Wavelet-Based Kernel Construction for Heart Disease Classification

Thanh-Nghia NGUYEN[1], *Thanh-Hai NGUYEN*[1], *Manh-Hung NGUYEN*[1], *Salvatore LIVATINO*[2]

[1]Department of Industrial Electronics and Biomedical Engineering, Faculty of Electrical and Electronic Engineering, Ho Chi Minh City University of Technology and Education, 01 Vo Van Ngan Street, Thu Duc District, Ho Chi Minh City, Vietnam
[2]School of Engineering and Computer Science, University of Hertfordshire, College Lane campus, AL10 9AB Hatfield, United Kingdom

nghiant@hcmute.edu.vn, nthai@hcmute.edu.vn, hungnm@hcmute.edu.vn, s.livatino@herts.ac.uk

DOI: 10.15598/aeee.v17i3.3270

Abstract. *Heart disease classification plays an important role in clinical diagnoses. The performance improvement of an Electrocardiogram classifier is therefore of great relevance, but it is a challenging task too. This paper proposes a novel classification algorithm using the kernel method. A kernel is constructed based on wavelet coefficients of heartbeat signals for a classifier with high performance. In particular, a wavelet packet decomposition algorithm is applied to heartbeat signals to obtain the Approximation and Detail coefficients, which are used to calculate the parameters of the kernel. A principal component analysis algorithm with the wavelet-based kernel is employed to choose the main features of the heartbeat signals for the input of the classifier. In addition, a neural network with three hidden layers in the classifier is utilized for classifying five types of heart disease. The electrocardiogram signals in nine patients obtained from the MIT-BIH database are used to test the proposed classifier. In order to evaluate the performance of the classifier, a multi-class confusion matrix is applied to produce the performance indexes, including the Accuracy, Recall, Precision, and F1 score. The experimental results show that the proposed method gives good results for the classification of the five mentioned types of heart disease.*

Keywords

Back-propagation neural network, Electrocardiogram signals, heart disease classification, wavelet-based kernel principal component analysis, wavelet coefficients.

1. Introduction

In the recent decades, a classifier of Electrocardiogram (ECG) signals has become a necessary and helpful tool for diagnosing different types of heart disease. ECG signals can provide information on people's heart status, while ECG signals' cardiac cycles (related to heartbeats) can directly be measured on human body surface through ECG sensors. The heartbeat characteristics can be observed through the analysis of the P, Q, R, S and T wave components, and different heart diseases' diagnosing can be made based on these components [1]. Even a small change in one of the components may represent an issue.

Noise reduction in ECG signals is relevant for better diagnosis. However, measured ECG signals often contain many noise sources such as: power-line interference, baseline wander, muscle noise, motion artifacts, and other types of interference. A number of methods have been applied in recent years to remove noise sources. For the correction of the baseline wander [2], Yang Xu et al. proposed the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and a wavelet threshold [1]. The CEEMDAN approach has been applied to decompose ECG signal original noise into a series of Intrinsic Mode Functions (IMFs) from the high to low frequency [3]. With other noises, researchers proposed different methods such as canceling the power-line interference noise using digital notch filters [4], and the muscle noise by combining a frequency filter and a time window [5] and [6].

For heart disease classification it is necessary to analysis features of ECG signals, such as the mentioned

P, Q, R, S and T wave components. Furthermore, we need to take into consideration other elements such as people nationality, age and sex, as these also affect the signal. ECG feature extraction for classification represents therefore a great challenge for researchers. The ECG features extraction is often based on studying the different parameters associated to the P, Q, R, S and T wave components, e.g. intervals and frequencies. Many methods have been proposed for the extraction of ECG features. They include the use of: adaptive threshold and Principal Component Analysis (PCA) [7] and [8] combination between features in the time and frequency domains using Random Forests [9]; normalized RR intervals and morphological features [10]; the Linear Discriminant Analysis (LDA) and the Independent Component Analysis (ICA) [11]; and fuzzy analysis [12] and [13].

ECG feature extraction has also been proposed by using wavelet analysis or the combination between wavelet transform and PCA algorithm. In [14], Qin Qin et al. presented a wavelet multi-resolution analysis, named as the ECG feature extraction tool, in which a number of unnecessary components were eliminated. Taiyong Li et al. [15] promoted a wavelet packet entropy to choose the heartbeat feature for ECG classification. In their work the use of Daubechies wavelet (db4) was proposed for classification leading to best performance. In other works [16], [17] and [18], a kernel Principal Component Analysis (kPCA) was combined with one kernel matrix, and the followings were applied: one PCA algorithm, the polynomial kernel function, the hyperbolic tangent (sigmoid) kernel function, the Radial Basis Function (RBF) and the Gaussian kernel function.

In human heart disease classification, high performance is a challenging problem as it may affect diagnosis. With the development of new technologies and high speed computers, it has become an attractive option to build a classification system based on big data. Many classifiers have then been proposed relying on the use of deep learning techniques. However, each ECG signals classifier is only suitable for one type of feature data. Because of this problem, different classifiers have been proposed, which have included: reservoir computing with logistic regression [19] support vector machine [14], [20] and [21], genetic algorithm-back propagation neural network [22], generalized regression neural network [23], discriminative hidden Markov model [24], fuzzy neural network [25], Gaussian mixture model of Karhunen-loève transform [26] and decision tree [27].

ECG signals classification relying on neural networks with deep learning have rapidly developed in the recent years with promising results. Meantime the manual feature extraction has become unnecessary. Serkan Kiranyaz et al. [28] proposed an ECG classifier based on 1D Convolutional Neural Networks (CNNs) and real-time patient-specific components. In their research, the two major blocks of ECG classification are: the feature extraction and the 1D CNNs with an adaptive part. Other research by Wei Li et al. [29], represented the local deep field method, which learns different deep models and the local manifold charts. The local data have been regionalized to concentrate on the local variation features. In [30] and [31] classification of heart diseases took place from a deep architecture in a convolution neural network for information, which was constructed within local distributions in hidden layer.

In the work presented in this paper, the Back-propagation Neural Network (BNN) is proposed for heart disease classification. In order to achieve a more effective classification, a kernel is proposed based on wavelet coefficients with Daubechies wavelet (db4) at level-4. This allows for calculating principal components using a PCA method, called "features of heartbeat signals". The proposed wavelet-based kernel PCA (wkPCA) allows for extracting features with high distribution useful to classify different types of heart disease. The heartbeat features after the wkPCA are applied to the BNN for classification with high performance. A confusion matrix is employed to evaluate the classifying performance over five types of heart disease in nine patients. Experimental results with high performance classification can support clinicians in diagnosing heart disease. This paper includes the following sections: Sec. 2. defines the proposed method related to the wavelet-based kernel; Sec. 3. describes experimental results and discusses them; and Sec. 4. draws the conclusions.

2. Proposed Method

The block diagram of the proposed method for heart disease classification is shown in Fig. 1. The ECG signals of patients are collected from the MIT-BIH database. A Wavelet Packet Decomposition (WPD) algorithm with the low-pass and high-pass filters is then used to produce the significant coefficients that are employed to calculate a kernel for extracting ECG features. A Principal Component Analysis (PCA) algorithm using the wavelet-based kernel is then utilized to extract features of the ECG signals. For the classification of heart disease, a Back-propagation Neural Network (BNN) is employed.

In the proposed method, the wavelet coefficients (Approximation and Details) obtained after applying the WPD algorithm are used to calculate the kernel parameters and to build kernel matrices for the extraction of heart disease features. Hence, the wavelet coefficients and the determined kernel parameters are projected onto a new feature space for determining one kernel matrix on which to apply a PCA algorithm.

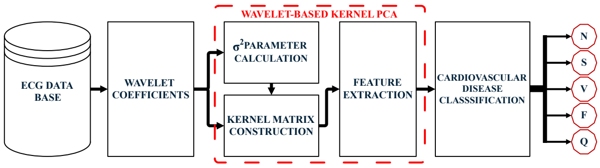


Fig. 1: Block diagram of the proposed heart disease classification with the wavelet-based kernel.

This process, called the wkPCA method, is applied to extract ECG signals main features and to reduce dimensions of the feature matrix. The features resulting from running the wkPCA method are the input of the classifier, which establishes the type of heart disease.

2.1. Wavelet Packet Decomposition Algorithm

An ECG signal records the heartbeats pattern, which has among its components the P, Q, R, S, and T waves. It may be difficult to recognize different types of heart disease from these components by only looking at the time domain. Therefore, we also look at the frequency domain. We apply a WPD algorithm with Daubechies wavelet (db4) to obtain wavelet coefficients (Approximation and Detail) of heartbeats [11] useful for calculating one kernel. In the WPD algorithm, a heartbeat signal, $\vec{x}[n]$ with n samples, is passed through high-pass and low-pass filters with down-sampling by two to obtain wavelet coefficients. It means that the heartbeat is divided into multiple of 2^i, in which i is the level of decomposition related to the Approximation and Detail coefficients, while $\vec{a}_i$ and $\vec{d}_i$ are defined as follows:

$$\vec{a}_i = \sum_{k=-\infty}^{\infty} \vec{x}\left[n\right] g\left[2n-k\right], \tag{1}$$

$$\vec{d}_i = \sum_{k=-\infty}^{\infty} \vec{x}\left[n\right] h\left[2n-k\right], \tag{2}$$

where $g[2n-k]$ is the low-pass filter with down-sampling by two and $h[2n-k]$ denotes the high-pass filter with down-sampling by two.

In general, at the first level of decomposition, the Approximation and Detail coefficients $\vec{a}_1$ and $\vec{d}_1$ are obtained when the heartbeat $\vec{x}[n]$ is passed through filters. The component $\vec{a}_1$ is the input of the next filters to produce $\vec{a}_2$ and $\vec{d}_1$. Therefore, this process will be performed to the decomposition at level-i, to produce the components $\vec{a}_1$ and $\vec{d}_1$. In this work, the WPD algorithm at level-4 is applied to ECG signals to produce the Approximation and Detail coefficients $\vec{a}_4$ and $\vec{d}_4$, in which each coefficient represents the frequency range of the ECG signal that we will use to calculate kernel for classifying heart disease.

In practice, a heartbeat $\vec{x}[n]$ has n samples. Therefore, in this paper, $\mathbf{X}[h] = \left[\vec{x}_j^1, \vec{x}_j^2, \ldots, \vec{x}_j^m\right]$ is the set of the heartbeat signals. When applying the WPD method to the set of $\mathbf{X}[h]$ to determine the wavelet coefficients of $\vec{a}_1$ and $\vec{d}_1$, one obtains the matrices $\mathbf{W}_{ai} = \left[\vec{a}_i^1, \vec{a}_i^2, \ldots, \vec{a}_i^m\right]$, $\mathbf{W}_{di} = \left[\vec{d}_i^1, \vec{d}_i^2, \ldots, \vec{d}_i^m\right]$, with $i = 1, 2, \ldots, l$ and $j = 1, 2, \ldots, n$.

2.2. Wavelet-Based Kernel Principal Component Analysis

In this paper, we focus on calculating a kernel that can allow for classifying different types of heart disease with the high accuracy. In particular, parameters of the kernel will be calculated based on the wavelet coefficients (Approximation and Detail) of heartbeat signals and then the wavelet-based kernel Principal Component Analysis (wkPCA) will be applied for producing features for classification. From the matrices $\mathbf{W}_{ai}$ and $\mathbf{W}_{di}$, one can calculate the standard deviations S_{ai} and S_{di} which are described as follows:

$$S_{ai} = \frac{1}{m}\sum_{k=1}^{m}\left(\frac{1}{l}\sum_{j=1}^{l}\left|a_j^k - \mu_{ai}^k\right|^2\right)^{\frac{1}{2}}, \tag{3}$$

$$S_{di} = \frac{1}{m}\sum_{k=1}^{m}\left(\frac{1}{l}\sum_{j=1}^{l}\left|d_j^k - \mu_{di}^k\right|^2\right)^{\frac{1}{2}}. \tag{4}$$

The μ_{ai}^k and μ_{di}^k are the mean values of k rows of the matrices $\mathbf{W}_{ai}$ and $\mathbf{W}_{di}$. Thus, the mean matrices are determined by the following equations:

$$\mu_{ai}^k = \frac{1}{l}\sum_{j=1}^{l} a_j^k, \tag{5}$$

$$\mu_{di}^k = \frac{1}{l}\sum_{j=1}^{l} d_j^k. \tag{6}$$

From the standard deviations S_{ai} and S_{di} of the heartbeat signals in Eq. (3) and Eq. (4), the wavelet-based kernel parameter σ^2 is defined as follows:

$$\sigma^2 = \frac{\alpha S_{ai} + \beta S_{di}}{\alpha + \beta}. \tag{7}$$

The hyper parameters α and β are chosen based on the values of typical classification systems for controlling the balance between the terms of the wavelet-based kernel.

In practice, ECG signals have non-linear distribution with noise sources and artifacts. Thus, it is applied a Principal Component Analysis (PCA) algorithm using a wavelet-based kernel, called the wkPCA, to project the ECG features extracted from the wavelet transform of the heartbeat signals. Assuming that $\mathbf{W}$ is the set of the wavelet coefficients and the non-linear mapping function $\phi(\mathbf{W})$, the $\mathbf{X} \to \phi(\mathbf{W})$ is determined. In addition, a Gaussian kernel function with the wavelet coefficients is applied for the nonlinear mapping functions, which calculates elements of the kernel as described in the following expressions:

$$\phi\left(a_i, a_j\right) = \exp\left(-\frac{\left\|a_i - a_j\right\|^2}{2\sigma^2}\right), \tag{8}$$

$$\phi\left(d_i, d_j\right) = \exp\left(-\frac{\left\|d_i - d_j\right\|^2}{2\sigma^2}\right). \tag{9}$$

The indicators a_i and a_j are the wavelet coefficients of the matrix $\mathbf{W}_{ai}$; while d_i and d_j are denote the matrix $\mathbf{W}_{di}$, and $\|a_i - a_j\| \, and \, \|d_i - d_j\|$ are Euclidean distances.

From the non-linear elements $\phi\left(a_i, a_j\right)$ and $\phi\left(d_i, d_j\right)$, we calculate the wavelet-based kernel elements $k\left(a_i, a_j\right)$ and $k\left(d_i, d_j\right)$ as in the following expressions:

$$k\left(a_i, a_j\right) = \phi\left(a_i, a_j\right) \phi\left(a_i, a_j\right)^T, \tag{10}$$

$$k\left(d_i, d_j\right) = \phi\left(d_i, d_j\right) \phi\left(d_i, d_j\right)^T. \tag{11}$$

Once the elements $k\left(a_i, a_j\right)$ and $k\left(d_i, d_j\right)$ are obtained, the matrices with the wavelet-based kernels $\mathbf{K}_{ai} = \vec{k}\left(a_i, a_j\right)$ and $\mathbf{K}_{di} = \vec{k}\left(d_i, d_j\right)$, $i, j = 1, 2, \ldots, m$ can be determined. One should note that the wavelet-based kernel matrices $\mathbf{K}_{ai}$ and $\mathbf{K}_{di}$, which represent covariance matrices, have high dimensions and their calculation have high computational cost. As a consequence, the wkPCA method is applied not only for extracting features of the heartbeats, but also for reducing dimensions of matrices. The wkPC eigenvalues and eigenvectors are calculated using the following equations:

$$\mathbf{K}_{ai} \mathbf{V}_{ai} = \lambda_{ai} \mathbf{V}_{ai}, \tag{12}$$

$$\mathbf{K}_{di} \mathbf{V}_{di} = \lambda_{di} \mathbf{V}_{di}. \tag{13}$$

The $\mathbf{V}_{ai}$ and $\mathbf{V}_{di}$ are the eigenvector matrices and each matrix has k eigenvectors. In particular, the matrices $\mathbf{V}_{ai}$ and $\mathbf{V}_{di}$ are described as $\mathbf{V}_{ai}^k = \left[\vec{v}_k^1(ai), \vec{v}_k^2(ai), \ldots, \vec{v}_k^m(ai)\right]^T$ and $\mathbf{V}_{di}^k = \left[\vec{v}_k^1(di), \vec{v}_k^2(di), \ldots, \vec{v}_k^m(di)\right]^T$, in which λ_{ai} and λ_{di} are the eigenvalues used to calculate the eigenvector for features extraction and dimension reduction.

The eigenvalues with the wavelet coefficients are chosen to calculate the eigenvectors with length d, depending on the chosen dimension reduction. They are: $\vec{\lambda}_{ai} = [\lambda_1(ai), \lambda_2(ai), \ldots, \lambda_m(ai)]$ and $\vec{\lambda}_{di} = [\lambda_1(di), \lambda_2(di), \ldots, \lambda_m(di)]$.

In each eigenvector, each element will represent feature components of heartbeat. In addition, the eigenvalues are arranged in a descent way, and the first components with the high eigenvalues often contain principal components used to be the input of the classifier. In this work, one may just choose r principal eigenvalues, corresponding to the following eigenvectors $\widetilde{\mathbf{V}}_{ai}$ and $\widetilde{\mathbf{V}}_{di}$ with $\widetilde{\mathbf{V}}_{ai} = \mathbf{V}_{ai}(:, 1:r)$ and $\widetilde{\mathbf{V}}_{di} = \mathbf{V}_{di}(:, 1:r)$. From the wavelet-based kernel, we can calculate the feature vectors for classification with high performance. Calling $\mathbf{F}_{ai}$ and $\mathbf{F}_{di}$ the matrices calculated with the wkPCA algorithm, they are as it follows:

$$\mathbf{F}_{ai} = \mathbf{K}_{ai} \widetilde{\mathbf{V}}_{ai}^T, \tag{14}$$

$$\mathbf{F}_{di} = \mathbf{K}_{di} \widetilde{\mathbf{V}}_{di}^T. \tag{15}$$

In addition, the matrix $\widetilde{\mathbf{F}}$ of the heartbeats is constructed using the wkPCA for extracting features as described below:

$$\widetilde{\mathbf{F}} = \begin{bmatrix} \mathbf{F}_{ai} & \mathbf{F}_{di} \end{bmatrix}. \tag{16}$$

The matrix $\widetilde{\mathbf{F}}$ of the heartbeats will be the input layer of the heart disease classifier.

2.3. Back-Propagation Neural Network for Classification of Heart Disease

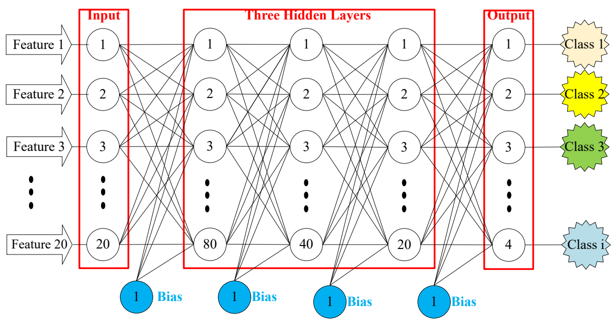


Fig. 2: Architecture of the BNN model for classification of heart disease.

The classifier of heart disease employed in this research is a Backpropagation Neural Network (BNN) [22], which is described in Fig. 2. The illustrated BNN

Tab. 1: Patient information in the MIT-BIH Arrhythmia Database.

Pat. No.	Pat. Rec.	Sex	Age	N (Normal beat)	S (Disease Type 1)	V (Disease Type 2)	F (Disease Type 3)	Q (Disease Type 4)	Heartbeats of each patient
1	104	F	66	163	0	2	0	2140	163
2	106	F	24	1505	0	518	0	69	1505
3	119	F	51	1539	0	443	0	106	1539
4	203	M	43	2525	2	444	1	130	2525
5	208	F	23	1584	2	990	372	86	1584
6	217	M	65	244	0	162	0	1868	244
7	221	M	83	2026	0	396	0	0	2026
8	223	M	73	2040	73	473	14	0	2040
9	232	F	76	396	1379	0	0	0	396
Total heartbeats				12022	1456	3428	387	4399	21692

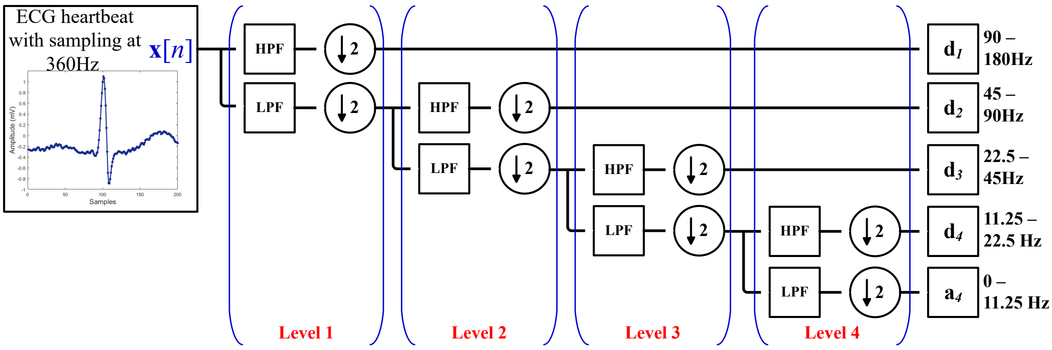


Fig. 3: Schema of the WPD algorithm at level-4 and the obtained coefficients.

model consists of an input layer, three hidden layers and an output layer. The number of nodes at the input layer corresponds to the number of features of a heartbeat after the wkPCA algorithm has run. The output nodes are designed to correspond to different types of heart disease, while the three hidden layers are applied to the BNN.

3. Results and Discussion

In this paper, the simulation results are shown to illustrate the high performance of the proposed method. In particular, we show the standardization of ECG signals for classification, with significant improvements compared to several previous research results. In order to extract wavelet coefficients, the WPD algorithm with Daubechies wavelet (db4) at level-4 was applied to heartbeat signals decomposition. Thus, the Approximation and Detail coefficients were used to calculate kernel for extracting heartbeat features. In particular, the wkPCA was employed to extract heartbeat features for classifying the type of heart disease. The heartbeat features were the inputs of the BNN with three hidden layers and two input-output layers for heart disease clssification.

3.1. Standardization of ECG Signals

The set of ECG signals from nine patients were collected from the MIT-BIH Database [32] in order to analyze the type of heart disease. In addition, the set of the ECG signals was assigned to patient records. The heartbeat signals from each patient's ECG are shown in Tab. 1.

Tab. 2: Introduction of character names of heartbeat groups in nine patients.

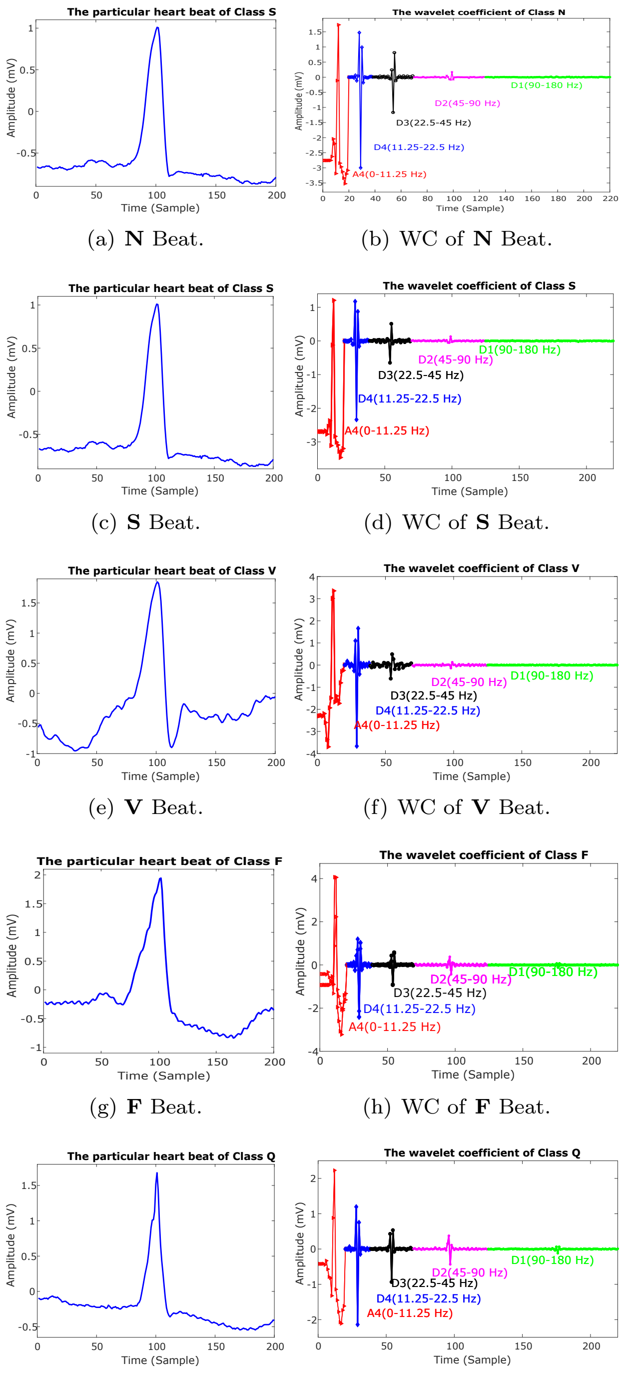
ANSI-AAMI EC57-2012	N, L, R, e, j	A, a, J, S	V, E	F	Q, /, f, !, ", +, [,], x, |, ~, B, n
MIT-BIT Recording No.	**N**	**S**	**V**	**F**	**Q**
Heartbeats	12022	1456	3428	387	4399

This paper introduces different types of heartbeat classes based on two standard groups such as the ANSI-AAMI and the MIT-BIH database as shown in Tab. 2. There are five types of heartbeat classes named as: **N** - Non-Ectopic Beat (or Normal beat); **S** - Supra-Ventricular Beat; **V** - Ventricular Ectopic Beat; **F** - Fusion beat and **Q** - Unknown Beat. In this work, 9 ECG signals of nine patients were collected from the MIT-BIH database. Each ECG signal was recorded for about 30 minutes at the sampling frequency of 360 Hz. For classification of heart disease, the nine ECG signals of nine patients were split into 21692 heartbeats, in which each heartbeat with one R peak of QRS complex, was set to be 99 points on the left side of the R peak [1].

3.2. Wavelet Packet Decomposition Algorithm of Heartbeats

In order to obtain wavelet coefficients for calculating kernels in the wkPCA method, the WPD algorithm was applied to ECG signals sampled at the frequency of 360 Hz. The sampling was performed using the Nyquist–Shannon sampling theorem with the maximum frequency set for the ECG signal set up at 180 Hz. In the WPD algorithm with Daubechies wavelet (db4) at level-4, heartbeats of the ECG signals were decomposed by passing the high-pass and low-pass filters to obtain the wavelet coefficients (Approximation and Detail), with the frequency ranges $\vec{a}_4$ (0–11.25 Hz), $\vec{d}_4$ (11.25–22.5 Hz), $\vec{d}_3$ (22.5–45 Hz), $\vec{d}_2$ (45–90 Hz) and $\vec{d}_1$ (90–180 Hz). These may contain the main features of the wavelet coefficients as shown in Fig. 3. The result is represented by the wavelet coefficients (Approximation and Detail) at the frequency of 180 Hz described in Fig. 4. The curve of the Approximation coefficient ($\vec{a}_4$) is shown in red color, while the detail coefficients $\vec{d}_4$, $\vec{d}_3$, $\vec{d}_2$, $\vec{d}_1$ are shown in blue, black, pink and green colors, respectively. For heart disease classification based on ECG and WPD processing, we considered the frequency range from 0 Hz to 20 Hz, in which the approximation ($\vec{a}_4$) and detail ($\vec{d}_4$) coefficients were chosen to obtain the heartbeat features.

Figure 4 represents five types of heartbeats and waveforms of the wavelet coefficients with the corresponding frequency bands. Figure 4(a), Fig. 4(c), Fig. 4(e), Fig. 4(g) and Fig. 4(i), show the heartbeats corresponding to five types of heart disease assigned to be **N**, **S**, **V**, **F** and **Q**. The remaining figures are the wavelet waveforms as described in Fig. 4(b), Fig. 4(d), Fig. 4(f), Fig. 4(h) and Fig. 4(j). The approximate waveform $\vec{a}_4$ is in red color, and the detail waveforms $\vec{d}_4$, $\vec{d}_3$, $\vec{d}_2$, $\vec{d}_1$ are in blue, black, pink and green colors.



(a) **N** Beat. (b) WC of **N** Beat.
(c) **S** Beat. (d) WC of **S** Beat.
(e) **V** Beat. (f) WC of **V** Beat.
(g) **F** Beat. (h) WC of **F** Beat.
(i) **Q** Beat. (j) WC of **Q** Beat.

Fig. 4: Representation of five types (**N**, **S**, **V**, **F**, and **Q**) of heart disease and the corresponding Wavelet Coefficient (WC) waveforms.

3.3. Principal Components of Heartbeats

In this paper, we proposed the wavelet-based kernel PCA (wkPCA) algorithm for increasing principal components of heartbeats, which can produce a classification with high performance. In particular, after applying the WPD to the heartbeat signals to produce the wavelet coefficients, the components of the Approximation coefficient $\vec{a}_4$ with the length of 19, and the components of the Detail coefficients $\vec{d}_4$ with the length of 19, were used to calculate the kernel parameters as well as the kernel matrices $\mathbf{K}_{a4}$ and $\mathbf{K}_{d4}$. Each heartbeat then produced 38 wavelet coefficients after running the WPD algorithm.

In practice, the distribution of ECG signals is quite chaotic and it is difficult to accurately classify different types of heart disease. For classification with high performance, heartbeat signals were extracted through the wavelet Approximation and Detail coefficients, which were then used to calculate a kernel for extracting ECG features using the wkPCA method. In this work, the Gaussian kernel function was calculated based on heartbeat features obtained from the wkPCA. In addition, the dissimilar samples of the heartbeat feature in five different types of heart disease were projected into the feature space dimensions for classification.

The feature space with large dimensions needs to be arranged in descending order based on eigenvalues, while the dense principal features of the heartbeats need to be chosen. Based on the cumulative sum of information in eigenvalues, the number of the principal components was $r = 10$, which is the column values of one kernel arranged in descending order. This was chosen to recover over 98 % of information on features of heart disease. It means that twenty features corresponding to the first ten principal components of the Approximation coefficients and the first ten principal components of the Detail coefficients of each heartbeat, were used to be the twenty inputs of the classifier using the BNN for the five different types of heart disease.

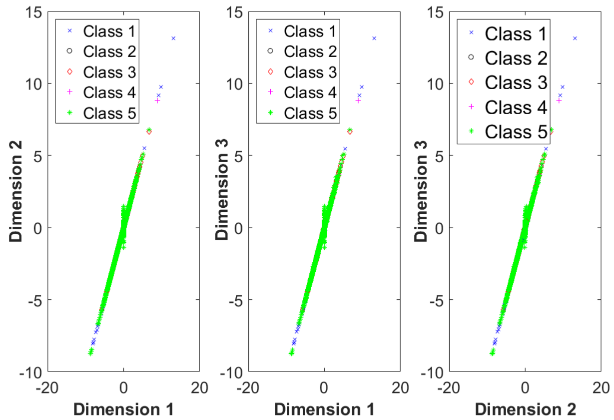


Fig. 5: Representation of the feature distribution after the WPD of five types of heart disease for training.

As for the features distribution of training datasets, the Fig. 5 shows the distribution of the heartbeat features five types after the WPD was run. They five types are: N_Class-1, S_Class-2, V_Class-3, F_Class-4, Q_Class-5. These five types are in Fig. 5 represented through the colors blue, black, red, pink and green. Thus, each heartbeat after the WPD at level-4 produces one Approximate vector $\vec{a}_4$ and one Detail vector $\vec{d}_4$ and both the vectors have the same size of 1 row and 19 columns. The distribution of these vectors is represented as three pairs of the first three dimensions of the training datasets (in coordinates). Each axis denotes one dimension of the 38 dimensions containing the wavelet feature coefficients (Approximation and Details). In particular, Fig. 5 shows the data distribution of the first three columns of the Approximation and Detail matrix. The first, second and third columns correspond to Dimension-1, Dimension-2 and Dimension-3. The feature distribution of five types of heart disease looks like sparse and chaotic, which can affect high performance classification.

Figure 6 represents the feature distribution of the five types of heartbeats in the new feature space dimensions after applying the wkPCA. The distributions of these heartbeat datasets show that the heartbeat features in the five types are dense. In Fig. 6 they have been arranged such that each type can be easily compared to the distributions of Fig. 5. The feature distributions of the dimensions increase the classifying performance over the types of heart disease. Similarly, the wkPCA was employed over the testing dataset and the feature distribution of the five types was arranged in groups. The wkPCA plays an important role in the high performance of a classifier.

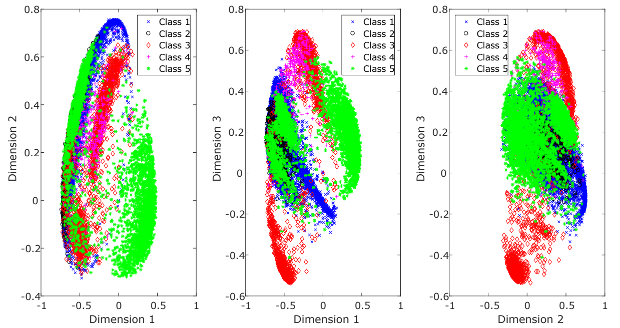


Fig. 6: Distribution of ECG training data in the new feature space dimensions after applying the wkPCA.

3.4. Classification of Heart Disease

The BNN is employed to classify five types of heart disease. In particular, the architecture of the BNN consisted of: an input layer with twenty neurons corresponding to the twenty features of one heartbeat signal; three hidden layers with the different number of neurons; and one output layer with five neurons corresponding to five types of heart disease (N_Class-1, S_Class-2, V_Class-3, F_Class-4, and Q_Class-5) as shown in Fig. 2. After the wkPCA was run, the heartbeat features were applied to the input of the BNN, in which the BNN weights were updated using the backpropagation method. Based on the training data, the MSE value was calculated between the actual and predicted values and the weights of the BNN stopped updating when the MSE reached a value under $1.00 \cdot 10^{-4}$.

In practice, the accurate classification of heart disease using the BNN model depends on choosing its layers. The accuracy of the BNN model can be low when the number of hidden layers chosen is one or two.

Tab. 3: Parameters used in the heart disease classifier.

Hidden layer number	Hidden layer size	Hidden layer active function	Stopping Epochs	Connection manner	Output active function
3	[80 40 20]	Tansig	400	fully	Softmax

Tab. 4: Training and testing data for validation of the classifier using the BNN.

Training data	%	10	20	30	40	50	60	70	80	90
	Heartbeats	2169	4338	6507	8676	10846	13016	15845	17354	19523
Testing data	%	90	80	70	60	50	40	30	20	10
	Heartbeats	19523	17354	15845	13016	10846	8676	6507	4338	2169

Moreover, the overfitting occurs when we have a large number of the hidden layers, and this affects accuracy of the classification. Therefore, when choosing three hidden layers of the BNN model, the performance of the classifier highly increased. In particular, the triple [80 40 20] denotes 80 nodes of the first layer, 40 nodes of the second layer and 20 nodes of the third one. In this research the best BNN model for parameters and functions was designed to best perform in experiments for classifying heart disease, as described in Tab. 3.

Table 4 shows statistics that evaluate the classifying performance, which considered the ratios reduced by 10 % to 90 % between two heartbeats training and testing datasets. All training data divided into the different percentage ratios were used into the BNN. Similarly, the testing percentage ratios were applied to evaluate the classifier performance to determine a pair of the best training and testing ratios for the five types of heart disease. The BNN architecture with the proposed layers and neural nodes produced the high classifying performance for the five types of heart disease. Moreover, for an accurate evaluation of the classifier, a confusion matrix method is employed for the five outputs of the BNN.

3.5. Multi-Class Confusion Matrix

A confusion matrix method [21] was employed to evaluate the classifying performance of different types of heart disease in nine patients. Table 5 shows a matrix with the five types of heart disease corresponding to the five letters: N, S, V, F, and Q. In order to perform a classification, four parameters were estimated using Eq. (17) to Eq. (30). They are: Recall (REC), Precision (PRE), F1 score ($F1S$) and Accuracy (ACC). The REC parameter identifies the actual classes of heartbeats in percentage; the PRE denotes the accurate recognition of predicted classes in the percentage ratio; the ACC is the percentage of accurate identification of all heartbeats; and the $F1S$ is the harmonic mean value of REC and PRE.

The high performance is determined when large values of these parameters are found. This shows the good classifier performance leading in our case to accurate results. The evaluated parameters of the confusion matrix for the classifier are calculated using the formulas Eq. (17), Eq. (18), Eq. (19), Eq. (20), Eq. (21), Eq. (22), Eq. (23), Eq. (24), Eq. (25), Eq. (26), Eq. (27), Eq. (28), Eq. (29) and Eq. (30).

Tab. 5: Confusion matrix applied for evaluation five types of heart disease.

	Predicted Classes					
		N	S	V	F	Q
Actual Classes	N	Nn	Ns	Nv	Nf	Nq
	S	Sn	Ss	Sv	Sf	Sq
	V	Vn	Vs	Vv	Vf	Vq
	F	Fn	Fs	Fv	Ff	Fq
	Q	Qn	Qs	Qv	Qf	Qq

Formulas for estimating classification performance:

$$REC_N = \frac{Nn}{Nn+Ns+Nv+Nf+Nq}, \tag{17}$$

$$REC_S = \frac{Sn}{Sn+Ss+Sv+Sf+Sq}, \tag{18}$$

$$REC_V = \frac{Vn}{Vn+Vs+Vv+Vf+Vq}, \tag{19}$$

$$REC_F = \frac{Fn}{Fn+Fs+Fv+Ff+Fq}, \tag{20}$$

$$REC_Q = \frac{Qn}{Qn+Qs+Qv+Qf+Qq}, \tag{21}$$

$$REC = \frac{REC_N+REC_S+REC_V+REC_F+REC_Q}{5}, \tag{22}$$

$$PRE_N = \frac{Nn}{Nn+Sn+Vn+Fn+Qn}, \tag{23}$$

$$PRE_S = \frac{Ss}{Ns+Ss+Vs+Fs+Qs}, \tag{24}$$

$$PRE_V = \frac{Vs}{Nv + Sv + Vv + Fv + Qv}, \tag{25}$$

$$PRE_F = \frac{Ff}{Nf + Sf + Vf + Ff + Qf}, \tag{26}$$

$$PRE_Q = \frac{Qq}{Nq + Sq + Vq + Fq + Qq}, \tag{27}$$

$$PRE = \frac{PRE_N + PRE_S + PRE_V + PRE_F + PRE_Q}{5}, \tag{28}$$

$$F1S = 2 \cdot \frac{REC \cdot PRE}{REC + PRE}, \tag{29}$$

$$ACC = \frac{Nn + Ss + Vv + Ff + Qq}{\sum}, \tag{30}$$

where $\sum = Nn + Ns + Nv + Nf + Nq + Sn + Ss + Sv + Sf + Sq + Vn + Vs + Vv + Vf + Vq + Fn + Fs + Fv + Ff + Fq + Qn + Qs + Qv + Qf + Qq$ and the values of Nn, Sn, Vn, Fn, Qn, etc. may, as shown in Tab. 5, be determined during the BNN classification.

Tab. 6: Kernel parameters for the Gaussian kernel function applied to produce the best classifier performance.

No.	α	β	REC	PRE	$F1S$	ACC
1	5	5	93.51	93.47	93.49	97.16
2	5	10	93.37	95.27	94.31	97.16
3	10	5	95.42	95.94	95.68	97.23
4	15	5	94.53	96.04	95.28	97.43
5	15	10	94.57	96.19	95.37	97.46
6	20	10	95.98	96.61	96.29	97.61
7	50	10	95.70	97.20	96.44	97.98
8	50	20	96.60	97.59	97.09	98.03
9	50	30	95.03	97.30	96.15	97.96
10	50	50	94.71	96.91	95.80	97.66

With the trained and tested datasets used to classify the five types of heart disease, a confusion matrix was employed to evaluate the five types of heart disease (N, S, V, F, and Q). In the confusion matrix, the accurate performance was evaluated based on the results from Eq. (17), Eq. (18), Eq. (19), Eq. (20), Eq. (21), Eq. (22), Eq. (23), Eq. (24), Eq. (25), Eq. (26), Eq. (27), Eq. (28), Eq. (29) and Eq. (30), in which the diagonal values (Nn, Ss, Vv, Ff, and Qq) were considered the correct classification, as shown in Tab. 5. The trained and tested datasets were divided into the different ratios, which led to accurate results.

From the proposed method to calculate the wavelet-based kernel parameters, we needed to determine the parameters α and β for an accurate classification. These parameters play an important role in the proposed classification method. They were chosen through various experiments aimed at finding the best values. In the WPD at level-4 with the wavelet coefficients $\vec{a}_4$ and $\vec{d}_4$, the parameters α and β were statistically chosen to find the best parameters. Table 6 shows the experimental results using the proposed method with $\alpha = 50$ and $\beta = 20$. The result of the classifier had the highest accuracy. In this experiment in order to evaluate the classification performance, it was proposed to use 90 % training data and 10 % testing data of the total dataset.

Table 7 represents ACC, REC, PRE and $F1S$ for evaluation of the classification performance related to the wavelet-based kernel with $\alpha = 50$ and $\beta = 20$. In the classification method, the chosen parameters $\alpha = 50$ and $\beta = 20$ were the best values, allowing the classifier to produce the best classification performance. The chosen heartbeats ratio is 90 % training data and 10 % testing data, the ACC performance of 98.03 % is the best classification.

Table 8 shows experimental results of testing the performance of the proposed method on four types of heart disease. In this case, the F type data in Tab. 2 was small compared to other disease types. We could then remove it and only analyze the four diseases N, S, V, and Q. Moreover, for the evaluation of the classifier performance, the confusion matrix method as described in Tab. 5, was employed with four classes. The performance of the classifier with the four diseases is described in Tab. 8, with the kernel parameters $\alpha = 50$ and $\beta = 20$, and the best accuracy being 98.41 %. In another case, we represented the classifier with three outputs corresponding to three types of heart disease (N, S, and Q) as described in Tab. 9. In this case, the high performance of the classifier was determined when we applied the confusion matrix method.

In addition, we estimated the experience for the classification of the three diseases N, S, and V to evaluate the performance of the proposed classifier. Two types (F and Q) of heart disease shown in Tab. 2 were removed and the remaining three types (N, S, and V) were used for the disease classification. In this case the BNN structure and the kernel parameters were similar to the cases that considered four or five types of heart disease.

The number of types of heart disease chosen for classification, affects the classifier performance. Figure 7 shows a comparison of classification accuracy in three cases of heart disease (three diseases, four diseases and five diseases). This is based on different rates for training datasets in percentage. In particular, the result shown through the blue line describing three diseases (N, S and V), achieves the best accuracy of 98.52 %. The green line shows the outcome of 4 disease classes, which produces an accuracy of 98.41 %, while the pink line representing the case of 5 diseases gets the low-

Tab. 7: Representation of the classifier performance for five types of heart disease (N, S, V, F and Q) with $\alpha = 50$ and $\beta = 20$, and the training and testing datasets in percentage.

Training	10	20	30	40	50	60	70	80	90
Testing	90	80	70	60	50	40	30	20	10
REC	92.26	94.27	95.16	95.39	96.01	96.15	96.53	96.53	96.60
PRE	93.36	95.17	95.45	96.21	96.20	96.31	96.70	97.47	97.59
F1S	92.81	94.72	95.30	95.80	96.10	96.23	96.61	97.00	97.09
ACC	95.48	96.75	96.96	97.25	97.34	97.64	97.73	97.84	98.03

Tab. 8: Representation of the classifier performance for four types of heart disease (N, S, V, and Q) with $\alpha = 50$ and $\beta = 20$, and the training and testing datasets in percentage.

Training	10	20	30	40	50	60	70	80	90
Testing	90	80	70	60	50	40	30	20	10
REC	95.63	96.13	96.27	96.74	96.95	97.16	97.00	97.78	97.82
PRE	95.38	96.31	96.66	97.15	97.59	97.34	97.74	98.13	98.29
F1S	95.50	96.22	96.46	96.94	97.27	97.25	97.37	97.95	98.05
ACC	96.57	97.11	97.38	97.59	97.89	98.03	98.02	98.25	98.41

Tab. 9: Representation of the classifier performance for three types of heart disease (N, S, and V) with $\alpha = 50$ and $\beta = 20$, and the training and testing datasets in percentage.

Training	10	20	30	40	50	60	70	80	90
Testing	90	80	70	60	50	40	30	20	10
REC	96.11	96.59	96.97	96.71	96.79	96.91	97.00	97.45	97.62
PRE	96.10	97.00	96.96	97.12	96.93	97.31	97.66	97.67	97.96
F1S	96.10	96.79	96.96	96.91	96.86	97.11	97.33	97.56	97.79
ACC	97.26	97.90	98.01	98.02	97.92	98.15	98.21	98.46	98.52

est accuracy of 98.03 %. Nonetheless, the error between them (3 and 5 types of heart disease) only is 0.49, being therefore small and negligible. Wavelet

Tab. 10: The accuracy comparison of the proposed method with other methods.

Authors	**Method for Feature**	**Classifier**	**Disease No.**	**Accuracy (%)**
Hongqiang Li [22]	WPCs	GA-BPNN	6	97.78
Taiyong Li [15]	WPE	RF	5	94.61
Proposed method	WPD + wkPCA	BNN	5	98.03

Packet Decomposition (WPD), Wavelet Packet Coefficients (WPCs), Genetic Algorithm-Back Propagation Neural Network (GA-NNPB), Wavelet Packet Entropy (WPE), Random Forests (RF), Wavelet-based kernel Principle Component Analysis (wkPCA), Backpropagation Neural Network (BNN).

As shown in Tab. 10, the result obtained from the proposed method is compared with other methods to illustrate its effectiveness. In particular, all previous research to our knowledge that used ECG signals of the MIT-BIH database differently designed the way to get accurate performance classification. In each research, authors proposed different methods for extracting features and classifying types of heart disease and the results produced different performance accuracy. In [15], authors used the dataset on MIT-BIH with 22 recordings for training and 22 recordings for testing, including types of heart disease including N, S, V, F, and Q. In the datasets, the number of heartbeats was different and there is a large difference with the five types where the obtained classification accuracy was of 94.61 %. In other research [22], authors employed the GA-BPNN method with extracting features using the WPCs algorithm for classifying six types of heart disease (N, L, R, P, V, A), in which ECG signals were collected from the MIT-BIH database. In particular, they used 30 training samples and 30 testing samples for each type of heart disease and then the GA-BPNN with one input layer, two hidden layers and one output layer produced an accuracy of 97.78 %. In this GA-BPNN model, the hidden layers used logistic functions as an active function. Moreover, a total of 48 input layer nodes, 50 hidden layer nodes and 6 output layer nodes were configured. As discussed above, despite our research used five types of heart disease for classification, while authors in [22] used six types of heart disease for classification using the same neural network, a main difference is that our research proposes the wkPCA for extracting features and that the result achieved of 98.03 % accuracy of classification is slightly higher than the 97.78 % obtained in [22].

In this paper, we proposed the feature extraction using the wkPCA and the classifier using the BNN method, in which the wavelet-based kernel was designed based on the wavelet coefficients at level-4 to arrange and to distribute features of five types of heart disease in nine patients. In particular, when the wkPCA was applied, the feature datasets of five types

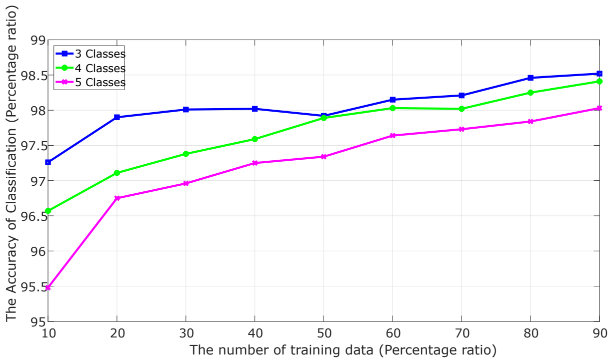


Fig. 7: Representatio of the accuracy comparison of 3 different cases of heart disease.

of heart disease were more easily classified due to the arranged distribution. In addition, in the wkPCA the features of each type of heartbeat are arranged in each group and this makes the classification of these five types of heart disease achieving higher performance. As a consequence, choosing the parameters for designing the wavelet-based kernel and the BNN with 3 hidden layers for the classifier represents the main element for producing the higher performance in the work presented in this paper. Wavelet Packet Decomposition (WPD), kernel Principle Component Analysis (kPCA), Support Vector Machine (SVM), Wavelet-based kernel PCA (wkPCA), Backpropagation Neural Network (BNN).

Tab. 11: Accurate comparison of the proposed method and the previous one.

Authors	Method for Feature	Method Classifier	Disease No.	Accuracy (%)
Maya Kallas [16]	kPCA	SVM	3	97.39
Proposed method	WPD + wkPCA	BNN	3	98.52

Table 11 shows the comparison between two classifiers with the same types of heart disease using kernels for extracting features. In a previous research [16], the kPCA method with the σ^2 constant was applied in the classifier using the multi-SVM algorithm without extracting features of heartbeats. This led to a classifying accuracy of 97.39 %. In the work presented in this paper, we have proposed a classifier for three types of heart disease, including Normal (N), Premature Ventricular Contraction (PVC-V) and Left Bundle Branch Block (LBBB-L). Moreover, the testing dataset was proposed with 34 ECG signals, including 10 N, 10 PVC and 14 LBBB, and the wavelet-based kernel has been built based on the wavelet coefficients of three types of heart disease, which resulted in a classifying accuracy of 98.52 %.

4. Conclusion

This paper proposed a new method for classification of heart disease where a wavelet-based kernel was used. ECG signals collected from the MIT-BIH database were split after heartbeat signals to calculate wavelet coefficients. Thereafter the heartbeats were decomposed to produce the wavelet coefficients (Approximation and Detail) using the WPD algorithm. From these wavelet coefficients, the wavelet-based kernel was constructed to determine features of heartbeats using the wkPCA method for classification with the high performance. This meant that the wkPCA was applied to obtain the highest quality features of the heartbeats. Therefore, the BNN was employed for training these features for the classification of five types of heart disease.

The experimental results, based on ECG signals of nine patients, showed that the proposed method relying on the wkPCA for the classification of five types of heart disease, is effective and has high accuracy. A multi-class confusion matrix was applied to evaluate the outcome, resulting in high classification performance of 98.03 %.

Acknowledgment

The authors would like to acknowledge the support of Ministry of Education and Training, Vietnam with Grand No. B2017.SPK.03 and HCMC University of Technology and Education, Vietnam.

References

[1] XU, Y., M. LUO, T. LI and G. SONG. ECG Signal De-noising and Baseline Wander Correction Based on CEEMDAN and Wavelet Threshold. *Sensors*. 2017, vol. 17, iss. 12, pp. 2754–2770. ISSN 1424-8220. DOI: 10.3390/s17122754.

[2] LUO, Y., R. H. HARGRAVES, A. BELLE, O. BAI, X. QI, K. R. WARD and K. NAJARIAN. A Hierarchical Method for Removal of Baseline Drift from Biomedical Signals: Application in ECG Analysis. *The Scientific World Journal*. 2013, vol. 2013, iss. 1, pp. 1–10. ISSN 2356-6140. DOI: 10.1155/2013/896056.

[3] JAIN, S., V. BAJAJ and A. KUMAR. Riemann Liouvelle Fractional Integral Based Empirical Mode Decomposition for ECG Denoising. *IEEE Journal of Biomedical and Health Informatics*. 2018, vol. 22, iss. 4, pp. 1133–1139. ISSN 2168-2208. DOI: 10.1109/JBHI.2017.2753321.

[4] ANEJA, M. K. and B. SINGH. Powerline Interference Reduction in ECG Using Combination of MA Method and IIR Notch. *International Journal of Recent Trends in Engineering*. 2009, vol. 2, iss. 1, pp. 125–129. ISSN 2455-1457. DOI: 10.1.1.381.8146.

[5] SUBRAMANIAM, S. R., B. W.-K. LING and A. GEORGAKIS. Motion artifact suppression in the ECG signal by successive modifications in frequency and time. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Osaka: IEEE, 2013, pp. 425–428. ISBN 978-1-4577-0216-7. DOI: 10.1109/EMBC.2013.6609527.

[6] NAGAI, S., D. ANZAI and J. WANG. Motion artefact removals for wearable ECG using stationary wavelet transform. *Healthcare technology letters*. 2017, vol. 4, iss. 4, pp. 138–141. ISSN 2053-3713. DOI: 10.1049/htl.2016.0100.

[7] RODRIGUEZ, R. J., A. MEXICANO, J. BILA, S. CERVANTES and R. PONCE. Feature Extraction of Electrocardiogram Signals by Applying Adaptive Threshold and Principal Component Analysis. *Journal of Applied Research and Technology*. 2015, vol. 13, iss. 2, pp. 261–269. ISSN 1665-6423. DOI: 10.1016/j.jart.2015.06.008.

[8] SHARMA, L. N., S. DANDAPAT and A. MAHANTA. Multichannel ECG Data Compression Based on Multiscale Principal Component Analysis. *IEEE Transactions on Information Technology in Biomedicine*. 2012, vol. 16, iss. 4, pp. 730–736. ISSN 1558-0032. DOI: 10.1109/TITB.2012.2195322.

[9] KROPF, M., D. HAYN and G. SCHREIER. ECG classification based on time and frequency domain features using random forests. *Computing in Cardiology (CinC)*. 2017, vol. 44, iss. 1, pp. 1–4. ISSN 2325-887X. DOI: 10.22489/CinC.2017.168-168.

[10] LIN, C. C. and C. M. YANG. Heartbeat Classification Using Normalized RR Intervals and Morphological Features. *Mathematical Problems in Engineering*. 2014, vol. 2014, iss. 1, pp. 1–11. ISSN 1563-5147. DOI: 10.1155/2014/712474.

[11] MARTIS, R. J., U. R. ACHARYA and L. C. MIN. ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform. *Biomedical Signal Processing and Control*. 2014, vol. 8, iss. 5, pp. 437–448. ISSN 1746-8094. DOI: 10.1016/j.bspc.2013.01.005.

[12] GACEK, A. Data structure-guided development of electrocardiographic signal characterization and classification. *Artificial Intelligence in Medicine*. 2013, vol. 59, iss. 3, pp. 197–204. ISSN 0933-3657. DOI: 10.1016/j.artmed.2013.09.004.

[13] KHUSHABAA, R. N., S. KODAGODA, S. LALB and G. DISSANAYAKE. Uncorrelated fuzzy neighborhood preserving analysis based feature projection for driver drowsiness recognition. *Fuzzy Sets and Systems*. 2013, vol. 221, iss. 1, pp. 90–111. ISSN 0165-0114. DOI: 10.1016/j.fss.2012.12.003.

[14] QIN, Q., J. LI, L. ZHANG, Y. YUE and C. LIU. Combining Low-dimensional Wavelet Features and Support Vector Machine for Arrhythmia Beat Classification. *Scientific Reports*. 2017, vol. 7, iss. 1, pp. 6067–6079. ISSN 2045-2322. DOI: 10.1038/s41598-017-06596-z.

[15] LI, T. and M. ZHOU. ECG Classification Using Wavelet Packet Entropy and Random Forests. *Entropy*. 2016, vol. 18, iss. 8, pp. 285–301. ISSN 1099-4300. DOI: 10.3390/e18080285.

[16] KALLAS, M., C. FRANCIS, L. KANAAN, D. MERHEB, P. HONEINE and H. AMOUD. Multi-class SVM classification combined with kernel PCA feature extraction of ECG signals. In: *2012 19th International Conference on Telecommunications (ICT)*. Jounieh: IEEE, 2012, pp. 1–5. ISBN 978-1-4673-0747-5. DOI: 10.1109/ICTEL.2012.6221261.

[17] GOSHVARPOUR, A., A. ABBASI and A. GOSHVARPOUR. An accurate emotion recognition system using ECG and GSR signals and matching pursuit method. *Biomedical Journal*. 2017, vol. 40, iss. 6, pp. 355–368. ISSN 2319-4170. DOI: 10.1016/j.bj.2017.11.001.

[18] WIDJAJA, D., C. VARON, A. C. DORADO, J. A. K. SUYKENS and S. V. HUFFEL. Application of Kernel Principal Component Analysis for Single-Lead-ECG-Derived Respiration. *IEEE Transactions on Biomedical Engineering*. 2012, vol. 59, iss. 4, pp. 1169–1176. ISSN 1558-2531. DOI: 10.1109/TBME.2012.2186448.

[19] MORAN, M. E., M. C. SORIANO, I. FISCHER and C. R. MIRASSO. Electrocardiogram Classification Using Reservoir Computing With Logistic Regression. *IEEE Journal of Biomedical and Health Informatics*. 2015, vol. 19, iss. 3, pp. 892–898. ISSN 2168-2208. DOI: 10.1109/JBHI.2014.2332001.

[20] HAI, N. T., N. Q. CUONG, T. Q. D. KHOA and V. V. TOI. Temporal hemodynamic classification of two hands tapping using functional near-infrared spectroscopy. *Journal of Frontiers in Human Neuroscience*. 2013, vol. 7, iss. 1, pp. 1–12. ISSN 1662-5161. DOI: 10.3389/fnhum.2013.00516.

[21] GARCIA, G., G. MOREIRA, D. MENOTTI and E. LUZ. Inter-Patient ECG Heartbeat Classification with Temporal VCG Optimized by PSO. *Scientific Reports*. 2017, vol. 7, iss. 1, pp. 1–11. ISSN 2045-2322. DOI: 10.1038/s41598-017-09837-3.

[22] LI, H., D. YUAN, X. MA, D. CUI and L. CAO. Genetic algorithm for the optimization of features and neural networks in ECG signals classification. *Scientific Reports*. 2017, vol. 7, iss. 1, pp. 1–12. ISSN 2045-2322. DOI: 10.1038/srep41011.

[23] LI, P., Y. WANG, J. HE, L. WANG, Y. TIAN, T.-S. ZHOU, T. LI and J.-S. LI. High-Performance Personalized Heartbeat Classification Model for Long-Term ECG Signal. *IEEE Transactions on Biomedical Engineering*. 2017, vol. 64, iss. 1, pp. 78–86. ISSN 1558-2531. DOI: 10.1109/TBME.2016.2539421.

[24] SONG, C., K. LIU, X. ZHANG, L. CHEN and X. XIAN. An Obstructive Sleep Apnea Detection Approach Using a Discriminative Hidden Markov Model From ECG Signals. *IEEE Transactions on Biomedical Engineering*. 2016, vol. 63, iss. 7, pp. 1532–1542. ISSN 1558-2531. DOI: 10.1109/tbme.2015.2498199.

[25] WANG, Y., Y. S. ZHU, N. V. THAKOR and Y. H. XU. A short-time multifractal approach for arrhythmia detection based on fuzzy neural network. *IEEE Transactions on Biomedical Engineering*. 2001, vol. 48, iss. 9, pp. 989–995. ISSN 1558-2531. DOI: 10.1109/10.942588.

[26] CRIPPA, P., A. CURZI, L. FALASCHETTI and C. TURCHETTI. Multi-class ECG beat classification based on a Gaussian mixture model of Karhunen-Loève transform. *International Journal of Simulation: Systems, Science and Technology*. 2015, vol. 16, iss. 1, pp. 1–10. ISSN 1473-8031. DOI: 10.5013/IJSSST.a.16.01.02.

[27] AROTARITEI, D., H. COSTIN, A. PASARICA and C. ROTARIU. A Proposal for Cardiac Arrhythmia Classification using Complexity Measures. *Advances in Electrical and Computer Engineering*. 2017, vol. 17, iss. 3, pp. 29–34. ISSN 1336-1376. DOI: 10.4316/AECE.2017.03004.

[28] KIRANYAZ, S., T. INCE and M. GABBOUJ. Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. *IEEE Transactions on Biomedical Engineering*. 2016, vol. 63, iss. 3, pp. 664–675. ISSN 1558-2531. DOI: 10.1109/TBME.2015.2468589.

[29] LI, W. and J. LI. Local Deep Field for Electrocardiogram Beat Classification. *IEEE Sensors Journal*. 2018, vol. 18, iss. 4, pp. 1656–1664. ISSN 1558-1748. DOI: 10.1109/JSEN.2017.2772031.

[30] POURBABAEE, B., M. J. ROSHTKHARI and K. KHORASANI. Deep Convolutional Neural Networks and Learning ECG Features for Screening Paroxysmal Atrial Fibrillation Patients. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2017, vol. 48, iss. 12, pp. 2095–2104. ISSN 2168-2232. DOI: 10.1109/TSMC.2017.2705582.

[31] SAVALIA, S. and V. EMAMIAN. Cardiac Arrhythmia Classification by Multi-Layer Perceptron and Convolution Neural Networks. *Bioengineering*. 2018, vol. 5, iss. 1, pp. 1–12. ISSN 2306-5354. DOI: 10.3390/bioengineering5020035.

[32] MIT-BIH Arrhythmia Database. In: *PhysioNet* [online]. 2019. Available at: `http://physionet.org/physiobank/database/mitdb/`.

About Authors

Thanh-Nghia NGUYEN was born in Binh Dinh, Viet Nam. He received the B.Sc. and M.Sc. degrees in Electrical-Electronics Engineering from the HCMC University of Technology and Education, Viet Nam, in 2007 and 2012, respectively. Currently, he is a Ph.D. student with Electronics Engineering at the HCMC University of Technology and Education. His research interests include biomedical signal processing, biomedical image processing and artifact intelligent.

Thanh-Hai NGUYEN received his B.Sc. degree with Electronics engineering from the HCMC University of Technology and Education (Vietnam, 1995), M.Sc. one with Telecommunication and Electronics Engineering from HCMC University of Technology (Vietnam, 2002), Ph.D. degree with Electronics Engineering from University of Technology, Sydney (Australia, 2010). Currently, he is an associate professor in the Department of Industrial Electronic-Bio-medical Engineering, Faculty of Electrical-Electronics Engineering, the HCMC UTE, Vietnam. His research interests are Bio-signal and -image processing, machine learning, smart wheelchairs and Artificial intelligence.

Manh-Hung NGUYEN received the B.Sc., M.Sc. degree in electrical engineering from National University of Technology and Education, Ho Chi Minh City, Vietnam, in 2009 and 2011, respectively; and the Ph.D. degree in electrical engineering from National Kaohsiung University of Applied Sciences, Taiwan, in 2016. He is currently an Assistant Professor with the Department of Electrical Engineering, HCMC University of Technology and Education, Vietnam; and a Postdoc researcher in the Department of Electrical Engineering, National Chung Cheng University, Taiwan. His research interests are in machine learning and data analysis.

Salvatore LIVATINO received the M.Sc. degree in Computer Science from the University of Pisa, Italy, in 1993 and the Ph.D. degree in Computer Science and Engineering from Aalborg University, Denmark, in 2003. He is currently a Reader in Virtual Reality and Robotics at the University of Hertfordshire, Hatfield, U.K. He is the author of several journal and conference papers, and has contributed to many European and U.K. projects. His research interests are in virtual and augmented reality user interfaces for tele-exploration and tele-operation with focus on the use of stereocopic-3D visualization and immersive technology, with applications in the field of telerobotics, telemedicine and immersive games.