Obstructive Sleep Apnea Diagnosis Using K-means Clustering Algorithm and Wavelet Transform

A. Aziz Kalteh[*1]

Abstract – The detection of obstructive sleep apnea (OSA) has turned out to be a hot study topic because of the excessive danger of this illness. In this paper, some effective and low-cost computational signal processing techniques are studied and their effects are compared with current achievements in OSA diagnosis. Two-tree complex wavelet transform (DT-CWT) is used to extract the feature coefficients. Multi-cluster feature selection (MCFS) algorithm has been used to select features. The remaining functions are implemented to a hybrid "okay-approach, RLS" RBF network. The obtained results show that the detection accuracy of the method presented in this article is about 96%, while the complexity of this method is much less than SVM-based techniques.

Keywords: Classification, Clustering, Obstructive sleep apnea, Wavelet-Transform, Symptom clusters.

1. Introduction

There is a close relationship between the pulse rate and the breathing pattern. Therefore, electrocardiogram (ECG) signals can be used to locate respiratory problems. Obstructive sleep apnea (OSA) is one of the common and dangerous breathing disorders that appear during moments of sleep, which can be detected using signal processing methods from ECG signals [1]. Using the ECG signals in OSA detection is only one way to try this task and the apnea may be detected by using the respiration and other types of alerts. However, in this paper, the simplest method is considered to OSA detection by using the single-lead ECG. Several various methods of diagnosing OSA have been proposed so far. Most of these techniques include feature extraction, feature selection and classification [2-12]. Khandokeret al. have proposed using the wavelet remodel for ECG function extraction [3]. furthermore, Rachimet al., Zarei et al., Avcı and Akbaş et al., and lots of different researchers have proposed the discrete wavelet transform (DWT)-based totally ECG decomposition for the OSA detection [4-9]. Moreover, the Tunable Q-factor wavelet transformation and twin-tree complex wavelet transformation (DT-CWT) have been proposed to extract the transform coefficients from the ECG signals [13-14]. After collecting the remodeling capabilities, the usual route is to extract the statistical features from those coefficients. In this research, DT-CWT method is used to extract features. To reduce the computational complexity of the proposed method, the characteristic reduction is vital. Zarei et al.have used the sequential ahead characteristic choice approach, while the principal factor analysis had been advised in Avcı and Akbaş et al and Rachim et al.[5,8-9]. In this paper, the multi-cluster function choice (MCFS) method for the characteristic reduction as lots as feasible for the quality consequences is proposed.

The final a part of the OSA detection method is the category. Many researchers have proposed the usage vector machines (SVMs) for classifying among the apnea and regular ECG signal. The proposed classifier in this article is the Hybrid Radial basis characteristic (RBF) community with the “okay-manner-recursive least-squares (RLS)” mastering set of rules. The selection of this community is to evaluate its results with that of the SVM networks within the OSA detection. These networks have been compared earlier than in other responsibilities, and it has been shown that the hybrid RBF network is advanced to the SVM community.

1* Corresponding Author , Department of Electrical Engineering, Aliabad Katoul Branch, Islamic Azad University, Aliabad Katoul, Iran. Email: Kalte@aliabadiau.ac.ir.......

Received: 2024.09.09; Accepted: 2024.10.04

2. The ECG Signal

In this section, preliminary processing of ECG signal for OSA diagnosis is considered. First, the database used in this article is introduced, then the pre-processing and signal preparation techniques are described.

2.1 Data Base

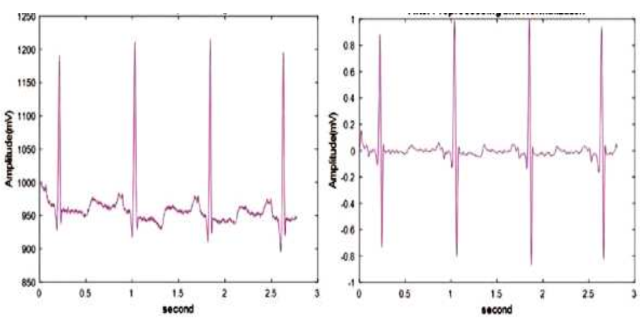
In this paper, a study has been done on ECG signals taken from the Physionet database, which includes information on 70 people during sleep. Of these data, 35 records are for training with 13 healthy subjects (normal with apnea-hypo-apnea index ≤ 5) and 22 participants with apnea. For the test set, we have 35 records with 12 healthy cases and 23 cases with apnea. Apnea or healthy condition of each part is also mentioned in this text file. Based on this, the results are presented in Table 1, assuming section by section (minute by minute).

Table 1. List of nonlinear features extracted from two-tree complex wavelet transform coefficients.

Features	Description
FE	Fuzzy entropy
APEN	Approximate Entropy
IQR	Inter Quartile Range
RP	Recurrence Plot
SD1, SD2, SD1/SD2	Poincare Plot

2.2 Signal Preparation

Before processing the ECG signal, the interference of the power line and the wandering of the base line must be removed first. For this purpose, a Chebyshev band-pass filter with a frequency range of 0.5-48 Hz has been used. Figure 2 shows a typical ECG signal before and after preprocessing and normalization.



(a) **(b)**

Fig. 1. The first second of normalized electrocardiogram signal (a)before and (b)after preprocessing

Not all parts of the recorded ECG signal are useful for diagnosing OSA. Some parts become useless due to patient movements and other parts may be contaminated by high noise levels. After filtering, weight calculation method is applied to remove noisy parts.

3. The twin-tree complicated wavelet transform

The main drawback of DWT-based feature extraction in 1D ECG signal analysis is the lack of shift change. This means that the amplitude of the wavelet coefficients changes significantly as the input signal is slightly shifted. This happens due to the down-sampling operation at each level. A better way to achieve change invariance is to implement a non-deterministic form of a binary filter tree. However, this method has heavy computational requirements and high redundancy in DT-CWT output. It solves this problem with a redundancy factor for a 1D signal that is significantly lower than non-deterministic DWT. DT-CWT implements two trees of real filters, tree A and tree B, as shown in Figure 3. The ECG signal, X(n), is run on the same data in parallel using two critical sampling DWTs. The filters are designed in such a way that the upper DWT sub-band signals can be interpreted as the real part of a CWT and the lower DWT sub-band signals as the imaginary part. When the transformation is designed in this way, the DT-DWT is almost invariant unlike the critical sampling DWT. The length of the filters implemented in each step is 10. The selected transformation coefficients are x1a, x01a, x001a, x000a, x1b, x01b, x001b, and x000b.

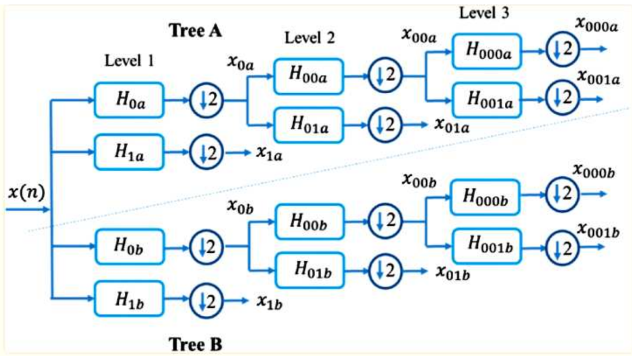


Fig. 2. The three level dual-tree complex wavelet transform

In Figures 3 and 4, the sub-band signals are depicted for three tree levels A and B, respectively. It is important to mention that all these signals are for record a01 of Physionet database.

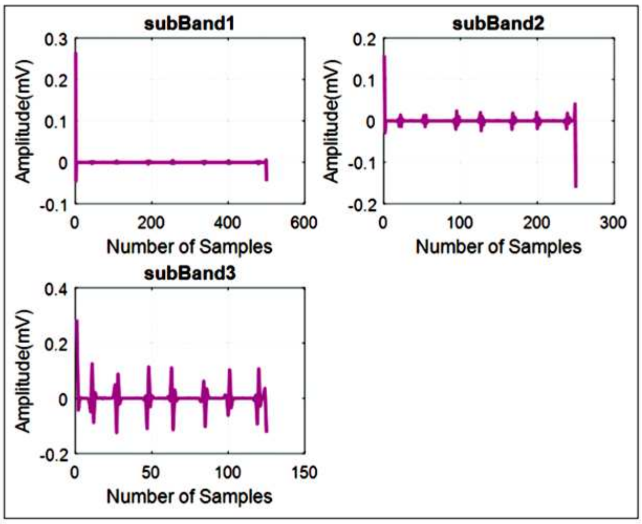


Fig. 3. The Electrocardiogram signal sub-bands for tree A

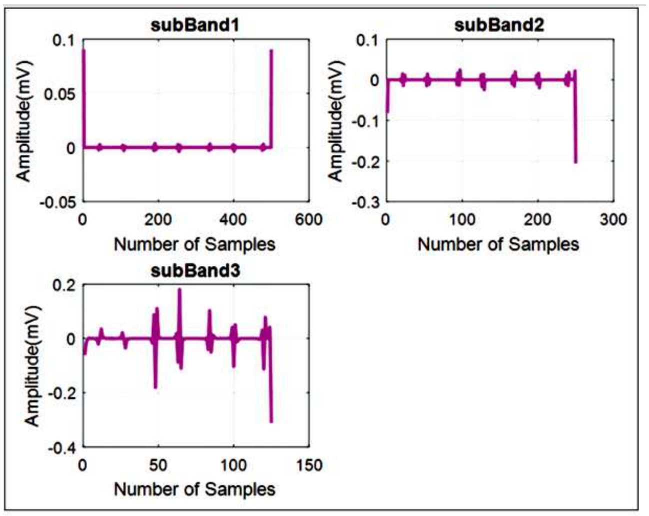


Fig. 4. The Electrocardiogram signal sub-bands for tree B

4. Radial base function classification

The SVM method is one of the most common classifiers in the field of diagnosis and classification of diseases. On the other hand, RBF networks are not used as much as SVM. The hybrid RBF network is a solution for this, because it can compete with SVM. This method consists of three layers, and the middle and output layers work with K-means and RLS algorithms, respectively, and for this reason, they are attributed the combined attribute. In this section, the RBF classifier with hybrid learning scheme, which is the proposed classification tool, is described. The proposed RBF method is called hybrid because it has a hybrid learning method with two steps as follows:

Step 1: Implements the okay-approach clustering set of rules to training the hidden layer in an unmanaged scheme. Usually, the number of clusters and computing units in the hidden layer is significantly smaller than the training sample.

Step 2: implements the RLS algorithm (or other adaptive algorithm) to determine the weight vector of the linear output layer.

The two-stage design method has some desirable features such as low computational complexity and fast convergence. As we mentioned, the RBF network consists of three layers.

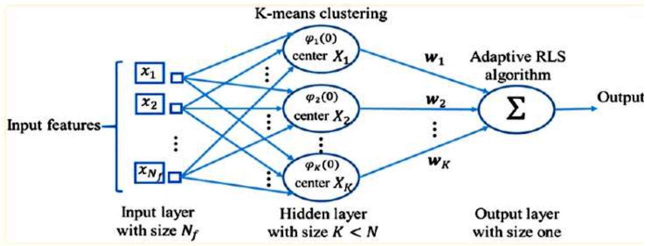


Fig. 5. The A combined radial basis function network for the diagnosis of obstructive sleep apnea

As shown in Figure 5, the input layer consists of the source nodes that connect the network to its inputs. The inputs to the network are the feature vectors for classification. The second layer, consisting of hidden units, performs a non-linear transformation from the input space to the hidden (feature) space. For most applications, the dimensions of the hidden layer of the network are large. In this layer, the data is trained in an unsupervised manner. Each unit in the hidden layer is mathematically described by an RBF:

$$\varphi_j(x) = \varphi(x - x_j)j = 1,2.....N \qquad (1)$$

The jth input data point xj defines the center of the RBF and the vector x is the signal applied to the input layer. Therefore, the links connecting the source nodes to the hidden units are direct and un-weighted links. There are several RBFs to use in the hidden layer, but we implement the Gaussian function to compare between SVM and RBF. The output layer is linear and provides the response of the network to the activation pattern implemented in the input layer. In this layer, the data is trained in a supervised manner using a hybrid method. There is no limit to the size of the output layer, except that, typically, the size of the output layer is much smaller than the hidden layer.

5. Results and Discussion

Evaluation of proposed methods for OSA detection is usually based on the accuracy and complexity of the signal processing techniques used in each part of the task. As previously emphasized, feature extraction, reduction and detection are the main parts of OSA diagnosis. Some of the proposed methods have poor results in diagnosis while having less computational complexity. Other methods have been proposed that have satisfactory results but also have high computational complexity. In this research, the comparison between computational complexity and accuracy in apnea diagnosis has been considered. We claim that the method proposed in this research is both accurate and has appropriate complexity.

Obviously, after feature extraction, we get 56 features from each ECG signal (7 sub-bands and 8 nonlinear features). One of the contributions of this article is the use of MCFS feature reduction algorithm in apnea diagnosis. By using this algorithm, the number of features of ECG signals is reduced to ten features. The results of OSA diagnosis based on the accuracy, sensitivity and specificity of the proposed method are presented as follows:

$$Accuracy\,(ACC) = \frac{Tp + TN}{TP + TN + FP + FN}$$

$$Sensititivity\,(Sen) = \frac{TP}{TP + FN} \qquad (2)$$

$$Specificity\,(Spec) = \frac{TN}{TN + FP}$$

Where TP, TN, FP and FN represent true positive, true negative, false positive and false negative respectively. The results of the proposed method show that, compared to other methods presented in various articles, this method has a more suitable performance and is a very suitable competitor for the CNN method, while it has less complexity in diagnosis. By comparing the results, it can be seen that an average of 3% improvement was achieved in all performance measures. In addition, the computational complexity of the proposed classifier is at least 30% lower than that of the SVM classifier.

5. Conclusion

In this paper, we considered OSA detection using different signal processing techniques to compare the results with previously proposed methods. Feature extraction in this paper is based on the nonlinear properties of DT-CWT coefficients. After extracting the feature, in order to reduce the computational complexity, we used the MCFS algorithm, which reduced the size of the feature vector to 10. In addition, the proposed method has less computational complexity, which makes it a powerful competitor to expensive but very computationally accurate DNNs and CNNs.

References

[1] Hilmisson H, Lange N, Duntley SP. "Sleep apnea detection: Accuracy of using automated ECG analysis compared to manually scored polysomnography (apnea hypopnea index)", *Springer, Sleep and Breathing*, p.23:125–33. 2019.

[2] Janbakhshi P, Shamsollahi MB "Sleep apnea detection from single-lead ECG using features based on ECG-derived respiration (EDR) signals", *Elsevier, IRBM*, vol.39: p.206-18, 2018.

[3] Khandoker AH, Gubbi J, Palaniswami M. "Automated scoring of obstructive sleep apnea and hypopnea events using short-term electrocardiogram recordings." *IEEE Trans Inf Technol Biomed*. Vol.13: p.1057–67. 2009.

[4] Zarei A, Asl BM. "Automatic detection of obstructive sleep apnea using wavelet transform and entropy-based features from single-lead ECG signal". *IEEE J Biomed Health Inform.*;vol.23:p.1011–21. 2019.

[5] Hassan AR, Bashar SK, Bhuiyan MI. "Computerized Obstructive Sleep Apnea Diagnosis from Single-Lead ECG Signals Using Dual-Tree Complex Wavelet Transform." *IEEE Region Humanitarian Technology Conference (R-HTC, Dhaka, Bangladesh)* 21-23 December. 2017

[6] A Nishad, RB Pachori, UR Acharya, "Application of TQWT based filter-bank for sleep apnea screening using ECG signals", *Journal of Ambient Intelligence and Humanized Computing, Springer*, 2024

[7] Avcı C, Akbaş A. "Sleep apnea classification based on respiration signals by using ensemble methods", *Bio-Medical Materials and Engineering*, vol. 26, no. s1, pp. S1703-S1710, 2015

[8] Rachim VP, Li G, Chung WY. "Sleep apnea classification using ECG-signal wavelet-PCA features", *Bio-Medical Materials and Engineering*, vol. 24, no. 6, pp. 2875-2882, 2014

[9] Hassan AR, Haque MA. "An expert system for automated identification of obstructive sleep apnea from single-lead ECG using random under sampling boosting", *Neurocomputing*, Vol. 235, P. 122-130. 2017

[10] Hassan AR, Haque MA. "Computer-aided obstructive sleep apnea screening from single-lead

electrocardiogram using statistical and spectral features and bootstrap aggregating". *Biocybernetics and Biomedical Engineering*.vol.36, p.256–66. 2016

[11] Wang T, Lu C, Shen G, Hong F. 2019. "Sleep apnea detection from a single-lead ECG signal with automatic feature-extraction through a modified LeNet-5 convolutional neural network". *PeerJ* 7:e7731

[12] Urtnasan, E., Park, JU., Joo, EY. et al. "Automated Detection of Obstructive Sleep Apnea Events from a Single-Lead Electrocardiogram Using a Convolutional Neural Network". *J Med Syst* 42, 104, 2018.

[13] Thomas M, Das MK, Ari S. "Automatic ECG arrhythmia classification using dual-tree complex wavelet based features". Int J Electron Commun. vol.69, p.715–21, 2015.

[14] Papini GB, Fonseca P, Margarito J, van Gilst MM, Overeem S, Bergmans JWM, et al. "On the generalizability of ECG-based obstructive sleep apnea monitoring: Merits and limitations of the Apnea-ECG database". *Conf Proc IEEE Eng Med Biol Soc*.2018:6022–5.