ATLANTIS PRESS
Atlantis Highlights in Computer Sciences, volume 3
21st International Scientific Workshop on Computer Science and Information Technologies (CSIT 2019)

Increasing Security of Telemedicine Service

Marat Bogdanov
Ufa State Aviation Technical University
M. Akmullah named after Bashkir State Pedagogical University
Ufa City, Russia
bogdanov_marat@mail.ru

Dajan Nasyrov
Ufa State Aviation Technical University
Ufa City, Russia
9373007333@mail.ru

Aleksander Dumchikov
Ural Federal University
Ekaterinburg City, Russia
redfoxufa@gmail.com

Artur Samigullin
M. Akmullah named after Bashkir State Pedagogical University
Ufa City, Russia
samigullin.art@gmail.com

***Abstract*—The paper is about the improving the safety of telemedicine services that carry out the diagnosis of cardiovascular diseases on ECG. There are 2 deep learning methods (Multilayer Perceptron and Recurrent Neural Network), 14 machine learning methods (Naive Bayes classifier for multivariate Bernoulli models, A decision tree classifier, An extremely randomized tree classifier, Classifier implementing the k nearest neighbors vote, Linear Discriminant Analysis, Linear Support Vector Classification, Logistic Regression, Nearest centroid classifier, A random forest classifier, Classifier using Ridge regression, Ridge classifier with built in cross validation, Gaussian Mixture Models, Support Vector Machines), 4 ECG digitization time intervals (5, 10, 15 and 20 seconds), and 3 databases of digitized electrocardiograms (The Physikalisch-Technische Bundesanstalt (PTB) Diagnostic ECG Database, European ST-T Database, St.-Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database). It was realized that The accuracy of biometric identification and diagnosis of cardiovascular diseases increases with an increase in ECG registration time to about 10 seconds, after which it reaches a plateau, biometric identification and diagnosis of cardiovascular diseases are possible with a signal registration time of 5 seconds and the most stable recognition results were given by such methods of classification of biometric features as fully connected neural network (MLP), An extremely randomized tree classifier and Classifier implementing the k-nearest neighbors vote.**

Keywords—Machine Learning, Deep Learning, ECG, Biometric authentication, Cardiovascular diseases diagnosis

I. INTRODUCTION

Modern electrocardiographs allows diagnosing of various disorders of the cardiovascular system. These devices are non-invasive, easy to use and widespread in the world. Recently, various wearable devices allowing to measure a person pulse or ECG are getting more popular. Such gadgets include, for example, smart watches from Apple or Samsung. Using of electrocardiograms, we can not only make a diagnosis, but also perform a person biometric identification. A piecewise aligned morphology of an electrocardiogram may be used as a biometric feature [1]. It is supposed to register the ECG using a smartphone and further remote person identification. During the research, biometric identification of 100 users was carried out. The recognition error was 1.34%. High noise immunity of the proposed technology is noted. In the past few years, researchers have shown interest in the problem of remote ECG registration using ultra-wideband radars. The goal of such work is to search for survivors of natural disasters buried under the rubble, as well as to identify terrorists and hostages in enclosed spaces. In 2012, the US Department of Defense Advanced Engineering Agency (DARPA) announced a challenge SB121-004: Biometrics-at-a-distance ($ 147,000,000) [3]. The purpose of the competition was to demonstrate the ability to register, localize and evaluate physiological signals (e.g., heartbeat) at a distance of more than 10 m, out of sight, through solid obstacles (walls, stones, concrete, etc.). According to the tender documentation It is necessary to organize remote detection, collection and evaluation of physiological signals of interest. Applications using this technology include, but are not limited to mopping up buildings, monitoring the health of soldiers or assessing and classifying personnel injuries sustained during a battle, situational awareness. Existing micro pulse radars (MIR) and ultra-wideband (UWB) radar technologies can detect heart rate and breathing at a distance of up to 8 meters, but are limited at large distances, as well as in difficult conditions, for example, in buildings and under radio frequency interference. There is interest in detecting and localizing sources of multiple physiological signatures in a cluttered environment. For example, in a building that has experienced a catastrophic event (fire, earthquake, etc.), the detection of survivors and an assessment of their health status, in addition to their location with an accuracy of 1 meter, will increase the likelihood of personnel being rescued. In addition, in a crowd, it is very difficult to uniquely identify a person based on a collection of physiological signatures, such as electrocardiograms (ECGs). It is possible that high-frequency ECGs or other signals can help increase confidence in identification. It is necessary to demonstrate the possibility of using a technology that can record human vital signs at distances of more than 10 meters, in the absence of visibility, non-invasively, non-contact. The technology should be able to uniquely identify 10 subjects with a confidence> 95% inside the building. VAWD Applied Science and Technology from Arizona won the challenge [4]. The company produces portable radars and car-based devices (VAWD STORM System). A portable radar detects people at a distance of up to 50 m, a car radar - over 100 m, underground, through walls. There is the possibility of biometric classification and simultaneous detection of several people, the tracking of people on the ground. Similar work is carried out by the

Israeli company Xaver, which produces tactical devices for the fast detection of signs of life [5]. In particular, one of its developments, the XaverTM 400 multi-channel ultra-wideband pulsed radar, allows to determine in real time the presence of living people, their number and movements.

When extracting cardiocycles from a radio signal reflected from a living object, low-pass filters are usually used (cutoff frequency 1-3 Hz), after which the principal component method is used [6].

The widespread use of telemedicine services is faced to security problems. Wearable devices that collect and transmit biomedical signals are Internet of things. Such gadgets usually have a relatively weak level of security, which limits the area of their medical use.

II. Motivation and aim

We consider the possibility of increasing the safety of telemedicine services based on electrocardiograms. It is supposed to use digitized electrocardiograms not only for the diagnosis of cardiovascular diseases, but also to confirm the identity of the user. Works on biometric identification and diagnosis of diseases include the following stages: signal preprocessing, extraction of biometric features, classification. During the preprocessing of the electrocardiogram, the widely used libraries WFDB [7] and Biosppy [8] were used. When extracting biometric features from an ECG, a number of authors prefer the statistical featuress of the signal using, for example, wavelet analysis or the autocorrelation function using, while other researchers prefer the geometric features of the PQRST complex. We decided to use wavelets, amplitude and time features of the cardiocycle, as well as a whole cardiocycle as biometric features. To classify biometric features, we used two deep learning methods and 14 machine learning methods. For biometric classification, the identifiers of the subjects were used as class labels, while in the diagnosis of diseases of the cardiovascular system, disease codes were used as the label column vector. The aim of the proposed work is to find the optimal combination of biometric features and classification methods that are equally well suited for biometric identification and for diagnosis.

Naive Bayes classifier for multivariate Bernoulli models, A decision tree classifier, An extremely randomized tree classifier, Classifier implementing the k nearest neighbors vote, Linear Discriminant Analysis, Linear Support Vector Classification, Logistic Regression, Nearest centroid classifier, A random forest classifier, Classifier using Ridge regression, Ridge classifier with built in cross validation, Gaussian Mixture Models, Support Vector Machines

III. Materials and methods

We used three databases of digitized electrocardiograms available for downloading at www.physionet.org.

The Physikalisch-Technische Bundesanstalt (PTB) Diagnostic ECG Database [9]

ECGs were obtained from healthy volunteers and patients with various cardiac diseases by professor of medicine Michael Oeff, who works in the cardiology department of the Benjamin Franklin University Hospital in Berlin [10].

Signal Characteristics:

16 input channels (12 traditional leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6), 3 Frank leads (vx, vy, vz), 1 for voltage control)

Input Voltage: ± 16 mV

Input Impedance: 100 Ω (DC)

Signal depth: 16 bit

Sampling Rate: 1 kHz

Registration time: 3 minutes

The database contains 549 records obtained from 290 subjects. The number of repetitions varies from 1 to 5.

Records are annotated. The following diagnoses were noted.

Myocardial infarction

Cardiomyopathy/Heart failure

Bundle branch block

Dysrhythmia

Myocardial hypertrophy

Valvular heart disease

Myocarditis

Healthy controls

European ST-T Database [11]

The European ST-T Database (European ST-T Database) is designed to evaluate algorithms for analyzing changes in the cardiocycle in the ST - region and changes in T-waves.

Signal Characteristics:

Number of channels: 2

Sampling Rate: 250 Hz

Bit depth: 12 bit

Input voltage: ± 16 mV.

Measurement time: 2 hours

The database contains 90 samples of digitized ECG obtained from 79 subjects. Each subject was diagnosed or suspected of myocardial ischemia. In addition, subjects observed such disorders as hypertension, ventricular dyskinesia, and the effects of drugs.

St.-Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database [9]

This database contains 75 ECG samples obtained from 32 subjects. This database was provided by the St. Petersburg Institute of Cardiology Technology (Incart). The database was originally developed by Viktor Tikhonenko and Alexander Haustov. An additional check was performed by Sergey Ivanov and Alexey Rivin (Incart).

Signal Characteristics:

Number of Leads: 12

Duration 30 minutes

Sampling rate 257 Hz.

Diagnoses:

Acute MI

Transient ischemic attack (angina pectoris)

Prior MI

Coronary artery disease with hypertension

Sinus node dysfunction

Supraventricular ectopy

Atrial fibrillation or SVTA

WPW

AV block

Bundle branch block

Signal Preprocessing

During the preprocessing of the electrocardiogram, the first lead was selected, the signal was filtered, and the cardiocycles were extracted using the WFDB and BioSPPy libraries. These libraries contain many functions for digitally processing biomedical signals. These libraries allow you to work with a variety of data sources, regardless of the bit depth, sampling rate, and the number of channels of digitized electrocardiograms [12].

Selection of biometric features

We used the complete cardiocycles, the number of elements in them depended on the sampling rate. For example, in the case of the PTB database (sampling rate is 1000 Hz), the cardiocycles consist of 600 points, in the case of the European ST-T Database (sampling rate is 250 Hz) the cardiocycles consist of 150 points, in the case of St.-Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database (sampling rate is 257 Hz) cardiocycles consist of 153 points. The selection of QRS complexes was performed with the BioSPPy library (Fig. 1).

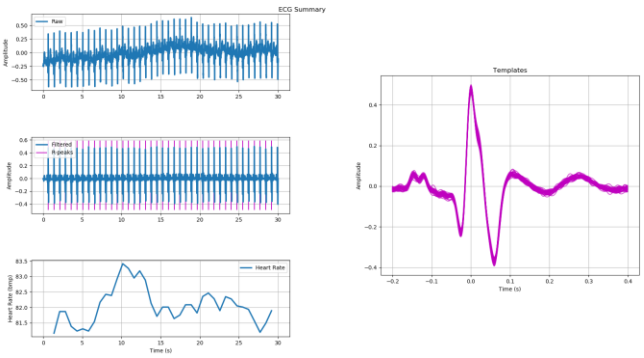

Fig. 1. Extraction of cardiocycles with BioSPPy library.

IV. Results and Discussion

Table 1 shows the effect of ECG registration time on the accuracy of ECG signals (PTB database). Here I correspond to biometric authentication and D correspond to diagnosis of cardiovascular disorders. Deep learning methods show good results. Both recurrent and multiplayer perceptron neural networks show good diagnostics and biometric identification with a reliability higher than 0.95. Machine learning methods such as An extremely randomized tree classifier and Classifier implementing the k-nearest neighbors vote show similar results (0.97 and higher). Label Propagation classifier and SVM show recognition accuracy greater than 0.89. High accuracy of biometric identification (0.93 and higher) and somewhat lower accuracy during diagnosis (0.74 and higher) were shown by such methods as Linear Discriminant Analysis and Logistic Regression (logit, MaxEnt) classifier.

Table 1

Effectiveness of ECG recognition for various periods of time, PTB database

Algorithm of recognition	10 sec		15 sec		20 sec	
	I	D	I	D	I	D
LSTM	0,98	0,97	0,99	0,97	0,99	0,97
MLP	0,98	0,97	0,98	0,95	0,99	0,97
Naive Bayes classifier for multivariate Bernoulli models	0,58	0,18	0,59	0,21	0,59	0,19
A decision tree classifier	0,81	0,88	0,82	0,91	0,86	0,92
An extremely randomized tree classifier	0,98	0,97	0,98	0,97	0,98	0,98
Classifier implementing the k-nearest neighbors vote	0,98	0,99	0,97	0,98	0,98	0,98
Label Propagation classifier	0,88	0,89	0,89	0,89	0,90	0,90
Linear Discriminant Analysis	0,93	0,74	0,93	0,73	0,94	0,74
Linear Support Vector Classification	0,88	0,75	0,90	0,73	0,91	0,73
Logistic Regression (aka logit, MaxEnt) classifier	0,97	0,78	0,97	0,76	0,98	0,76
Nearest centroid classifier	0,61	0,28	0,63	0,28	0,61	0,27
A random forest classifier	0,11	0,70	0,12	0,69	0,10	0,69
Classifier using Ridge regression	0,46	0,71	0,44	0,70	0,46	0,70
Ridge classifier with built-in cross-validation	0,46	0,71	0,44	0,70	0,46	0,70
Gaussian Mixture Models	0,75	0,24	0,78	0,24	0,77	0,24
SVM	0,89	0,88	0,94	0,87	0,97	0,89

From a practical point of view, it is desirable to minimize the signal recording time. Table 2 shows the results of measuring the accuracy of biometric identification and diagnostics on different databases. Very good results were obtained in case of using a fully connected neural network (in two cases, the result is higher than 0.9, in one - above 0.88). When using the methods An extremely randomized tree classifier and Classifier implementing the k-nearest neighbors vote, in almost all cases the recognition coefficient was higher than 0.9. The Label Propagation classifier method gave slightly smaller results (0.86 and higher).

Table 2

ECG recognition results on different databases, signal registration time 5 seconds

Algorithm of recognition	PTB		European ST-T		St.-Petersburg	
	I	D	I	D	I	D
LSTM	0,02	0,69	0,97	0,78	0,85	0,71
MLP	0,96	0,88	1,00	0,99	0,93	0,90
Naive Bayes classifier for multivariate Bernoulli models	0,51	0,12	0,70	0,33	0,33	0,18
Decision tree classifier	0,67	0,84	0,90	0,88	0,76	0,78
Extremely randomized tree classifier	0,95	0,93	0,99	1,00	0,91	0,90
Classifier implementing the k-nearest neighbors vote	0,92	0,96	0,99	0,99	0,88	0,92
Label Propagation	0,86	0,87	0,99	0,99	0,89	0,92
Linear Discriminant Analysis	0,90	0,71	0,99	0,81	0,70	0,66
Linear Support Vector Classification	0,77	0,69	0,98	0,79	0,76	0,66
Logistic Regression	0,94	0,72	0,94	0,65	0,70	0,63
Nearest centroid classifier	0,61	0,28	0,83	0,32	0,39	0,28
Random Forest	0,08	0,72	0,03	0,10	0,20	0,54
Classifier using Ridge regression	0,39	0,73	0,26	0,22	0,49	0,64
Ridge classifier with built-in cross-validation	0,39	0,73	0,77	0,64	0,49	0,64
Gaussian Mixture Models	0,56	0,24	0,80	0,66	0,36	0,25
SVM	0,65	0,80	0,94	0,46	0,20	0,44

V. Conclusions

Summing up the above, we can draw the following conclusions. The accuracy of biometric identification and diagnosis of cardiovascular diseases increases with an increase in ECG registration time to about 10 seconds, after which it reaches a plateau. Biometric identification and diagnosis of cardiovascular diseases are possible with a signal registration time of 5 seconds. The most stable recognition results were given by such methods of classification of biometric features as fully connected neural network (MLP), An extremely randomized tree classifier and Classifier implementing the k-nearest neighbors vote.

Acknowledgment

The reported study was funded by RFBR according to the research project № 19-07-00780

References

[1] Md Saiful Islam. Heartbeat Biometrics for Remote Authentication Using Sensor Embedded Computing Devices. International Journal of Distributed Sensor Networks. Volume 2015, Article ID 549134, 7 pages. http://dx.doi.org/10.1155/2015/549134

[2] Hui-Sup Cho, Young-Jin Park. Detection of Heart Rate through a Wall Using UWB Impulse Radar. Journal of Healthcare Engineering. Volume 2018, Article ID 4832605, 7 pages. https://doi.org/10.1155/2018/4832605

[3] Konkursnaya documentaciya DARPA. http://www.acq.osd.mil/osbp/sbir/solicitations/sbir20121/darpa121.htm

[4] VAWD Applied Science and Technology. http://vawdengineering.com/products.html

[5] Official site of Xaver. www.camero-tech.com

[6] Hui-Sup Cho, Young-Jin Park. Detection of Heart Rate through a Wall Using UWB Impulse Radar. Journal of Healthcare Engineering. Volume 2018, Article ID 4832605, 7 pages. https://doi.org/10.1155/2018/4832605

[7] The research resource for complex physiological signals. https://physionet.org/physiotools/wpg/wpg.pdf

[8] Biosignal processing in Python https://biosppy.readthedocs.io/en/stable/index.html

[9] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/content/101/23/e215.full]; 2000 (June 13).

[10] Bousseljot R, Kreiseler D, Schnabel, A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik, Band 40, Ergänzungsband 1 (1995) S 317. doi:10.13026/C28C71

[11] Taddei A, Distante G, Emdin M, Pisani P, Moody GB, Zeelenberg C, Marchesi C. The European ST-T Database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. European Heart Journal 13: 1164-1172 (1992). doi:10.13026/C2D59Z

[12] Christov I.I. Real time electrocardiogram QRS detection using combined adaptive threshold. BioMedical Engineering OnLine 2004, vol. 3:28. 2004. https://doi.org/10.1186/1475-925X-3-28