Classification of electrocardiogram and auscultatory blood pressure signals using machine learning models

Manjeevan Seera [a,*], Chee Peng Lim [b], Wei Shiung Liew [c], Einly Lim [c], Chu Kiong Loo [a]

[a] Faculty of Computer Science and Information Technology, University of Malaya, Malaysia
[b] Centre for Intelligent Systems Research, Deakin University, Australia
[c] Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, Malaysia

ARTICLE INFO

Article history:
Available online 20 December 2014

Keywords:
Machine learning
Data classification
Medical signals
Electrocardiogram
Auscultatory blood pressure

ABSTRACT

In this paper, two real-world medical classification problems using electrocardiogram (ECG) and auscultatory blood pressure (Korotkoff) signals are examined. A total of nine machine learning models are applied to perform classification of the medical data sets. A number of useful performance metrics which include accuracy, sensitivity, specificity, as well as the area under the receiver operating characteristic curve are computed. In addition to the original data sets, noisy data sets are generated to evaluate the robustness of the classifiers against noise. The 10-fold cross validation method is used to compute the performance statistics, in order to ensure statistically reliable results pertaining to classification of the ECG and Korotkoff signals are produced. The outcomes indicate that while logistic regression models perform the best with the original data set, ensemble machine learning models achieve good accuracy rates with noisy data sets.

1. Introduction

Data classification constitutes one of the fundamental requirements in undertaking many decision-making tasks (Örkcü & Bal, 2011). A classification task involves building a model that depicts a mapping from the input feature space to the target output space (Oza & Tumer, 2008). In general, there are a number of classification methods, which include statistical methods, mathematical programming methods, and a variety of machine learning methods (Örkcü & Bal, 2011). Researchers in the medical domain have used many methods to perform data classification. Methods with higher classification accuracy are desirable to correctly identify potential diseases; therefore improving diagnosis accuracy (Fan, Chang, Lin, & Hsieh, 2011).

The main contribution of this study is a comprehensive performance evaluation and analysis pertaining to a number of machine learning models for undertaking real medical data classification problems. Specifically, we use two sets of real data collected from patients, i.e., the electrocardiogram (ECG) and auscultatory blood pressure (Korotkoff) signals. ECGs are signals related to electrical activity of the heart, which can be recorded by placing surface electrodes on a patient's body (Mitra, Mitra, & Chaudhuri, 2006). It is an effective non-invasive clinical tool for the diagnosis of certain cardiovascular diseases, and it provides useful information pertaining to pathological physiology of heart activity (Chen & Yu, 2012). ECG signals carry valuable information about the heart function, and provide a cardiologist with useful insight about the rhythm and functioning of the heart (Chen & Yu, 2012).

As stated in Mele (2008), an estimated 300 million ECGs are performed each year. As such, there is a clear need for reliable and accurate interpretation tools of ECG readings. Although trained cardiologists can discover different cardiac abnormalities in ECG recordings, it is time-consuming and laborious for them to examine a large number of ECG recordings (Kiranyaz, Ince, Pulkkinen, & Gabbouj, 2011). Moreover, visual inspection can take considerable time, and some vital information can be neglected due to fatigue in carrying out the tedious manual procedure (Sun, Lu, Yang, & Li, 2012). As such, automated tools to help accurately analyze a large number of ECG data samples are required. Similarly, blood pressure (BP) is an established link in determining coronary heart disease and cardiovascular incidents (Mendiola, Luna, Guerra, & Ramírez, 2013). The most commonly used methods for measuring BP in clinical activities is the use of a manual sphygmomanometer and a stethoscope to detect the Korotkoff sounds (Kurl et al., 2001). Korotkoff waveforms are one of the most reliable means for monitoring blood pressure (Mendiola et al., 2013). To exploit the advantages of using computerized tools to help medical prognosis

* Corresponding author.
E-mail addresses: mseera@gmail.com (M. Seera), chee.lim@deakin.edu.au (C.P. Lim), liew.wei.shiung@gmail.com (W.S. Liew), einly_lim@um.edu.my (E. Lim), ckloo.um@um.edu.my (C.K. Loo).

http://dx.doi.org/10.1016/j.eswa.2014.12.023
0957-4174/

and diagnosis tasks, an empirical study to evaluate a variety of machine learning models for classification of both ECG and Korotkoff signals is undertaken in this study.

In respect to machine learning models, the artificial neural networks (ANNs) are popular methods for tackling medical diagnosis problems (Al-Shayea, 2011). There are a number of advantages of using ANNs for medical data classification, e.g. a detailed mathematical model that relates the input features and the target outputs is not necessary. In addition, ANNs have the ability to learn complex relationship in data samples. Recent advancements in ANNs have shown their usefulness in analyzing signals, which has opened up the possibility of solving problems typically not possible with some existing signal processing techniques (Übeyli & Güler, 2005). On the other hand, model-based signal processing is a new method to describe physiological systems (Porta, Baselli, & Cerutti, 2006). The method is useful for short-term cardiovascular control and analysis of cardiovascular regulation mechanisms (Porta et al., 2006). Besides that, other medical signal processing methods have been used in many different application, e.g. in synthesized and real biomedical signals using frequency-domain methods (Mitov, 1998), medical ultrasound signals using a fast wavelet-based edge method (Nes, 2012), and a multi-sensor fetal movement detection system based on a time–frequency signal processing method (Boashash, Khlif, Ben-Jabeur, East, & Colditz, 2014). A review of different machine learning and related methods for medical applications is presented in Section 2.

The organization of this paper is as follows. After a literature review in Section 2, a description on signal pre-processing, experimental setup, and the background of different classifiers used for experimentation is presented in Section 3. In Section 4, two real medical case studies consisting of ECG and Korotkoff signals are detailed. Finally, conclusions and suggestions for further work are presented in Section 5.

2. Literature review

Biomedical signal processing is becoming an essential feature in many advanced medical equipment, and is widely used in clinical and biomedical research (Simpson, De Stefano, Allen, & Lutman, 2005). In this review, a number of statistical methods, machine learning models, and other related techniques for medical signal processing are reviewed. The details are as follows.

2.1. Statistical models

A Bayesian-based classifier was used in classifying patients according to statistical features extracted from their ECG signals (Wiggins, Saad, Litt, & Vachtsevanos, 2008). Based on 12 extracted features, the Bayesian network produced an accuracy rate of 86.25% (Wiggins et al., 2008). Another Bayesian-based method was utilized in analyzing interval ECG signals (Lee, McManus, Bourrell, Sörnmo, & Chon, 2013). A high resolution time–frequency spectral method was developed to process atrial activities. The method produced an accuracy rate of 88% (Lee et al., 2013). In Atoui, Fayn, and Rubel (2010), multiple regression-based methods were used to process ECGs from cardiac patients, while in Iacoviello et al. (2007), a linear regression analysis was conducted for ECG monitoring. No accuracy rates were reported for the studies by Atoui et al. (2010) and Iacoviello et al. (2007).

2.2. Artificial neural networks

Benchmark ECG data sets from the Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) repository is popularly used in the literature, as explained in the section. A multilayer perceptron (MLP) model with backpropagation was used in classification of ECG arrhythmias in Özbay, Ceylan, and Karlik (2011). Based on the MIT-BIH data sets, wavelet transform was performed for feature extraction (Özbay et al., 2011). The MLP model was able to produce a high accuracy (99%) rate (Özbay et al., 2011). A new ANN model with adaptive activation functions to classify ECG arrhythmias was proposed in Özbay and Tezel (2010). The activation functions were used in hidden neurons, in an attempt to improve the performance of the classical MLP model (Özbay & Tezel, 2010). Using the MIT-BIH data sets, an accuracy rate of 98.19% was obtained (Özbay & Tezel, 2010).

An evolvable block-based ANN was used for personalized ECG heartbeat classification in Jiang and Kong (2007). The input comprised the Hermite transform coefficients and the time interval between two neighboring R-peaks of ECG signals (Jiang & Kong, 2007). Based on the MIT-BIH data sets, accuracy rates of 96.6–98.1% were achieved (Jiang & Kong, 2007). Feedforward and fully connected ANN models were deployed in a classification system designed for robust and accurate detection of ECG heartbeat patterns (Ince, Kiranyaz, & Gabbouj, 2009). The MIT-BIH data sets were evaluated with a morphological wavelet transform method for feature extraction (Ince et al., 2009). Average accuracy rates of 97.4–98.3% were produced from the experiments (Ince et al., 2009).

A radial basis function network, evolved using particle swarm optimization, was employed for ECG beats classification (Korürek & Doğan, 2010). Using the MIT-BIH data sets, the proposed method was able to classify ECG beats with a smaller network size (Korürek & Doğan, 2010). A novel ECG arrhythmia classification model with a modular mixture of experts and negatively correlated learning neural network was proposed in Javadi, Arani, Sajedin, and Ebrahimpour (2013). Using ECG records from the MIT-BIH repository, an accuracy rate of 96.02% was achieved from the modular network (Javadi et al., 2013). A probabilistic ANN classifier was used to discriminate eight types of arrhythmia from ECG beats (Wang, Chiang, Hsu, & Yang, 2013). The classifier achieved accuracy rates of 99.71% (Wang et al., 2013) with the use of the MIT-BIH data sets.

An ANN with backpropagation training was applied to prediction of transfusion requirements of trauma patients (Walczak, 2005). Based on a series of experiments, the highest recorded accuracy rate was 91.42% (Walczak, 2005). For the diagnosis of arterial diseases and internal carotid arterial disorders, an MLP network with the Levenberg–Marquardt (LM) algorithm was utilized in Übeyli and Güler (2005). Wavelet transform was performed on the Doppler signals, and accuracy rates ranging from 95.52% to 97% were achieved (Übeyli & Güler, 2005). A moving average self-organizing map (MA-SOM) was utilized for segmenting medical images (Torbati, Ayatollahi, & Kermani, 2014). A two-dimensional discrete wavelet transform was used to build the input features from computerized tomography (CT) and magnetic resonance (MR) head images (Torbati et al., 2014). The experimental results indicated that MA-SOM was able to determine the input patterns properly while preserving its robustness against noise (Torbati et al., 2014).

The probabilistic ANN was used for diagnosing patients with urinary tract infections (Mantzaris, Anastassopoulos, & Adamopoulos, 2011). A genetic algorithm was employed to search for potential redundancy in the diagnostic factors (Mantzaris et al., 2011). A prognosis rate of up to 100% was achieved (Mantzaris et al., 2011). In Irigoyen and Miñano (2013), data samples from machine resistance and patients' heart rates were first collected from participants performing exercise on a cyclo-ergometer. A non-linear autoregressive exogenous (NARX) neural network was then used to obtain the optimal training configuration (Irigoyen & Miñano, 2013). To undertake prediction of advanced bladder cancer in patients, a neural network optimized by a genetic algorithm (GA) was proposed in Vukicevic, Jovicic, Stojadinovic, Prelevic, and Filipovic (2014). The genetic algorithm was deployed to optimize the best prognostic performances of clinicians (Vukicevic et al., 2014). The best accuracy rate reported was 95.9% (Vukicevic et al., 2014).

2.3. Decision trees and ensemble models

Based on ECGs from patients with heart failures, Chi-square-based decision trees were produced to differentiate patients with varying levels of risk (Zhang, Goode, Rigby, Balk, & Cleland, 2013). The resulting models were concise, and could be easily understood by clinicians (Zhang et al., 2013). A classification tree based on condition combination competition was proposed for ischemia detection of spatiotemporal ECGs, and an accuracy rate of 98% was achieved (Fayn, 2011).

An ensemble classifier based on extremely randomized decision trees was used for classification of ECG signals (Scalzo, Hamilton, Asgari, Kim, & Hu, 2012). A total of twenty-four features were extracted from records of patients suffering from various intracranial pressure related conditions (Scalzo et al., 2012). On the other hand, ANN ensembles for patients visiting an emergency department with chest pain were presented in Green et al. (2006). The *k*-fold cross validation procedure was used to estimate the performance of the classifier using ECG data inputs (Green et al., 2006). However, no accuracy rates were reported in both Scalzo et al. (2012) and Green et al. (2006).

2.4. Support vector machine

Data sets from MIT-BIH and BIDMC Congestive Heart Failure Database (CHFD) were widely used with different support vector machine (SVM) models, as reviewed in this section. The SVM model was used for ECG arrhythmia analysis in Luz, Nunes, De Albuquerque, Papa, and Menotti (2013). Feature extraction was accomplished using six different methods for comparison purposes (Luz et al., 2013). Using the MIT-BIH data sets, the highest accuracy rate achieved was 92.2% (Luz et al., 2013). To enhance the accuracy rate of ECG signals classification, a statistical method for segmenting heartbeats from ECG signals was used (Wu & Zhang, 2011). Based on the MIT-BIH data sets, independent component analysis and temporal features were extracted (Wu & Zhang, 2011). The highest accuracy rate acquired by the SVM classifier was 99.45% (Wu & Zhang, 2011).

A fast least square SVM model for classification of ECG beats was proposed (Acır, 2005). Five different feature extraction methods were compared, with a total of fifteen features extracted (Acır, 2005). Using the MIT-BIH data sets, the best accuracy rate achieved was 95.2% (Acır, 2005). The SVM classifier was utilized for discriminating ECG beats in Daamouche, Hamami, Alajlan, and Melgani (2012). A wavelet filter was used for feature extraction. Based on the MIT-BIH data sets, the highest recorded accuracy rate was 96.19% (Daamouche et al., 2012). ECG beats were classified using the SVM model in Zidelmal, Amirou, Ould-Abdeslam, and Merckle (2013). The QRS complexes were segmented after pre-processing the ECG signals obtained from the MIT-BIH repository (Zidelmal et al., 2013). Accuracy rates from 97.2% to 98.8% were reported (Zidelmal et al., 2013).

The SVM model was used for automatic classification of ECG beats (Melgani & Bazi, 2008). Two input sets were provided to the SVM model, i.e., with and without feature selection based on the MIT-BIH data sets (Melgani & Bazi, 2008). Compared with other classifiers, the SVM model yielded the highest accuracy rate of 85.98% (Melgani & Bazi, 2008). Based on the MIT-BIH data sets, a combination of thirteen metrics was tested, and the best combination was selected for further evaluation (Li, Rajagopalan, & Clifford, 2014). The SVM model was used for classifying the data samples, which led to an accuracy rate of 99.34% (Li et al., 2014).

The CHFD data set was used in Abawajy, Kelarev, and Chowdhury (2013) for automatic classification of ECG data. A multi-stage algorithm which aimed to reduce data dimension was first used. Then, a sequential minimal optimization method with SVM was used for classification. However, no accuracy rates were reported (Abawajy et al., 2013). A least-square SVM (LS-SVM) model was used to categorize epileptic seizure and seizure-free electroencephalography (EEG) signals (Sharma & Pachori, 2015). Using EEG time series data sets from University Hospital of Bonn, an accuracy rate of 98.67% was achieved (Sharma & Pachori, 2015).

2.5. Other models

A fuzzy expert system was used for arrhythmic beat classification pertaining to ECG recordings (Exarchos et al., 2007). Beat detection with the MIT-BIH data sets was performed, and the accuracy rate achieved by the fuzzy expert system was 96% (Exarchos et al., 2007). In mitigating label noise in ECG signal classification, a genetic optimization method was proposed in Pasolli and Melgani (2014). Specifically, the non-dominated sorting genetic algorithm was deployed to process the ECG signals from the MIT-BIH data sets, and the results showed the effectiveness of the proposed solution (Pasolli & Melgani, 2014).

Classification of normal and abnormal cardiac patterns was conducted using a cross wavelet transform (XWT) model in Banerjee and Mitra (2013). The data samples were first de-noised before feature extraction was carried out. The XWT model as then used in the analysis and classification of ECG signals. An accuracy rate of 97.6% was achieved (Banerjee & Mitra, 2013). A data set comprising scalp EEG signals from 16 children (7 control and 9 pediatric epilepsy patients) was used in a study by Sargolzaei, Cabrerizo, Goryawala, Eddin, and Adjouadi (2015). The study aimed to classify pediatric subjects with epilepsy. Using the *k*-means clustering algorithm, the accuracy rate achieved was 96.87% (Sargolzaei et al., 2015).

An offline data acquisition system was developed in Mitra et al. (2006). Digitized ECG signals were de-noised, and useful time-domain features were extracted. Accuracy rates ranging from 95.8% to 100% were achieved using a rule-based rough-set decision system (Mitra et al., 2006). A multiple instance learning algorithm was used for ECG classification (Sun et al., 2012). The proposed algorithm was able to automatically detect ECG with myocardial ischemia without labeling any heartbeats. An accuracy rate of 90% was reported from the experiments (Sun et al., 2012). In arrhythmia detection from ECG signals, a fuzzy classifier with the genetic algorithm was employed (Vafaie, Ataei, & Koofigar, 2014). Based on the data set obtained from the PhysioBank database, an accuracy rate of 98.67% was achieved (Vafaie et al., 2014).

2.6. Summary

A comprehensive literature survey on biomedical signal processing with a total of 37 papers covering statistical models, ANNs, decision trees, ensemble models, SVMs, and other methods has been conducted. A summary of the reviews is given in Table 1. It can be noticed that the MIT-BIH repository appears to be a popular source of ECG data sets, i.e., in almost half of the ECG related publications. The ANN and SVM models are effective for tackling the MIT-BIH data sets, whereby all models (except one) have been reported to achieve more than 90% accuracy.

Out of the 37 papers reviewed, a total of 16 publications are recently published papers, i.e., between 2013 and 2015. Again,

Table 1
Summary of the literature review.

Type	Reference	Data set	Classifier	Accuracy (%)
Statistical models	Wiggins et al. (2008)	Clinic patients	Bayesian	86.25
	Lee et al. (2013)	Interval ECG signals	Bayesian	88
	Atoui et al. (2010)	Cardiac patients ECGs	Multiple regression-based methods	–
	Iacoviello et al. (2007)	ECG monitoring	Linear regression	–
Artificial neural networks	Özbay et al. (2011)	MIT-BIH	MLP	99
	Özbay and Tezel (2010)	MIT-BIH	MLP with adaptive activation	98.19
	Jiang and Kong (2007)	MIT-BIH	Evolvable block-based ANN	96.6–98.1
	Ince et al. (2009)	MIT-BIH	Feedforward ANN	97.4–98.3
	Korürek and Doğan (2010)	MIT-BIH	Radial basis function	–
	Javadi et al. (2013)	MIT-BIH	ME with NCL	96.02
	Wang et al. (2013)	MIT-BIH	Probabilistic ANN	99.71
	Walczak (2005)	Trauma patients transfusion	ANN backpropagation	91.42
	Übeyli and Güler (2005)	Internal carotid arterial disorders	MLP with LM	95.52–97
	Torbati et al. (2014)	CT and MR images	MA-SOM	–
	Mantzaris et al. (2011)	Urinary tract infections	Probabilistic ANN	100
	Irigoyen and Miñano (2013)	Heart rate	NARX	–
	Vukicevic et al. (2014)	Bladder cancer	ANN with GA	95.9
Decision trees and ensemble models	Zhang et al. (2013)	Heart failure patient ECGs	Chi-square-based decision trees	–
	Fayn (2011)	Spatiotemporal ECGs	Classification tree	98
	Scalzo et al. (2012)	ECG signals	Randomized decision trees	–
	Green et al. (2006)	Chest pains	ANN ensembles	–
Support vector machine	Luz et al. (2013)	MIT-BIH	SVM	92.2
	Wu and Zhang (2011)	MIT-BIH	SVM	99.45
	Acır (2005)	MIT-BIH	Fast least square SVM	95.2
	Daamouche et al. (2012)	MIT-BIH	SVM	96.19
	Zidelmal et al. (2013)	MIT-BIH	SVM	97.2–98.8
	Melgani and Bazi (2008)	MIT-BIH	SVM	85.98
	Li et al. (2014)	MIT-BIH	SVM	99.34
	Abawajy et al. (2013)	CHFD	SVM	–
	Sharma and Pachori (2015)	Seizure EEG signals	LS-SWM	98.67
Other methods	Exarchos et al. (2007)	MIT-BIH	Fuzzy expert system	96
	Pasolli and Melgani (2014)	MIT-BIH	NSGA-II	–
	Banerjee and Mitra (2013)	Cardiac patterns	Cross wavelet transform	97.6
	Sargolzaei et al. (2015)	Scalp EEG signals	*k*-Means clustering	96.87
	Mitra et al. (2006)	Digitized ECG signals	Rough-set decision system	95.8–100
	Sun et al. (2012)	Myocardial ischemia ECGs	Multiple instance learning algorithm	90
	Vafaie et al. (2014)	PhysioBank	Fuzzy with GA	98.67

the MIT-BIH repository appears to be the most commonly used source of data sets. While some researchers use their own data sets, most of the studies utilize data sets available in the public domain. The reported accuracy rates of all models are above 90%, with exception of one model at 88%. ANN and SVM models appear to be the most commonly used methods.

In this study, we focus on two real data sets. In addition to the real ECG data samples, we examine another data set of real Korotkoff signals. Based on these two data sets, an extensive study using a total of nine different classifiers covering classical and state-of-the-art models is presented. An extended analysis and discussion of the results is included (as in Section 4).

3. Data pre-processing and classification

In this section, the acquisition and pre-processing steps of the ECG and Korotkoff signals are detailed. This is then followed by an explanation on the types of classifiers used in the study.

3.1. ECG signals

In the first experiment, a total of 300 single lead-I ECG recordings were collected using a remote monitoring system, i.e., TeleMedCare Health Monitor (TMC-HM) (as shown in Fig. 1), from 288 home-dwelling patients. The participants were patients suffering from chronic obstructive pulmonary disease and/or congestive heart failure. They were trained and then asked to record their ECG measurements daily using the TMC-HM system in an unsupervised manner. From a total of 300 ECG recordings, 250 were selected

Fig. 1. TeleMedCare Health Monitor.

randomly from 100 subjects in Australia and 20 subjects from the UK. The remaining 50 records were manually selected to obtain a larger representation of poor-quality ECG recordings, which would not be useful for determining heart rate, but which were not so bad as to be completely obscured by excessive movement artifacts. These 50 recordings were selected from approximately 1000 ECG signals obtained from 168 subjects based in the UK, but excluding previously selection (20) recordings.

To have an accurate clinical diagnosis, it is important to remove noise from ECG signals because a contaminated signal can result in an incorrect diagnosis (Zidelmal et al., 2013). An efficient technique for denoising ECG signals is wavelet transform, and the discrete wavelet transform (DWT) is useful in this aspect (Poungponsri & Yu, 2013). Different wavelets are generated from a single basic wavelet $\psi(t)$ known as the mother wavelet. Based on the mother wavelet, the shifted and dilated versions can be expressed as (Poungponsri & Yu, 2013)

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-\tau}{s}\right) \tag{1}$$

where s is the scale factor and τ is the translation factor. The wavelet transform of a signal $x(t)$ with the mother wavelet of $\psi(t)$ is given as (Poungponsri & Yu, 2013)

$$T(s,\tau) = \int_{-\infty}^{\infty} x(t)\psi * \left(\frac{t-\tau}{s}\right)dt \tag{2}$$

where the asterisk represents the complex conjugate of the wavelet function. The DWT family is given as (Poungponsri & Yu, 2013)

$$\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n) \tag{3}$$

where m and n are integers for indices. A series of high-pass and low-pass filters is considered for DWT of a signal to pass through (Kim, Shin, Shin, & Lee, 2009). In this study, the noisy signal was decomposed into six levels by DWT using the Daubechies wavelet (Zidelmal et al., 2013). The baseline drift was removed by zeroing the scaling coefficients of DWT at level 6 (Zidelmal et al., 2013). Fig. 2 shows a sample ECG signal in its raw form and the de-noised form using DWT at level 6.

Using the Augsburg Biosignal Toolbox (AuBT) (Wagner, 2009), a total of 79 features were extracted based on the de-noised ECG signals. The extracted features are shown in Table 2. The features include the mean, median, standard deviation (std), minimum (min), maximum (max), and a range of one cardiac cycles which consists of a P-wave, a QRS-complex, a T-wave, with some additional features. The class labels were obtained based on averaged values of the scores from three experts, whom classified the signals into two categories, i.e., good or bad ECG signals.

3.2. Korotkoff signals

In the second experiment, Korotkoff signals pertaining to a blood pressure measurement system were obtained. The auscultatory waveform captured by a stethoscope was transduced by a built-in microphone. During the course of blood pressure measurement, an electronically controlled mechanical pump first raised the air pressure in the cuff, followed by releasing the valve to reduce the pressure, and finally the recording process stopped automatically. A pressure transducer calibrated with a manometer was used to continuously measure the air pressure in the cuff, while a computer electronically controlled the mechanical pump and stored all the recorded signals. A total of 100 recordings were obtained from 25 healthy subjects, 16 men and 9 women aged between 23 and 33 years old. All subjects were advised not to take any caffeine 12 h prior to signal acquisition. The subjects were seated with their hands comfortably laid on a pillow, at the heart level, during the measurement period. A cuff was placed around the upper arm with a stethoscope positioned over the brachial artery. In order to generate various signal qualities, a variety of activities with intensities of noise caused by common artifact including movements were designed.

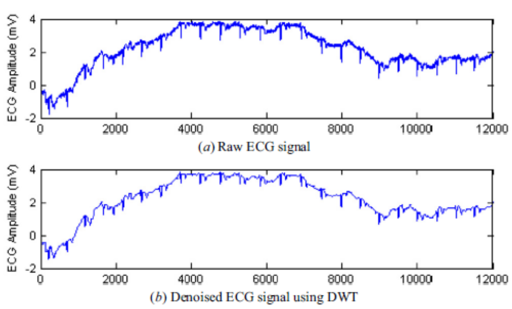

Fig. 2. Raw and denoised ECG signals.

Table 2
Variables extracted from ECG signal.

No	Variable	No	Variable	No	Variable
1	ecgR-mean	28	ecgT-min	54	ecgPampl-range
2	ecgR-median	29	ecgT-max	55	ecgRampl-mean
3	ecgR-std	30	ecgT-range	56	ecgRampl-median
4	ecgR-min	31	ecgPQ-mean	57	ecgRampl-std
5	ecgR-max	32	ecgPQ-median	58	ecgRampl-min
6	ecgR-range	33	ecgPQ-std	59	ecgRampl-max
7	ecgP-mean	34	ecgPQ-min	60	ecgRampl-range
8	ecgP-median	35	ecgPQ-max	61	ecgSampl-mean
9	ecgP-std	36	ecgPQ-range	62	ecgSampl-median
10	ecgP-min	37	ecgQS-mean	63	ecgSampl-std
11	ecgP-max	38	ecgQS-median	64	ecgSampl-min
12	ecgP-range	39	ecgQS-std	65	ecgSampl-max
13	ecgQ-mean	40	ecgQS-min	66	ecgSampl-range
14	ecgQ-median	41	ecgQS-max	67	ecgHrv-mean
15	ecgQ-std	42	ecgQS-range	68	ecgHrv-median
16	ecgQ-min	43	ecgST-mean	69	ecgHrv-std
17	ecgQ-max	44	ecgST-median	70	ecgHrv-min
18	ecgQ-range	45	ecgST-std	71	ecgHrv-max
19	ecgS-mean	46	ecgST-min	72	ecgHrv-range
20	ecgS-median	47	ecgST-max	73	ecgHrvDistr-mean
21	ecgS-std	48	ecgST-range	74	ecgHrvDistr-median
22	ecgS-min	49	ecgPampl-mean	75	ecgHrvDistr-std
23	ecgS-max	50	ecgPampl-median	76	ecgHrvDistr-min
24	ecgS-range	51	ecgPampl-std	77	ecgHrvDistr-max
25	ecgT-mean	52	ecgPampl-min	78	ecgHrvDistr-range
26	ecgT-median	53	ecgPampl-max	79	ecgHrvDistr-triind
27	ecgT-std				

In this study, the Welch method (Welch, 1967) was used for feature extraction of Korotkoff signals. It estimates the power spectrum of a signal at different frequencies, i.e., a technique for spectral density estimation. This method is based on periodogram spectrum estimates, which converts time-domain signals into those in the frequency-domain. The Welch method is able to reduce noise in the estimated power spectrum in exchange of reducing the frequency resolution. The spectral estimate $\hat{I}^{\ell}_M$ is defined as (Power Spectra Estimation, 1995)

$$\hat{I}^{\ell}_M(f) = \frac{1}{L}\sum_{\ell=1}^{L} I^{\ell}_M(f) \tag{4}$$

and its expected value is given by

$$E\left[I^{\ell}_M(f)\right] = \int_{-1/2}^{1/2} S_{N_{xx}}(\eta)w(f-\eta)d\eta = S_{N_{xx}}(\eta) * w(\eta) \tag{5}$$

where

$$W(f) = \frac{1}{UM}\left|\sum_{n=0}^{M-1} w(n)e^{-j\omega nT}\right|^2 \tag{6}$$

The correlations between individual periodograms increase in proportion to the increase in the data segments. A sample of the raw and processed Korotkoff signals using the Welch method is shown in Fig. 3.

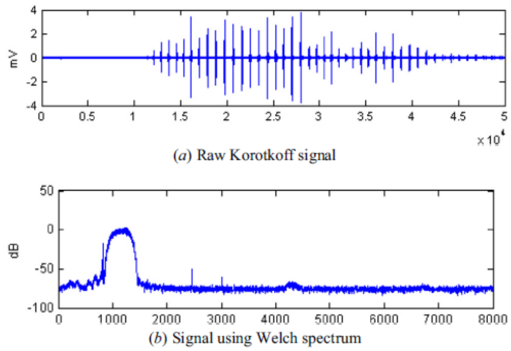

Fig. 3. Raw and processed Korotkoff signals.

Table 3
Variables extracted from Korotkoff signal.

Variable	Definition
fmax	Maximum of frequency spectrogram
fmin	Minimum of frequency spectrogram
p75	Percentile 75 of frequency spectrogram
med	Median of frequency spectrogram
p25	Percentile 25 of frequency spectrogram

A total of five variables, as shown in Table 3, from the extracted frequency spectrogram were used (Mendiola et al., 2013). The class labels were obtained based on the error information provided by domain experts, where the signal was classified as either error-free or erroneous.

3.3. Classifiers

In this study, a total of nine different classifiers, consisting of statistical, ANN, decision tree, and ensemble models, as shown in Table 4, are evaluated. These classifiers are commonly used in the literature. As an example, in Kim et al. (2009), a comparison between different machine learning models such as the backpropagation ANN, radial basis function network, and support vector machine was conducted pertaining to arrhythmia classification of ECG. The logistic regression model was used to provide predictions of coronary artery disease based on a set of independent variables (either continuous, categorical, or both) (Kurt, Ture, & Kurum, 2008).

Naïve Bayes (NB) is a simple probabilistic classifier, which assumes the presence/absence of a particular feature is unrelated to that of any other features with the given target class. The logistic regression (LR) model, a type of statistical method, predicts a binary response, based on one or more input features. The multilayer perceptron (MLP) is a feedforward ANN model, which has an input layer, an output layer, and one or more hidden layers in between. The radial basis function (RBF) network uses the radial activation function in its learning dynamics, and its output comprises a linear combination of radial-based functions of the input features, parameterized by the network variables.

Decision trees are commonly used in data mining. They are useful for disclosing the underlying reasoning in classifying data samples. The classification and regression tree (CART) uses branches to represent conjunction of features, which lead to leafs that represent the target classes. Decision stump (DT), on the other hand, consists of a one-level decision tree. It makes a prediction based on single input feature. In terms of ensemble models, Bagging (BG) improves accuracy of machine learning algorithms and reduces variances while avoiding over-fitting. In AdaBoost (Adaptive Boosting), the outputs of a number of learning algorithms are combined with a weighted sum representing the final output. On the other hand, random forest (RF) operates by constructing multiple decision trees and providing the target class estimate based on the mode of the classes produced by individual trees.

Table 4
Types of classifiers used.

Method	Label	Classifier
Statistical	NB	Naïve Bayes
	LR	Logistic regression
Neural network	MLP	Multilayer perceptron
	RBF	Radial basis function
Decision tree	CART	Classification and regression trees
	DT	Decision stump
Ensemble	BG	Bagging
	AB	AdaBoost
	RF	Random forest

4. Results and discussion

The experimental study was conducted using Weka (Waikato Environment for Knowledge Analysis) (Hall et al., 2009), a popular suite of machine learning and other models. The *k*-fold cross validation technique, where $k = 10$, was employed. As such, the data samples were split into ten equal subsets. Each subset contained approximately the same proportion of data samples from each target class. Nine subsets were used for training, while the remaining was used for performance evaluation. This procedure was repeated 10 times, each time with a previously unused subset for performance evaluation. The performance scores were averaged across 10 runs.

The common performance metrics in medical applications, i.e., accuracy, sensitivity, specificity, and area under the Receiver Operating Characteristic curve (AUC), were computed to quantify the performance (Loo, 2005). Sensitivity measures the proportion of positive and negative cases that are correctly identified, respectively. The ROC curve depicts the discrimination ability of a classifier subject to different threshold settings (Hanley, 1982). Using normalized units, the AUC indicates the probability that a classifier ranks a randomly chosen positive data sample higher than a randomly chosen negative one.

To further evaluate the robustness of the classifiers against noisy data, noise was injected into the input features by using the *AddNoise* function in Weka. In this study, 10%, 30%, and 50% of the training data samples were corrupted by noise, while the test data samples were unaffected, to evaluate the classifiers' learning capabilities in noisy environments.

4.1. ECG signals

Fig. 4 summarizes the noise-free and noise-induced (10, 30%, and 50%) test results from nine different classifiers. With the noise-free data set, LR and RF yielded the highest accuracy rate of 94%, respectively, while CART produced the lowest accuracy rate (89.3%). With 10% noise, RF and LR showed the highest (88.3%) and lowest (76.6%) accuracy rates, respectively. When half of the data samples were corrupted by noise, RF maintained its ranking with the highest accuracy rate of 50.2%. In general, ensemble methods achieved the best accuracy rates across both noise-free and

Link to Full-Text Articles:

http://ac.els-cdn.com/S095741741400801X/1-s2.0-S095741741400801X-main.pdf?_tid=bacd29d8-e7cf-11e4-a4fa-00000aab0f6c&acdnat=1429584190_a9c034bfc0490d7fb60fc36edb602f9c