Chapter

Biometric Authentication Based on Electrocardiogram

M.R. Bogdanov, A.S. Filippova, G.R. Shakhmametova and Nikolai N. Oskin

Abstract

The life of modern society is impossible without trust. To ensure trust in the digital world, various encryption algorithms and password policies are used. Passwords are used in a variety of applications from banking applications to email. The advantages of passwords include ease of use and widespread distribution. Forgotten password can be restored or changed. Password weaknesses are largely related to the human factor. Many users use passwords such as "1234" or "qwerty," and they are also willing to share passwords with friends and colleagues. Vulnerabilities are also associated with software and hardware manufacturers. Many Wi-Fi routers preset very simple passwords, which many users leave unchanged. There are questions for manufacturers of mobile applications. Due to the imperfection of their software, personal data of users often leak. Due to the prevalence of social networks, new authentication methods have appeared. On many websites, you can use accounts from Facebook or Gmail.com for authentication. If hackers manage to break into large IT vendors, then millions of accounts will be leaked. Many common password problems can be overcome with biometric identification. In particular, biometric data are very difficult to fake; they usually do not change over time. Widespread methods of biometric identification, such as fingerprinting, retina recognition, and voice recognition have various vulnerabilities unfortunately.

Keywords: biometric authentication, electrocardiogram, information security

1. Fingerprinting

This identification method is widely used by the FBI.

The FBI has managed the nation's collection of fingerprints since 1924, but it went fully electronic in 1999 when launched the Integrated Automated Fingerprint Identification System, or IAFIS. This national repository of fingerprints and criminal histories enables law enforcement at every level to quickly match up criminal evidence with criminal identities [1].

On the other hand, the Department of Homeland Security's IDENT—the Automated Biometric Identification System that houses fingerprint records and limited biographic information—was created in 1994 to help U.S. border and immigration officials keep criminals and terrorists from crossing US borders.

In this post-9/11, globalized world, the Department of Justice (DOJ) and FBI, the Department of Homeland Security (DHS), and the Department of State have worked hard in recent years to establish interoperability between these two fingerprint databases.

2. Retina recognition

Retina recognition is a biometric technique that uses the unique patterns on a person's retina for person identification. The retina is the layer of blood vessels situated at the back of an eye. The eye is positioned in front of the system at a capture distance ranging from 8 cm to 1 m. The person must look at a series of markers, viewed through the eyepiece, and line them up. The eye is optically focused for the scanner to capture the retina pattern. The retina is scanned with the near infrared (NIR 890 nm) irradiation, and the unique pattern of the blood vessels is captured. Retina recognition makes use of the individuality of the patterns of the blood vessels. It has been developed commercially since the mid-1970s. Sandia Laboratory reported a false rejection rate of lower than 1.0% [2].

3. Voice recognition

Voice biometrics is the science of using a person's voice as a uniquely identifying biological characteristic in order to authenticate them. Also referred to as voice verification or speaker recognition, voice biometrics enables fast, frictionless, and highly secure access for a range of use cases from call center, mobile, and online applications to chatbots, IoT devices, and physical access.

Like other biometric modalities, voice offers significant security advantages over authentication methods that are based on something you know (like a password or answer to a "secret" question) or something you have (like your mobile phone). Voice biometrics also improves the customer experience by removing frustration associated cumbersome login processes and lost and stolen credentials [3].

Many banks use the voice recognition technologies for person identification.

Unfortunately, the above biometric identification technologies are not free from some disadvantages. In particular, there are methods to fake fingerprints [4], retina [5], and voice [6].

Currently, biometric identification technologies such as ECG, EEG, and DNA are considered resistant to hacking.

In this chapter, we would like to talk about biometric identification using ECG and related problems.

4. What is an ECG?

There are several scenarios for using biometric identification using ECG.

1. Contact ECG

2. Remote ECG sensing

For contact ECG, you can use medical electrocardiographs or sensors installed in smartphones or, for example, in the steering wheel of a car. An interesting direction

in the development of computer technology is wireless sensor networks. A special case of this technology is Body Area Networks (BAN), which include one or more sensors that record the medical parameters of the human body. BANs are Internet of Medical Things (IoMT). ECG sensors in this case are mounted on the patient's body. Medical parameters are transmitted to the server for analysis and storage. An ECG can identify the patient.

Remote ECG sensing is carried out using ultra-wideband radars. An example is the instrument of the Israeli military company XAVER, which allows special forces to detect living people through a brick wall, as well as determine their location, gender, and age (**Figure 1**) [7].

More advanced is the technology of WAVD Technology, Arizona. Its ultra-wideband radar allows not only to detect, but also to identify people buried under the rubble of buildings.

Figure 1.
Xaver 800. Wall-through 3D imagine system.

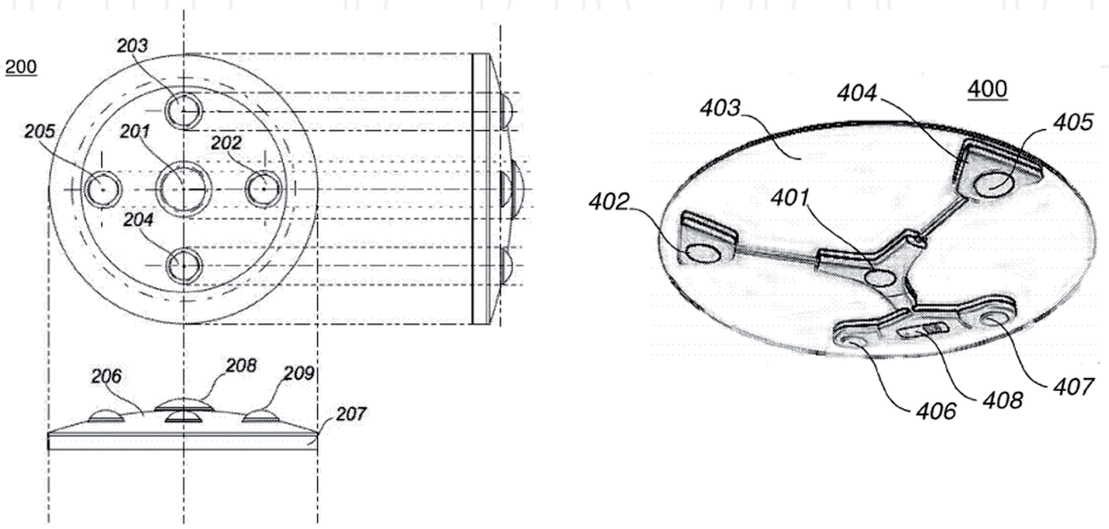

Figure 2.
The use of ultra-wideband radars for biometric identification. Left: banking sector; right: security systems.

Several patents have been published that propose the use of ultra-wideband radars for conducting banking operations without a credit card, as well as for controlling premises during confidential meetings (**Figure 2**). As it turned out, with the help of ultra-wideband radars, it is possible to restore not only the ECG, but also speech.

Xiaolin et al. described using of ultra wide band radars for detecting of vital signs [8, 9].

5. ECG-based biometric identification structure

Biometric identification involves the stage of user registration and the stage of user recognition. At the registration stage, the user takes biometric features and writes them to the database. At the recognition stage, biometric features are taken from an unknown person and consequently compared with the features stored in the database. If the features received from an unknown person by a certain criterion coincided with the features from the database, then a decision is made on the success of the identification. Biometric identification is a complex multi-stage process in which each stage can affect the final recognition accuracy.

While performing the project, we investigated the influence of various factors on the accuracy of biometric identification using electrocardiograms. To do this, a large-scale computational experiment was carried out using our programs written in Python. We used the popular libraries such sklearn, scipy, and matplotlib. Most digitalized electrocardiogram samples were taken from www.physionet.org website. When performing the digital signal processing, we used the biosppy and wfdb libraries. When classifying electrocardiograms, we used Multilayer Perceptron and Convolutional Neural Networks using TensorFlow technology.

The following main stages of biometric identification are follows: signal registration, signal preprocessing, biometric feature extraction, assessment of the informativeness of biometric features and selection of the most informative features (this is done to reduce dimensionality of input data), and classification of features. Consider each of the steps.

6. Registration of an electrocardiogram

Electrocardiogram (ECG or EKG [a]) is a graph of voltage versus time – of the electrical activity of the heart using electrodes placed on the skin. These electrodes detect the small electrical changes that are a consequence of cardiac muscle depolarization followed by repolarization during each cardiac cycle (heartbeat) (**Figure 3**).

From a technical point of view, Electrocardiographs are multichannel voltmeters that record electrical potentials in various areas of human surface. These devices differ in such characteristics as sampling frequency, bit depth, input voltage range, etc. A valuable resource for researchers in the field of analysis of biomedical signals is the website https://www.physionet.org/. PhysioNet is a repository of freely available medical research data, managed by the MIT Laboratory for Computational Physiology. The project is supported by the National Institute of General Medical Sciences (NIGMS) and the National Institute of Biomedical Imaging and Bioengineering (NIBIB) under NIH grant number 2R01GM104987-09.

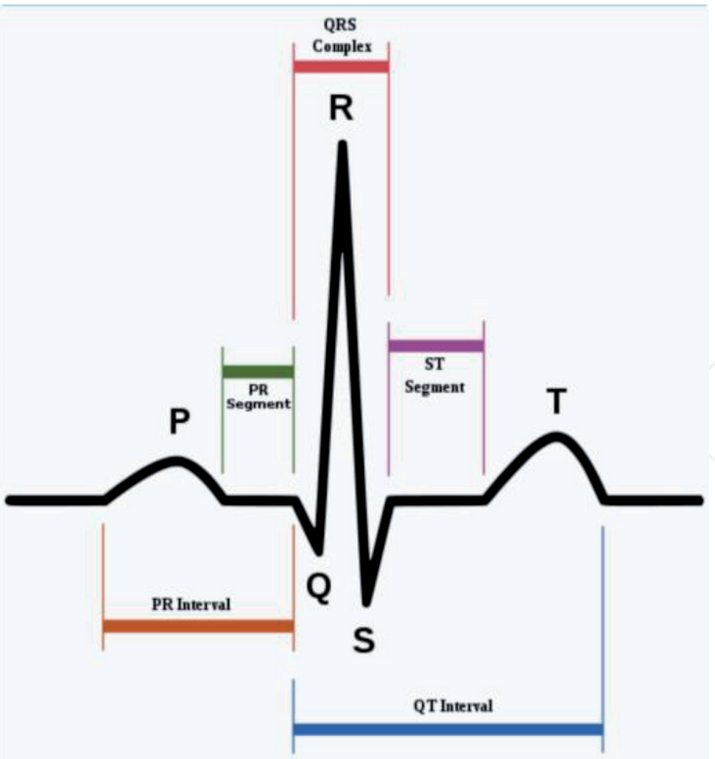

Figure 3.
ECG of a heart in normal sinus rhythm (https://en.wikipedia.org/wiki/Electrocardiography).

This site presents a large number of digitized electrocardiograms, for example, PTB Diagnostic ECG Database [10]

- 16 input channels,
- Input voltage: ±16 mV,
- Input resistance: 100 Ω (DC),
- Resolution: 16 bit with 0.5 μV/LSB (2000 A/D units per mV),
- Bandwidth: 0–1 kHz (synchronous sampling of all channels, time of registration is 3 minutes).

European ST-T Database [11]

- each record is two hours in duration and contains two signals,
- each sampled at 250 samples per second with 12-bit resolution over a nominal 20 millivolt input range.

ECG-ID Database [12, 13]

- ECG lead I, recorded for 20 s,
- digitized at 500 Hz with 12-bit resolution over a nominal ±10 mV range

As we can see, all the above databases used different electrocardiographs.

Previously, we investigated factors that influence the accuracy of biometric identification using an ECG. We have shown that the quality of an electrocardiograph affects the accuracy of biometric identification. Thus, the recognition accuracy during ECG classification using mixed Gaussian models of subjects from the ECG-ID database was 0.66, while for PTB this indicator was 0.8 [14].

7. Signal preprocessing

Signal preprocessing is carried out in order to reduce noise, reduce data dimension, find R-peaks, and cut ECG into cardiocycles. In this case, R-peak synchronization is usually performed. Noises are usually removed using a low-pass filter, while the cutoff frequency is selected experimentally. There is still no consensus on how best to find R-peaks. This is due to the fact that cardiocycles in different people are

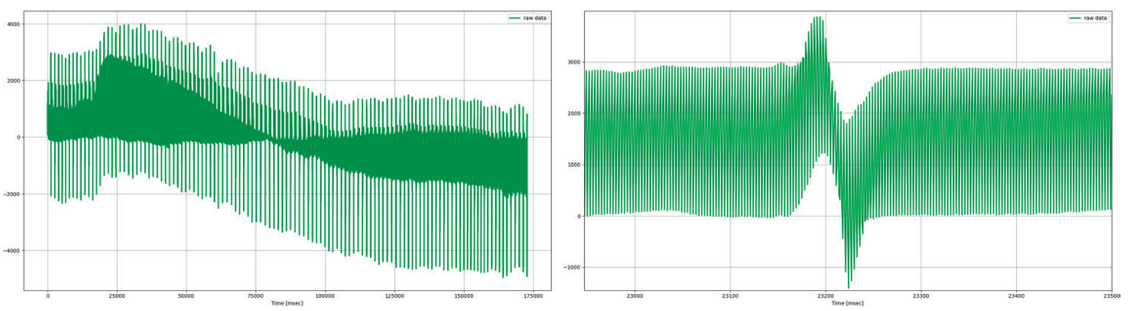

Figure 4.
The original digitized electrocardiogram has a high frequency.

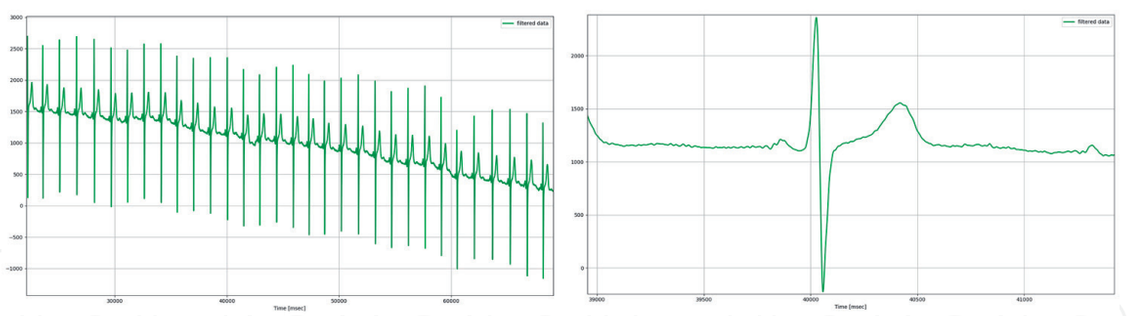

Figure 5.
Using the low-pass filter, the envelope of the cardiac signal is extracted.

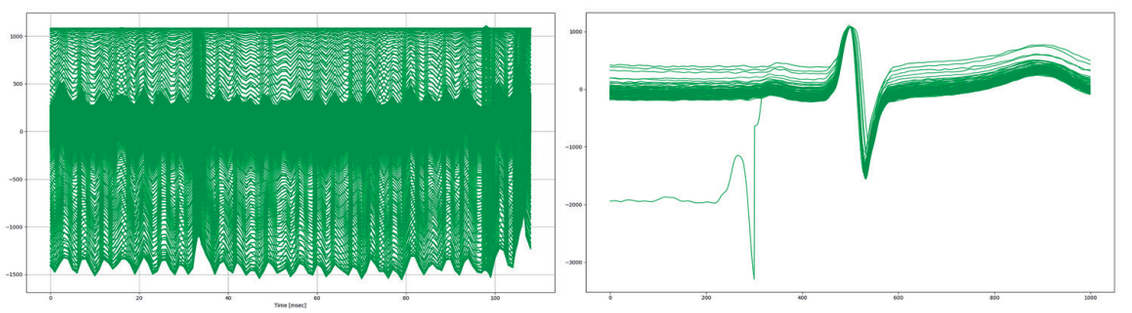

Figure 6.
Further, R-peaks are detected in the ECG, with their help the signal is cut into cardiocycles, after which the R-peak is synchronized.

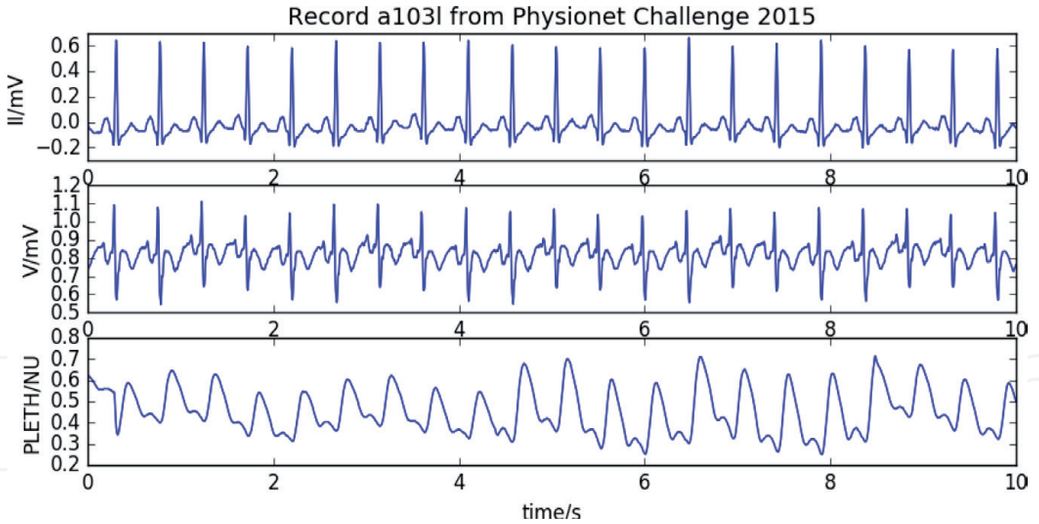

Figure 7.
Wfdb library example.

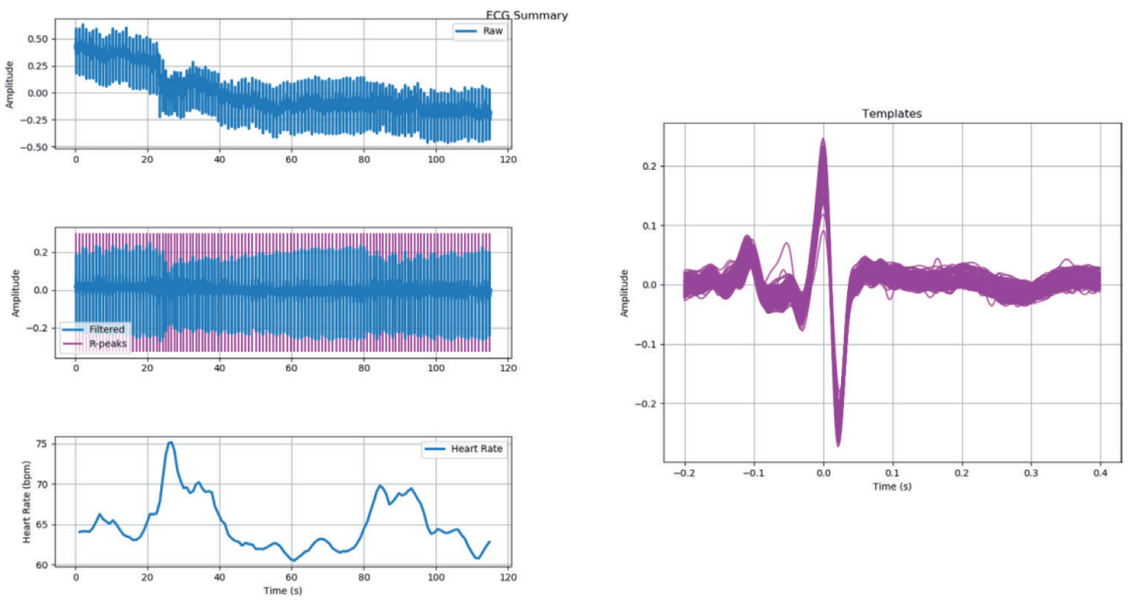

Figure 8.
Example of using the biosppy library.

distinguished by a rather high degree of great variability. In **Figure 4**, a digitized electrocardiogram is shown (sample rate is 1000 Hz).

To reduce noises, we can use a low-pass filter (**Figure 5**).

After noise removal, the procedure for finding R-peaks, slicing an ECG into cardiocycles and synchronizing cardiocycles by R-peaks follows (**Figure 6**).

There are several popular computer libraries for ECG preprocessing. Among them, libraries for the python wfdb [15] and biosppy [16] languages are very popular (**Figures 7** and **8**).

8. Biometric features extraction

There are several opinions as to which ECG features are best used for biometric identification. Some authors propose using the geometric characteristics of the cardiocycle, such as the amplitude and time characteristics of the cardiocycle peaks.

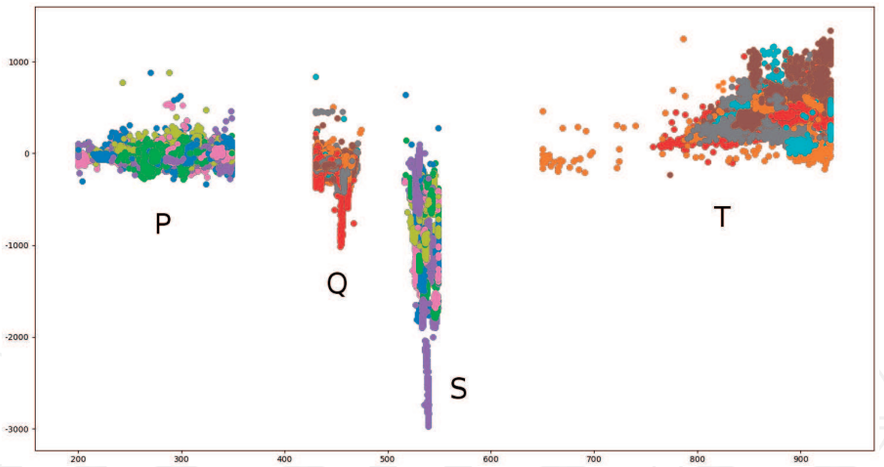

Figure 9.
Amplitude and time characteristics of P, Q, S, T peaks. Cardiocycles are synchronized in amplitude and time of onset of R-peak.

We can see a feature cloud in **Figure 9**, which are the amplitude and time characteristics of P, Q, S and T peaks of cardiac cycles.

Other authors suggest working with the frequency characteristics of the signal. For example, biometric features can be obtained using a discrete wavelet transform. Previously, we explored wavelets such as Haar wavelets, Daubechies wavelets (from db1 to db38), Symlets (from sym2 to sym20), Coiflets (from coif1 to coef17), Biorthogonal (from bior1.1 to bior6.8), Reverse biorthogonal wavelet (from rbio1.1 to rbio6.8), and Discrete Meyer (FIR Approximation) [17]. We have shown that wavelets such as Haar, Daubechies, and Symlets are best suited for biometric identification. We have shown that good results can be obtained if the entire cardiocycle is used as biometric features [18]. The number of features in this case depends on the sampling frequency of the signal. We used data from the following databases. 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, and in the case of St.Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database (sampling rate is 257 Hz), cardiocycles consist of 153 points.

9. Assessment of the informative value of biometric features and the selection of the most informative features

Experience shows that not all biometric features have the same information content. If you remove of uninformative features, you can significantly increase the speed of data processing. We investigated the informativeness of analytical features (amplitude and time characteristics of P, Q, S, T peaks) obtained from 51 subjects from the PTB database [19]. To do this, we determined the significance of differences between the clouds of points P, Q, S and T regions of the electrocardiograms of the subjects using Student's criterion at a significance level of 95%. Matrices of significance of differences are given below (**Figure 10**).

It can be seen from the figure that the overlap of the points is much smaller in the S and T regions. When using all eight signs together, the overlap of the points is not observed (**Figure 11**).

Conclusion: the most informative analytical features are the amplitude values in the S and T regions.

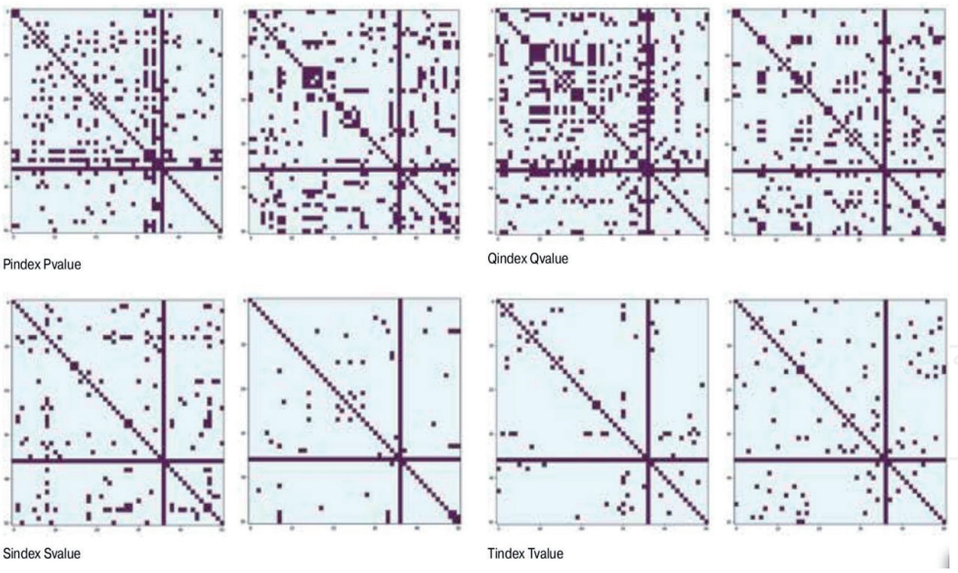

Figure 10.
Matrices of significance of differences according to 8 characteristics for 51 subjects (P value < 0.05). Note: in the bright areas of the figures, the differences are significant, in the dark areas – unreliable. The figures are symmetrical with respect to the diagonals passing through the upper left and lower right corners. The abscissa and ordinate axes show the numbers of subjects (1–51).

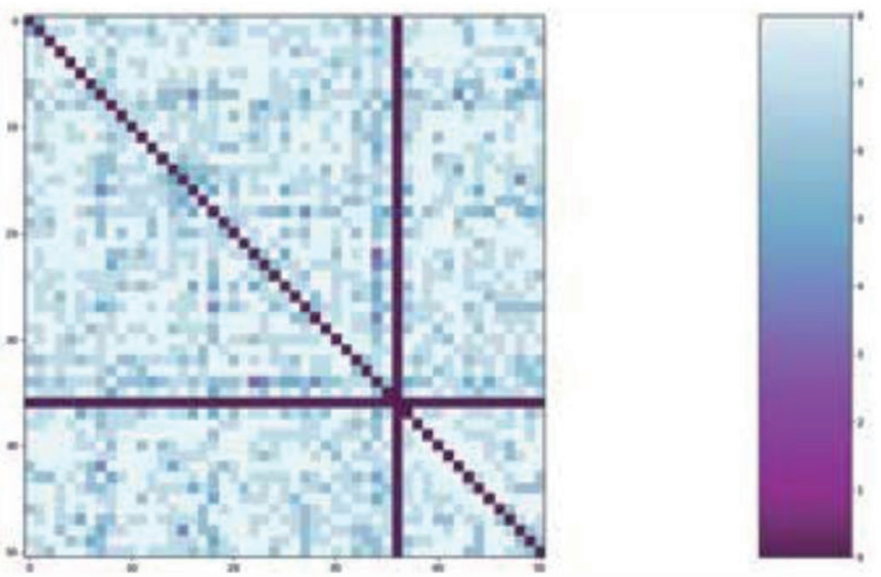

Figure 11.
Matrix of significance of differences when sharing eight features. Note: the color shows the number of cases of significance of differences from eight (the lightest area) to zero (the darkest area). The pattern is symmetrical with respect to the diagonal passing through the upper left and lower right corners. The abscissa and ordinate axes show the numbers of subjects (1–51).

10. Feature classification

The problem of person biometric identification concerns classification problems. To solve it, we have to consider algorithms from some finite set and choose an algorithm that gives the least error of the forecast. Let's introduce some notation. Let us suppose X is a space of objects.

Y is a set of answers.

$$X^l = (xi, yi)^l{}_{i=1} \tag{1}$$

is a training set, l is a sample size.

$$y_i = y^*(x_i), \tag{2}$$

$$A_t = \{a:X \to Y\} \tag{3}$$

are a model of algorithms, t ⊆ T, T is a number of algorithms under consideration.

$$\mu t:(X \times Y)^l \to A_t \tag{4}$$

are learning methods. It is required to find a method μ t with the best generalizing power.

When finding a method μ_t, we often have to solve the following subtasks:

- Choice of the best model A_t (model selection).
- Choice of learning method μ_t for a given model A_t (in particular, optimization of hyperparameters).
- Features selection:

$$F = \{f_j:X \to Dj:j = 1,\ldots,n\} \tag{5}$$

is a set of features. The method of learning μ j uses only features J ⊆ F.

It is used to assess the quality of learning by precedents.

$L(a, x)$ is a cost function of algorithm a on the object x.

$$Q(a, X^l) = \frac{1}{l}\sum_{i=1}^{l} L(a, x_i) \tag{6}$$

is a functional of accuracy a on X. In this case, we consider an internal quality criterion that is measured on the training set X^l:

$$Q_\mu(X^l) = Q(\mu(X^l), X^l) \tag{7}$$

and an external criterion evaluating the quality of learning on hold-out set Xk [2]:

$$Q_\mu(X^l, X^k) = Q(\mu(X^l), X^k) \tag{8}$$

Recognition accuracy will be affected by both the choice of the classification method and its implementation, in particular, the selection of hyperparameters. We tested 14 methods of Machine Learning for classification (Naive Bayes classifier for multivariate Bernoulli models, A decision tree classifier, An extremely randomized tree classifier, Classifier implementing the k-nearest neighbors vote, Label Propagation classifier, Linear Discriminant Analysis, Linear Support Vector Classification, Logistic Regression (aka logit, MaxEnt) classifier, Nearest centroid classifier, A random forest classifier, Classifier using Ridge regression, Ridge classifier with built-in cross-validation, and Gaussian Mixture Models, SVM) [20]. We found that the most accurate methods of classification are Label Propagation classifier (accuracy of recognition is 0.94), an extremely randomized tree classifier (accuracy is 0.92), and a Classifier implementing the k-nearest neighbors vote (accuracy is 0.90) [21].

By selecting model hyperparameters, it is possible to significantly increase recognition accuracy. So, in our previous study, it was shown that using the Support Vector Machine classifier for ECG classification uses as default following hyper parameters: C = 1.0, kernel = "rbf," gamma = "auto." When using of default parameters while performing of classification of electrocardiograms, we had an accuracy score equal to 0.93. We tuned hyper parameters of classification with Grid Search procedure varying C parameter in range of [1, 10, 100, 1000], kernel in range of ["linear," "rbf"], and gamma in range of [1e-3, 1e-4]. After performing of tuning, we had the following best parameters set: "kernel": "rbf," "C": 10, "gamma": 0.001. Using these parameters, we had an accuracy score equal to 0.99.

11. Conclusion and perspectives

Traditional password-based authentication methods have a number of disadvantages related primarily to the human factor. Biometric methods of identification and authentication are much more reliable, although they have some disadvantages. Some of them (fingerprints, retina, and voice) were compromised. It is not clear what to do if hackers gain access to a biometric database, because a person cannot change fingerprints as easily as a forgotten password. The development of wireless technologies and technologies of the Internet of medical things makes possible the emergence of new biometric identification scenarios. Here, first of all, I would like to note the biometric authentication of the patient in the Body area network. In this case, ECGs are not used to generate biometric keys confirming the patient's authenticity [22]. The second important area is contactless ECG recording using ultra-wideband radars.

Acknowledgements

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

Author details

M.R. Bogdanov[1,2*], A.S. Filippova[2], G.R. Shakhmametova[1] and Nikolai N. Oskin[3]

1 Ufa State Aviation Technical University, Ufa City, Russia

2 M.Akmullah Bashkir State Pedagogical University, Ufa City, Russia

2 Siberian Telecommunication Company, Moscow, Russia

*Address all correspondence to: bogdanov_marat@mail.ru

References

[1] Fingerprint Technology. Making Two Systems Work as One. The web site of Federal bureau of Investigation [Internet]. Available from: https://archives.fbi.gov/archives/news/stories/2010/july/fingerprints [Accessed: 07 January 2010]

[2] Seto Y. Retina recognition. In: Li SZ, Jain A, editors. Encyclopedia of Biometrics. Boston, MA.: Springer; 2009

[3] What is Voice Biometrics and why should you use it? The web-site of ID R&D Co. [Internet]. Available from: https://www.idrnd.ai/voice-biometrics/

[4] Software for fingerprint spoof and liveness detection. The web-site of Precise Biometrics Co. [Internet]. Available from: https://precisebiometrics.com/products/fingerprint-spoof-liveness-detection/software/

[5] Lai CL, Tai CY. A smart spoofing face detector by display features analysis. Sensors. 2016;**16**:1136. DOI: 10.3390/s16071136

[6] Evans N, Kinnunen T, Yamagishi J, Wu Z, Alegre F, De Leon P. Anti-spoofing for speaker recognition. In: Marcel S, Nixon MS, Li SZ, editors. Handbook of Biometric Anti-Spoofing. Trusted Biometrics under Spoofing Attacks; 2014

[7] XAVER™ 800. High performance imaging system. The web-site of Camero, part of the SK Group. Available from: https://www.camero-tech.com/xaver-products/xaver-800/

[8] Xiaolin L, Jianqin D, Hao Z, Thomas AG. Ultra-wideband impulse radar through-wall detection of vital signs. Scientific Reports. 2018;**8**:13367. DOI: 10.1038/s41598-018-31669-y

[9] Chocorresponding H-S, Park Y-J. Detection of heart rate through a wall using UWB impulse radar. Journal of Healthcare Engineering. 2018;**2018**:4832605. DOI: 10.1155/2018/4832605

[10] Bousseljot R, Kreiseler D, Schnabel A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik. 1995;**40**(1):317

[11] Taddei A, Distante G, Emdin M, Pisani P, Moody GB, Zeelenberg C, et al. The European ST-T database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. European Heart Journal. 1992;**13**:1164-1172

[12] Lugovaya TS. Biometric human identification based on electrocardiogram [Master's thesis]. Faculty of Computing Technologies and Informatics, Electrotechnical University "LETI", Saint-Petersburg, Russian Federation; 2005

[13] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation. 2003;**101**(23):e215-e220

[14] Bogdanov MP et al. Factors influencing accuracy of biometrical personal identification based on cardiograms. Pattern Recognition and Image Analysis. 2018;**28**(3):421-426

[15] The WFDB Python Toolbox [Internet]. Available from: https://pypi.org/project/wfdb/

[16] BioSPPy is a toolbox for biosignal processing written in Python [Internet]. Available from: https://biosppy.readthedocs.io/en/stable/

[17] Bogdanov M et al. Processing of biomedical data with machine learning. Atlantis highlights in computer sciences.

21st International Scientific Workshop on Computer Science and Information Technologies (CSIT 2019). 2019;**3**:6-16

[18] Bogdanov M et al. Increasing security of telemedicine service. Atlantis highlights in computer sciences. 21st International Scientific Workshop on Computer Science and Information Technologies (CSIT 2019). 2019;**3**:162-165

[19] Bogdanov MP et al. Statistical assessment of informativeness of biometric features extracted from electrocardiograms. Russian Cardiological Journal. 2018;**23**(7):84-91

[20] Bogdanov MR et al. Diagnosis of heart diseases with machine learning. Journal of Mathematics and Statistical Science. 2019;**5**:81-84

[21] Bogdanov MR et al. Optimizing Factors Influencing on Accuracy of Biometrical Cardiometry. In: Belim S, editor. Biometrical Cardiometry. Omsk, Russia, published at http://ceur-ws.org: OPTA-SCL; 2018. pp. 61-64

[22] Nima K, Zimu G, Fatemeh T, Damon W, Mark T, Domenic F. Secure and Reliable Biometric Access Control for Resource-Constrained Systems and IoT. Available from: https://www.researchgate.net/publication/324055574_Secure_and_Reliable_Biometric_Access_Control_for_Resource-Constrained_Systems_and_IoT