and timely prevention of heart disease.

We used raw data on 303 patients from the Hungarian Institute of Cardiology, the University Hospital of Zurich, the University Hospital of Basel, the Long Beach Medical Center, and the Cleveland Clinic for our research. The patient database has fourteen attributes. Information about the attributes of the data set is presented in Table 1. To demonstrate the data structure, Table 2 shows a part of the database used in this study.

We used coding for categorical variables such as age, blood pressure, and cholesterol level, since these are independent discrete values that were prenormalized. Some attributes were considered as fixed, such as age and cholesterol levels. Some attributes were considered as variables, such as pain in the heart.

The aim of this study was to use the first thirteen signs to predict the fourteenth, the presence of heart disease in a patient. For this study, various methods of machine learning were used with their subsequent comparative analysis and study of their effectiveness.

RESULTS AND DISCUSSION

The correlation relationship between each pair of attributes was analyzed and revealed. In Fig. 2, plots on the main diagonal are histograms of each attribute compared to the classification score (presence or

Table 2. Fragment of the data set for testing (by attribute numbers)

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
64	1	0	110	211	0	2	144	1	1.8	1	0	0	0
59	1	0	160	273	0	2	125	0	0	0	0	0	1
52	1	0	118	186	0	2	190	0	0	1	0	1	0
45	1	0	110	264	0	0	132	0	1.2	1	0	2	1
40	1	0	140	199	0	0	178	1	1.4	0	0	2	0
38	1	0	120	231	0	0	182	1	3.8	1	0	2	1
34	1	0	118	182	0	2	174	0	0	0	0	0	0
74	0	1	120	269	0	2	121	1	0.2	0	1	0	0
77	1	3	125	304	0	2	162	1	0	0	3	0	1

absence of heart disease). Plots that are not on the main diagonal show correlations between two different attributes according to identifiers:

• age;
• gender;
• chest pain type;
• resting blood pressure in mm/Hg;
• serum cholesterol in mg/dl;
• fasting blood pressure>120mg/dl;
• resting electrocardiographic results (ECG at rest);
• maximum heart rate achieved (maximum heart rate);
• exercise induced angina;
• ST depression induced by exercise relative to rest;
• the slope of the peak exercise ST segment (number of major vessels colored by flourosopy);
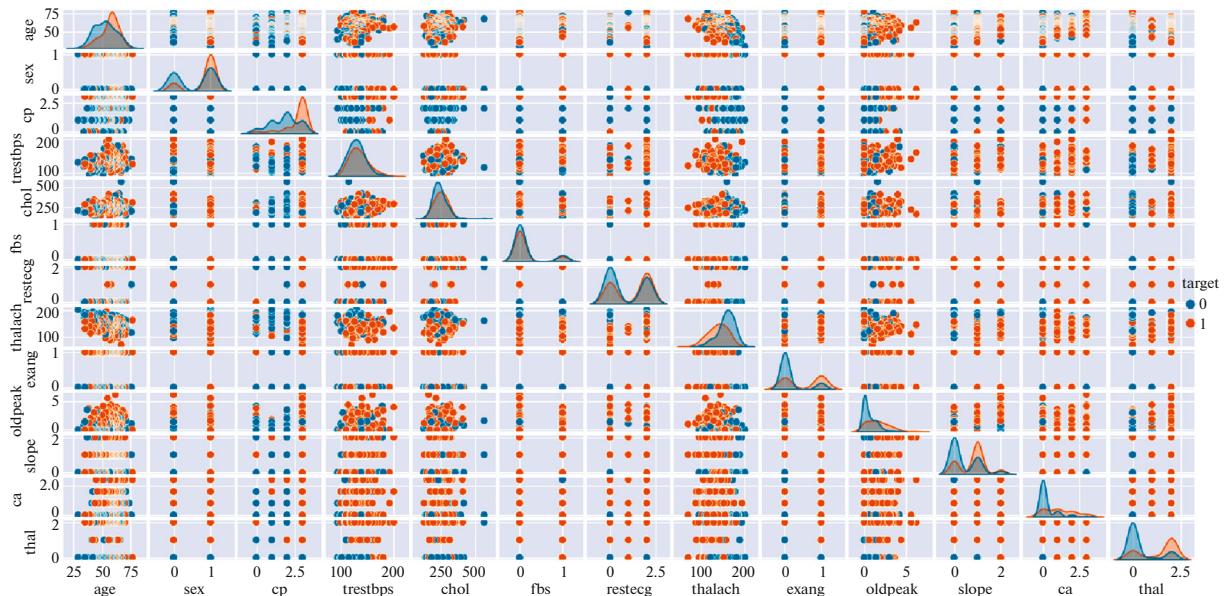• thalassemia, the presence or absence of heart disease.

Fig. 2. Relationship between data attributes (explanations are in the text).

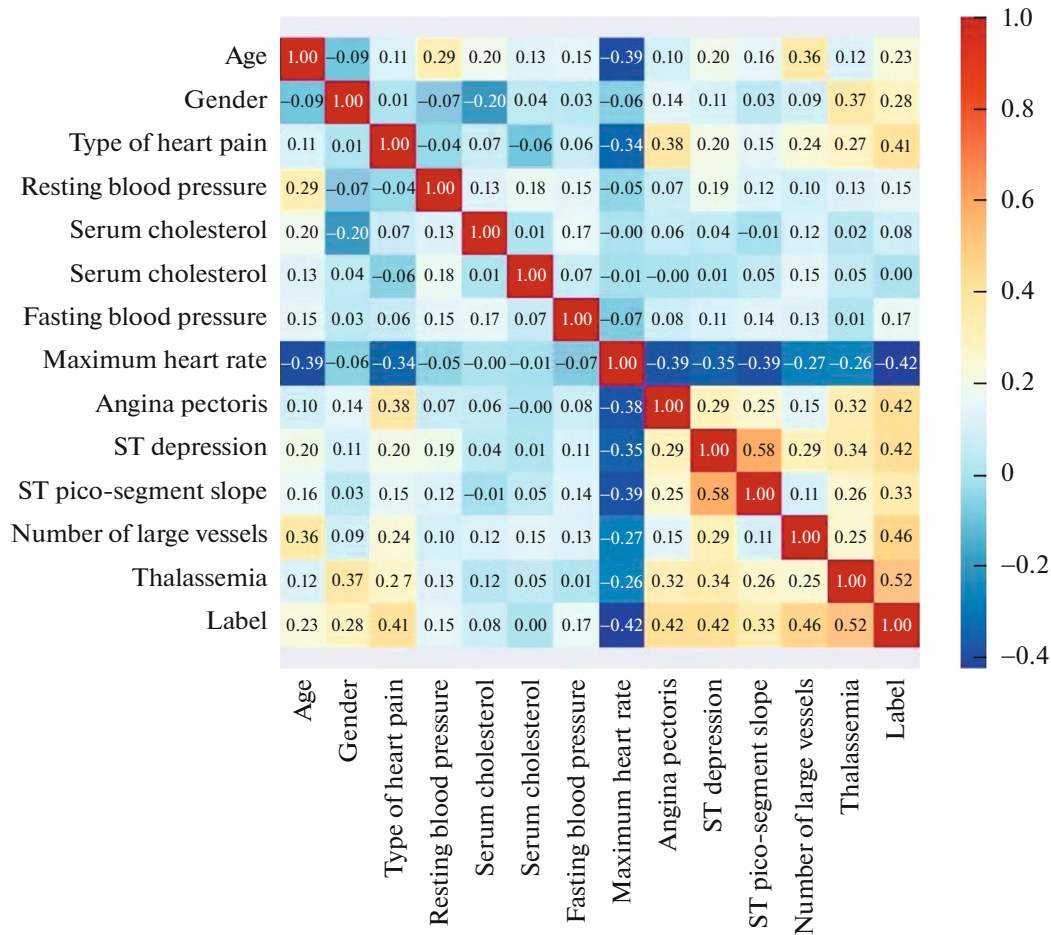|	Age	Gender	Type of heart pain	Resting blood pressure	Serum cholesterol	Serum cholesterol	Fasting blood pressure	Maximum heart rate	Angina pectoris	ST depression	ST pico-segment slope	Number of large vessels	Thalassemia	Label
Age	1.00	−0.09	0.11	0.29	0.20	0.13	0.15	−0.39	0.10	0.20	0.16	0.36	0.12	0.23
Gender	−0.09	1.00	0.01	−0.07	−0.20	0.04	0.03	−0.06	0.14	0.11	0.03	0.09	0.37	0.28
Type of heart pain	0.11	0.01	1.00	−0.04	0.07	−0.06	0.06	−0.34	0.38	0.20	0.15	0.24	0.27	0.41
Resting blood pressure	0.29	−0.07	−0.04	1.00	0.13	0.18	0.15	−0.05	0.07	0.19	0.12	0.10	0.13	0.15
Serum cholesterol	0.20	−0.20	0.07	0.13	1.00	0.01	0.17	−0.00	0.06	0.04	−0.01	0.12	0.02	0.08
Serum cholesterol	0.13	0.04	−0.06	0.18	0.01	1.00	0.07	−0.01	−0.00	0.01	0.05	0.15	0.05	0.00
Fasting blood pressure	0.15	0.03	0.06	0.15	0.17	0.07	1.00	−0.07	0.08	0.11	0.14	0.13	0.01	0.17
Maximum heart rate	−0.39	−0.06	−0.34	−0.05	−0.00	−0.01	−0.07	1.00	−0.39	−0.35	−0.39	−0.27	−0.26	−0.42
Angina pectoris	0.10	0.14	0.38	0.07	0.06	−0.00	0.08	−0.38	1.00	0.29	0.25	0.15	0.32	0.42
ST depression	0.20	0.11	0.20	0.19	0.04	0.01	0.11	−0.35	0.29	1.00	0.58	0.29	0.34	0.42
ST pico-segment slope	0.16	0.03	0.15	0.12	−0.01	0.05	0.14	−0.39	0.25	0.58	1.00	0.11	0.26	0.33
Number of large vessels	0.36	0.09	0.24	0.10	0.12	0.15	0.13	−0.27	0.15	0.29	0.11	1.00	0.25	0.46
Thalassemia	0.12	0.37	0.2 7	0.13	0.12	0.05	0.01	−0.26	0.32	0.34	0.26	0.25	1.00	0.52
Label	0.23	0.28	0.41	0.15	0.08	0.00	0.17	−0.42	0.42	0.42	0.33	0.46	0.52	1.00

Fig. 3. Representation of correlation between attributes.

It is seen from the correlation diagram in Fig. 2 that the presence or absence of heart disease has significant differences in the distribution of attributes such as heart pain (cp); exercise-induced angina (exang); exercise-induced ST depression compared to rest (oldpeak); the slope of the peak ST segment under load (slope); the number of large vessels stained by fluoroscopy (ca), and thalassemia (thal).

Thus, the most statistically significant attributes were identified, which contain information about the probability of having heart disease. To display correlations between attributes, a color-coded matrix was constructed, in which the color gradation corresponds to the degree of correlation (Fig. 3). It can be seen from this matrix that the slope of the peak ST segment during exercise positively correlates with the exercise-induced ST depression indicator. Compared to the resting state, the correlation coefficient is 0.58. This means that if the value of the ST segment slope increases, then the peak of depression will also increase, and vice versa.

The heart disease target was found to have the highest positive correlation with the indicator of thalassemia (at 0.52), followed by the number of large vessels colored in roentgenoscopy, indicators of exercise-induced angina, ST depression, and cardiac pain.

All data was analyzed using seven models: logistic regression, *k*-nearest neighbors, decision tree, support vector machine, naive Bayes classifier, random forest, and deep neural networks. To evaluate machine learning methods, multiple random division of preprocessed data into training and test subsets was used. Since the result may vary depending on the random initial values of this split, two series of experiments were carried out with 20 and 15% of the data as a test series, respectively. The results of two series of experiments for all models with 20 and 15% data redundancy as a test set are shown in Table 3.

To analyze the results obtained, plots were constructed to compare the accuracy of the models (Fig. 4). The main parameters of all machine learning methods studied are given in Table 4.

Table 3. Results of two series of experiments

Machine learning model	Accuracy of models with a test series of 20%	Accuracy of models with a test series of 15%	Model quality parameter values (AUC)
Logistic regression	0.852	0.911	0.95
k-Nearest neighbor method	0.836	0.867	0.94
Support vector machines	0.852	0.911	0.94
Naive bayes classifier	0.869	0.889	0.95
Decision trees	0.754	0.822	0.89
Random forest	0.836	0.933	0.94
Neural network method	0.902	0.911	0.95

The three-dimensional curves of the accuracy surface obtained as a result of a series of experiments are shown in Fig. 5. On the basis of the developed procedural interactive program, plots of the values of the parameters of the test series size as an independent variable X and the values of the argument of the pseudo-random generation function were constructed to evaluate each of the Y models. The accuracy of the model is plotted along the Z axis.

When constructing three-dimensional surfaces for each type of model, models with test series sizes from 5 to 55% were generated along the X axis. Further, for each example, 250 machine learning models were generated from the data set with values of the initialization argument of the pseudo-random number generator from 0 to 250. This parameter uniquely determines the composition of the test and training series and is responsible for reproducibility of the results. After setting values for independent variables,
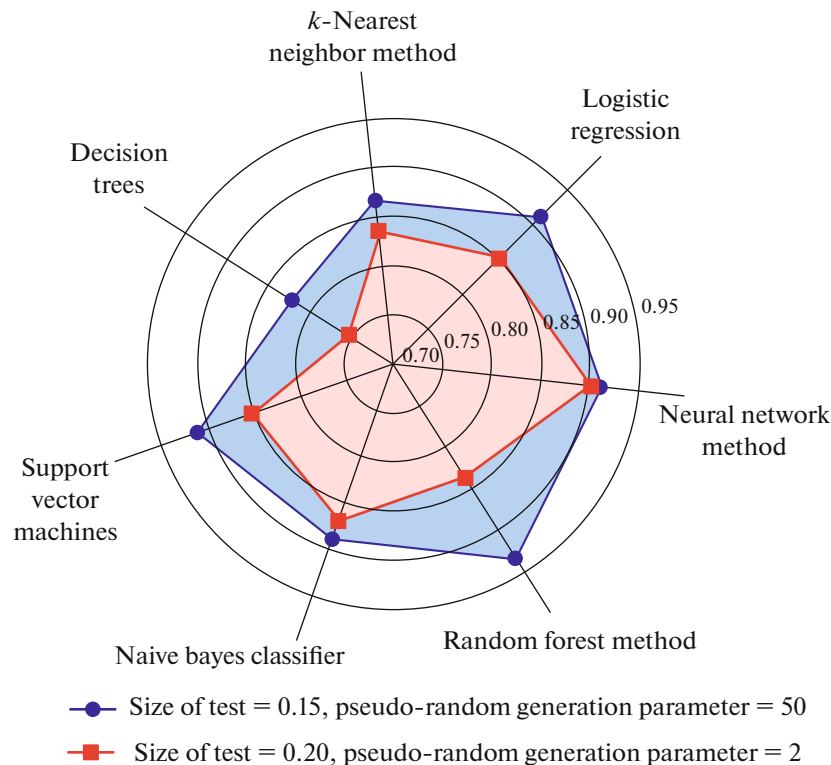
Fig. 4. Comparison of the efficiency of machine learning models with different distributions between test and training samples.
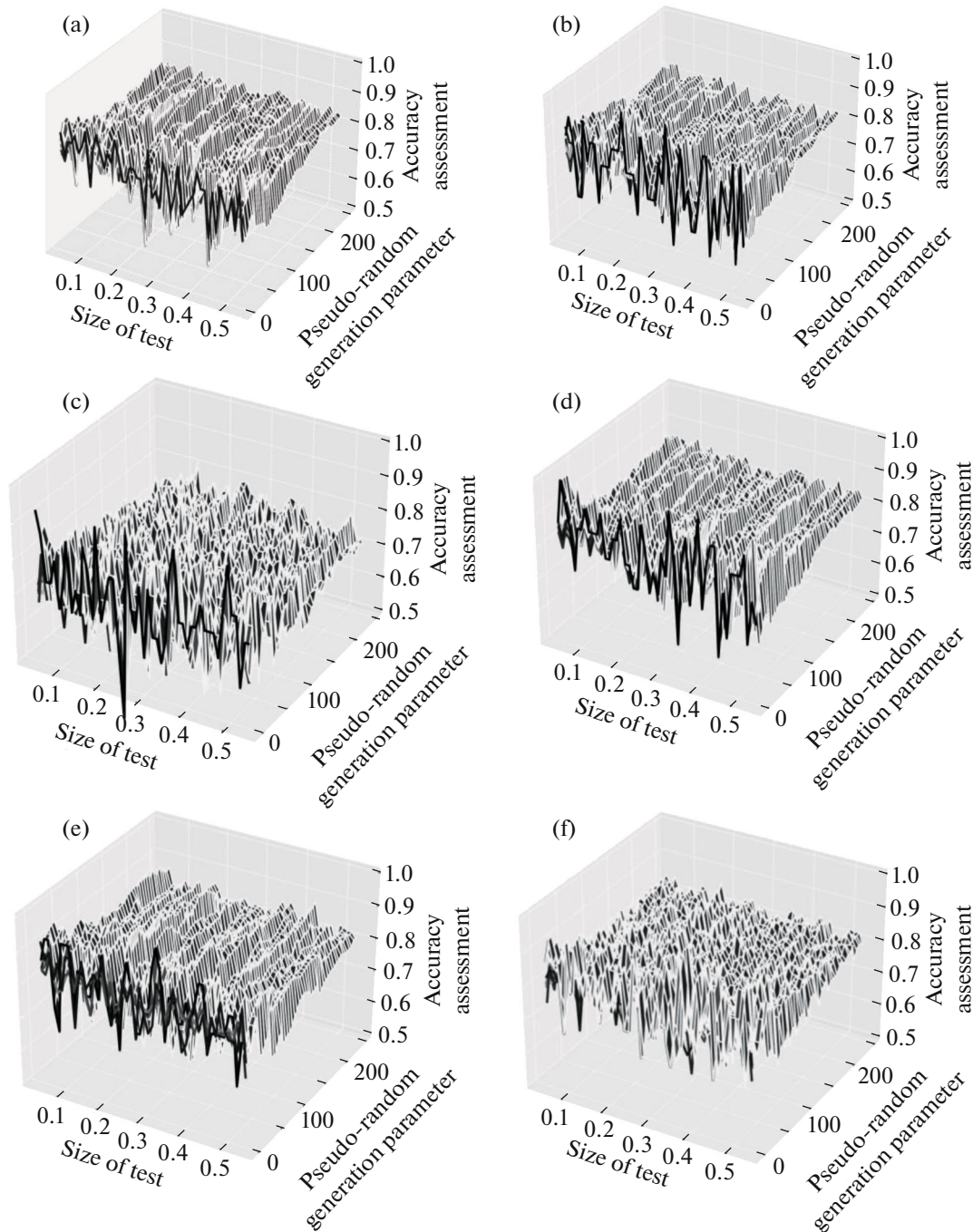
Fig. 5. Accuracy surface curves for machine learning models: (a) logistic regression; (b) *k*-nearest neighbor method; (c) decision tree; (d) support vector machine; (e) naive Bayes classifier; (f) random forest.

two corresponding arrays were iteratively entered into each of the machine learning models with fixed parameters from the given ranges of values.

As a result, 12 500 experiments were conducted for each type of machine learning model; in total 75 000 computational experiments were performed with various parameters. Each resulting graph in Fig. 5 is a precision curved surface. The axes (Fig. 5) show the relative size of the test series, the value of the initiating argument of the pseudorandom number generator, and accuracy of the model in arbitrary units. In the graphs, darker lines (peak values) correspond to a higher forecast accuracy.

Table 4. Parameters of supervised machine learning methods

Machine learning method	Main parameters of the method
Logistic regression	Regularization type: L2 (contributes to the appearance of small weight coefficients of the model)
	Optimization algorithm: Broyden–Fletcher–Goldfarb–Shanno algorithm with limited memory
	Maximum number of iterations for solver convergence: 100
k-Nearest neighbor method	Number of neighbors used by default for requests: 5
	Weighting function used in prediction: the weight of each neighbor is equivalent to
	Metric: Minkowski distance
Support vector machines	Kernel: radial basis function
	Degree of polynomial kernel function: 3
	Cache size: 200 mb
Naive bayes classifier	Prior probabilities of classes: not adjusted according to the data
	Variable smoothing factor to ensure calculation stability: 10^9 (part of the largest variance of all objects, which is added to the variances)
Decision trees	Separation quality measurement function: Ginny coefficient
	Minimum number of samples required to split an internal tree node: 2
	Minimum number of samples required to be at the end node: 1
Random forest method	Number of decision trees: 100
	Function for measuring separation quality: Gini impurity
	Minimum number of samples required to split an internal node: 2
	Minimum number of samples required to be at the end node: 1
Neural network method	Number of hidden layers: 100
	Activation function of each hidden layer: truncated linear transform
	Weight optimization algorithm: adaptive moment estimation method
	Maximum number of iterations: 200
	Percentage of training data to be set aside as a validation set for early stopping: 0.1
	Value for the numerical stability of the adaptive estimation method: 10^{-8}

By points on the plots, one can trace the main patterns and changing trends between them. It can be seen that the magnitude of the fluctuation in the accuracy of forecasting models is related to the value of the argument of the pseudorandom generation function. At the same time, the accuracy of forecasts decreases as the value of this parameter increases. The model achieves relatively high accuracy (local maximum), with the value of the argument of the pseudo-random generation function in the range of 5–30 units. It was found that the accuracy of model prediction can be optimized by choosing the size of the test set and parameter of pseudorandom generation of separation into test and training series. This result is useful for obtaining a stable separation method of dataset and debugging methods of machine learning.

In some ranges of values, the change in the accuracy parameter of the models is relatively stable, while in other areas, the change in the accuracy parameter is characterized by significant fluctuations. Such relatively high but unstable accuracy values are associated with model overfitting. Thus, the accuracy of each model is not constant.

This is due to the fact that, when normalizing data, dividing the series into test and training sets, as well as in the controlled construction of machine learning models, the prediction accuracy largely depends on the numerical parameters of algorithms (Table 4).

For comparison among the models constructed for predicting heart diseases, ROC analysis [23] with the construction of an ROC diagram was used (Fig. 6). To assess the quality of machine learning models,
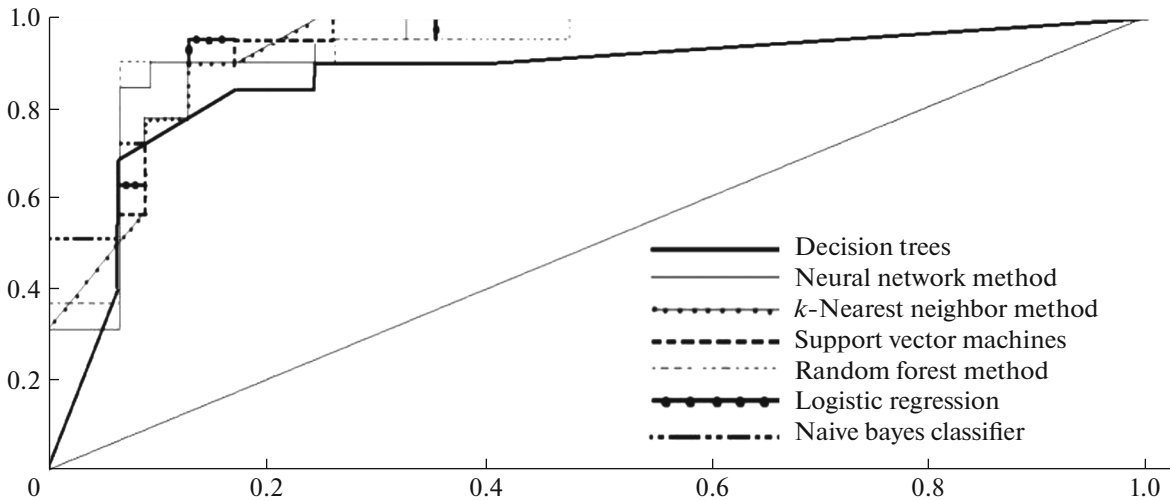
Fig. 6. ROC moduli curve for machine learning methods. The *X*-axis shows a false positive rate, FPR, along the *Y*-axis is the true positive rate, TPR.

the value of the area under the ROC curve AUC (area under the curve) was used. The AUC values calculated from the models that participated in the series of experiments are shown in Table 3.

CONCLUSIONS

Due to the intensive development of machine learning methods, cloud technologies, and neural networks and the high social significance of cardiovascular diseases, the relevance of this study is beyond doubt. At the same time, the healthcare system requires the introduction of highly accurate and reliable methods for supporting medical decision-making, methods for collecting and consolidating big data, as well as methods for evaluating and validating predictive methods. At the same time, machine learning methods, including artificial neural networks, play an increasingly important role in the diagnosis of cardiovascular diseases, especially at an early stage, contributing to timely prevention and increasing the duration and quality of human life.

In the present study, a correlation analysis of medical parameters was carried out, and machine learning models for predicting the state of the human cardiovascular system have been considered and analyzed. As a result of this research, estimates of the effectiveness of each method were obtained. The presented methods have confirmed their effectiveness in solving practical problems of public health. As a result of a series of computational experiments, it was found that, for predicting heart disease, it is important not only to choose one machine learning method or another, but also to select its parameters, including chaotic characteristics of testing algorithms.

When solving biomedical problems using machine learning methods, one should always take into account the fact that only a doctor can diagnose, and a decision support system based on machine learning algorithms can only play an advisory role. In this regard, the developments made in this study can be recommended for the practical implementation of the developed software and mathematical tools for prevention of heart attacks and preventive measures of heart diseases. It should be noted that the implementation of such systems in healthcare will have a positive impact on the improvement of measures to preserve and improve the health of the population.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. Yerdessov, S., Kadyrzhanuly, K., Sakko, Ye., Gusmanov, A., Zhkhina, G., Galiyeva, D., Bekbossynova, M., Salustri, A., and Gaipov, A., Epidemiology of arterial hypertension in Kazakhstan: Data from unified nation-

wide electronic healthcare system 2014−2019, *J. Cardiovasc. Dev. Disease*, 2022, vol. 9, no. 2, p. 52.
https://doi.org/10.3390/jcdd9020052

2. Gorny, B.E., Kalinina, A.M., and Drapkina, O.M., Prognostic significance of the integral index of the alcohol situation in assessing regional differences in mortality from cardiovascular diseases in the Russian Federation, *Ratsion. Farmakoterapiya Kardiol.*, 2022, vol. 18, no. 1, pp. 36−41.
https://doi.org/10.20996/1819-6446-2022-02-05

3. Wadhawan, S. and Maini, R., A systematic review on prediction techniques for cardiac disease, *Int. J. Inf. Technol. Syst. Approach*, 2021, vol. 15, no. 1, pp. 1−33.
https://doi.org/10.4018/ijitsa.290001

4. Śmigiel, S., Pałczyński, K., and Ledziński, D., Deep learning techniques in the classification of ECG signals using r-peak detection based on the PTB-XL dataset, *Sensors*, 2021, vol. 21, no. 24, p. 8174.
https://doi.org/10.3390/s21248174

5. Zhu, X., Wei, Ya., Lu, Yu, Zhao, M., Yang, Ke, Wu, Sh., Zhang, H., and Wong, K.K.L., Comparative analysis of active contour and convolutional neural network in rapid left-ventricle volume quantification using echocardiographic imaging, *Comput. Methods Programs Biomed.*, 2021, vol. 199, p. 105914.
https://doi.org/10.1016/j.cmpb.2020.105914

6. Eltrass, A.S., Tayel, M.B., and Ammar, A.I., Automated ECG multi-class classification system based on combining deep learning features with HRV and ECG measures, *Neural Comput. Appl.*, 2022, vol. 34, no. 11, pp. 8755−8775.
https://doi.org/10.1007/s00521-022-06889-z

7. Nadeem, M.W., et al., Fusion-based machine learning architecture for heart disease prediction, *CMC-Comput. Mater. Continua*, 2021, vol. 67, no. 2, pp. 2481−2496.
https://doi.org/10.32604/cmc.2021.014649

8. Ahmad, M., et al., Data and machine learning fusion architecture for cardiovascular disease prediction, *Comput., Mater. Continua*, 2021, vol. 69, no. 2, pp. 2717−2731.

9. Gola, D., Erdmann, J., Müller-Myhsok, B., Schunkert, H., and König, I.R., Polygenic risk scores outperform machine learning methods in predicting coronary artery disease status, *Genetic Epidemiol.*, 2020, vol. 44, no. 2, pp. 125−138.
https://doi.org/10.1002/gepi.22279

10. Pan, Yu., Fu, M., Cheng, B., Tao, X., and Guo, J., Enhanced deep learning assisted convolutional neural network for heart disease prediction on the internet of medical things platform, *IEEE Access*, 2020, vol. 8, pp. 189503−189512.
https://doi.org/10.1109/ACCESS.2020.3026214

11. Desai, F., Chowdhury, D., Kaur, R., Peeters, M., Arya, R.Ch., Wander, G.S., Gill, S.S., and Buyya, R., HealthCloud: A system for monitoring health status of heart patients using machine learning and cloud computing, *Internet Things*, 2022, vol. 17, p. 100485.
https://doi.org/10.1016/j.iot.2021.100485

12. Qiu, P.L., Liu, Shu-Yan, Bradshaw, M., Rooney-Latham, S., Takamatsu, S., Bulgakov, T.S., Tang, Shu-R., Feng, J., Jin, D.-Ni, Aroge, T., Li, Yu, Wang, Li-L., and Braun, U., Multi-locus phylogeny and taxonomy of an unresolved, heterogeneous species complex within the genus *Golovinomyces* (Ascomycota, Erysiphales), including *G. Ambrosiae, G. Circumfusus* and *G. Spadiceus, BMC Microbiol.*, 2020, vol. 20, no. 1, p. 51.
https://doi.org/10.1186/s12866-020-01731-9

13. Sharma, A., Pal, T., and Jaiswal, V., Heart disease prediction using convolutional neural network, *Cardiovascular and Coronary Artery Imaging*, El-Baz, A.S. and Suri, J.S., Eds., Academic Press, 2022, vol. 1, pp. 245−272.
https://doi.org/10.1016/B978-0-12-822706-0.00012-3

14. Wang, J., Rao, C., Goh, M., and Xiao, X., Risk assessment of coronary heart disease based on cloud-random forest, *Artif. Intell. Rev.*, 2022.
https://doi.org/10.1007/s10462-022-10170-z

15. Riyaz, L., Butt, M.A., Zaman, M., and Ayob, O., Heart disease prediction using machine learning techniques: a quantitative review, *International Conference on Innovative Computing and Communications*, Khanna, A., Gupta, D., Bhattacharyya, S., Hassanien, A.E., Anand, S., and Jaiswal, A., Eds., Advances in Intelligent Systems and Computing, vol. 1394, Singapore: Springer, 2022, pp. 81−94.
https://doi.org/10.1007/978-981-16-3071-2_8

16. Qiao, S., Pang, Sh., Luo, G., Pan, S., Yu, Z., Chen, T., and Lv, Zh., RLDS: An explainable residual learning diagnosis system for fetal congenital heart disease, *Future Gener. Comput. Syst.*, 2022 vol. 128, pp. 205−218.
https://doi.org/10.1016/j.future.2021.10.001

17. Bihri, H., Nejjari, R., Azzouzi, S., and El Hassan Charaf, M., An artificial neural network-based system to predict cardiovascular disease, *Advances in Information, Communication and Cybersecurity. ICI2C 2021*, Maleh, Y., Alazab, M., Gherabi, N., Tawalbeh, L., and Abd El-Latif, A.A., Eds., Lecture Notes in Networks and Systems, vol. 357, Cham: Springer, 2021, pp. 393−402.
https://doi.org/10.1007/978-3-030-91738-8_36

18. Mohapatra, D., Bhoi, S.K., Mallick, Ch., Jena, K.K., and Mishra, S., Distribution preserving train-test split directed ensemble classifier for heart disease prediction, *Int. J. Inf. Technol.*, 2022, vol. 14, pp. 1763−1769.
https://doi.org/10.1007/s41870-022-00868-2

19. Chitra, S. and Jayalakshmi, V., Prediction of heart disease and chronic kidney disease based on internet of things using RNN algorithm, *Proceedings of Data Analytics and Management*, Gupta, D., Polkowski, Z., Khanna, A., Bhattacharyya, S., and Castillo, O., Eds., Lecture Notes on Data Engineering and Communications, vol. 90, Singapore: Springer, 2022, pp. 467−479.
https://doi.org/10.1007/978-981-16-6289-8_40

20. Fathima, K. and Vimina, E.R., Heart disease prediction using deep neural networks: A novel approach, *Intelligent Sustainable Systems*, Raj, J.S., Palanisamy, R., Perikos, I., and Shi, Y., Eds., Lecture Notes in Network and Systems, vol. 213, Singapore: Springer, 2022, pp. 725−736.
https://doi.org/10.1007/978-981-16-2422-3_56

21. Rani, S. and Dutta, M.K., Heart anomaly classification using convolutional neural network, *Proceedings of International Conference on Data Science and Applications*, Saraswat, M., Roy, S., Chowdhury, C., and Gandomi, A.H., Eds., Lecture Notes in Networks and Systems, vol. 288, Singapore: Springer, 2022, pp. 541−550.
https://doi.org/10.1007/978-981-16-5120-5_41

22. Suresh, T., Assegie, T.A., Rajkumar, S., and Kumar, N.K., A hybrid approach to medical decision-making: diagnosis of heart disease with machine-learning model, *Int. J. Electr. Comput. Eng.*, 2022, vol. 12 no. 2, pp. 1831−1838.

23. Hernández-Orallo, J., ROC curves for regression, *Pattern Recognit.*, 2013, vol. 46, no. 12, pp. 3395−3411.
https://doi.org/10.1016/j.patcog.2013.06.014

Translated by Sh. Galyaltdinov