A two-stage Feature Extraction Approach for ECG signals

Conference Paper · November 2016

4 authors:

Essam H Houssein
Minia University

Moataz Kilany
Minia University

Aboul Ella Hassanien
Cairo University

Vaclav Snasel
VŠB-Technical University of Ostrava

A two-stage Feature Extraction Approach for ECG signals

Essam H. Houssein[1,4], Moataz Kilany[1,4], Aboul Ella Hassanien[2,4], and Vaclav Snasel[3]

[1] Faculty of Computers and Information, Minia University, Egypt
[2] Faculty of Computers and Information, Cairo University, Egypt
[3] Dept. of CS and IT4Innovations, VSB-TU of Ostrava, Czech Republic
[4] Scientific Research Group in Egypt (SRGE) http://www.egyptscience.net

Abstract. This paper investigate various techniques of extracting features from the electrocardiogram (ECG) signal in order to analyze the ECG signals to detect the heart disease. Feature extraction, is a one of the widespread process of decompose the ECG data. This paper introduce a two-stage feature extraction approach to extract features from ECG signals for different types of arrhythmias. Firstly, Modified Pan-Tomkins Algorithm (MPTA) is implemented to remove noise and extract nine features. Then the proposed Improved Feature Extraction Algorithm (IFEA) is applied to extract additionally ten different features from the ECG signal. The MIT-BIH arrhythmia database have been used to test the proposed approach. It is obvious from the results that the proposed approach shows a high classification in terms of the following four statistical measures: Accuracy (Ac)98.37%, Recall 48.29%, Precision 43.91%, F Measure 45.31%, and Specificity (Sp) 93.30%, respectively.

Keywords: *ECG, Feature Extraction, Wavelet Transform, Pan-Tompkins Algorithm.*

1 Introduction

The scope and recrudescence of cardiovascular diseases (CVDs) have increased in recent years. World Health Organization (WHO) refer to CDVs as the main cause of death around the world as per the recent WHO report, the overall death rates due to CVD have declined, however the encumbrance of disease still remains high [1,2]. ECG is widely used as a tool by cardiologists for determining the abnormalities of human heart. An expert medical practitioner may fail to diagnose the heart abnormalities due to the non-stationary nature of ECG signals findings in life threatening situations. In the literature, many feature extraction techniques are applied to decompose and classify ECG beats [3,4]. The three main events presented in the ECG signal of each heartbeat are: the P wave, the QRS complex, and the T wave [5]. Computer-based ECG analysis requires an accurate detection of QRS complex and R wave. Therewith, this is a non-easy

task since a real ECG signal usually contain artifacts (i.e. muscular noise, motion artifacts, and baseline drifts changes). Other components of an ECG, such as P and T waves, are also found to be high in some cases, and these waves must be identified from the QRS waves. This increases the complexity of QRS detection. False R-wave detection or the failure to detect R-waves may lead to undesired results in computer-based ECG analysis [6].
The rest of the paper is organized as follows: Section 2 presents the literature review. The materials and methods are described in Section 3. The proposed two-stage feature extraction approach is provided in Section 4. Experimental results and discussions are introduced in Section 5. The conclusion of this paper is reported in Section 6.

2 Literature Review

Feature extraction plays an important function in any classification task. The aim of feature extraction is to select and retain pertinent information from the original signals. Many researchers have expatiated the problem of feature extraction ECG heartbeat signals. The most important step in classification problems is the feature extraction since a good classifier may fail to classify the beats, if the extracted features are not proper. In the context of ECG features extraction which represents the focus of this paper, several interesting works can be found in the literature. Table 1 shows some of the previous research techniques for the feature extraction. This paper aim to introduce an optimal feature extraction algorithm (IFEA) seems useful for ECG signals.

Table 1
Review of previous studies on the feature extraction techniques (Part1).

Ref.	**Feature Extraction**	**Remark**
[7]	Wavelet transform	Proposed a feature extraction technique to extract the morphological information from ECG.
[8]	Discrete wavelet and morphology	Proposed a technique to classify ECG signal into two classes using various neural classifier.
[9]	Wavelet coefficients	Proposed an approach to detect PVCs using a neural network with weighted fuzzy functions.
[10]	Discrete wavelet transform	Introduced a supervised noise-artifact-robust heart arrhythmia fusion classification solution.
[11]	Wavelet transform	Introduced a method to discriminate between the singular points of the maternal and fetal ECGs, both present in the composite abdominal signal.

Table 1
Review of previous studies on the feature extraction techniques (Part2).

Ref.	Feature Extraction	Remark
[12]	Wavelet coefficient - Wavelet decomposition	Extracted the features heart rate variability (HRV) from RR intervals and ECG derived respiration (EDR) from R waves of QRS.
[13]	Different wavelets	Introduced the most appropriate wavelet and the choice must be done empirically by comparing results of different wavelets.
[14]	Wavelet transform	The heart beat intervals and RR intervals as feature and classifier model based on linear discriminants are used for classifying five types of ECG beats.
[15]	Discrete wavelet transform	Introduced that is the DWT features provide better recognition performance when compared to the LPC and MFCC based features.
[16]	Discrete wavelet transform	Eliminate the negative impacts of material, time and environment, these selected features are normalized with respect to the emotional mode.
[17]	Higher order statistics of wavelet packet coefficients	The results show that HOS of WPC as features are highly discriminative for the classification of different arrhythmic ECG beats.
[18]	Power spectral, Autoregressive	Presented an efficient embedded approach for feature selection and linear discrimination of EEG signals.
[19]	Discrete cosine transform	Five types of ECG beats (ANSI/AAMI EC57:1998 standard) of MITBIH arrhythmia database were automatically classified.

3 Materials and Methods

3.1 ECG Data Acquisition

The PhysioNet website is dedicated for medical data corresponding to various diseases. PhysioNet databases are built up of hundreds of medical records for digitized ECG, EEG and other types of signals. Each ECG record is annotated and revised by a number of cardiologists [20]. Many research efforts depend on MIT-BIH arrhythmias database [21] provided by PhysioNet. It consists of several ECG signal records for patients that indicate different types of diseases and abnormalities of the heart rhythms. The amplitude and duration of normal ECG is tabulated below in Table 2.
In this paper The MIT/BIH Arrhythmia Database contains 48 half-hour records obtained from 25 male and 22 female subjects studied by the BIH Arrhythmia

Table 2
ECG Waveform Parameters.

Amplitude(mV)	Duration (Seconds)
P Wave - 0.25mV	PR interval- 0.12s to 0.20s
R Wave -1.60mV	QT interval- 0.35s to 0.44s
Q Wave -25% of R wave	ST interval- 0.05s to 0.15s
T Wave -0.1 to 0.5mV	QRS interval-0.09s

Laboratory is used. The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. Two or more cardiologists independently annotated each record; disagreements were resolved to obtain the computer-readable reference annotations for each beat (approximately 110,000 annotations in all) included with the database. Our experiment conducted on the basis of ECG data (five patients), Table 3 displays the datasets employed in our experiment. Five patients that define a considerably sufficient number of beat types in order to make classification results more valid and describe several types of heartbeats.

Table 3
ECG dataset description.

N	Patient Record	Gender	Age	Types of beats (PhysioNet Standard)
1	202	Male	68	N-A-a-V-F
2	205	Male	59	N-A-V-F
3	214	Male	53	L-V-F
4	215	Male	81	N-A-V-F
5	219	Male	Unknown	N-A-V-F

3.2 Modified Pan-Tompkins Algorithm (MPTA)

The Modified Pan-Tompkins Algorithm (MPTA) [22,23] consists of derivative, moving average, squaring and threshold operations. MPTA follow these steps: first of all there is a band-pass filter which is composed of low pass and a high pass filter and it reduces noise .After this a derivative filter is used in order to get the slope information. After that an amplitude squaring is done and then the signal is passed to a moving- window integrator. Then thresholding is done to locate R-peaks. Figure 1 shows the steps of MPTA to remove different kinds of artifacts and noise.

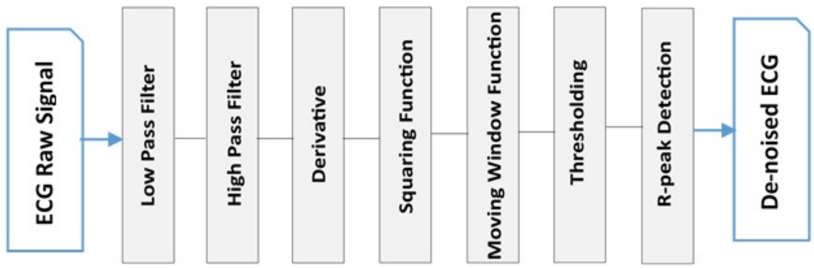

Figure 1: Modified Pan-Tompkins Algorithm Process.

3.3 Support Vector Machines (SVMs)

The classification process of ECG signals utilize many classification techniques such as Artificial Neural Network (ANN), Fuzzy Logic, and Support Vector Machine (SVM). Most researchers focused on SVM and Fuzzy Logic system for CVDs classification of ECG signals. Previous research efforts showed great performance of Support Vector Machines (SVMs). In SVM data is represented by a P-dimensional vector [24]. It makes classifications by means of Optimal Separating Hyper-planes (OSH) which achieve the greatest margin between closest data points that belong to separate classes. SVMs depend on kernels in classification process. Kernel choice greatly affect the classification performance. It is a difficult task to identify a suitable kernel for a given problem. SVM algorithm aims to find the greatest distance around a hyper-plane, which separates a positive class from a negative class.

4 Proposed Two-stage Feature Extraction Approach

The proposed approach consists of two main stages: 1) pre-processing stage, 2) two-phase feature extraction stage. Block diagram of the proposed approach is shown in Figure 2. The pre-processing stage consists of amplitude normalization and filtering of ECG signals. The embedded noises in ECG signals are removed using a band pass filter with a cutoff frequency of 422 Hz. MTPA and the IFEA are applied to extract the ECG heartbeat features.

4.1 Pre-processing Stage

Removal of different kind of artifacts from the ECG signal is the main objective of the pre-processing stage. In this paper MPTA is used to remove different kinds of artifacts and noise. First of all there is a band-pass filter which is composed of low pass and a high pass filter and it reduces noise .After this a derivative filter is used in order to get the slope information. After that an amplitude squaring is done and then the signal is passed to a moving window integrator. Finally, a thresholding technique is applied and the peaks are detected.

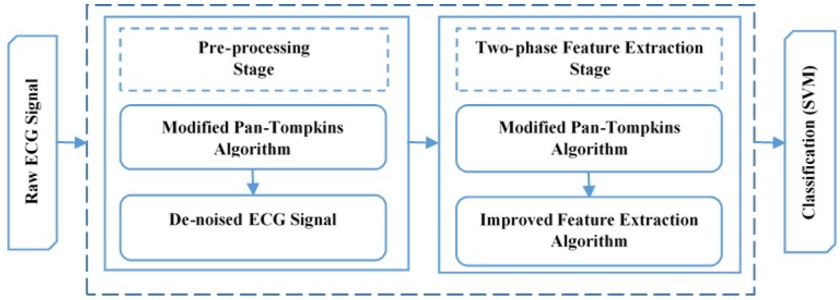

Figure 2: Block diagram of the two-stage feature extraction approach.

4.2 Two-phase Feature Extraction Stage

Phase One using MPTA: QRS complex was detected using MPTA on the de-noised ECG signal. After the detection of QRS mid-point, the ECG data was segmented such that each segment consists of 99 samples before QRS mid-point, 100 samples after QRS mid-point. Each of these 200 samples of ECG beats of five classes were used. In this paper the ECG signal parameters are extracted from the QRS complex, the ST segment, and the statistical characteristics of the signal using the MPTA.

Phase Two using IFEA: ECG signals were decomposed to the approximate information called low frequency components and detailed information called high frequency components. Decomposition of the signal is done up many steps. If these steps are high, the low-frequency components of the original signal are better conserved. Low frequency band of the ECG signal is used for the detection of QRS, T, and P waves. Algorithm 1, poses the improved feature extraction algorithm (IFEA)steps. IFEA algorithm task is to take the output of MPTA, pinpoint wave components from the results and calculate new features. MPTA output takes the form of textual annotations along with their time locations; P Start, P Wave, P End, Q, R, S, T start, T Wave and T End. The annotations take the form of "(P) (N) (t)" where "(" denotes the start of a wave form, ")" ends a wave form. P, N, T represent the P Wave, Normal Beat and a T wave respectively. N is replaced by any meaningful notation such as V. The algorithm also resolves some defects in the P-QRS-T extractor applied where some patterns of waveforms are not consistent, not complete or does not exist for their corresponding beats. In algorithm 1, step 8 extract locations of heart beat components (discussed in previous sections) defined by complete wave forms in annotation file ("(p)(N)(t)") without any missing waveform components by means of regular expressions. Steps 12 through 23 extracts wave form components from incomplete wave forms by means of regular expressions such as a QRS ("(N")) component without a "(p)" or "(t)" components. Steps 27 – 28 builds a list of heart beat types (Classification Labels), convert beat labels to numeric representations that SVM can deal with, and produce a complete data set.

Algorithm 1
Improved Features Extraction Algorithm (IFEA).

```
Input: SigTime, Ann, Sig
Output: Dataset (PQRST Time values + types)
Annotations ← "NLRBAaJSVrFejnE/fQ\?"
FullWaves ← RegExp("(p)(Annotations])(tt*)", Ann)
Current ← WaveLocs(1).
PS, P, PE, Q, R, S, TS, T, TE ← 0.
for I ← 2tosize(FullWaves) do
    Add FullWaves.Time to PS, P, PE, Q, R, S, TS, T, TE
    if FullWaves(i) − Loc > 10 or WaveLocs(i) − Loc < 9 then
        PWave ← RegExp(" \W?p\W?", Current : WaveLocs(i))
        QRS ←RegExp("\W?Annotations\W?)", Current : WaveLocs(i))
        TWave ←RegExp("\W?tt?\W?)",Current : WaveLocs(i))
        for j ← 1 to size(PWaves) do
            Add Pwaves.Time to PS, P, PE
        end for
        for j ← 1 to size(QRS) do
            Add QRS.Time to Q, R, S
        end for
        for j ← 1 to size(TWave) do
            Add TWave.Time to TS, T, TE
        end for
    end if
end for
for I ← 1 to size(R) do
    Add Ann(R) to Labels.
end for
LabelsNum ← ConvertToNumeric(Labels).
DataSet ← Combine(PS, P, PE, Q, R, S, TS, T, TE, Labels).
EXIT
```

5 Experiments and Results

The simulation results have been drawn using MATLAB R2014a.The experimental setup of dataset, their training and testing data beats are explained. SVMs was trained and tested using Radial Basis Function (RBF), Penalty Optimization Range [1–100], Gamma Optimization Range [0–1000], Data Scaling Factor [-1 – 1] and we applied 3-fold Leave One Out cross validation on all dataset.

5.1 Performance Measurements

Classification performance was evaluated in terms of five measures, which are: (1) the accuracy, (2) specificity (Sp), (3) F Measure, (4) Precision, and (5) Recall. Performance measures in general depend on four main metrics of a binary classification result (positive / Negative); True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN). Performance measures are calculated from the following Equations.
Accuracy (Ac): Overall accuracy of the classifier has been defined as:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} * 100 \tag{1}$$

Specificity (Sp), also known as true negative rate, measures the proportion of negative predictions that are correctly identified as negative:

$$Sp = \frac{TN}{TN + FP} * 100 \tag{2}$$

Recall known as the true positive rate (TPR) or Sensitivity (Se), measures the proportion of positives that are correctly identified as positive.

$$TPR = \frac{TP}{TP + FN} * 100 \tag{3}$$

Precision, known as Positive Predictive Value (PPV), is the fraction of retrieved instances that are relevant.

$$PPV = \frac{TP}{TP + FP} * 100 \tag{4}$$

F Measure, the harmonic mean of precision and recall.

$$F = 2 * \frac{PPV * TPR}{PPV + TPR} \tag{5}$$

The problem being solved in this paper is not a binary classification, so we extract the TP, FP, TN, and FN measures by means of a confusion matrix that is constructed for a classification test.

5.2 Results and Discussion

MPTA is used to remove different kinds of artifacts and noise in the pre-processing stage. The noise of ECG signal contains power line interference, baseline wandering effect, and muscle noise. The common noise of raw ECG signals is shown in Figure 3. In the second stage the common feature extraction of ECG signals is shown in Figure 4 using the MPTA. MPTA was employed to extract nine heartbeat wave characteristic features (1 through 9) as shown in Table 4, which represent detailed features for the previously described P, Q, R, S, T wave forms.

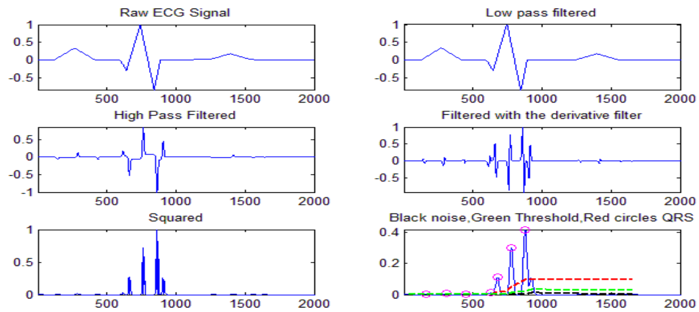

Figure 3: Shows Step by step output of MPTA.

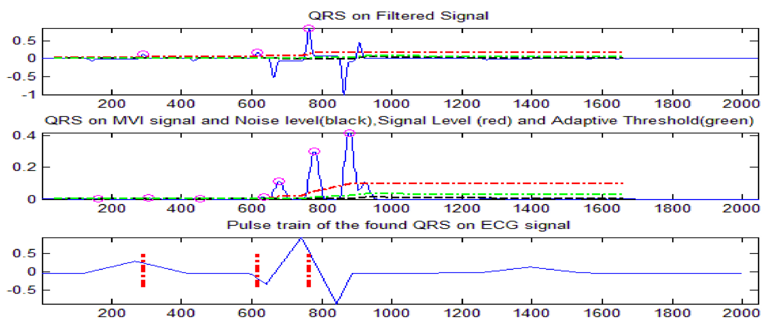

Figure 4: Shows Feature Extraction of MPTA.

The additional ten features (10 through 19) are extracted based on IFEA. However, the nineteen features extracted using MPTA and IFEA are depicts in Table 4. Table 5 summarizes for each patient record along with classification accuracy, precision, recall, F-measure, specificity abbreviated to "Acc", "Prec", "Rec","F" and "Spec" respectively. Also for each patient the best accuracy obtained (presented in bold face font) on the test beats by feeding the SVM classifier with the features generated by the proposed approach.
Figure 5,shows the results for each record and the visual comparison between the best results obtained for each patient. In general,the results show that the proposed approach have achieved a classification accuracy of 98.37%. These results show that, compared to reported results in the literature, the proposed approach designed in this study provides above satisfactory performance. However, it should be noted due to the varieties in the related works in the literature as mentioned previously, providing a completely fair and objective comparison is very difficult. The proposed approach can be put into practice easily in any classification system. This approach also has the advantage of working with a small dataset.

Table 4
Heartbeats features extracted by MPTA and IFEA.

N	Feature	Meaning
1	PS	Beginning location of P wave form.
2	P	Peak location of P wave form.
3	PE	End location of P wave form.
4	Q	Beginning of QRS complex.
5	R	R peak of QRS complex.
6	S	End of QRS complex.
7	TS	Beginning of T wave form.
8	T	Peak of T wave form.
9	TE	End of T wave form.
10	QRS Complex	QRS = S Q.
11	P-R segment	P_RSeg = Q PE.
12	P-R Interval	P_RInt = Q PS.
13	S-T Segment	S_TSeg = TS S.
14	Q-T Interval	TE Q
15	R-R Interval	RNext R
16	P-P Interval	PNext P.
17	R-R / P-P similarity	RR-PPSim = ABS(R-R P-P).
18	R-R Variance	Var (R-R).
19	Heartbeat	60/R-R

Table 5
Classification Results.

Patient	Acc%	Prec%	Rec%	F%	Spec%
202	97.28%	49.52	34.03%	38.59%	86.49%
205	98.79%	47.77%	45.12%	45.92%	92.31%
214	97.70%	48.20%	45.98%	46.96%	96.00%
215	98.81%	47.41%	46.08%	46.71%	95.85%
219	**99.26%**	48.57%	48.36%	48.37%	95.84%

According to [18], the accuracy obtained on the test beats by applying the SVM classifier with the features generated by the Continuous Wavelet Transform (CWT), Principal Component Analysis (PCA), Power Spectral Density (PSD), Autoregressive (AR) and Hjorth parameters are 71.42%, 80%, 73.57%, 72.86% and 65.71% respectively. For comparison, the accuracy obtained by the SVM classifier with the features generated by the proposed approach is 99.26%, we note that the proposed approach outperforms these standard wavelets feature extraction techniques. In [17], the performance in term of specificity is 98% using a k-Nearest neighbor (k-NN) classifier based on Wavelet packet decomposition (WPD) and Discrete Wavelet Transform (DWT) feature extraction techniques outperforms the proposed approach in this paper by 2%.

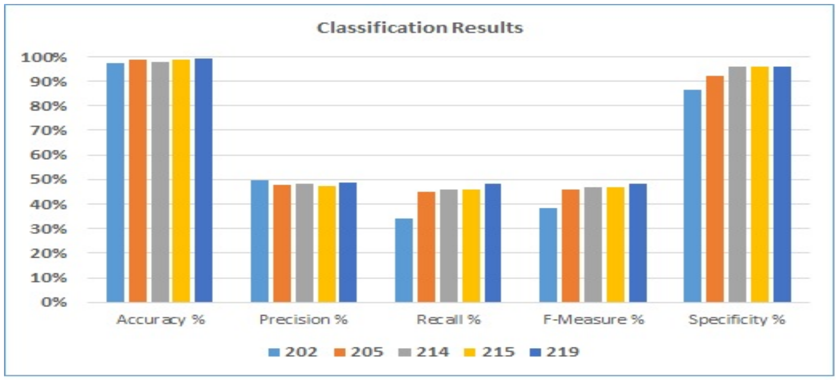

Figure 5: Shows summary of experimental results.

6 Conclusion

Two-stage Feature extraction approach consists of pre-processing stage and two-phase feature extraction stage have been proposed. In the first stage, applying MPTA to remove the noise of ECG signals. The second stage contains two phases, in the first phase, MPTA extract nine features and in the second stage, the proposed Improved Feature Extraction Algorithm (IFEA) implemented to extract additional features than what the MPTA has been able to extract. It can be drawn that the proposed algorithm has come up with more than twice the features the MPTA can extract. Results show that the suggested algorithm is more successful in extracting different types of heartbeats.

References

1. S. S. Anand and S. Yusuf, "Stemming the global tsunami of cardiovascular disease," *The Lancet*, vol. 377, no. 9765, pp. 529–532, 2011.
2. M. Elgendi, B. Eskofier, S. Dokos, and D. Abbott, "Revisiting qrs detection methodologies for portable, wearable, battery-operated, and wireless ecg systems," *PloS one*, vol. 9, no. 1, p. e84018, 2014.
3. B. Mali, S. Zulj, R. Magjarevic, D. Miklavcic, and T. Jarm, "Matlab-based tool for ecg and hrv analysis," *Biomedical Signal Processing and Control*, vol. 10, pp. 108–116, 2014.
4. D. Cvetkovic, E. D. Übeyli, and I. Cosic, "Wavelet transform feature extraction from human ppg, ecg, and eeg signal responses to elf pemf exposures: A pilot study," *Digital signal processing*, vol. 18, no. 5, pp. 861–874, 2008.
5. M. Hasan and M. Mamun, "Hardware approach of r-peak detection for the measurement of fetal and maternal heart rates," *Journal of applied research and technology*, vol. 10, no. 6, pp. 835–844, 2012.
6. B. Kohler, C. Hennig, and R. Orglmeister, "Qrs detection using zero crossing counts," *Applied genomics and proteomics*, vol. 2, no. 2, pp. 138–145, 2003.
7. T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ecg signals," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 5, pp. 1415–1426, 2009.

8. H. M. Rai, A. Trivedi, and S. Shukla, "Ecg signal processing for abnormalities detection using multi-resolution wavelet transform and artificial neural network classifier," *Measurement*, vol. 46, no. 9, pp. 3238–3246, 2013.
9. J. S. Lim, "Finding features for real-time premature ventricular contraction detection using a fuzzy neural network system," *IEEE Transactions on Neural Networks*, vol. 20, no. 3, pp. 522–527, 2009.
10. M. R. Homaeinezhad, S. Atyabi, E. Tavakkoli, H. N. Toosi, A. Ghaffari, and R. Ebrahimpour, "Ecg arrhythmia recognition via a neuro-svm–knn hybrid classifier with virtual qrs image-based geometrical features," *Expert Systems with Applications*, vol. 39, no. 2, pp. 2047–2058, 2012.
11. A. Khamene and S. Negahdaripour, "A new method for the extraction of fetal ecg from the composite abdominal signal," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 4, pp. 507–516, 2000.
12. A. H. Khandoker, M. Palaniswami, and C. K. Karmakar, "Support vector machines for automated recognition of obstructive sleep apnea syndrome from ecg recordings," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 1, pp. 37–48, 2009.
13. L. Senhadji, G. Carrault, J. Bellanger, and G. Passariello, "Comparing wavelet transforms for recognizing cardiac patterns," *IEEE Engineering in Medicine and Biology Magazine*, vol. 14, no. 2, pp. 167–173, 1995.
14. P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ecg morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1196–1206, 2004.
15. N. S. Nehe and R. S. Holambe, "Dwt and lpc based feature extraction methods for isolated word recognition," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2012, no. 1, pp. 1–7, 2012.
16. H.-Y. Lin, S.-Y. Liang, Y.-L. Ho, Y.-H. Lin, and H.-P. Ma, "Discrete-wavelet-transform-based noise removal and feature extraction for ecg signals," *IRBM*, vol. 35, no. 6, pp. 351–361, 2014.
17. Y. Kutlu and D. Kuntalp, "Feature extraction for ecg heartbeats using higher order statistics of wpd coefficients," *Computer methods and programs in biomedicine*, vol. 105, no. 3, pp. 257–267, 2012.
18. G. Rodríguez-Bermúdez, P. J. García-Laencina, J. Roca-González, and J. Roca-Dorda, "Efficient feature selection and linear discrimination of eeg signals," *Neurocomputing*, vol. 115, pp. 161–165, 2013.
19. R. J. Martis, U. R. Acharya, C. M. Lim, and J. S. Suri, "Characterization of ecg beats from cardiac arrhythmia using discrete cosine transform in pca framework," *Knowledge-Based Systems*, vol. 45, pp. 76–82, 2013.
20. Z. Long, G. Liu, and X. Dai, "Extracting emotional features from ecg by using wavelet transform," pp. 1–4, 2010.
21. G. B. Moody, R. G. Mark, and A. L. Goldberger, "Physionet: a web-based resource for the study of physiologic signals," *IEEE Eng Med Biol Mag*, vol. 20, no. 3, pp. 70–75, 2001.
22. J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE transactions on biomedical engineering*, no. 3, pp. 230–236, 1985.
23. P. S. Kritika Bawa, "R-peak detection by modified pan-tompkins algorithm," *International Journal of Advancements in Research & Technology*, vol. 3, no. 5, pp. 30–33, 2014.
24. S. Karpagachelvi, M. Arthanari, and M. Sivakumar, "Classification of electrocardiogram signals with support vector machines and extreme learning machine," *Neural Computing and Applications*, vol. 21, no. 6, pp. 1331–1339, 2012.