Maciej GOŁGOWSKI[1], Stanisław OSOWSKI[1,2]

Military University of Technology,Faculty of Electronic Engineering (1), Warsaw University of Technology, Faculty of Electrical Engineering (2)

doi:10.15199/48.2021.06.13

Classical versus deep learning methods for anomaly detection in ECG using wavelet transformation

Abstract. *The paper describes and compares two forms of wavelet transformation: discrete (DWT) and continuous (CWT) in the analysis of electrocardiograms (ECG) to detect the anomaly. The anomalies have been limited to two types: cardiac and congestive heart failure. Two independent approaches to the problem have been considered. One is based on discrete wavelet transformation and feature generation based on statistical parameters of the results of the transformed ECG signals. These descriptors, after selection, are delivered as the input attributes to different classifiers. The second approach applies continuous wavelet transformation of ECG signals and the resulting two-dimensional image formed in time-frequency dimensions represents the input to the convolutional neural network, which is responsible for the generation of the diagnostic features and final classification. The experiments have been performed on the publically available database Complex Physiologic Signals PhysioNet. The calculations have been done in Python. The results of both approaches: DWT and CWT have been discussed and compared.*

Streszczenie. *Artykuł predstawia dwa podejścia do wykrywania anomalii w sygnalach ECG. Jako anomalie rozważane są: arytmia i zastoinowa niewydolność serca. Podstawą analizy jest sygnał ECG poddany transformacji falkowej w dwu postaciach: transformacja dyskretna oraz transformacja ciągła. W przypadku transformacji dyskretnej sygnał ECG poddany jest dekompozycji falkowej na kilku poziomach a wyniki tej dekompozycji (sygnały szczegółowe i sygnał aproksymacyjny ostatniego poziomu) podlegają opisowi statystycznemu tworząc zbiór deskryptorów numerycznych – potencjalnych cech diagnostycznych. Po przeprowadzonej selekcji stanowią one atrybuty wejściowe dla zespołu 9 klasyfikatorów. W drugim podejściu sygnał ECG jest poddany ciągłej transformacji falkowej generując dwuwymiarową macierz w postaci obrazu. Zbiór takich obrazów podawany jest na wejście głębokiej sieci neuronowej CNN, która w jednej strukturze dokonuje jednocześnie generacji cech diagnostycznych i klasyfikacji. Eksperymenty numeryczne przeprowadzone zostały na ogólnie dostępnej bazie danych Complex Physiologic Signals PhysioNet. Wyniki eksperymentów wykazały przewagę podejścia wykorzystujacego dyskretną transformację falkową. (**Porównanie metod klasycznych i uczenia głębokiego w problemie wykrywania zaburzeń ECG wykorzystując analizę falkową.**)*

Keywords: anomaly detection, wavelet transform, diagnostic features of ECG, classification, CNN.
Słowa kluczowe: wykrywanie anomalii, transformacja falkowa, cechy diagnostyczne ECG, klasyfikacja, CNN

Introduction

An electrocardiogram (ECG) is a simple test that measures the heart's electrical activity using electrodes allocated on the skin. It shows whether the rhythm of the heartbeats is steady or irregular, and the strength and timing of the electrical impulses passing through your heart is normal. Changes in normal ECG patterns can be caused by numerous cardiac abnormalities, like cardiac rhythm disturbances, inadequate coronary artery blood flow, and electrolyte disturbances. The computer-aided analysis of ECG results assists physicians to detect cardiovascular diseases.

Many signal processing methods are applied in the analysis of ECG signals [1-5]. The most efficient tool now seems to be the wavelet transformation of the ECG signal, producing results on different levels of signal resolution. The statistical numerical descriptors following from such transformation are usually combined with different classification systems, which are responsible for the decision, whether the anomaly is present or not. The numerical results reported in the published papers depend on the database used in experiments. The reported accuracy and sensitivity values are changing from 95% up to 100%, depending on the applied method and used database [1-6].

An additional advantage of the wavelet transformation of ECG signals is the application of this technique in different pre-processing tasks, such as de-noising, extraction of basic parameters used by medical experts in abnormality detection, compression of data, etc. In ECG waveform we recognize five basic waves: P, Q, R, S, T (sometimes also U). The P wave represents atrial depolarization, Q, R, S are commonly known as QRS complex which represents the ventricular depolarization, and the T wave describes the repolarization of the ventricle. The most significant in ECG signal analysis is the shape of the QRS complex.

The problem is that ECG signal sequences may differ for the same person and represent different types of anomalies [2]. Hearth beat without anomalies is very regular, and atrial depolarization is always followed by ventricular depolarization. In the case of arrhythmia, heart rhythm becomes irregular, which is either too slow or too fast.

The ECG waveform is not smooth, with many sudden transitions. If we analyze its spectrum the noisy and normal signal transitions cannot be separated by using the classical filter approach. This is, where wavelet transformation can be successfully applied [2,4,].

This paper presents the computer system to detect the anomaly in the ECG waveform. It applies the wavelet transformation on the stage of pre-processing of the signal and application of the Convolutional Neural Network (CNN) network in final anomaly detection.

Database used in experiments

The database used in experiments contains 162 ECG recordings from Research Resource for Complex Physiologic Signals PhysioNet (publicly available) [6]. The ECG time series have been obtained from three groups (classes) of people with different hearth phenomena. They represent samples with 1) cardiac arrhythmia (ARR) in which the heartbeat is irregular, too fast, or too slow, 2) samples with congestive heart failure (CHG) when the heart is unable to pump sufficiently to maintain blood flow to meet the body's needs and 3) samples with normal sinus rhythms (NSR). The recordings represent part of MIT-BIH Arrhythmia Database, MIT-BIH Normal Sinus Rhythm Database, and BIDMC Congestive Heart Failure Database [6]. Every recording is sampled with 128 Hz and has a length of 216 samples. In total, there are 96 recordings of ARR, 36 of NSR, and 30 of CHG. These recordings contained in total 6144 ARR samples, 1920 samples representing CHG, and 2304 normal samples.

Fig. 1 shows the exemplary ECG recordings representing these three classes of data. We aim to build an automatic system, which can recognize the class of samples belonging to any of these three groups.

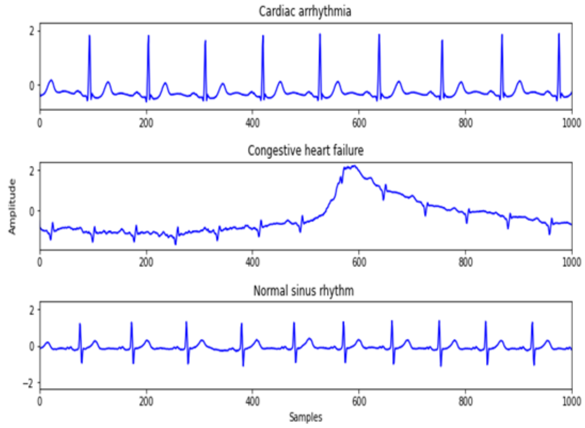

Fig. 1 The exemplary ECG representations of three classes of ECG: arrhythmia (ARR), congestive heart failure (CHG), and normal sinus rhythm (NSR)

Frequency analysis of signals

As the ECG signal is recorded directly from a patient's body, the recording will be disturbed by some noise. The noise signals are due to factors such as baseline wandering, motion artifacts, supply-line interference within the signal, electrode contact, or attenuation losses [7]. Baseline wandering noise is caused by respiration, electrode impedance variation, or excessive body movements.

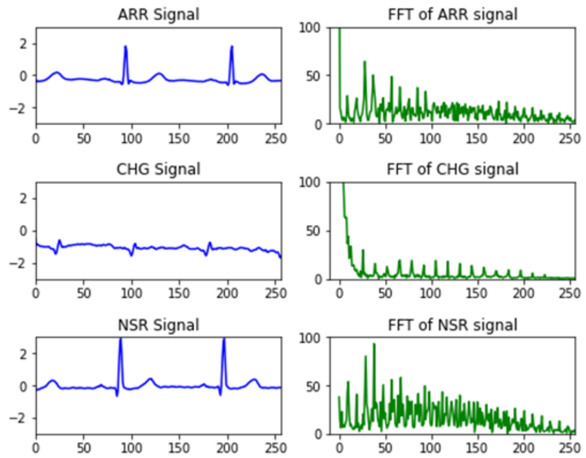

Fig. 2 The representation of three types of signals (left) and their Fourier transforms (right)

One of the typical approaches to feature generation of ECG signals is the application of Fourier transformation. Fig. 2 presents three types of signals and their Fourier transforms. As it is seen power distribution differs significantly in different ranges of frequency, as the basic waves of the ECG signal are different. The most different is the spectrum of the CHG signal.

However, it should be observed, that the Fourier transform is representative only when ECG segment variation over time is negligibly low since it assumes the stationarity of the analyzed signal. Moreover, it applies the basis function of infinite support, hence in the frequency domain, time dependency is irreversibly lost. Therefore, the Fourier approach to ECG analysis is of limited usage. Instead, we will use wavelet transformation in its continuous and discrete form.

Applied methods

In our approach, we will apply the wavelet transformation instead of Fourier. The wavelet transform provides simultaneously high resolution in the frequency domain and also in the time domain. This is obtained thanks to the application of the wavelet function of limited support, which is subject to scaling and shifting. This is in contrast to the sinusoidal function of infinite support applied in Fourier transformation [8-10].

Application of continuous wavelet transform (CWT) maps the signal x(t) to 2-dimensional space represented by scale *a* and shift *b* using the following equation [9]

$$W_x(a,b) = |a|^{-1/2} \int_{-\infty}^{\infty} x(t)\psi\left(\frac{t-b}{a}\right)dt \quad (1)$$

where ψ represents the mother wavelet function used in the analysis, *a* – time scale, and *b* – the time shift. The result of CWT may be presented as the smooth image in (*a*,*b*) coordinate system.

DWT in application to feature generation

In the case of discrete wavelet transform (DWT) the values of *a* and *b* are discrete and taken in a dyadic system as follows $a \to a_m = 2^m$ and $b \to b_{m,n} = 2^m nT$, *T* – the sampling period. The original discrete signal $x(t)$ is now represented as the weighted sum of wavelet functions of different scale and shift values

$$x(t) = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} W_x\left(2^m, 2^m nT\right)\psi_{m,n}(t) \quad (2)$$

where $\psi_{m,n}(t)$ is the wavelet function at *m*th level, shifted by *n* samples.

The DWT transformation is most often used in practice since it is very easily implemented using the Mallat algorithm [9, 11]. In this approach the integration is replaced by the digital filtering operations, applying FIR low-pass and high-pass filters [8,10,11]. The low-pass filter is responsible for generating the lower resolution approximation of signal and the high-pass filter generates the so-called details, i.e., the difference between approximated signals of two neighboring levels of resolution. The detailed signals of the highest resolution are usually associated with the high-frequency components, often treated as the noise. The interesting point is that small detail values might be dropped without affecting the major features of the data set. Therefore, reconstructing signals deprived of such removed details results in the operation of de-noising [10,11].

The idea of DWT signal decomposition is to split the original signal into different frequency sub-bands. Their number is dictated by the user. It should be observed, that the succeeding levels of decomposition reveal separate frequency characteristics in particular sub-bands. Different statistical parameters characterizing these distributions in different levels of decomposition, such as mean, median, standard deviation, skewness, kurtosis, different percentile values, etc., may be used as the numerical descriptors, well characterizing the analyzed ECG waveform. Additionally, zero-crossings, mean-crossing, and entropy may be also calculated for each level of DWT decompositions and used as additional descriptors of the analyzed waveform [12,13,14]. In our approach, every coefficient of decomposition is assigned to the bin with a width of 0.2 standard deviations. Entropy was estimated using the following equation:

$$H = -\frac{1}{N}\sum_i p_i \log p_i \quad (3)$$

where *N* is the total number of coefficients, and p_i is the probability of the *i*th coefficient.

The next step of processing is the selection of the most significant features. The selected descriptors create the set of diagnostic features, serving as the input attributes to the classifier, which is responsible for the final recognition of classes. Different types of classifiers have been used in this work: Random Forest (RF), Gradient Boosting (GB), Multilayer Perceptron (MLP), Support Vector Machine (SVM), KNN, Naive Bayes, etc. [15,16]. All of them belong to the most efficient set of classification systems.

CWT in the transformation of the signal to an image

The continuous wavelet transformation converts the 1D signal into a 2D image of the coordinate system represented by (*a*,*b*), in which the scale factor *a* represents scale in a time domain and *b* – the time shift (both of continuous form). Scale *a* is inversely proportional to frequency, i.e., the higher value of the scale, the lower frequency range.

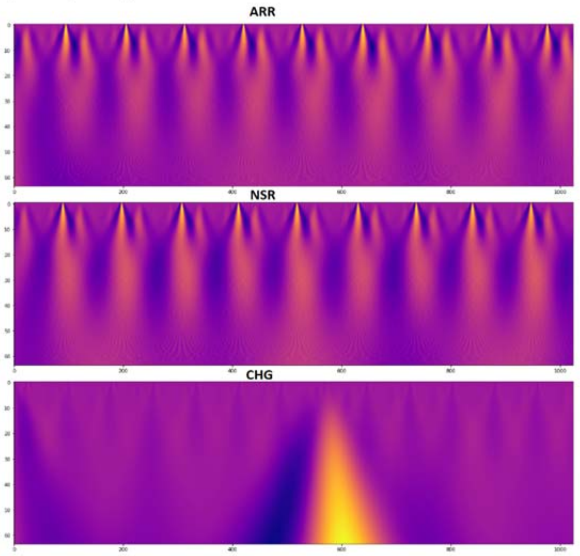

Fig. 3 The image of the CWT transformation of ECG signals from Fig. 1. The upper image corresponds to cardiac arrhythmia, the middle one – to normal sinus rhythm, and the bottom one – to congestive heart failure

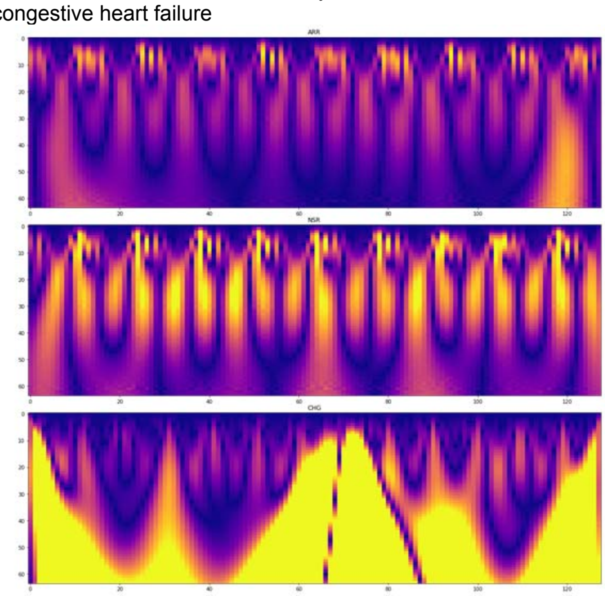

Fig 4 Compressed CWT coefficient matrices corresponding to their original images presented in Fig. 3

Fig. 3 presents the CWT graphical results of the wavelet transformation of three types of ECG signals presented in Fig. 2. The scale was changing up to 64 and the Mexican hat has been applied as the mother wavelet function. The analyzed ECG signals were first split into sections of the length equal to 1024. Therefore, the resulting image, which is used in the analysis is of the size 64x1024. Analyzing the original signals of Fig. 2 it is evident, that the CHG signal is significantly distinguishable from the others. However, ARR and NSR signals are very similar. These observations are very well depicted in the obtained images of the transformed signals. It is observed, that the largest differences are seen in the range of high scales (low frequencies).

To save computation resources, the resizing process was applied in further experiments. The initial matrices of size 64×1024 were resized to 64×128. The resizing process is done to reduce the complexity of the CNN network while preserving the details of “images” with satisfactory accuracy. The graphical results of the compressed images are presented in Fig. 4.

Convolutional Neural Network

The convolutional neural network is a deep network structure that is specially designed to analyze the two-dimensional data [17,18]. Nowadays, CNN is not restricted only to two-dimensions and can be used with multi-dimensional data, including also 1-dimensional. Many hidden layers applied in the structure CNN perform locally connected convolution process.

The convolution is a linear operation that includes the multiplication of a set of weights of a kernel (linear filter) with the input signals representing pixel values of an image, like in traditional neural networks. Assuming, the system is designed for two-dimensional data, the multiplication is done for a defined array of input data and a two-dimensional array of weights called a filter or a kernel.

The filter is of a much smaller size than the size of input data (typical 3×3 for small images to 15×15 for large images). The multiplication applied between a filter-sized patch of the input and the filter is a dot product. As the moving filter is applied multiple times to the input array, the result is a two-dimensional image of output values called a feature map. Once a feature map is created, every value of this map is passed through a nonlinearity. Usually, the rectified linear unit (ReLU) is used. The next layers are created by reducing the size of the convolved images through the pooling operation (max or average pooling is most often used). The final CNN structure used in experiments is presented in Fig. 5.

The structure of input data is organized in the tensor form of the dimension 1x128x128 (the greyscale image of the size 128x128). The width and height of it represent the size of the image and the depth – the number of images subject to processing in each layer.

Three convolutional, locally connected layers applying ReLU activation and MAX pooling have been used in the circuit structure. The first convolution layer has used 128 filters of the size 2x2, creating the same number of output images of the size 64x64 reduced do 32x32 after MAX pooling.

The second layer has reduced the size of images to 8x8 (after convolution and MAX pooling) and the number of output images was increased to 256.

The third convolution followed by the MAX pooling layer produced 256 images of the size 2x2 pixels. Every image was converted to the vector form and formed the concatenated vector of the length 1024, which performed the role of input attributes to the fully connected network (the dense connection between neurons) performing the role of final softmax classifier.

The fully connected part of CNN has the structure: 1024-64-3. The network was trained in the mode of softnet to recognize three classes of ECG beats. In the final scoring, both ARR and CHG represent the anomaly and NSR - the normal class. The learning algorithm was based on ADAM with the application of mini-batches of the size 32. The experiments have been performed in Python.

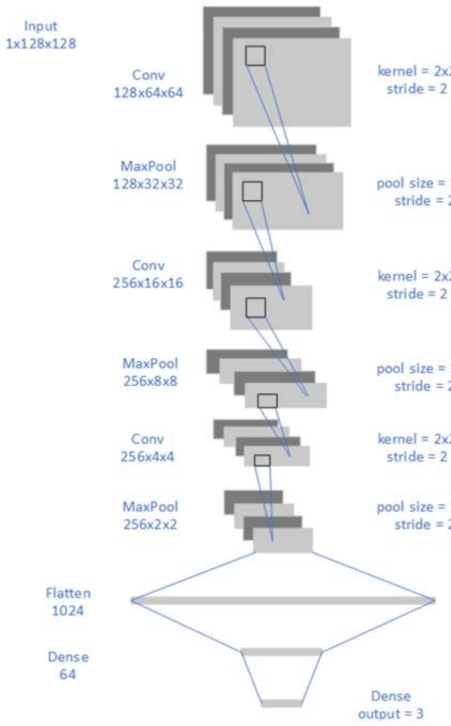

Fig. 5 CNN structure used in experiments

Numerical experiments

Applied recognition systems

The original ECG signals are split into 10-second interval frames and then CWT or DWT are applied. Such a split of data is suggested by the authors of the database and interprets results of anomaly detection resembling the way the medical experts do. The resulting wavelet coefficients were resized from 1024 to 128. Two types of experiments have been performed:

- Application of discrete wavelet transformation to create a set of numerical descriptors defined based on statistics of all decomposition levels. Six levels of decomposition have been applied. The performed selection process assesses the class discrimination ability of each descriptor and the set of the descriptors of the highest discrimination ability form input attributes to the family of 9 classifiers.
- The second set of experiments has been done at the application of continuous wavelet transformation. As a result of this transformation, the 1-dimensional ECG waveform is changed to the 2-dimensional data (images). These data are delivered in the form of tensor to the CNN network, which performs also the role of classifier [16,17,18]. In the case of CWT, the obtained images were directly supplied to the CNN network of the structure shown in Fig.5. This network is responsible for two tasks at the same time: the creation of diagnostic features and the final recognition of the class.

In the case of DWT, the ECG signals transformed to wavelet coefficients were described by different statistical parameters. They include: variance, standard deviation, mean, median, 5th, 25th, 75th 95th percentile value, RMS value; skewness, kurtosis; zero-crossing rate (the number of times a signal crosses the level of y = 0), mean crossing rate, i.e. the mean value of the number of times of zero crossings within the 10-second interval frames [12].

These parameters were estimated for all 6 levels of the wavelet decomposition. The total number of descriptors created in this way was 78. Among these descriptors, there are many irrelevant, not well representing the particular class. Therefore, the selection process was needed to form a smaller number of diagnostic features, well correlated with a class. There are many different selection procedures developed in the past [12,13]. In this work, we have applied the method based on chi-squared statistics. The score from computed chi-squared between each feature and class was used to select top *n* the most important features. The importance is measured by the value of the test chi-squared statistic concerning the considered class.

Different values of *n* have been tried in experiments and the selected quality measures of anomaly detection are depicted in Table 1. The results are presented for recognizing the abnormal cases and are limited to accuracy, sensitivity, and precision. Accuracy is the ratio of the number of true positive decisions of abnormal rhythms to the total number of rhythms, i.e.

$$ACC = \frac{TP}{TP+TN+FP+FN} \quad (4)$$

where TP represents true positive, TN – true negative, FP – false positive, and FN – false-negative cases. Precision is defined as the ratio of true positive to the total of true positive and false positive cases, i.e.,

$$PREC = \frac{TP}{TP+FP} \quad (5)$$

Sensitivity (called also recall) is the ratio of true positive to the sum of true positive plus false-negative cases. i.e.,

$$SENS = \frac{TP}{TP+FN} \quad (6)$$

Both abnormal beats have been jointed in one abnormal class and represented true positive cases. All normal beats represent the true negative in such a statement of the recognition problem.

Results of DWT application

Table 1 depicts the values of the presented above quality measures at four populations of selected diagnostic features: n=78 (maximum number of features), n=48, n=24, and n=12. The reduced number of features has been chosen as the set of the top important features indicated in the selection process [19]. 70% of the available data have been used in learning and the remaining 30% only in testing. The results of testing are given for different classifiers of the names shown in the upper row of the table. All experiments have been performed using Python and Keras libraries [18].

The best results have been obtained at the application of an ensemble of classifiers named Extra Random Forest and using 24 of the best diagnostic features selected by chi-squares statistics. In almost all cases the maximum number of descriptors used as input attributes was the least efficient. To the best classifiers belong random forest, K-nearest neighbor, and SVM classifier. The SVM and K-nearest neighbor classifiers happened to be the least sensitive to the number of input attributes. The least effective was the naïve Bayes classifier (irrespective of the number of input attributes).

Results at CWT application

The second set of experiments has been performed at the application of CNN and CWT. Table 2 presents the numerical results of abnormality detection. In this case, different types of mother wavelet and the different number of scales have been checked [19]. The results corresponding to different types of mother wavelets are differing. However, the best seems to be the Mexican hat wavelet function. These results have been obtained at the application of 128 scales in CWT transformation. The CNN network seems to be less efficient in comparison to DWT. In our opinion, the main reason is the too-small number of learning samples, especially those corresponding to CHG and ARR. The other reason may be the not fully optimal structure of the CNN network, designed from the scratch. Future investigations will be directed to find the optimized number of convolution layers and their parameters, including the size and number of filters in each layer.

Table 1 The results of numerical experiments at the application of DWT at the different number of diagnostic features for different implementation of the classification system.

Features selected		Gradient boosting	ADA classifier	Extra Random Forests	Random Forests	SVM classifier	Gaussian Process	MLP with SGD	KNN	Naïve Bayes
12	ACC [%]	94.19	83.79	96.07	94.99	93.42	89.76	81.91	95.27	81.3
	PREC [%]	92.68	81.78	94.93	93.55	91.78	87.55	81.8	93.53	78.53
	SENS [%]	93.84	86.18	95.75	94.56	92.82	88.26	76.53	94.57	75.41
24	ACC [%]	95.12	90.18	**97.78**	97.61	95.82	92.61	84.5	97.37	80.05
	PREC [%]	95.09	88.56	**97.54**	97.13	94.28	91.54	81.42	96.3	73.92
	SENS [%]	93.49	89.62	**97.20**	96.49	94.44	90.34	81.09	96.07	72.32
48	ACC [%]	93.00	91.88	96.14	92.81	90.79	93.85	85.51	95.85	73.71
	PREC [%]	93.25	93.38	96.58	94.37	93.22	93.50	86.64	95.30	69.08
	SENS [%]	88.90	87.37	94.27	88.74	84.95	89.43	80.05	94.45	65.86
78	ACC [%]	92.61	92.10	94.27	91.40	90.18	94.08	88.97	96.32	50.65
	PREC [%]	90.53	90.20	94.78	92.70	93.51	94.67	89.50	96.03	70.95
	SENS [%]	91.18	89.47	91.29	86.80	84.71	91.06	84.54	95.11	52.28

Table 2 The results of numerical experiments at the application of CNN and CWT at different types of wavelet mother function

Levels	16			32			64			128		
Wavelet	ACC [%]	PREC [%]	SENS [%]	ACC [%]	PREC [%]	SENS [%]	ACC [%]	PREC [%]	SENS [%]	ACC [%]	PREC [%]	SENS [%]
Gauss8	69.62	70.53	68.33	71.93	72.37	71.16	81.00	81.31	80.55	81.42	81.69	81.03
Mex hat	72.96	73.47	72.19	75.98	76.38	75.57	80.16	80.51	79.81	**82.06**	**82.28**	**81.51**
Morlet	72.16	72.58	71.32	75.15	75.63	74.83	77.27	77.76	76.43	80.00	80.46	79.55

onclusions

The paper has presented the application of wavelet transformation in recognition of anomaly in ECG waveforms. Two types of anomaly have been considered: the ARR and CHG. The results of wavelet transformation have been used as the source of the ECG waveform descriptor forms, either as vector (DWT) or matrix (CWT).

In the case of DWT, the statistical parameters (mean, median, skewness, kurtosis, etc.) associated with each decomposition level have formed the initial set of potential diagnostic features. After the selection process was performed by the chi-square method, their number was significantly reduced and then the selected features have been applied as the input attributes to different types of classifiers.

In the case of CWT, the ECG signals have been converted to images and these images formed the input to CNN performing at the same type the role of feature generation and selection, as well as the classification.

The obtained results have shown the superiority of the DWT approach over CWT. The best results of the proposed methodology give an accuracy of 97.78% in anomaly detection, the sensitivity of 97.20%, and precision of 97.54%. These results have been obtained for testing data (not taking part in learning) for the analyzed database of ECG.

***Autorzy**: prof. dr hab. inż. Stanisław Osowski, Politechnika Warszawska, Instytut Elektrotechniki Teoretycznej i Systemów Informacyjno-Pomiarowych, 00-661 Warszawa, Koszykowa 75 oraz Wojskowa Akademia Techniczna, Instytut Systemów Elektronicznych, 00-908 Warszawa, Kaliskiego 2, E-mail: stanislaw.osowskii@pw.edu.pl; mgr inż. Maciej Gołgowski, Wojskowa Akademia Techniczna, Instytut Systemów Elektronicznych, 00-908 Warszawa, Kaliskiego 2, E-mail: maciej.golgowski@wat.edu.pl.*

REFERENCES

[1] Singh R., Mehta R., Rajpal N., Efficient wavelet families for ECG classification using neural classifiers, *Procedia Computer Science*, vol. 132. Pp. 11-21, 2018
[2] Rai H. M., Trivedi A., Shukla S., ECG signal processing for abnormalities detection using multi-resolution wavelet transform and Artificial Neural Network classifier, *Measurement*, vol. 46, 3238-3246, 2013
[3] Patil D. D., Singh R. P., ECG classification using wavelet transform and wavelet network classifier, *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, vol. 668, pp 289-303, 2018
[4] Arumugam M., Sangaiah A. K., Arrhythmia identification and classification using wavelet centered methodology in ECG signals, *Concurrency and Computation, Practice and Experience*, 2019, http,//doi.org/10.1002/cpe.5553
[5] Lenis G., Pilia N., Loewe A., Schulze W. H., Dössel O., Comparison of Baseline Wander Removal Techniques considering the Preservation of ST Changes in the Ischemic ECG, A Simulation Study, *Comput. & Mathematical Methods in Medicine*, vol. 2017, https,//doi.org/10.1155/2017/9295029
[6] PhysioNet - The Research Resource for Complex Physiologic Signals, https,//physionet.org/
[7] Zheng B. Murugappan S., Frequency Band Analysis of Electrocardiogram (ECG) Signals for Human Emotional State Classification Using Discrete Wavelet Transform (DWT), *J. Phys. Ther. Sci..*, vol. 25, No 7, pp. 753–759, 2013
[8] Daubechies I., *Ten lectures on wavelets*, SIAM, Philadelphia, 1992
[9] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, Academic Press, London, 2011.
[10] Teolis A., *Computational signal processing with wavelet*s, Springer Science& Business Media,2012
[11] *Matlab user manual*, MathWorks, 2020, Natick, USA.
[12] Llamedo M., Martınez J. P., Heartbeat classification using feature selection driven by database generalization criteria, *IEEE Trans. Biomedical Eng*., vol. 58, No 3, pp.616-25, 2011
[13] Tan P. N., Steinbach M., Kumar V., *Introduction to data mining*, 2014, Pearson Education Inc., Boston
[14] Bulmer M. G., *Principles of statistics*, Courier Corporation, 2012
[15] Breiman L., Random forests, *Machine Learning,* vol. 45, No 11, pp. 5–32, 2001
[16] Osowski S., Sieci neuronowe do przetwarzania informacji, OWPW, 2020, Warszawa
[17] Goodfellow I., Bengio Y., Courville A, *Deep learning*, MIT Press, Massachusetts, 2016
[18] Chollet F., *Deep Learning with Python*, Manning Publications Co. 2017
[19] Golgowski M., Osowski S., Anomaly detection in ECG using wavelet transformation, Int. on-line Conf. Computational Problems of Electrical Engineering, Warsaw, 2020, doi: 10.1109/CPEE50798.2020.9238709