

UDC 616.12-073.7

Myocardial Ischemia Detection Using a Reduced Number of ECG Leads

Mnevets A. V., Ivanushkina N. G., Ivanko K. O.

Electronic Engineering Department, Faculty of Electronics,
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

E-mail: amnevets-ee22@iit.kpi.ua, niva-ee@iit.kpi.ua, ivanko-ee@iit.kpi.ua

The study is devoted to the investigation of the electrocardiographic (ECG) features to distinguish norm and myocardial ischemia in reduced set of electrocardiographic leads. In particular, for myocardial ischemia detection the spectral features of the electrocardiographic signal and characteristics of the shape of ECG waves are considered. The main features commonly used for myocardial ischemia detection are described in the paper, as well as more reliable analogs are proposed for the considered task. The approach for ECG signal preprocessing, identification of the necessary signal segments and subsequent calculation of features is described in detail. The considered features are based on the areas under the characteristic waves of the ECG signal and the spectral distribution of these waves. The most informative features for myocardial ischemia detection are identified and selected from the initial set of parameters which led to a two-fold reduction in number of ECG leads comparing to the standard 12-lead electrocardiogram. The techniques for determining the proposed features, namely the ratio of the area under T wave to the area under the P wave, as well as the ratio of the area under T wave to the area of the entire cardiac cycle, are considered. These features together with other calculated parameters are assumed to describe the majority of pathology cases and gave a high accuracy of the classification ECG to norm and ischemic myocardial diseases since they reflect the bioelectrical processes that occur in the presence of myocardial ischemia and manifest themselves on the surface ECG. Based on the analysis of principal components and the method t-distributed stochastic neighbor embedding, the distribution of data in the space of features that characterize the classes of norm and pathology was shown. Raw ECG data in norm and with cases of myocardial ischemia were obtained from the "PTB Diagnostic ECG Database" used in "The PhysioNet/Computing in Cardiology Challenge 2020". This database contains 22353 ECG records from 290 persons with 12 ECG leads (I, II, III, aVR, aVL, aVF, and V1–V6). The database contains the high-resolution ECG signals, which enabled to obtain 10,000 cardio cycles presenting norm and myocardial ischemia pathology for the subsequent training the machine learning algorithms. Based on the obtained features, various machine learning algorithms were trained and the accuracy was compared on different combinations of ECG leads. As a result of cross-validation, the accuracy of myocardial ischemia detection was 99% with a standard deviation of 0.4% for 6 leads (I, II, III, AVR, AVL, AVF) and 93% with a standard deviation of 0.12% for one lead (I). Thus, it was shown, that with machine learning methods it is possible to recognize ischemic myocardial disease with high accuracy and stability using six standard ECG leads or only one ECG lead.

Keywords: myocardial ischemia; machine learning; cardiocycle; cardio interval; wavelet analysis; area T wave; cross-validation; heart diseases detection

DOI: [10.20535/RADAP.2022.89.39-47](https://doi.org/10.20535/RADAP.2022.89.39-47)

Introduction

Heart diseases are the main reason for mortality in the world [1] and myocardial ischemia and infarct are among them.

To avoid severe consequences, it is necessary to find out the initial reason of a heart attack [2]. Especially it is extremely important to reveal the initial signs of myocardial ischemia in the early stages. So automatic detection of myocardial infarct and ischemia is an important part of timely identification of pathology and saving human lives.

Today leading IT companies develop smart technologies for healthcare, which eventually take the format of portable devices. Nowadays portable technologies and devices give an opportunity to register ECG at home, calculate heart rate and heart-rate variability [3]. So it makes it possible to detect ischemia and myocardial infarct before occurring of obvious symptoms. Today for detection of myocardial infarct and heart ischemia six ECG leads are commonly used or three leads at least. So it makes sense to detect this pathology with a lower number of ECG leads with the use of machine learning methods.

The reason for myocardial infarction is a shortage of oxygen in the myocardium with the subsequent death of heart muscle cells. The shortage of oxygen can be caused by obesity, diabetes mellitus, blood pressure disorders, smoking, alcoholism, drug addiction, peripheral vascular disease. Also, these pathologies can be manifestations of respiratory diseases, complications or viral infections such as COVID-19, and others. The big amount of possible causes may complicate immediate diagnosis, so it is very important to detect myocardial ischemia and infarction automatically, using only data that can be measured (like ECG, blood pressure, photoplethysmogram etc).

Big amount of biomedical data published in world-famous resources gives an opportunity for developers to create new ideas and improve methods of diagnosis and treatment. Therefore machine learning is a promising way to improve disease recognition.

More often for myocardial ischemia detection a doctor makes a decision of myocardial ischemia type, localization and degree using a 12-lead ECG, depending on a lead where pathology manifests itself most clearly [4]. But for urgent diagnostics (for fast detection myocardial infarction) much fewer leads can be used, which can give the opportunity to detect myocardial ischemia using portable ECG devices.

In recent years there are many types of research linked with automatic heart disease detection in particular myocardial ischemia and infarction [6]. There are several verified algorithms that can detect myocardial ischemia using 12 and 6 leads [21].

The study is focused on the automated detection of myocardial ischemia using one, two, or three ECG leads. Manifestations of myocardial ischemia on the ECG can be found in standard, amplified, and chest leads [25].

The main aim of the research is to build reliable models for recognizing ischemic myocardial disease using machine-learning methods including decision trees, random forest classifier, support vector machines, k -nearest neighbors' classifier, and gradient boosting classifier. Principal component analysis and t -distributed stochastic neighbor embedding are used as methods for visualizing high-dimensional data.

Machine learning methods are applied to solve the problem of decreasing the number of the ECG leads used for myocardial infarction detection. The expected outcome of the proposed research is accurate detection of myocardial ischemia and infarction using one, two or three ECG leads.

The first stage of the study is preparing the dataset with the sufficient number of the norm and ischemic myocardial disease cases. The next step is determining the set of the most informative features for classification, which allow us not only to distinguish between norm and pathology but also to decrease the number of the used ECG leads. The final step is training and test-

ing of the developed models based on the considered machine learning methods and features.

1 Materials and methods

For the study we used the publicly open "PTB (Physikalisch-Technische Bundesanstalt) Diagnostic ECG database" from the PhysioNet/Computing in Cardiology Challenge 2020 [5]. This database contains 22353 ECG records from 290 persons with 12 conventional ECG leads (I, II, III, aVR, aVL, aVF, and V1–V6) together with 3 Frank leads (vx, vy, vz).

The average length of signals is 120 seconds. Age range of the patients is from 17 to 87 years, with mean value of 57.2. Among the patients 209 men (average age is 55.5 years) and 81 women (average age is 61.1 years). Signals are digitized at 1000 samples per second, with 16-bit resolution over a range of ± 16.384 mV. The distribution of illness shown in the Table 1.

Table 1 DISTRIBUTION OF ILLNESS IN THE DATABASE

#	Diagnostic class	Number of subjects
1	Myocardial ischemia	148
2	Heart failure	18
3	Bundle branch block	15
4	Dysrhythmia	14
5	Myocardial hypertrophy	7
6	Valvular heart disease	6
7	Myocarditis	4
8	Miscellaneous	4
9	Healthy controls	52

The result of the analysis of diagnostic classes shows that the biggest amount of illnesses in the database takes myocardial ischemia. 148 subjects with myocardial ischemia provide data representative enough for building the machine-learning models for myocardial ischemia detection. In the study we used only myocardial ischemia and healthy controls diagnoses. The imbalance between the "normal" and "ischemia" classes was eliminated by equalizing the set of observations. Each ECG record was split into cardiocycles. Each observation in the training and testing datasets was formed by the features extracted from the currently considered cardiocycle. The data sets were balanced by removing extra observations in a random order. Thus, we received 10,000 observations for the each class of norm and pathology.

Detection of myocardial ischemia is usually based on the evaluation of the following parameters [6]:

- ST segment elevation;
- QRS complex changes;
- changes in T wave;
- changes in QT interval;
- elevation angle of the ST segment;

- elevation or depression of Q-onset;
- tangent of the elevation angle of the ST segment;
- ST segment area;
- other parameters presented in Fig. 1.

The mentioned above parameters are shown in Fig. 1, where all the deviations are calculated relatively to the isoelectric line. As we can see, the main features for the ischemia detection are the parameters, associated with the ST segment of ECG.

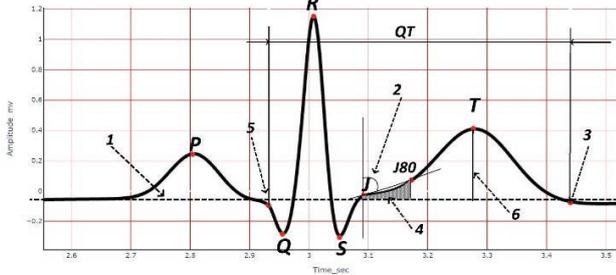


Fig. 1. Basic features for ECG myocardial ischemia detection, where 1 — isoelectric line, 2 — elevation angle of the ST segment, 3 — T wave offset, 4 — ST segment area, 5 — Q onset, 6 — T wave amplitude

Detection of myocardial ischemia using the considered features has low accuracy due to the problem of accurate detection of the angles, QRS points, slopes, shifts, elevations [22]. Also the problem of accurate definition of characteristic points of the cardiocycles is a separate difficult task. Therefore to improve reliability of the myocardial ischemia detection, we should choose relative and integral indicators.

To bring the parameters of the different cardiocycles of different patients into one range of values, it was decided to use the primary processing implemented by the scheme presented in Fig. 2.

ECG signals were filtered using a 5th order low-pass Butterworth filter in the bandwidth of the necessary ECG components 0-30 Hz. The removal of the baseline wander was carried out by high-frequency filtering with the cutoff frequency 0.01 Hz. The delay after filtering the baseline drift with the Butterworth filter was eliminated using the forward-backward filtering method [24]. Thus, the filtered signal has no delay relative to the original. To increase the signal to noise ratio (SNR) in noise recordings, the averaging of the consecutive cardiocycles can be also applied.

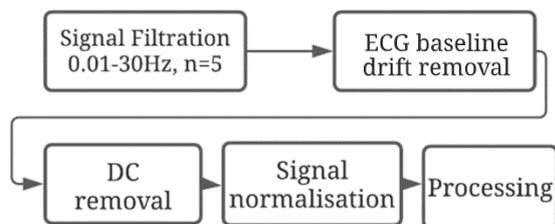


Fig. 2. ECG signal standardization procedure

The removal of the DC component is carried out by subtracting the average value from the signal. Signal normalization is done by dividing each sample by the maximum value in the sample:

$$y_{i_{new}} = \frac{y_i}{MAX[y_1, y_2, \dots, y_n]},$$

where y_i — original value, $y_{i_{new}}$ — new normalized value.

Also, when normalizing a long signal, noise or spikes in amplitudes can destroy the normalization result. To avoid distortions, the maximum value is taken as a value in the range from 0.01 to 0.99 quantiles of the distribution of signal sample values:

$$y_{i_{new}} = \frac{y_i}{MAX(Quantile(0.01-0.99)[y_1, y_2, \dots, y_n])}.$$

This method can give an unexpected spread in the calculated parameters, therefore, in short sections of the signal, it is permissible to use normalization by the maximum value [7].

The area under the curve of the ST segment was calculated using the trapezoidal rule method [8], where the total area is calculated as the sum of elementary areas taken after a certain period of time Δt .

To evaluate the ST segment area contribution, a parameter calculated as the ratio of the area under the ST segment to the area under all the waves of the QRS complex was used.

To express the balance of the phases of depolarization and repolarization in the ECG signal, we used a parameter calculated as the ratio of the area under the T wave to the area under the P wave.

Also, the contribution of the T wave can be evaluated as the spectral power of T wave. Thus, a pathological increase or decrease in the contribution of ventricular repolarization to the spectral power of the signal shows the development of the pathological activity in the myocardium.

To increase the stability of the considered parameter, the spectral power of the ST segment was normalized in its frequency range to the spectral power of the filtered ECG signal in its entire frequency band. Thus, the spectral contribution of the power of the ST segment to the entire signal can be expressed by the following formula:

$$P_{ST} = 10 \left(\frac{P_{spec}(f_1, f_2)}{P_{spec}(0, f_s/2)} \right),$$

where f_1, f_2 are the boundaries of the spectral range of the ST segment, P_{spec} is sum of the power spectrum values over a certain frequency range, $f_s/2$ is the Nyquist frequency [9].

To obtain the spectral characteristics of the signal, and to identify the spectral range of the T wave, it was decided to use also a continuous wavelet transform (CWT) as a suitable approach to extracting the power in frequency bands of the EEG signal [26]. CWT takes advantage of the fact that in the EEG signal low-frequency signals are propagated over time while high-frequency components occur at short intervals. To

decompose the ECG signal into spectral components, the Morlet wavelet was used, which makes it possible to more accurately estimate the frequency distribution of the ECG components on the scalogram [10]. To improve the temporal resolution of the wavelet, the sigma parameter, which is the scaling parameter that affects the width of the Morlet wavelet window, was chosen equal to 2.

Figure 3 shows the CWT scalogram of the ECG signal in the norm. It can be noted that in the time domain of the T wave, there is a power increase of the spectrum in the range of 2-8 Hz. This range is the range of the contribution of the T wave to the spectral power of the signal.

Thus, for each cardiocycle the following list of the features was calculated:

- area under the T wave;
- the ratio of the area under the T wave to the area under the cardiac cycle;
- the ratio of the area under the T wave to the area under the P wave;
- the ratio of the spectral power of a T wave to the total spectral power of the cardiac cycle.

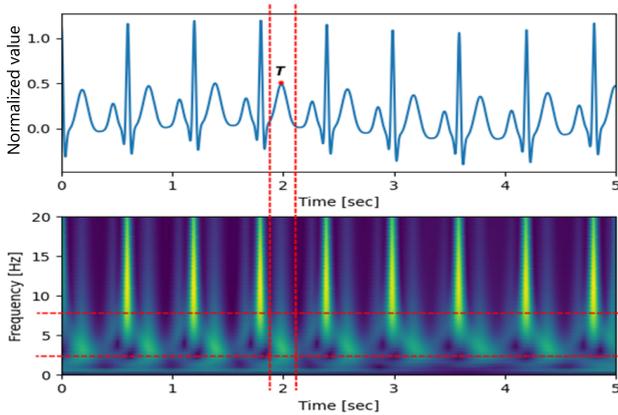


Fig. 3. Time-spectral representation of the ECG signal in norm using the Morlet wavelet, and detection of the frequency range of the T wave

According to the methods mentioned above, preprocessing and normalization of the ECG signal was carried out, which allows obtaining the features for myocardial ischemia detection in the expected and acceptable range of values. The resulting set of the features should include the above parameters for each analyzed lead. So for 12 leads $12 \times 4 = 48$ features, for 6 leads – 24 features for each cardio cycle are obtained.

The considered PTB Diagnostic ECG database contains ECG recordings of some patients with overlapping pathologies, for example, one patient may have a myocardial infarction and one type of arrhythmia, or even one or more comorbidities. For more accurate training of the models for myocardial ischemia detection, it was decided to drop out recordings of the patients, who have other comorbidities apart from myocardial ischemia.

For the separation of cardiac cycles, an ECG detector was used, which detects the position of the R peaks according to the Pan–Tompkins algorithm [20]. Detection of the redundant peaks in case of R wave splitting is excluded by an additional check of the RR interval length. Location of the characteristic points of P, T waves in ECG was calculated according to the scheme in Fig. 4. The intervals $d1$ - $d6$ are used in identifying the necessary components of the QRS complex.

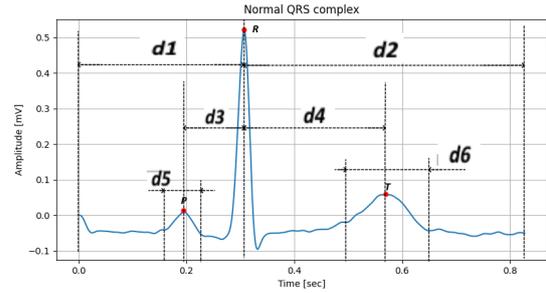


Fig. 4. Splitting the ECG signal into characteristic intervals and the allocation of P and T waves

After detecting the R peaks in the ECG, the cardiocycle selection procedure takes place, where starting and ending points of the cardiocycle are determined in accordance with the intervals $d1, d2$, where $d1 = 0.5 \cdot RR$, $d2 = 0.95 \cdot RR$, where RR is the interval between adjacent R peaks.

The next step is the determination of the P and Q peaks. The approximate location of the P and Q peaks is determined at a distance $d3, d4$ from the R peak, respectively. Next, the positions of the P and Q waves are defined by finding the maximum values in the intervals $d5$ and $d6$. At this stage, the intervals $d3 - d6$ are calculated as: $d3 = 0.12 \cdot RR$, $d4 = 0.3 \cdot RR$, $d5 = 0.07 \cdot RR$, $d6 = 0.1 \cdot RR$.

The ranges equal to $d5/2$ and $d6/2$, respectively, are set aside in two directions from P and T peaks and the minima for each of the ranges are searched and these points are taken as the edges of the P and T waves. Relative to the position of the P and T peaks as well as their boundaries, the required areas are calculated as features for myocardial ischemia recognition.

Thus, an intermediate database that contains normal and pathological cardiocycles in the case of myocardial ischemia was formed. As a result of the signal processing 70,000 cardiocycles were obtained: 60,000 cases of pathology and 10,000 cases of norm. To avoid skew in the characteristics of sensitivity and specificity for the pathology and make equal the size of the two classes, data were mixed and 50,000 of cardiocycles were randomly removed from the pathology group, thereby we obtained a balanced dataset of cardiocycles with normal and pathological states.

2 Machine learning for detection of myocardial ischemia with searching for the most informative combinations of features

After calculating the cardiocycle's features in leads I, II, III, AVR, AVL, AVF, V1-V6, the recursive feature extraction was carried out in order to reduce the number of features without losing a significant amount of information. Support vector machine classifier was used and expected reduction in features was from 48 to 12.

The following features were identified as the most important after the recursive feature extraction:

- ratio of the spectral power of the T wave to the total spectral power of the cardiac cycle (lead I);
- ratio of the area under the T wave to the area under the P wave (lead I);
- ratio of the area under the T wave to the area under the cardiac cycle (lead I);
- area under the T wave (lead II);
- ratio of the spectral power of the T wave to the total spectral power of the cardiac cycle (lead II);
- area under the T wave (lead III);
- ratio of the spectral power of the T wave to the total spectral power of the cardiac cycle (lead III);
- area under the T wave (lead III);
- the ratio of the spectral power of the T wave to the total spectral power of the cardiac cycle (lead AVL);
- the ratio of the spectral power of the T wave to the total spectral power of the cardiac cycle (lead AVF);
- the ratio of the area under the T wave to the area under the cardiac cycle (lead AVF);
- area under T wave (lead AVF).

As we can see, all the features that were determined by the recursive feature extraction algorithm relate to the first three leads (I, II, III), and enhanced leads (AVF, AVL), and the largest number of informative features among all the selected features belongs to the first three leads, of which it follows that it makes sense to reduce the number of features by reducing the number of analyzed leads from 12 to 6, removing leads V1-V6 and leaving I, II, III, AVR, AVL, AVF. Thus, the number of features was reduced from $12 \times 4 = 48$ to $6 \times 4 = 24$.

To visualize the differences between norm and myocardial ischemia groups, as well as to analyze the distinguishability of the two groups, a principal component analysis (PCA) was carried out [12]. As a result of PCA analysis, the number of features was reduced from 24 to 3 for the possibility of visualization in three-dimensional space. In Fig. 5 the results of the PCA method and visualization in three-dimensional space are presented.

Visualization of the principal components in three-dimensional space shows that the two groups have the sets of features that can be quite clearly separated from each other; however, clusters of points belonging to a certain class are not far apart in the feature space. The points in the class "norm" are scattered more widely in feature space than in the class "pathology".

After reducing the number of features using PCA analysis and obtaining the dispersion coefficient [23], we found that the total variance of the three principal components set does not exceed 35%. This shows that three principal components are not enough to describe the total variance ratio of the twenty-four feature set.

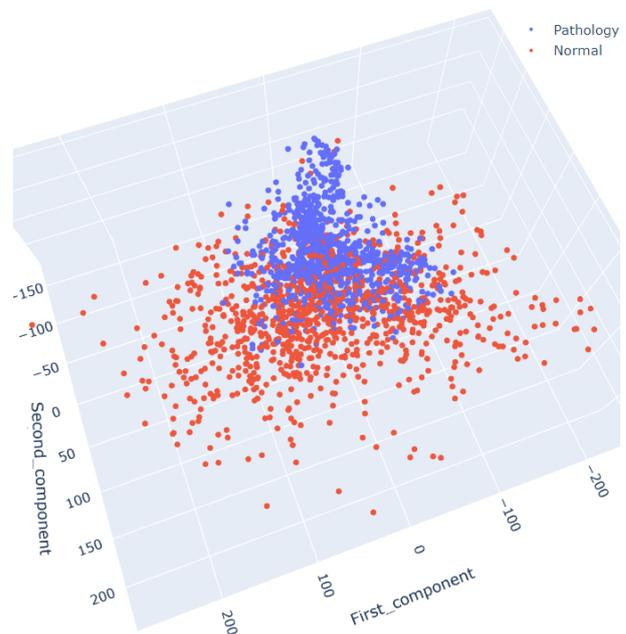


Fig. 5. Visualization of 3 principal components (myocardial ischemia is shown in blue and norm in red)

Other way to visualize the norm and myocardial ischemia groups is the t-distributed stochastic neighbor embedding (TSNE) method, which non-linearly projects a set of 24 features into a three-dimensional space (Fig. 6).

After applying the TSNE analysis, we see that the distribution of the points for the class "norm" partially intersects with the distribution of the points for the class "pathology". But at the same time, they also form distinct clusters that can describe significant differences between classes. Some pathology cases are located at a distance from the majority of the other pathology cases and form separate small islands, which may indicate additional undetected pathologies.

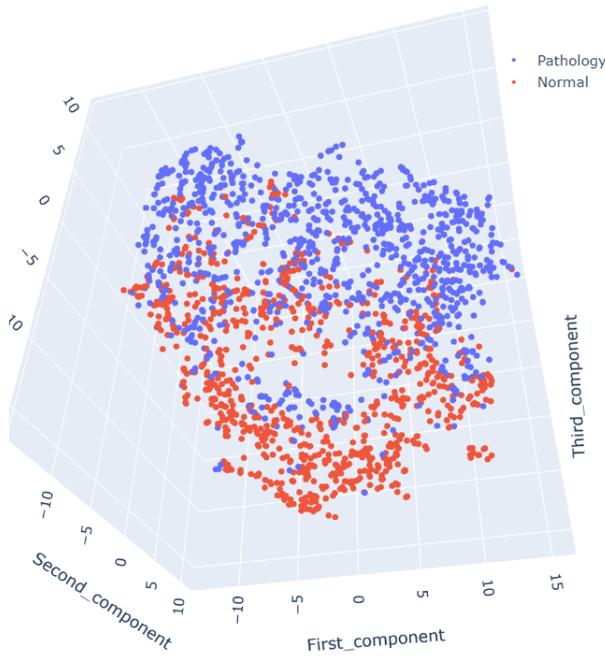


Fig. 6. Visualization of the features set after TSNE analysis (myocardial ischemia is shown in blue and norm in red)

To train the algorithms of machine learning, the dataset was divided into two parts - training set (80%) and test set (20%) with random data mixing.

To determine the best classification models for myocardial ischemia detection, we implemented supervised machine learning based on logistic regression, discriminant analysis, k-nearest neighbors method, support vector machines (polynomial, Gaussian, sigmoidal kernels), decision trees, random forest classifiers, gradient boosting [15–17]. Each method was evaluated on a different set of features to assess the possibility of reducing the number of leads.

To reduce the number of features at first we should pay attention to the possibility of excluding the enhanced leads, since they correlate with standard leads as obtained by their combinations. Then we evaluated the performance of classification algorithms for decreasing number of the standard leads.

The main purpose of such a reduction is to check whether the classification algorithm can show classification ability using only one, two or three ECG leads. The evaluation of the classification algorithms shown in the Table 2.

As a result of analyzing the accuracy of the classification results of algorithms on different sets of features we found, that with a feature set from the first three leads and amplified leads, classification accuracy is 99% with using the gradient boosting method and 98% with the random forest method. Using of only three standard leads shows a slight decrease in accuracy comparing to 6 leads, but still remains to be high. With a decreasing in the number of leads, removing the features associated with amplified leads, we found that the maximum accuracy for gradient boosting and

the random forest classifier dropped by only 1% and amounted to 98% and 97%, respectively. When removing the third lead from the set of features, we also got a drop in classification accuracy by 1%, at which the accuracy for gradient boosting was 97% and for the random forest classifier was 96%. But when switching from using two leads to just one, the classification accuracy deteriorated stronger, by 4% on average and amounted to 92% for gradient boosting and 93% for the random forest classifier.

The dependence of the classification accuracy of different algorithms on the number of features varies from 50% to 99% accuracy throughout the study. This fact demonstrates that the features have a complex distributed structure, and the groups are at a close distance from each other in the feature space. This can be seen on PCA and TSNE analyses distributions, therefore, linear classification methods showed low accuracy. Regression models showed the lowest accuracy. The best classifiers for the task of distinguishing between norm and myocardial ischemia are gradient boosting, random forest classifier, decision trees, and k-nearest neighbors classifier.

Analysis of the classification accuracy shows that when moving from two leads (I, II) to one (I), the drop in classification accuracy is the strongest. This indicates that II lead contains important information for classification. After reducing the leads number from six (I, II, III, AVR, AVL, AVF) to three (I, II, III), a weaker decrease in accuracy is observed, in spite of a sharp halving in the number of features. Reducing the leads number from three (I, II, III) to two leads (II, III) demonstrates the smallest drop in accuracy, which suggests that either II, III leads carry similar information for classification or lead III contains little information for detecting ischemic illness.

Two best estimators, random forest classifier and gradient boosting classifier, were also combined in a stacking algorithm. As a result, a classification accuracy of 98.6% was obtained, which is slightly lower than with the direct use of gradient boosting.

To compare the stability of the models on a set of features from (I, II, III, AVR, AVL, AVF), a five-step cross-validation was carried out [19], which gave the following results:

- for the stacking algorithm, 99% accuracy was obtained with a standard deviation of 0.004% (0.9858, 0.9819, 0.98645, 0.98795, 0.98705)

- for the gradient boosting algorithm, 99% accuracy was also obtained with a standard deviation of 0.0047% (0.98856, 0.98253, 0.98525, 0.9858, 0.9834)

- for stacking and direct random forest use in the case of using a single lead (I), cross-validation results show the average accuracy of 93% with the same low standard deviation of 0.0012%.

As a result, we can conclude that the obtained models are quite stable on different combinations of data, and demonstrate high accuracy with a low standard deviation.

Table 2 EVALUATION OF THE CLASSIFICATION ALGORITHMS PERFORMANCE DEPENDING ON THE NUMBER OF LEADS SELECTED (TOTAL CLASSIFICATION ACCURACY AND TRUE POSITIVE RATE FOR NORM AND MYOCARDIAL ISCHEMIA CLASSES IN PARENTHESES)

Machine learning method	I, II, III, AVR, AVL, AVF	I, II, III	I, II	I
Logistic regression	0.782 (0.792, 0.772)	0.755 (0.791, 0.7189)	0.737 (0.751, 0.723)	0.736 (0.745, 0.727)
Discriminant analysis	0.778 (0.787, 0.768)	0.749 (0.799, 0.698)	0.733 (0.765, 0.700)	0.732 (0.755, 0.709)
k-nearest neighbors (n=3)	0.945 (0.913, 0.977)	0.942 (0.915, 0.968)	0.941 (0.898, 0.983)	0.881 (0.856, 0.906)
Support vector machines (polynomial kernel)	0.868 (0.792, 0.944)	0.821 (0.707, 0.935)	0.802 (0.628, 0.976)	0.724 (0.499, 0.949)
Support vector machines (Gaussian kernel)	0.914 (0.996, 0.832)	0.957 (0.970, 0.944)	0.948 (0.944, 0.952)	0.884 (0.810, 0.958)
Decision tree (depth=4)	0.830 (0.735, 0.925)	0.812 (0.733, 0.891)	0.822 (0.690, 0.954)	0.778 (0.671, 0.885)
Decision tree (depth=7)	0.883 (0.828, 0.938)	0.882 (0.817, 0.947)	0.870 (0.763, 0.976)	0.817 (0.737, 0.897)
Decision tree (depth=10)	0.914 (0.871, 0.957)	0.909 (0.862, 0.956)	0.908 (0.832, 0.984)	0.861 (0.817, 0.904)
Random forest (10 trees in the forest)	0.972 (0.968, 0.975)	0.969 (0.971, 0.967)	0.961 (0.956, 0.966)	0.927 (0.908, 0.946)
Random forest (20 trees in the forest)	0.978 (0.979, 0.978)	0.974 (0.977, 0.971)	0.962 (0.958, 0.966)	0.927 (0.905, 0.950)
Random forest (30 trees in the forest)	0.983 (0.979, 0.987)	0.973 (0.975, 0.971)	0.962 (0.954, 0.970)	0.929 (0.898, 0.960)
Gradient boosting	0.988 (0.986, 0.99)	0.980 (0.982, 0.978)	0.972 (0.964, 0.980)	0.920 (0.885, 0.955)
Stacking	0.988 (0.986, 0.992)	0.980 (0.981, 0.979)	0.972 (0.965, 0.979)	0.920 (0.911, 0.929)

To visualize the errors of the algorithm, a confusion matrix [18] is shown in Fig. 7.

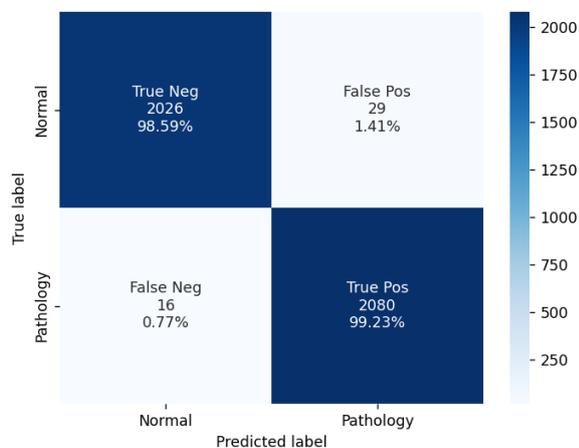


Fig. 7. Confusion matrix for the myocardial ischemia recognition algorithm for six leads I, II, III, AVR, AVL, AVF

As can be seen from the confusion matrix, the model has a fairly balanced number of true positive (99.23%), true negative (98.59%), and false positive (1.41%) and false negative (0.77%) predictions. As for confusions, the bias towards false positive predictions plays into our hands, since it is more important to predict the pathology and exclude it if necessary than to miss it. Sensitivity (0.985), Specificity (0.988), and F1 (0.987) scores indicate high classification accuracy and low type I and type II errors.

Conclusion

In this work, methods and approaches for identifying, applying and reducing features for classification of myocardial ischemia were investigated, and a study of the classification accuracy for different classifiers for different combinations of features was carried out. As a result, it was shown that using machine learning methods makes it possible to classify ischemic myocardial disease with 98% accuracy using

only three standard leads and 93% accuracy using only one lead. The methods for obtaining the features for myocardial ischemia detection were described in details. The technique for determining the proposed features, namely: the ratio of the area under T wave to the area under the P wave, as well as the ratio of the area under T wave to the area of the entire cardiac cycle, was considered. These features together with other calculated parameters gave a high accuracy of the classification of the norm and pathology since they arise from the bioelectrical processes that reflect on the surface ECG. Analysis of confusion matrices obtained for the best classifiers and feature sets showed a low probability of classification errors of the normal and pathology classes. The number of correct decisions of the obtained models turned out to be quite balanced for norm and myocardial ischemia cases. Therefore the probability of making type I and type II errors is low. PCA and TSNE analysis showed that the calculated features do not have sufficient spatial distance in the distribution between groups. The obtained results demonstrate the ability to detect the ischemic myocardial disease using the reduced number of leads. Obtained classification accuracy of 99% for 6 leads (I, II, III, AVR, AVL, AVF) and 93% for one lead (I) makes it possible to implement the obtained models and the proposed features in the medical devices.

References

- [1] Deal B. J. (2009). *Arrhythmias in congenital heart disease. Adult Congenital Heart Disease*, pp. 221–236.
- [2] Heart attack. (2022). *National Heart Lung and Blood Institute*.
- [3] Take an ECG with the ECG app on Apple Watch. (2022). *Apple Support*.
- [4] Glazunov M., Aranda A., and Galuzzi C. (2021). Optimal ECG Lead System for Automatic Myocardial Ischemia Detection. *2021 Computing in Cardiology (CinC)*. DOI: 10.23919/CinC53138.2021.9662763.
- [5] Wagner P., et al. (2020). PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data*, Vol. 7, Article number: 154. DOI:10.1038/s41597-020-0495-6.
- [6] Brundage J. N., et al. (2021). Myocardial Ischemia Detection Using Body Surface Potential Mappings and Machine Learning. *2021 Computing in Cardiology (CinC)*. DOI: 10.23919/CinC53138.2021.9662808.
- [7] Dash S. R., Sheeraz A. S., and Samantaray A. (2018). Filtration and Classification of ECG Signals. In Handbook of Research on Information Security in Biomedical Signal Processing. *IGI Global*, pp. 72–94. DOI: 10.4018/978-1-5225-5152-2.ch005.
- [8] Bohara R. (2014). Trapezoidal Method Algorithm and Flowchart. *CODEWITHC. The way to Programming*.
- [9] Kannan P., Maheswari S., Pon Bharathi A., and Wilson A. J. (2021). Spectral and performance measures analysis of ECG signal using various transforms and different types of iir and fir filters with different orders. *International Journal of Electrical Engineering and Technology (IJEET)*, Vol. 12, Iss. 5, pp. 96–108. DOI: 10.34218/IJEET.12.5.2021.009.
- [10] Verma A. K., Saini I., and Saini B. S. (2018). The baseline wandering noise removal from ECG signal using forward-backward Riemann Liouville fractional integral-based empirical wavelet transform approach. *International Journal of Wavelets, Multiresolution and Information Processing*, Vol. 16, No. 06, 1850049. doi:10.1142/S0219691318500492.
- [11] Ramkumar, M., Ganesh Babu, C., Manjunathan, A., Udthayan, S., Mathankumar, M., and Sarath Kumar, R. (2021). A Graphical User Interface Based Heart Rate Monitoring Process and Detection of PQRST Peaks from ECG Signal. In: Smys, S., Balas, V.E., Kamel, K.A., Lafata, P. (eds) *Inventive Computation and Information Technologies*. Lecture Notes in Networks and Systems, Vol 173. Springer, Singapore. DOI:10.1007/978-981-33-4305-4_36.
- [12] Jolliffe I. (2022). A 50-year personal journey through time with principal component analysis. *Journal of Multivariate Analysis*, Vol. 188, 104820. doi:10.1016/j.jmva.2021.104820.
- [13] Sklearn.Manifold.TSNE. *scikit-learn*. [Online].
- [14] Kobak D. and Berens P. (2019). The art of using t-SNE for single-cell transcriptomics. *Nature Communications*, Vol.10, Article number: 5416. DOI:10.1038/s41467-019-13056-x.
- [15] Machine learning. *W3schools.com*. [Online].
- [16] Varoquaux G., Buitinck L., Louppe G., Grisel O., Pedregosa F., and Mueller A. (2015). Scikit-learn: Machine Learning Without Learning the Machinery. *GetMobile: Mobile Computing and Communications*, Vol. 19, Iss. 1, pp. 29–33. doi:10.1145/2786984.2786995.
- [17] Sarkar D., Bali R., and Sharma T. (2018). *Practical Machine Learning with Python: A Problem-Solver's Guide to Building Real-World Intelligent Systems*. APress, 530 p. DOI: 10.1007/978-1-4842-3207-1.
- [18] Kerrigan G., Smyth P., and Steyvers M. (2021). Combining Human Predictions with Model Probabilities via Confusion Matrices and Calibration. *Cornell University*, arXiv:2109.14591 [cs.LG]. doi:10.48550/arXiv.2109.14591.
- [19] Wasnik A. (2020). K-Fold Cross-Validation in Python Using SKLearn. *AskPython*.
- [20] Palaniappan Y., Vishanth V. A., Santhosh N., Karthika R., and Ganesan M. (2020). R - Peak Detection using Altered Pan-Tompkins Algorithm. *2020 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India. DOI: 10.1109/ICCSP48568.2020.9182298.
- [21] Cho Y., et al. (2020). Artificial intelligence algorithm for detecting myocardial infarction using six-lead electrocardiography. *Scientific Reports*, Vol. 10, Article number: 20495. DOI:10.1038/s41598-020-77599-6.
- [22] Pueyo E., Sornmo L., and Laguna P. (2008). QRS Slopes for Detection and Characterization of Myocardial Ischemia. *IEEE Transactions on Biomedical Engineering*, Vol. 55, No. 2, pp. 468–477. DOI: 10.1109/TBME.2007.902228.
- [23] Fernandes L. Explanation of the coefficient of variance PCA. *OpenClassrooms*.
- [24] Forward-Backward Filtering. Introduction to Digital Filters with Audio Applications, by Julius O. Smith III. (2007). *Center for Computer Research in Music and Acoustics (CCRMA)*, Stanford University.
- [25] Burns E. and Cadogan M. (2022). Myocardial Ischaemia. *Life in the FastLane*.

- [26] Akin M. (2002). Comparison of Wavelet Transform and FFT Methods in the Analysis of EEG Signals. *Journal of Medical Systems*, Vol. 26, Iss. 3, pp. 241–247. DOI:10.1023/A:1015075101937.

Виявлення ішемії міокарду за допомогою зменшеної кількості відведень ЕКГ

Мневець А. В., Іванушкіна Н. Г., Іванько К. О.

Дослідження присвячено аналізу електрокардіографічних (ЕКГ) ознак для розпізнавання станів норми і ішемії міокарда у разі зменшеного набору електрокардіографічних відведень. Зокрема, для виявлення ішемії міокарда розглядаються спектральні ознаки електрокардіографічного сигналу та характеристики форми ЕКГ хвиль. В статті описані основні ознаки, які зазвичай використовуються для виявлення ішемії міокарда, а також запропоновані інші більш надійні показники для застосування у класифікаційних моделях. Детально описано підхід до попередньої обробки ЕКГ сигналу, ідентифікації необхідних сегментів сигналу і подальшого розрахунку ознак. Розглянуті ознаки засновані на площях під характерними хвилями ЕКГ сигналу і спектральних параметрах цих хвиль. Описано та обрано найбільш інформативні ознаки для виявлення ішемії міокарда, отримані з початкового набору параметрів, що призвело до зниження кількості ЕКГ відведень до 6 у порівнянні з 12 відведеннями стандартної електрокардіограми. Запропоновано методики визначення нових ознак, а саме: відношення площі під хвилею Т до площі під хвилею Р, а також відношення площі під Т хвилею до площі кардіоциклу. Ці ознаки разом з іншими

розрахованими параметрами показали високу точність класифікації сигналів на норму та патологію, оскільки вони відображають біоелектричні процеси, що протікають за наявності ішемії міокарда та проявляються на поверхневій ЕКГ. Також було проведено аналіз візуалізації даних за допомогою методу головних компонент та t-розподіленого стохастичного вбудовування сусідів. Це дозволило показати розподіл даних у просторі ознак, що характеризують класи норми та патології. Дані ЕКГ сигналів у нормі та у випадку ішемії міокарда були отримані з бази даних «PTB Diagnostic ECG Database». Ця база даних містить 22353 ЕКГ сигнали з наявністю 12 ЕКГ відведень (I, II, III, aVR, aVL, aVF та V1–V6), зареєстрованих з високою роздільною здатністю у 290 осіб. З використанням розглянутої бази даних отримано по 10 000 кардіоциклів для класів норми та ішемії міокарда, які застосовано для побудови моделей машинного навчання. На основі отриманих ознак було виконано дослідження алгоритмів машинного навчання та розрахована точність для різних комбінацій відведень ЕКГ. В результаті перехресної перевірки, точність виявлення ішемії міокарда склала 99% зі стандартним відхиленням 0,4% для 6 відведень (I, II, III, aVR, aVL, aVF) та 93% зі стандартним відхиленням 0,12% для одного відведення (I). Таким чином, було показано, що за допомогою методів машинного навчання можна розпізнавати ішемічну хворобу міокарда з високою точністю, використовуючи шість стандартних відведень ЕКГ або лише одне відведення ЕКГ.

Ключові слова: ішемія міокарда; машинне навчання; кардіоцикл; кардіоінтервал; вейвлет-аналіз; площа Т хвилі; кросс-валідація; виявлення захворювань серця