IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 16 September 2023, accepted 11 November 2023, date of publication 20 November 2023, date of current version 1 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3335176

RESEARCH ARTICLE

Multimodal Neural Network for Recognition of Cardiac Arrhythmias Based on 12-Load Electrocardiogram Signals

MARIYA R. KILADZE[1], ULYANA A. LYAKHOVA[1,2], PAVEL A. LYAKHOV[1,2], NIKOLAY N. NAGORNOV[1], AND MOHSEN VAHABI[3]

[1]Department of Mathematical Modeling, North-Caucasus Federal University, 355017 Stavropol, Russia
[2]North-Caucasus Center for Mathematical Research, North-Caucasus Federal University, 355017 Stavropol, Russia
[3]Faculty of Electrical Engineering, Shahrood University of Technology, Shahrud, Semnan 3619995161, Iran

Corresponding author: Mariya R. Kiladze (merchali@mail.ru)

This work was supported in part by the Russian Science Foundation under Project 22-71-00009, in part by the Russian Science Foundation under Project 23-71-10013, and in part by the Council for Grants of President of Russian Federation under Project MK-371.2022.4.

ABSTRACT Automatic classification of heart rhythm disturbances using an electrocardiogram is a reliable way to timely detect diseases of the cardiovascular system. The need to automate this process is to increase the number of electrocardiogram signals. Classification methods based on the use of neural networks provide a high percentage of arrhythmia recognition. However, known classification methods do not take into account patient characteristics. The work proposes a multimodal neural network that takes into account the age and gender characteristics of the patient. It includes a Long short-term memory (LSTM) network for feature extraction on twelve-channel electrocardiogram signals and a linear neural network for processing patient metadata such as age and gender. Extraction of electrocardiogram signal features occurs in parallel with metadata processing. The last unifying layer of the proposed multimodal neural network integrates heterogeneous data and features of electrocardiogram signals obtained using an LSTM network. The developed multimodal neural network was verified using the PhysioNet/Computing in Cardiology Challenge 2021 ECG database. The simulation results showed that the proposed multimodal neural network achieves a recognition accuracy of 63%, which is 2 percentage points higher compared to state-of-the-art methods.

INDEX TERMS Neural network classification, metadata, linear perceptron, LSTM network, PhysioNet/Computing in Cardiology Challenge 2021.

I. INTRODUCTION

Cardiovascular disease is the most common cause of death in people worldwide. Ischemic heart disease [1], rheumatic heart disease [2], cerebrovascular disease [3] are the main ones. According to the World Health Organization (WHO), every year between 2018 and 2020, an average of 17.9 million people died due to problems related to the cardiovascular system. One third of deaths is premature (among people under 70) [4]. Detection of diseases of the cardiovascular system is carried out mainly in the diagnosis of the results of the electrocardiogram (ECG) [5]. Highly accurate classification of ECG signals can prevent diseases, thereby significantly reducing the mortality rate [6].

An ECG is a recording of registration of electric fields that occur during the work of the heart [7]. The number of ECG data is steadily increasing worldwide. This is due both to an increase in the number of patients and the need for a more detailed ECG. This makes the diagnostic process much more difficult. There is a need to automate the signal processing process. Different types and wavelengths complicate this process. It is often necessary to immediately classify the heart rhythm during the ECG process. The process of classifying cardiac arrhythmias also requires automation in

The associate editor coordinating the review of this manuscript and approving it for publication was Hongli Dong.

this regard [8]. Developers are faced with the problem of increasing the number of functions that affect the classification result when creating automation systems. This problem is solved by size reduction or feature selection [9], as well as the use of deep learning models that learn ECG patterns during training [10]. The advantage of using automated ECG classification systems is the accurate and rapid classification of ECG signals, however, this will lead to an increase in their number. Such systems can be implemented in ambulances, emergency rooms, telemedicine systems, and so on.

Modern classifiers based on signal preprocessing and neural networks show diagnostically significant results. So in [11] a method is shown using signal preprocessing based on the SMOTE algorithm for class synthesis with a smaller amount of signal and the Edited Nearest method for removing incorrectly classified samples for subsequent classification by a TD-CNN convolutional neural network. In [12], a method is shown using signal preprocessing based on the SMOTE algorithm and the Edited Nearest method to remove misclassified samples with subsequent classification by an LSTM neural network. In [13], signal preprocessing consists of segmentation of heart beats using the Pan-Tomrkins algorithm, feature extraction is carried out by wavelet transform, and classification of ECG signals by a recurrent neural network (RNN). In [13] and [14], segmentation is used for signal preprocessing. In [14], signal classification is carried out using the Spike-timing-dependent plasticity (STDP) method and the R-STDP reward modulation method. In [15] and [16], feature extraction is performed using a convolutional neural network (CNN), signal classification using k-nearest neighbors, support vector machines, and a multilayer perceptron. In [17], signal preprocessing is the extraction of Q, R, T, and U waves on the ECG signal. Backpropagation Neural Network (BPNN) is used to extract and classify features on the ECG stream. In [18], signal preprocessing is performed by segmentation, feature extraction is performed by spectral analysis, and signal classification is performed by a long short-term memory (LSTM) network. In [19], preprocessing is performed by a short-time Fourier transform, classification and feature extraction using a 2D-CNN neural network. The classifier from [20] can be singled out separately. A striking difference of this method is the formation of ECG signal histograms as signal preprocessing, followed by feature extraction by a convolutional neural network (CNN) and classification with an LSTM network. In [21] and [22] methods for obtaining ECG signals by a neural network CNN with signal preprocessing are presented. There are also classification methods without signal preprocessing based on convolutional neural network (CNN) [23], [24], recurrent neural network (RNN) [25], [26], [27] networks with long short-term memory [28], [29]. Deep learning methods are used to handle all types of signal comparisons [30], [31], [32], [33], [34]. All the considered methods do not take into account the individual metadata of the patient, which is an important source of additional information.

The use of metadata can significantly improve the quality of processing and classification accuracy of ECG signals in modern automated medical diagnostic systems. Combining signals and multivariate statistical data on patients makes it possible to create heterogeneous databases of medical information that can be used to build intelligent diagnostic systems and decision support for specialists, doctors, and clinicians [35]. Combining heterogeneous data provides additional information and increases the efficiency of systems for analyzing and classifying neural networks [36]. This approach, based on the use of multimodal neural networks specialized in the processing of heterogeneous biomedical data, has shown its effectiveness in solving the problem of classifying skin cancer from a photograph [37] and diagnosing liver failure in surgery [38].

Gender [39] and race [40], age [41] and other patient history can be taken into account by the cardiologist when making a diagnosis based on ECG signals [42]. The pre-collected case history data is the metadata for the proposed neural network system. However, public ECG databases only have information about the patient's age and gender. The more individual patient data is included in the metadata, the more accurate the classification result will be [43]. Information about race [44], bad habits [41], obesity, mental state [41], and other patient history can be used as metadata.

The aim of this work is to classify cardiac arrhythmias using a multimodal system of neural networks. We propose to classify cardiac arrhythmias using a multimodal neural network consisting of an LSTM with long-term short-term memory and a linear perceptron. The LSTM network processes the ECG signals and the linear perceptron processes individual patient data such as gender and age. The novelty of the study lies in the simultaneous processing of ECG signals and metadata (individual data) of the patient. This approach allows taking into account personal data, which improves classification results. The "One-Hot Code" algorithm was used to encode the metadata. The simulation was carried out in the PhysioNet/Computing in Cardiology Challenge 2021 database.

The rest of the paper is organized as follows: section II presents the methodology of the proposed approach, section III presents the simulation results, IV analyzes the simulation results and compares them with known methods. Section V draws conclusions from the work and outlines directions for further research.

II. MULTIMODAL NEURAL NETWORK WITH METADATA PROCESSING

We propose a multimodal neural network (MM-NN) for recognition of cardiac arrhythmias based on ECG signals and patient metadata, including age and gender. The MM-NN scheme is shown in Figure 1.

The multimodal neural network consists of two neural network architectures. ECG signals are not pre-processed. The decision on the need for pre-processing of signals is one of the

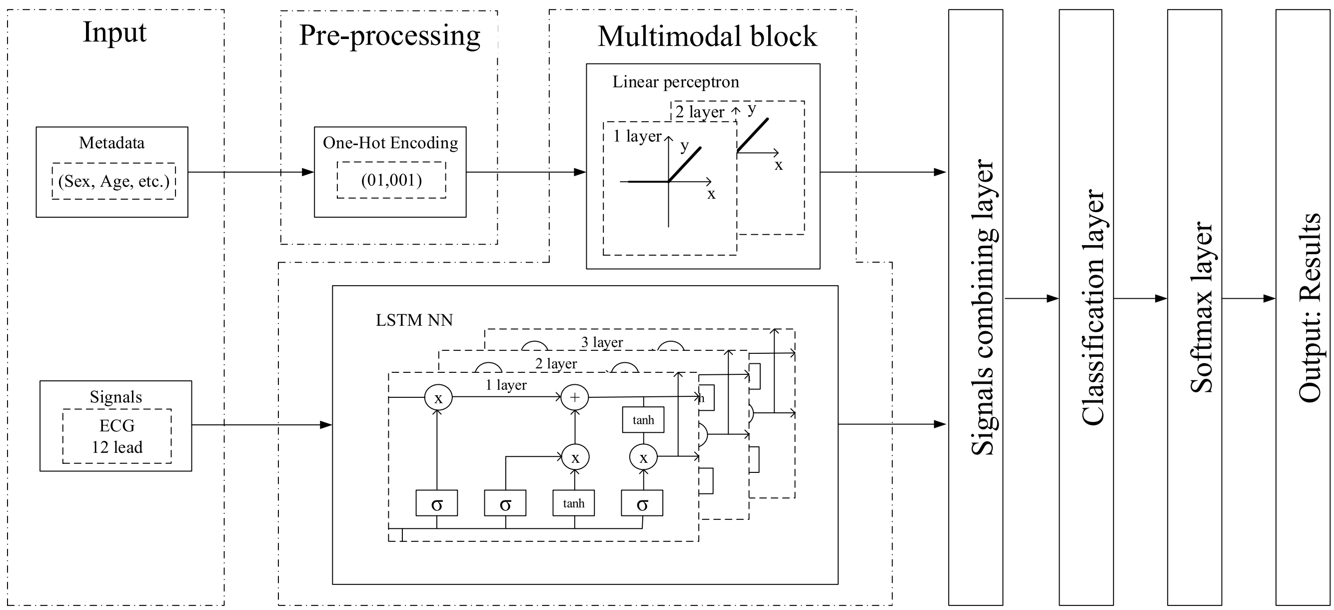

FIGURE 1. **Architecture of the proposed multimodal neural network for classifying ECG signals.**

further directions of research. Patient statistics are subjected to an one-hot encoding process to create a feature vector. ECG signals are processed using a three-layer recurrent LSTM network. Statistical metadata, consisting of data on the gender and age of the patient, is processed using a linear two-layer neural network. The resulting feature vector at the output of the LSTM and the output of the linear perceptron are combined at the signal combining layer. The combined signal is fed to the layer for classification. ECG signals are classified into five types of arrhythmias: sinus rhythm, sinus bradycardia, sinus tachycardia, atrial fibrillation, atrial flutter.

In deep learning, multimodal fusion or heterogeneous synthesis combines different types of data obtained from different sources [46]. ECG signals and patient statistics such as age and gender are the most common types of data in the field of ECG signal diagnostics.

Modeling of the proposed method was carried out on the PhysioNet / Computing in Cardiology Challenge 2021 database described in paragraph "A. DATABASE". Item "B. METADATA PREPROCESSING" describes the method data preprocessing method. The metadata classification method is described in paragraph "C. LINEAR PERCEPTRON". The method for classifying ECG signals is described in "D. LSTM". Paragraph "E. SIGNAL COMBINATION LAYER" describes a method for combining ECG signals and metadata. Paragraph "F. Loss Function" describes the loss function used to evaluate simulation results of the proposed MM-NN.

A. DATABASE

The normal ECG signal usually makes up the majority of ECG databases, which affects the classification result [47]. This type of signal is not available in the PhysioNet/Computing in Cardiology Challenge 2021 database [48], so it was chosen for modeling. This 12-lead ECG signal database is from Chempeng University, Shaoxing People's Hospital, and Ningbo First Hospital. The database contains ECG signals taken from 45,152 patients with a sampling rate of 500 Hz [48]. This database was chosen by us because it includes personal information about patients and does not have any particular problems or limitations that could affect the results of the classification. Signals from 3000 patients with cardiac arrhythmias: Sinus Bradycardia, Atrial Fibrillation, Sinus Rhythm, Sinus Tachycardia, Atrial Flutter. It was decided to divide the parameter "Age" into four groups in accordance with the classification adopted by WHO, to reduce the input categories to a linear perceptron. The first group of "young age" includes patients under the age of 44 years. The second group "middle age" includes patients aged 45 to 59 years. The third group "elderly" includes patients aged 60 to 74 years. The fourth group "long-livers" includes patients aged 75 years and older. When modeling the proposed MM-NS, the database signals were divided into training (60%), test (20%) and test (20%) samples.

Researchers or practitioners should use an ECG database with a large number of patients when implementing the proposed system. This is because each patient's ECG is unique. Using ECG signals from the same person can distort the simulation results.

B. PRE-PROCESSING METADATA

The increase in the volume of digital information in medicine occurs due to the accumulation of data from the results of laboratory and instrumental studies, data from monitoring devices, digitized past medical data, etc. [49]. Biomedical statistics are structured patient data, including gender, age, allergies, bad habits, etc. This information is necessary for the correct interpretation of the results of the analysis and medical examination [50].

The input data is presented as a feature vector for the correct operation offered by MM-NN. One-hot encoding is used to convert metadata into a set of binary variables [51]. Let $M = \{M_1, M_2, \ldots M_n\}$ be the patient's statistical data and $m = \{m_1, m_2, \ldots m_n\}$ the vector of statistical features, where $m_n \in M_n$ is a pointer to a specific parameter. For each set M_n, which is one of the patient's indicators, its cardinality is calculated $\mu_n = |M_n|$. Cardinality indicates the number of elements in m_n. After that, the binary code 100 …0, consisting of μ_n elements, is reserved for the first element of the set M_n. For the second element M_n of the set, the binary code 010 …0 is reserved, consisting of μ_n elements, etc. The expression $\sum \mu_n = \dim |M_n|$ calculates the dimension of the output encoded vector M. For example, for a database consisting of four ECG signals and such statistical data as gender and age, the vector M consists of data on gender and age $M = \{Sex, Age\}$, where the Sex parameter is represented by two genders, and the Age parameter, for example, by four age values 4, 15, 20, 30. Then the vector m_1 and m_2 for M_1 will take the form $m_1 = male$, $m_2 = female$. The vector M_2 will take the form $M_2 = (4, 15, 20, 30)$. So for M_1 the cardinality is $\mu_1 = |M_1| = 2$, and for M_2 is $\mu_2 = |M_2| = 4$ then the binary code for M_1 encoding will take the form 10, for M_2 encoding it will take the form 1000. Thus, the vector of encoded statistics for a 15-year-old male patient would be: $M = \{10, 0100\}$.

Age and gender are used as patient metadata. The cardinality of the statistical factor "gender" is 2, "age" is 86.

C. LINEAR PERCEPTRON

The transformed metadata is sent to the Linear perceptron block. The block consists of a multilayer perceptron with two hidden layers. The block "Linear perceptron" (Fig. 1), which is a multilayer perceptron with two hidden layers, receives a vector of encoded metadata consisting of 6 categories. The signal has 6 categories of values. Expression (1) describes the neuron y coming out of the k-th hidden layer [52]

$$y_k = f\left(\sum_{i=0}^{k} w_k^{(2)} f\left(\sum_{j=0}^{k} w_{ki}^{(1)} x_j\right)\right), \tag{1}$$

where: $w_k^{(2)}$, $w_{ki}^{(1)}$ are the weight coefficients; x_j are the input values. Expression (2) describes the signal coming from the last layer u_k:

$$u_k = \sum_{j=0}^{N} w_{kj} x_j, \tag{2}$$

where: w_{kj} are the weight coefficients, x_j are the input values, N is the last layer number. The error backpropagation algorithm [53] was used to train this network. A modified cross-entropy loss function was used to evaluate the simulation. The loss function used in the simulation is described in the paragraph "F. Loss function". The modified cross-entropy loss function was chosen due to the imbalance of the database.

D. LSTM

In each layer of the LSTM recurrent network, consisting of 64 LSTM blocks (Figure 2), the ECG signal goes through four stages: determining the "needed" information (sigmoid layer), updating information (sigmoid layer), creating and saving a new candidate vector (hyperbolic layer), definition and output of the necessary information (hyperbolic and sigmoidal layers) [54]. ECG signal x_t, hidden meaning h_{t-1} and the state of the cell at the previous time C_{t-1}, enter the LSTM block. Then these signals enter the input filter layer i_t (3) and the "forgetting" layer f_t (4), then a candidate vector for the output is determined $\tilde{C}_t$ (5) and the output vector is calculated o_t (6).

$$i_t = \sigma\left(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}\right) \tag{3}$$

$$f_t = \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}) \tag{4}$$

$$\tilde{C}_t = \tanh(W_{i\tilde{C}}x_t + b_{i\tilde{C}} + W_{h\tilde{C}}h_{t-1} + b_{h\tilde{C}}) \tag{5}$$

$$o_t = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}), \tag{6}$$

where σ and the sigmoidal and hyperbolic tangent layers, respectively, W_{ii}, W_{hi}, W_{if}, W_{hf}, $W_{i\tilde{C}}$, $W_{h\tilde{C}}$, W_{io}, W_{ho} are the weight matrices, b_{ii}, b_{hi}, b_{if}, b_{hf}, $b_{i\tilde{C}}$, $b_{h\tilde{C}}$, b_{io}, b_{ho}- free vectors. Then C_t are the cell state vector, h_t are the hidden value vector:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{7}$$

$$h_t = o_t \tanh(C_t), \tag{8}$$

where $\times$ is the term-by-term multiplication. For the last block, the hidden value vector h_t becomes the output vector from the LSTM layer.

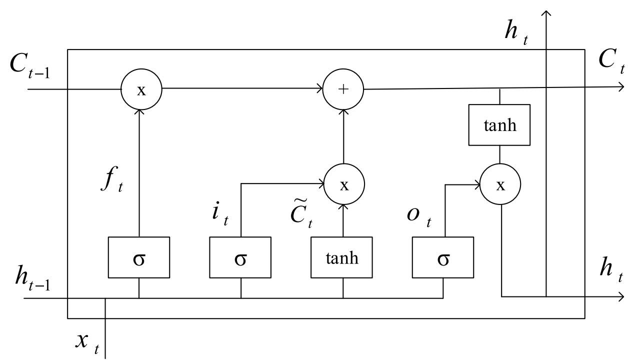

FIGURE 2. Scheme of operation of one LSTM block.

A modified cross-entropy loss function was used to evaluate the simulation. The loss function used in the simulation is described in the paragraph "F. Loss function". The modified cross-entropy loss function was chosen due to the imbalance of the database. The simulation results are presented in the next section.

E. SIGNALS COMBINING LAYER

The signal-combining layer receives as input the vector h_t, which was obtained on the last layer of the LSTM network, and the vector u_k, obtained on the last layer of the multilayer perceptron. The operation of combining heterogeneous data on the concatenation layer can be represented as follows

$$F = \sum_i \sum_j \sum_l h_t w_{ijl} + \sum_{i=1}^{l} u_k w_{il}^{(1)}, \tag{9}$$

where w_{ijl}– set of scales for processing ECG signals, $w_{il}^{(1)}$ is a set of weights for processing metadata vectors. A modified cross-entropy loss function was used to evaluate the simulation. The loss function used in the simulation is described in the next paragraph.''

The simulation results are presented in the next section.

F. LOSS FUNCTION

The last layers of the multimodal neural network are activated using the softmax function. The cross entropy function compares the probability distribution between the classified categories and the original true distribution. Imbalance in the data affects the normal loss function, using unequal misclassification costs between categories solves this problem:

$$\delta_k = \frac{H}{K \sum_{x=1}^{H} q_{yk}}, \tag{10}$$

where H - number of ECG signals, K – the number of categories over which the data is distributed q_{yk} - signal membership indicator y to category k.

Modified cross entropy loss function E'_δ for the combined data will take the form:

$$E'_\delta = -\frac{1}{H} \sum_{k=1}^{K} \sum_{y=1}^{H} \delta_k \times l_y^k \times \log(m_\mu(f_i, k)), \tag{11}$$

where l_y^k – true label for example y from category k, m_μ – neural network model with weights μ; δ_k – category weighting k, f_i - the result of combined dissimilar signals.

Function specifies the distance between the output distribution and the original probability distribution. There is a gradual memorization of true vectors and minimization of losses during training. Modifying the cross-entropy loss function with weighting factors minimizes the impact of imbalanced data.

The simulation results are presented in the next section.

III. SIMULATION RESULTS

The simulation was run using Python 3.11.0 on a PC with a 3.00 GHz Intel™ Core™ i5-8500 processor, 16 GB of RAM, and a 64-bit Windows 10 operating system. MM-NN training was carried out using a GPU based on the NVIDIA GeForce GTX 1050TI video chipset. The Pytorch machine learning framework was used to model neural network systems. The NumPy, Pandas and ScikitLearn libraries were used to process statistical data. The Matplotlib library was used to visualize the data.

Part of the PhysioNet/Computing in Cardiology Challenge 2021 database consisting of 3000 signals and metadata ''Gender'' and ''Age'' participated in the simulation of the proposed method MM-NN. The ''age'' parameter at the stage of preliminary processing of statistical data of patients is divided into four groups in accordance with the classification adopted by WHO. The ''young'', ''middle age'', ''elderly'', and ''centenarians'' groups include patients under 44 years of age, 45 to 59 years of age, 60 to 74 years of age, and 75 years of age and older, respectively. Thus, the variability of the ''Age'' parameter was reduced from 86 to 4 possible values. Figure 3 shows the graphs of the distribution of selected ECG signals according to the statistical factors of patients.

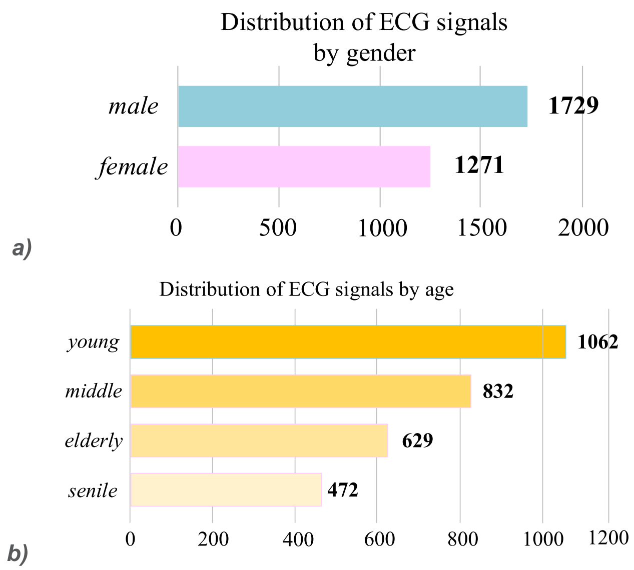

FIGURE 3. Graph of the distribution of ECG signals from the PhysioNet / Computing in Cardiology Challenge 2021 database, participating in the simulation, according to the statistical factors of patients: a) by gender, b) by age.

Further processing of the statistical data was to create an input vector using the one-hot encoding method. Tables 1 and 2 present the cardinality of each pre-processed statistical factor by the one-hot encoding method.

TABLE 1. Encoding the statistical parameter ''Gender'' of patients by the one-hot encoding method.

The gender of the patient	ONE-HOT CODE	
male	0	1
female	1	0

Each multimodal neural network was trained for 15 epochs. A pronounced overfitting was observed when using more epochs in each of the proposed blocks. The input batch size was 32. SGD was used as an optimizer with a standard learning rate of 0.001 and a moment of 0.9.

As a result of the simulation (Table 3), it was found that the combined use of heterogeneous data and ECG signals

TABLE 2. Coding the statistical parameter "Age" of patients by the one-hot encoding method.

The age of the patient	ONE-HOT CODE			
young	0	0	0	1
middle	0	0	1	0
elderly	0	1	0	0
senile	1	0	0	0

increases the accuracy of the neural network recognition of cardiac arrhythmias and reduces the value of the loss function. The loss function is used to calculate the error between the actual and received responses. The highest accuracy of recognition of ECG data and signals was 63.00% and was obtained when testing heterogeneous cardiac data with the MM-NN system. The smallest loss function index was 1.0533 and was also obtained when testing MM-NN. A decrease in the loss function index and a simultaneous increase in the accuracy of recognition of ECG signals prove the effectiveness of the proposed method.

TABLE 3. Simulation results of various neural networks.

NN architecture	Accuracy,%	Loss	Time, s
Linear perceptron	47.67	1.2961	**84**
LSTM	55.55	1.2160	191
Proposed MM-NN	**63.00**	**1.0533**	189

IV. EVALUATION OF SIMULATION RESULTS

The classification results were evaluated in two stages. The first stage is the analysis of the MM-NN statistical estimates, the second stage is the comparison of the MM-NN classification accuracy and statistical estimates with the corresponding state-of-the-art values.

A. STATISTICAL SCORE

Specificity, sensitivity, F-1 score, Matthews correlation coefficient (MCC), false negative rate (FNR), false positive rate (FPR) were selected for statistical evaluation of the trained models (Table 4). All scores were measured as a single overall metric. Sensitivity measures how well the multimodal neural network is able to detect the presence of a disease in actually sick patients. Specificity determines how the multimodal neural network detects the absence of disease in healthy individuals. The higher the sensitivity, the more reliable the intelligent classification multimodal neural network for cardiac arrhythmias. The F-1 score is the harmonic mean of the positive predictive value and sensitivity. The statistical metric F-1 score is dependent on the ratio of data in categories and cannot always correctly evaluate systems in which there is a clear imbalance of data. MCC is a more reliable measure of the statistical evaluation of systems with unbalanced data. A high

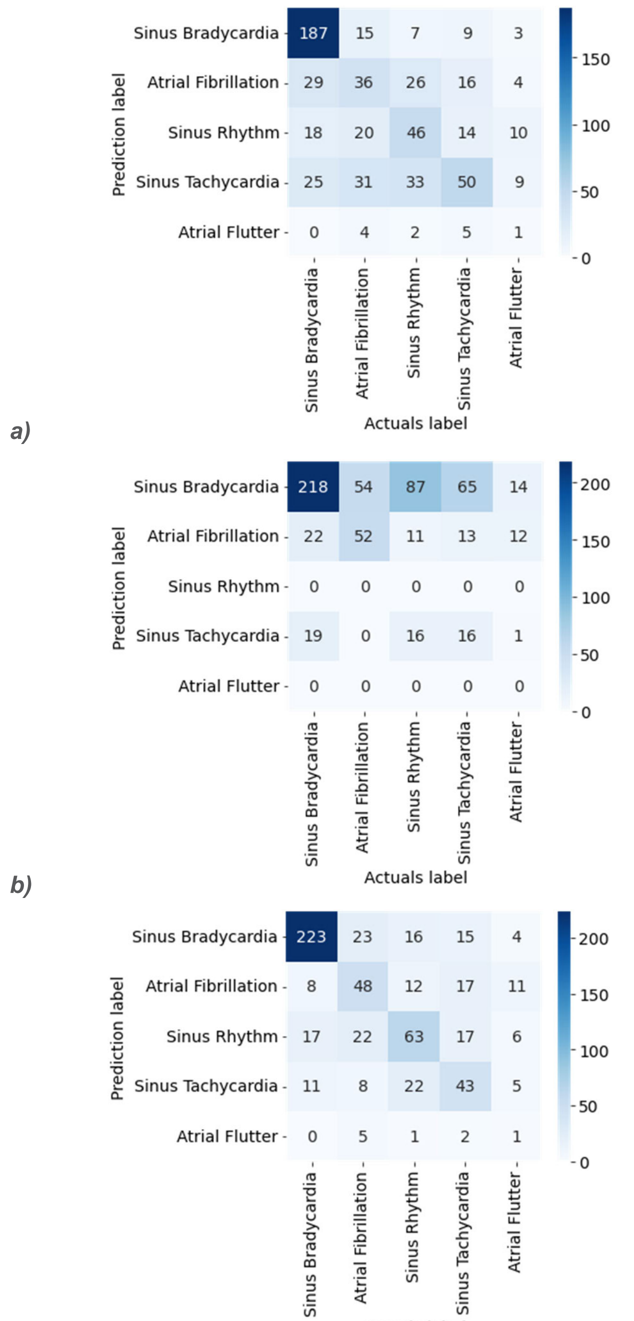

FIGURE 4. A matrix of tests for confusion of the developed architecture of a multimodal neural network for classifying ECG signals: (a) LSTM for processing ECG signals, (b) a linear perceptron for processing patient statistics, (c) a multimodal neural network for processing heterogeneous data from a selected base for modeling.

MCC score indicates that the multimodal neural network performs well in all four categories of the confusion matrix in proportion to the amount of data in the categories [55]. False positive rate (FNR) and true positive rate (FPR) are the probability of false and true rejection of the null hypothesis as a result of testing a neural network system. When testing the proposed neural network systems for recognition of cardiac

TABLE 4. Results of the test evaluation of various neural networks.

NN architecture	Specificity	Sensitivity	F-1 score	MCC	FNR	FPR
Linear perceptron	0.8691	0.4766	0.4766	0.2087	0.5233	0.1308
LSTM	0.8833	0.5333	0.5333	0.3679	0.4666	0.1166
MM-NN	**0.9075**	**0.6300**	**0.6300**	**0.4770**	**0.3700**	**0.0925**

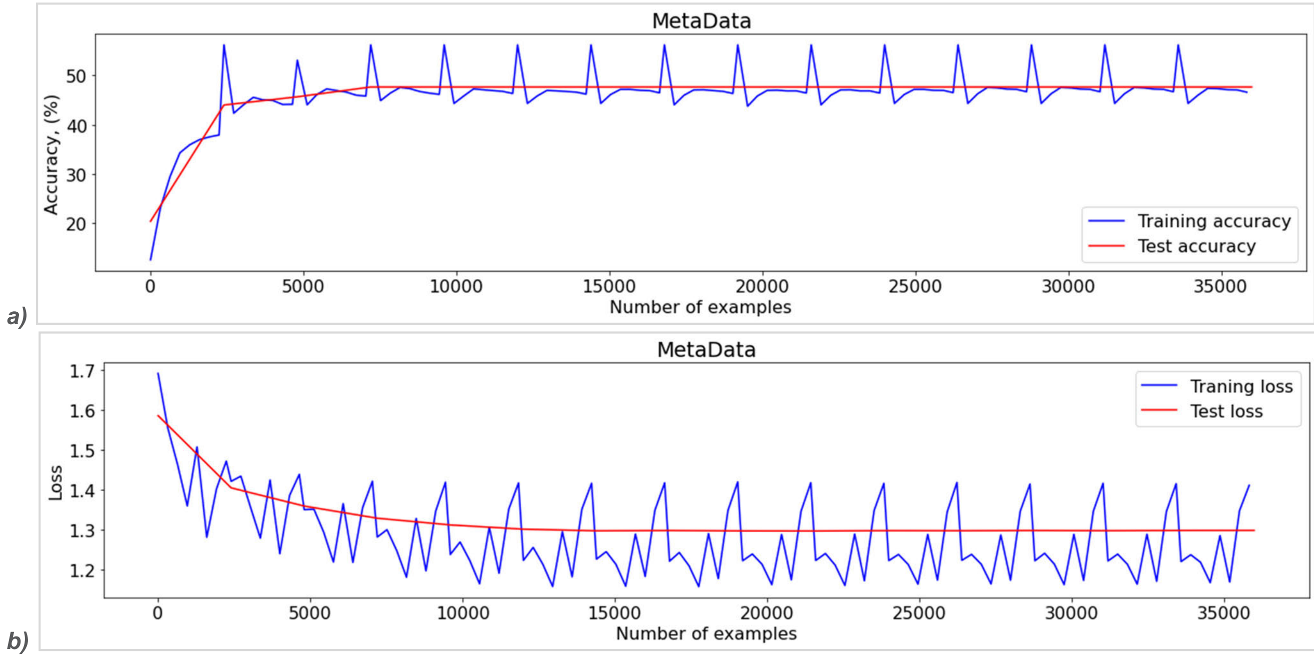

FIGURE 5. Graphs for modeling the architecture of Linear perceptron (MetaData), a part of the developed multimodal neural network for processing statistical data of patients: (a) recognition accuracy values; (b) error function values.

arrhythmias, it was found that the highest sensitivity index belongs to the MM-NN analysis of heterogeneous data based on the LSTM architecture and is 0.6300. The highest estimation index F-1 belongs to MM-NN analysis of heterogeneous data and is 0.6300. The best MCC score was 0.4770 and was obtained by estimating MM-NN based on the LSTM architecture. As a result of testing all trained neural network systems, the best result in terms of statistical evaluation measures FNR and FPR was obtained from MM-NN analysis of heterogeneous data based on the LSTM architecture and amounted to 0.3700 and 0.0925, respectively.

Figures 4(a)-4(c) present confusion matrices for testing the developed neural network systems, according to which the simultaneous analysis of heterogeneous data when training a multimodal neural network multimodal neural network can reduce the number of false predictions. When modeling a multimodal neural network based on the Linear perceptron architecture for processing patient statistical data, recognition shifted towards the most common categories. Figures 5-7 show simulation graphs of the developed systems for analyzing cardiological data for recognizing heart diseases. The graphs confirm the results presented in Table 3. An increase in recognition accuracy and a decrease in the value of the loss function are shown. The merging of the recognition graphs of the test and training bases in Figure 5 indicates that the best simulation result has been obtained. The graphs in figures 6 and 7 show the possibility of improving the classification.

On Figure 8 shows comparative graphs of test accuracy values and error functions for the developed multimodal neural network and its components. As shown in Figure 8(a), MM-NN gives a better classification result than classification using only the LSTM network or only the linear perceptron. In Figure 8(b), when testing the model with only a linear network, the loss function remains almost unchanged. When testing only with the LSTM network, the loss function, having reached its minimum value at a certain moment, begins to grow. This may indicate model overfitting due to missing any data or due to data imbalance. However, when testing the proposed MM-NN neural network system, the loss function decreases, and the accuracy graph (Fig. 8(a)) grows. This means that adding information about individual patient data to the MM-NN allows the MM-NN to learn more accurately even from a limited amount of data.

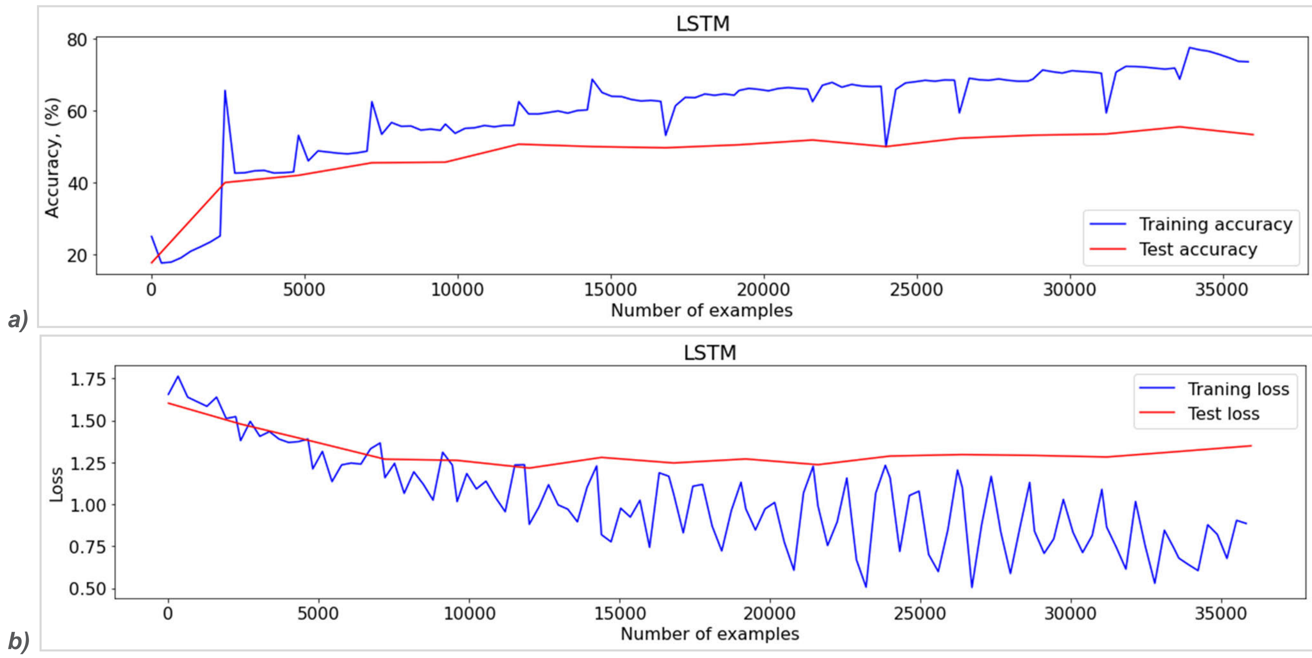

FIGURE 6. LSTM network modeling graphs are part of the developed neural network architecture for processing ECG signals: (a) recognition accuracy values; (b) error function values.

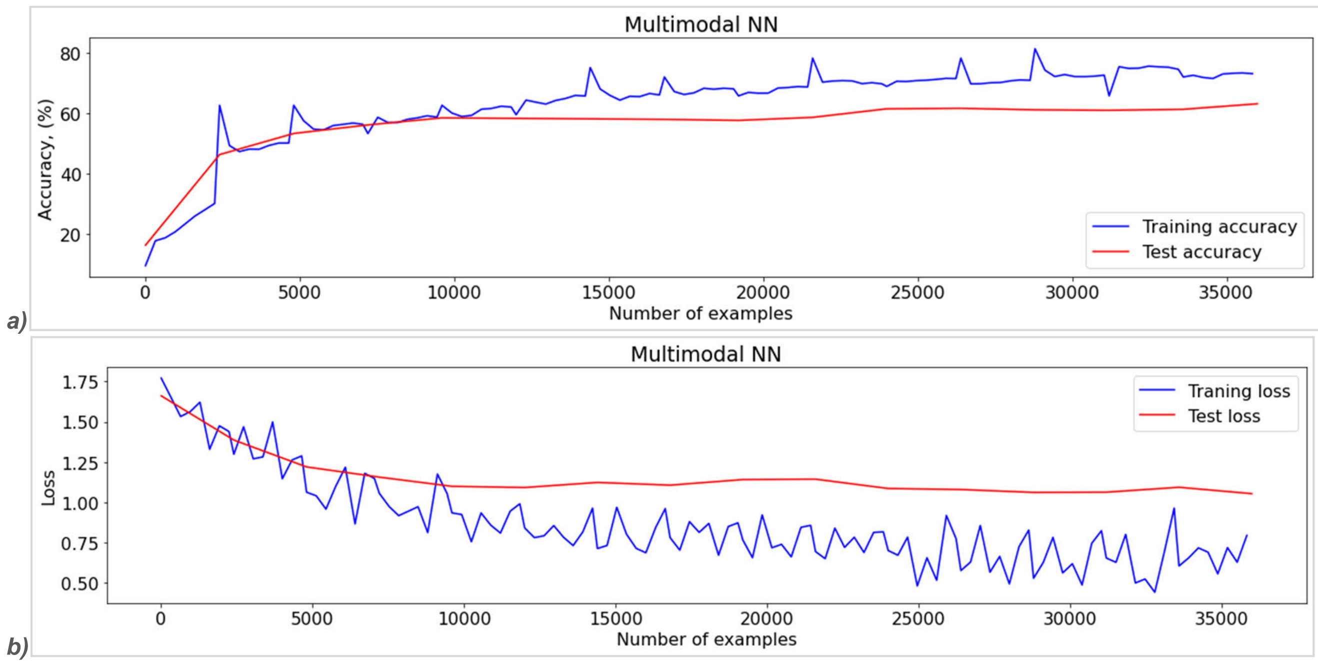

FIGURE 7. Graphs of modeling a multimodal neural network for classifying ECG signals: (a) recognition accuracy values; (b) error function values.

B. COMPARISON WITH STATE-OF-THE-ART

A comparative analysis of the simulation results with already known methods for detecting cardiac arrhythmias based on the classification of recurrent networks was carried out.

In [16], the MIT-BIH database was used to model the proposed method. This database includes the results of taking a two-load electrocardiogram from 47 patients for 48 hours. The authors used a modification of the base, in which the ECG stream is divided into signals containing one heartbeat each. Signal preprocessing in this method includes the synthetic minority resampling (SMOTE) method to add new information and the Edited Nearest method to remove misclassified samples. The LSTM network is used to classify ECG signals. Combining the first two methods led to the most accurate separation of classes, which made it possible to improve the classification result. However, the use of the database by a relatively small number of patients could skew the results of the study. The authors of [12] simulated two models based on classification by the LSTM network, but to solve the problem of data imbalance, they pre-processed

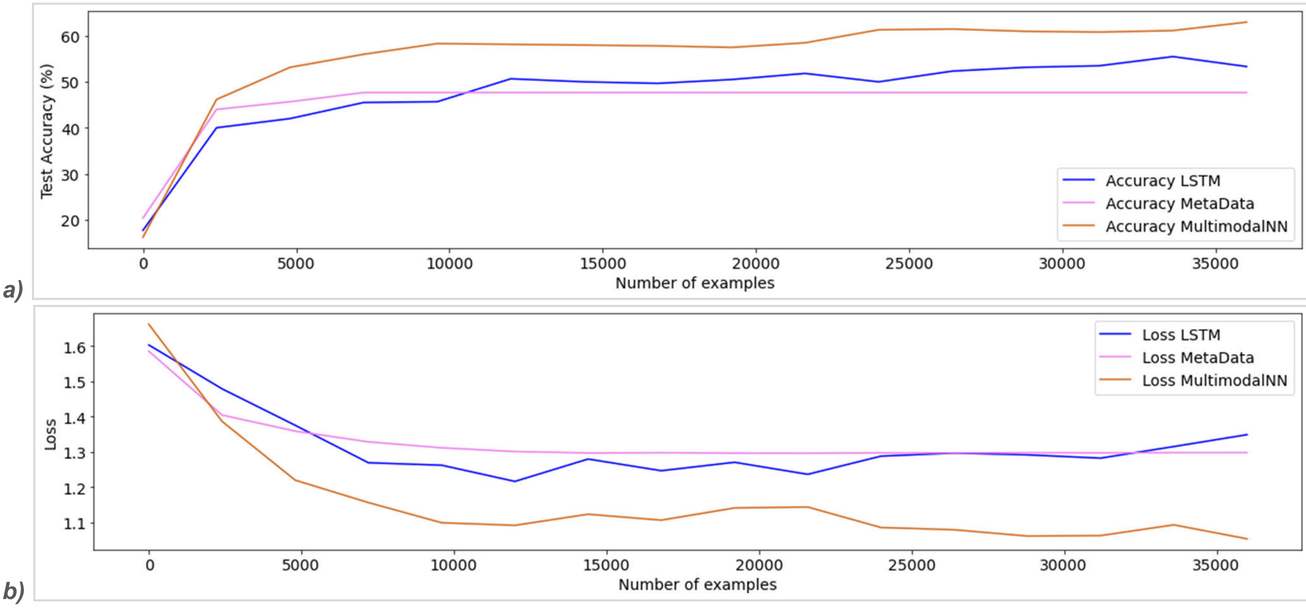

FIGURE 8. Graphs of modeling parts of a multimodal neural network for classifying ECG signals: (a) test accuracy values; (b) error checking function values.

the signals using random resampling (ROS) and SMOTE methods. Data imbalance can lead to retraining of the neural network, which leads to incorrect classification of signals. The ROS and SMOTE resampling methods correct the data imbalance by adding new data. The simulation was carried out on the 12-channel PhysioNet/Computing in Cardiology Challenge database. Such low rates are due to the lack of personalization of signals. In [25], the PhysioNet/Computing in Cardiology Challenge database was also used for modeling. The simulation was carried out on a three-layer architecture of the ResNet recurrent network. The opportunity to improve performance lies in the pre-processing of data and their personalization. Comparison of simulation results with previous works are presented in Table 5.

TABLE 5. Comparison of simulation results with previous works.

Method	Accuracy	F-1 score
LSTM ROS [12]	0.38	0.39
LSTM SMOTE [12]	0.39	0.39
three layers of ResNet [25]	0.61	NA
Propose MM-NN	**0.63**	**0.63**

The proposed MM-NN showed a result 2 percentage points higher than the method proposed in [39] and 45 and 44 percentage points higher than the LSTM network [35].

Pre-processing of ECG signals is necessary to improve classification results by methods such as wavelet transform, synthetic minority resampling, random resampling, P, R, S, T, and U wave extraction from the ECG signal stream. Also, adding a feature extraction stage using CNN, RNN, pre-trained neural networks will achieve a better result. Using ensemble methods, for example [56], to classify ECG signals will also improve modeling results. Combining all the proposed methods for improving the MM-NN will possibly achieve the maximum result of signal classification.

V. CONCLUSION

The paper proposes a multimodal neural network that improves the accuracy of cardiac arrhythmias recognition on ECG using patient data. Patient metadata is processed by a linear perceptron. ECG signals classification by LSTM network. The simulation results showed that the proposed multimodal neural network achieves a recognition accuracy of 63%, which is 2 percentage points higher compared to state-of-the-art methods. The accuracy of the classification is increased due to the use of metadata Methods for improving classification results include pre-processing of signals and pre-feature extraction on an ECG. The introduction of data balance methods, signal preprocessing methods, ensemble methods for classifying ECG signals is the direction of future research.

The proposed method is not an independent method for diagnosing heart diseases. The use of the proposed method is possible only as an aid to a cardiologist. The proposed method is not an independent method for diagnosing heart diseases. The use of the proposed method is possible only as an aid to a cardiologist. The main limitation of the system is the simultaneous availability of patient metadata and their ECG signals.

ACKNOWLEDGMENT

The authors express their gratitude to the North-Caucasus Federal University for supporting the competition of projects of research teams and individual scientists of the North-Caucasus Federal University.

REFERENCES

[1] G. Mattsson and P. Magnusson, Eds., *Cardiomyopathy—Disease of the Heart Muscle*. IntechOpen, 2021, doi: 10.5772/intechopen.91489.

[2] G. V. Santalova, A. A. Garanin, P. A. Lebedev, M. E. Kuzin, and O. V. Tereshina, "Promising approaches to early diagnosis of chronic rheumatic heart disease," *Sci. Innov. Med.*, vol. 7, no. 3, pp. 170–175, Sep. 2022.

[3] L. R. Caplan, R. P. Simon, and S. Hassani, "Cerebrovascular disease-stroke," in *Neurobiology of Brain Disorders*, 2nd ed. New York, NY, USA: Academic, 2023, pp. 457–476.

[4] WHO List of Priority Medical Devices for Management of Cardiovascular Diseases and Diabetes. (2021). *WHO Medical Device Technical Series*. World Health Organization, Geneva, Switzerland. Accessed: Jan. 3, 2023. [Online]. Available: https://www.who.int/publications/i/item/9789240027978

[5] H. Ding and J. He, "Cardiovascular system," in *Clinical Molecular Diagnostics*, S. Pan and J. Tang, Eds. Singapore: Springer, Jul. 2021, pp. 221–227.

[6] H. Thomas, J. Diamond, A. Vieco, S. Chaudhuri, E. Shinnar, S. Cromer, P. Perel, G. A. Mensah, J. Narula, C. O. Johnson, G. A. Roth, and A. E. Moran, "Global atlas of cardiovascular disease 2000–2016: The path to prevention and control," *Global Heart*, vol. 13, no. 3, p. 143, Sep. 2018.

[7] B. Campbell, D. Richley, C. Ross, and C. J. Eggett. (2017). *Clinical Guidelines by Consensus: Recording a Standard 12-Lead Electrocardiogram*. An Approved Method by the Society for Cardiological Science and Technology (SCST). Accessed: Jan. 3, 2023. [Online]. Available: http://www.scst.org.uk/resources/SCST_ECG_Recording_Guidelines_2017

[8] M. R. Fikri, I. Soesanti, and H. A. Nugroho, "ECG signal classification review," *IJITEE Int. J. Inf. Technol. Electr. Eng.*, vol. 5, no. 1, p. 15, Jun. 2021.

[9] F. Murat, O. Yildirim, M. Talo, U. B. Baloglu, Y. Demir, and U. R. Acharya, "Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review," *Comput. Biol. Med.*, vol. 120, May 2020, Art. no. 103726, doi: 10.1016/j.compbiomed.2020.103726.

[10] Y. H. Awni, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, pp. 65–69, Jan. 2019.

[11] F. Liu, X. Zhou, J. Cao, Z. Wang, H. Wang, and Y. Zhang, "A LSTM and CNN based assemble neural network framework for arrhythmias classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1303–1307.

[12] D. M. J. Gutierrez, *Application of Federated Learning Techniques for Arrhythmia Classification Using 12-Lead ECG Signals*. Amsterdam, The Netherlands: Elsevier, 2022.

[13] S. Saadatnejad, M. Oveisi, and M. Hashemi, "LSTM-based ECG classification for continuous monitoring on personal wearable devices," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 515–523, Feb. 2020.

[14] A. Amirshahi and M. Hashemi, "ECG classification algorithm based on STDP and R-STDP neural networks for real-time monitoring on ultra low-power personal wearable devices," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 6, pp. 1483–1493, Dec. 2019.

[15] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 48, no. 12, pp. 2095–2104, Dec. 2018.

[16] E. Y. Shchetinin and A. G. Glushkova, "Arrhythmia detection using resampling and deep learning methods on unbalanced data," *Comput. Opt.*, vol. 46, no. 6, pp. 980–987, 2022.

[17] E. Al-Masri, "Detecting ECG heartbeat abnormalities using artificial neural networks," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 5279–5281.

[18] P. Lyakhov, M. Kiladze, and U. Lyakhova, "System for neural network determination of atrial fibrillation on ECG signals with wavelet-based preprocessing," *Appl. Sci.*, vol. 11, no. 16, p. 7213, Aug. 2021.

[19] J. Huang, B. Chen, B. Yao, and W. He, "ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network," *IEEE Access*, vol. 7, pp. 92871–92880, 2019.

[20] R. Banerjee, A. Ghose, and K. Muthana Mandana, "A hybrid CNN-LSTM architecture for detection of coronary artery disease from ECG," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.

[21] R. L. B. Buana and I. Hudati, "Review: Analisis fitur deteksi aritmia dan metode deep learning untuk wearable devices," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 11, no. 1, pp. 61–69, Feb. 2022.

[22] T. Zheng, W. Feng, C. Yu, and Q. Wu, "Feature extraction analysis method of pre-trained CNN model for SAR target recognition," *Int. J. Remote Sens.*, vol. 44, no. 7, pp. 2294–2316, Apr. 2023.

[23] A. Rajkumar, M. Ganesan, and R. Lavanya, "Arrhythmia classification on ECG using deep learning," in *Proc. 5th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Mar. 2019, pp. 365–369.

[24] S. Savalia and V. Emamian, "Cardiac arrhythmia classification by multi-layer perceptron and convolution neural networks," *Bioengineering*, vol. 5, no. 2, p. 35, May 2018.

[25] S. Hong, Y. Zhou, J. Shang, C. Xiao, and J. Sun, "Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review," *Comput. Biol. Med.*, vol. 122, Jul. 2020, Art. no. 103801, doi: 10.1016/j.compbiomed.2020.103801.

[26] X. Xu, S. Jeong, and J. Li, "Interpretation of electrocardiogram (ECG) rhythm by combined CNN and BiLSTM," *IEEE Access*, vol. 8, pp. 125380–125388, 2020.

[27] Z. Zhao, H. Fang, S. D. Relton, R. Yan, Y. Liu, Z. Li, J. Qin, and D. C. Wong, "Adaptive lead weighted ResNet trained with different duration signals for classifying 12-lead ECGs," in *Proc. Comput. Cardiol.*, Sep. 2020, pp. 1–4.

[28] A. Rana and K. K. Kim, "ECG heartbeat classification using a single layer LSTM model," in *Proc. Int. SoC Design Conf. (ISOCC)*, Oct. 2019, pp. 267–268.

[29] M. R. Kiladze, "Application of bidirectional LSTM neural networks and noise pre-cleaning for feature extraction on an electrocardiogram," presented at the APAMCS, Oct. 2022.

[30] D. Garg, G. K. Verma, and A. K. Singh, "A review of deep learning based methods for affect analysis using physiological signals," *Multimedia Tools Appl.*, vol. 82, no. 17, pp. 26089–26134, Jul. 2023, doi: 10.1007/s11042-023-14354-9.

[31] S. A. Park, H. C. Lee, C. W. Jung, and H. L. Yang, "Attention mechanisms for physiological signal deep learning: Which attention should we take?" in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds, vol. 13431. Cham, Switzerland: Springer, Sep. 2022, pp. 613–622.

[32] Y. Wu, S. Krishnan, and B. Ghoraani, "Computational methods for physiological signal processing and data analysis," *Comput. Math. Methods Med.*, vol. 2022, pp. 1–4, Aug. 2022.

[33] F. Genz, C. Hufeld, and D. Kranzlmuller, "Processing physiological sensor data in near real-time as social signals for their use on social virtual reality platforms," in *Extended Reality*, vol. 13446, L. T. De Paolis, P. Arpaia, and M. Sacco, Eds. Cham, Switzerland: Springer, Aug. 2022, pp. 44–62.

[34] T. Arias-Vergara, P. Klumpp, J. C. Vasquez-Correa, E. Nöth, J. R. Orozco-Arroyave, and M. Schuster, "Multi-channel spectrograms for speech processing applications using deep learning methods," *Pattern Anal. Appl.*, vol. 24, no. 2, pp. 423–431, May 2021.

[35] J. Li and Q. Wang, "Multi-modal bioelectrical signal fusion analysis based on different acquisition devices and scene settings: Overview, challenges, and novel orientation," *Inf. Fusion*, vol. 79, pp. 229–247, Mar. 2022.

[36] Z. Xu, "Adversarial Uni- and multi-modal stream networks for multimodal image registration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 12263. Berlin, Germany: Springer, 2020, pp. 222–232.

[37] W. Yin, J. Huang, J. Chen, and Y. Ji, "A study on skin tumor classification based on dense convolutional networks with fused metadata," *Frontiers Oncol.*, vol. 12, Dec. 2022, Art. no. 989894.

[38] A. A. Solomakha and V. I. Gorbachenko, "Neural network system for diagnosing liver failure in surgery," *J. Experim. Clin. Surg.*, vol. 14, no. 4, pp. 303–306, Dec. 2021.

[39] F. Moaddab, A. Ghanbari, Z. Taheri-Ezbarami, A. Salari, and E. Kazemnezhad-Leyli, "Clinical parameters and outcomes in heart failure patients based on gender differences," *Int. J. Afr. Nursing Sci.*, vol. 18, Jan. 2023, Art. no. 100525.

[40] B. Waitzfelder, L. Palaniappan, and A. Varga, "Prevalence of cardiovascular disease among Asian, Pacific islander and multi-race populations in Hawai'i and California," *BMC Public Health*, vol. 23, p. 885, May 2023, doi: 10.1186/s12889-023-15795-5.

[41] R. A. Wild, K. M. Hovey, C. Andrews, J. G. Robinson, A. M. Kaunitz, J. E. Manson, C. J. Crandall, R. Paragallo, C. Shufelt, and C. N. B. Merz, ''Cardiovascular disease (CVD) risk scores, age, or years since menopause to predict cardiovascular disease in the women's health initiative,'' *Menopause*, vol. 28, no. 6, pp. 610–618, Jun. 2021, doi: 10.1097/gme.0000000000001753.

[42] F. L. J. Visseren, ''2021 ESC guidelines on cardiovascular disease prevention in clinical practice,'' *Eur. Heart J.*, vol. 42, no. 34, pp. 3227–3337, Sep. 2021, doi: 10.1093/eurheartj/ehab484.

[43] Y. Harada, S. Tomiyama, T. Sakamoto, S. Sugimoto, R. Kawamura, M. Yokose, A. Hayashi, and T. Shimizu, ''Effects of combinational use of additional differential diagnostic generators on the diagnostic accuracy of the differential diagnosis list developed by an artificial intelligence–driven automated history–taking system: Pilot cross-sectional study,'' *JMIR Formative Res.*, vol. 7, Aug. 2023, Art. no. e49034, doi: 10.2196/49034.

[44] J. Miller, L. Williams, A. Alhurani, Z. Saleh, A. Bailey, A. Connell, M. Hammash, M. Chung, and D. Moser, ''Race matters: Cardiovascular disease risk in male U.S. prisoners,'' *Eur. J. Cardiovascular Nursing*, vol. 20, Jul. 2021, Art. no. zvab060, doi: 10.1093/eurjcn/zvab060.060.

[45] F. Ciccirillo, M. G. Abrignani, P. L. Temporelli, G. Binaghi, C. Cappelletto, V. Lopriore, A. Cesaro, A. Maloberti, D. Cozzoli, C. Riccio, P. Caldarola, F. Oliva, D. Gabrielli, and F. Colivicchi, ''The key role of a psychoactive substance use history in comprehensive cardiovascular risk assessment, diagnosis, treatment, and prevention,'' *Cardiology*, vol. 148, no. 3, pp. 257–268, Apr. 2023, doi: 10.1159/000530246.

[46] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, ''Multimodal fusion for multimedia analysis: A survey,'' *Multimedia Syst.*, vol. 16, no. 6, pp. 345–379, Nov. 2010.

[47] H. Haibo and M. Yunqian Eds., *Imbalanced Learning: Foundations, Algorithms, and Applications*. Hoboken, NJ, USA: Wiley, 2013, pp. 13–41.

[48] *PhysioNet/Computing in Cardiology Challenge 2021*. Accessed: Oct. 22, 2022. [Online]. Available: moody-challenge.physionet.org/2021/

[49] J. Andreu-Perez, C. C. Y. Poon, R. D. Merrifield, S. T. C. Wong, and G.-Z. Yang, ''Big data for health,'' *IEEE J. Biomed. Health Informat.*, vol. 19, no. 4, pp. 1193–1208, Jul. 2015.

[50] Y. Murakami and Y. Takeuchi, ''Medical statistics as a foundation of biomedical research,'' *Brain Nerve Shinkei Kenkyūno Shinpo*, vol. 74, no. 4, pp. 341–351, Apr. 2022.

[51] P. Waldmann, ''Approximate Bayesian neural networks in genomic prediction,'' *Genet. Selection Evol.*, vol. 50, no. 1, pp. 1–9, Dec. 2018.

[52] A. R. Barron, ''Approximation and estimation bounds for artificial neural networks,'' *Mach. Learn.*, vol. 14, no. 1, pp. 115–133, Jan. 1994.

[53] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, ''Learning internal re presentations by error propagation,'' *Parallel Distrib. Process.*, vol. 1, pp. 318–362, Jan. 1986.

[54] S. Hochreiter and J. Schmidhuber, ''Long short-term memory,'' *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[55] D. Chicco and G. Jurman, ''The advantages of the Matthews correlation coefficient (MCC) over f1 score and accuracy in binary classification evaluation,'' *BMC Genomics*, vol. 21, no. 1, pp. 1–13, Dec. 2020.

[56] E. Essa and X. Xie, ''An ensemble of deep learning-based multi-model for ECG heartbeats arrhythmia classification,'' *IEEE Access*, vol. 9, pp. 103452–103464, 2021.

MARIYA R. KILADZE received the bachelor's degree in mathematics and computer science and the master's degree in mathematics and applied informatics from North-Caucasus Federal University, Stavropol, Russia, in 2017 and 2019, respectively. She is a Postgraduate Student and a Junior Researcher with the Department of Mathematical Modeling, Faculty of Mathematics and Computer Science, named after Prof. N. I. Chervyakov North-Caucasus Federal University. She is the author of more than ten scientific papers. Her research interests include digital visual data processing, artificial intelligence, medical data processing, digital signal processing, and physiological signal processing.

ULYANA A. LYAKHOVA received the bachelor's degree in biology and the master's degree in applied mathematics and informatics from North-Caucasus Federal University, in 2018 and 2020, respectively.

She is a Postgraduate Student and a Junior Researcher with the Department of Mathematical Modeling, Faculty of Mathematics and Computer Science, named after Prof. N. I. Chervyakov North-Caucasus Federal University. She is the author of more than 20 scientific papers, four certificates of state registration of computer programs, and one invention patent. Her research interests include digital processing of visual data, artificial intelligence, and medical data processing.

PAVEL A. LYAKHOV graduated in mathematics from Stavropol State University, in 2009. He received the Ph.D. degree in mathematics from Stavropol State University, in 2012. He has been with North-Caucasus Federal University, since 2012, where he is currently the Head of the Department of Mathematical Modeling. He is the author of more than 170 scientific papers and educational and methodical works. His research interests include high-performance computing, quantum-dot cellular automata, residue number systems, digital signal processing, image processing, and machine learning.

NIKOLAY N. NAGORNOV received the master's degree in applied mathematics and informatics in 2016. In 2020, he defended his dissertation for the degree of candidate of technical sciences with St. Petersburg State Electrotechnical University ''LETI'' named after V. I. Ulyanov.

Since 2020, he has been a Junior Researcher with the Department of Mathematical Modeling, North-Caucasus Federal University. He is the author of more than 40 scientific papers and nine copyright certificates. His research interests include digital imaging, medical imaging, convolutional neural networks, artificial intelligence, digital filtering, discrete wavelet transform, modular arithmetic, parallel computing, high performance computing, digital devices, and hardware accelerators.

MOHSEN VAHABI received the B.Sc. degree from the Azad University of Kerman, Kerman, Iran, in 2014, and the M.Sc. degree (Hons.) from the Science and Research Tehran Branch, Azad University, Iran, in 2017, both in electrical engineering-electronic. He is currently pursuing the Ph.D. degree in electrical engineering-electronic with the Faculty of Electrical Engineering, Shahrood University of Technology. His research interests include computer arithmetic, reversible circuits, digital signal processing, image processing, machine learning, nano-electronics, and quantum and molecular electronic devices, with an emphasis on quantum-dot cellular automata technology and atomic silicon quantum dot circuits.