

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/292341987>

# Feature Selection And Classification Methodology For The Detection of Knee-Joint Disorders

Article in *Computer Methods and Programs in Biomedicine* · January 2016

DOI: 10.1016/j.cmpb.2016.01.020

CITATIONS

45

READS

522

4 authors, including:



**Saif Nalband**

DY Patil International University

10 PUBLICATIONS 79 CITATIONS

[SEE PROFILE](#)



**Aditya Sundar**

Texas Instruments Inc.

13 PUBLICATIONS 69 CITATIONS

[SEE PROFILE](#)



**A Amalin Prince**

BITS Pilani, K K Birla Goa

34 PUBLICATIONS 142 CITATIONS

[SEE PROFILE](#)

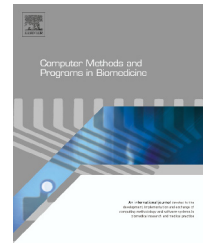
Some of the authors of this publication are also working on these related projects:



Development of FMCW Ka-Band Reflectometer at IPR [View project](#)



Design and development of reconfigurable hardware accelerator for the analysis of reflectometry data [View project](#)



# Feature selection and classification methodology for the detection of knee-joint disorders

Saif Nalband, Aditya Sundar, A. Amalin Prince\*, Anita Agarwal

Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani, KK Birla Goa Campus, Goa 403726, India

## ARTICLE INFO

### Article history:

Received 23 May 2015

Received in revised form

6 January 2016

Accepted 21 January 2016

### Keywords:

Vibroarthographic signal

Biomedical signal processing

Feature selection

Apriori algorithm

Genetic algorithm

Wavelets

## ABSTRACT

Vibroarthographic (VAG) signals emitted from the knee joint disorder provides an early diagnostic tool. The nonstationary and nonlinear nature of VAG signal makes an important aspect for feature extraction. In this work, we investigate VAG signals by proposing a wavelet based decomposition. The VAG signals are decomposed into sub-band signals of different frequencies. Nonlinear features such as recurrence quantification analysis (RQA), approximate entropy (ApEn) and sample entropy (SampEn) are extracted as features of VAG signal. A total of twenty-four features form a vector to characterize a VAG signal. Two feature selection (FS) techniques, apriori algorithm and genetic algorithm (GA) selects six and four features as the most significant features. Least square support vector machines (LS-SVM) and random forest are proposed as classifiers to evaluate the performance of FS techniques. Results indicate that the classification accuracy was more prominent with features selected from FS algorithms. Results convey that LS-SVM using the apriori algorithm gives the highest accuracy of 94.31% with false discovery rate (FDR) of 0.0892. The proposed work also provided better classification accuracy than those reported in the previous studies which gave an accuracy of 88%. This work can enhance the performance of existing technology for accurately distinguishing normal and abnormal VAG signals. And the proposed methodology could provide an effective non-invasive diagnostic tool for knee joint disorders.

© 2016 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

The human knee joint is one of the most complicated and largest joints in a human body. It not only provides flexible movements but also sustains the weight of the body and withstands pressure loads during routine/sports activities. These activities could lead to softening of the articular cartilage, tendons and ligaments of the kneecap which leads to substantial injuries and various kinds of knee joint

disorders [1]. This has lead to exploring the early diagnosis of the knee joint disorders. Both, invasive and non-invasive procedures are available for the evaluation of the knee joint. Imaging techniques such as MRI, fMRI, Computer Tomography (CT) offer non-invasive detection but fail to provide an earlier diagnosis [2]. Whereas a semi-invasive technique such as Arthroscopy, offer considerable information about the diagnosis. But a conventional Arthroscopy would not be feasible for a patient's repetitive or periodic checkups. Both of these procedures, are cost ineffective and are inefficient for a

\* Corresponding author. Tel.: +91 832 2580155.

E-mail address: [amalinprince@gmail.com](mailto:amalinprince@gmail.com) (A.A. Prince).

<http://dx.doi.org/10.1016/j.cmpb.2016.01.020>

0169-2607/© 2016 Elsevier Ireland Ltd. All rights reserved.

routine physician's checkups. Their inability to showcase the dynamic characteristic is one of the major disadvantages of these procedures [3]. These limitations lead us to explore for an alternative low-cost non-invasive detection of knee joint disorders.

During the active movements of the legs such as the flexion and extension, the vibration or auditory signals emits from the mid-patella are called Vibroarthrographic (VAG) signals [4]. VAG signals are characterized by nonlinearity and nonstationary. The analysis of VAG signals could provide an early diagnostic tool for detecting knee joint disorders. The parameters computed from the VAG signal as features could provide a discriminant characteristic. Various signal processing techniques for feature extraction and classification have been carried out in the previous literature [5,6]. The techniques based on the time domain, frequency domain, time-frequency and nonlinear signal processing have been proposed for the analysis of VAG signal [4,7]. Recent works include an analysis of VAG signal by fractional analysis using power spectral analysis by Rangagayan [8]. The classification technique has been carried out using the  $k$ -nearest neighbor method. Wang Yu et al. have recently carried out the work by representation of fluctuation feature using kernel density modeling [9]. The literature review revealed that improvisation in signal processing technique would improve the feature extraction and classification techniques. This, in turn, could aid the medical experts to make accurate decisions on knee joint disorders. Therefore, to build an effective diagnostic system, it is essential to have effective distinct feature extraction techniques and a highly accurate classification algorithm. Hence, we propose an effective feature extraction and classification technique for the diagnosis of knee joint disorder by analysis of VAG signals.

In the recent studies, it has been concluded that wavelet decomposition based sub-bands signals may render substantial information about neuronal activities, especially in the case of EEG signals [10]. The study concluded that EEG signals are complex for analysis and could not differentiate between seizure and non-seizure signals effectively. These distinctions were quite visible in the sub-band signals obtained from wavelet decomposition of EEG signals. Hence, we propose a novel technique of wavelet decomposition for the analysis of VAG signals. With respect to nonlinearity and nonstationary nature of VAG signal, we propose three nonlinear features specified as recurrence quantification analysis (RQA), approximate entropy (ApEn) and sample entropy (SampEn). These three features are extracted from each of the sub-bands signals which are obtained from the wavelet decomposition of the main VAG signal. Wavelet based energy is also considered in this study. Once these features are extracted, a classifier technique is used for the decision-making stage of VAG signal.

A feature selection algorithm has also been proposed in order to identify the most stable, significant and discriminate features. Different techniques have been utilized for improving the accuracy of the classifier by discarding redundant and irrelevant features. These techniques have been employed in the domain of medical imaging, biomedical signals and power-quality disturbance evaluations. Techniques used for feature selection in the previous studies include the sequential forward selection method [11], wrapper method [12] and sequential backward feature selection method [13]. In our

study, the  $K$  nearest neighbor based apriori algorithm and genetic algorithm has been used for selecting the most stable and significant features from the extracted features. The feature sets obtained by feature selection algorithms are given as input to two different classifiers, i.e., LS-SVM and random forest. Classification results are compared between the two feature selection algorithm. Hence, we propose a novel technique for feature extraction based on wavelet decomposition. Feature selection algorithms have been proposed in order to select the most significant, relevant and stable features. The selected features are then fed to the classifier as input in order to compare the performance. With the aim of improving the effectiveness and efficiency of the classification accuracy for knee joint disorders using VAG signals, a Computer Aided Diagnosis (CAD) system based on LS-SVM and random forest is introduced.

The article is organized as follows: Section 2 includes methodology adopted for the study that includes the feature extraction techniques, feature selection algorithms and classification methods. Section 3 shows the results obtained after computing feature extraction and classification. Section 4 concludes with the detailed discussion.

## 2. Methods

### 2.1. Block diagram of the system

The basic block diagram of the VAG detection system is shown in Fig. 1. The original main VAG signals are decomposed into sub-band signals of different frequencies. For the quantification of the VAG signal, features such as RQA, ApEn, SampEn and wavelet-based energy is computed for the original main VAG signals as well as for the decomposed sub-band signals. In order to evaluate the predictability or repeatability within a time series mathematical algorithms such as ApEn and SampEn are used. Both of these algorithms are sensitive to input parameters:  $m$  (length of the data segment being compared),  $r$  (similarity criterion), and  $N$  (length of data).

To quantify the nonlinear information in a time series, Eckmann et al gave the concept of recurrence plots [14]. Recurrence plots are graphical techniques which visualize the recurrence behavior of the phase space trajectory of dynamical systems. This method was proposed by Eckmann et al and was later developed by Webber and Zbilut [15]. It aims to quantify differences appearing towards recurrence plots (RPs) based on the small-scale structures. Several recurrence quantification measures have been introduced in the recent past and have been successfully applied to analyze noisy and nonstationary data. In this research, we have studied three primary measures defined by Webber et al. [15]. Thus, RQA has been successfully applied for the analysis of physiological signals.

Pincus introduced the concept of ApEn to determine the information containing in time series data [17]. A time series is deterministic if computed ApEn value is low and high value of ApEn indicates randomness. SampEn is the modification of approximate entropy. It overcomes the drawback of the ApEn. It has been used extensively for diagnosing the physiological signal in the time-series [18,19]. Features such as RQA,

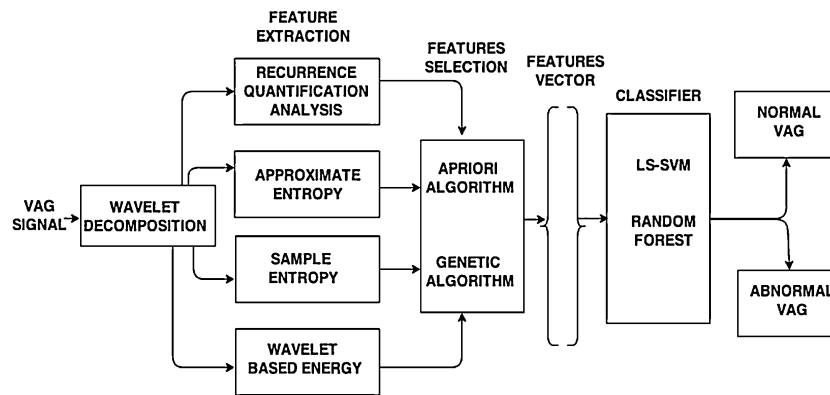


Fig. 1 – Block diagram of VAG detection system.

ApEn, SampEn and wavelet based energy would be extracted from each of the wavelet decomposed sub-band signals. The extracted features are given as inputs to the build the classification model. Since most of the extracted features do not contribute for building an effective classification model, we propose a feature selection algorithm. The feature selection algorithm selects the most significant, relevant and stable features. The irrelevant or redundant features are discarded as these features would serve no purpose and may degrade the classification model. By retaining the most significant, relevant and stable features, a highly accurate classification model can be built. In this work, an apriori algorithm [20] and genetic algorithm [21] as feature selection algorithm has been proposed. The selected features would be provided as inputs to the classifier. LS-SVM and random forest are considered as the classifier and their performances are observed.

## 2.2. Wavelet decomposition

Since the VAG signal has been sampled at 2000Hz, therefore the maximum available frequency would be 1000Hz. The VAG signal is decomposed into eleven wavelets based sub-band signals (D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, A10). The number of levels of decomposition and selection of suitable wavelet is very important in the analysis of signals. The selection of 11 sub-bands provides sufficient resolution for analyzing the frequency range of interest.

This is due to fact that wavelet decomposition into more than 11 levels was found to be redundant and a smaller number of decomposition levels does not provide enough information to distinguish between the signals. Table 1 shows the decomposition of the VAG signal into different frequencies. Based upon the prominent frequency component of the VAG signal, the number of levels of decomposition is chosen. The levels are chosen in such a manner that, the frequencies required for the classification of the signal are correlated to the parts of the signal which are retained by wavelet coefficients. Slow baseline variation mainly caused by different artifacts are represented by D8, D9, D10, A10 sub-band signals, while D1 represents high frequency components caused by environmental noise or muscle contraction [22]. These sub-band signals D1, D8, D9, D10, A10 are not considered for computing the feature extraction. The Daubechies 4(db4) wavelet which is regarded as an efficient wavelet in the analysis of biomedical signal processing was selected as the mother wavelet function. Important information on the event types was obtained from coefficients of multi-resolution analysis.

## 2.3. Feature extraction

### 2.3.1. Recurrence plots and recurrence quantification analysis

Recurrence plots are advanced techniques to carry out the nonlinear data analysis. Recurrence plots are graphical

Table 1 – Decomposition of the VAG signal into sub-band signals.

Sub-band	D1	D2	D3	D4
Frequency	1000–500 Hz	500–250 Hz	250–125 Hz	125–62.5 Hz
Sub-band	D5	D6	D7	D8
Frequency	62.5–31.25 Hz	31.25–15.62 Hz	15.62–7.81 Hz	7.81–3.90 Hz
Sub-band	D9	D10	A10	
Frequency	3.90–1.95 Hz	1.95–1 Hz	1–0 Hz	

techniques which represents the recurrence of the phase space trajectory of the dynamic system. Recurrence plots (RPs) are binary symmetric of  $N \times N$  arrays in which a spot is placed at  $(i, j)$  when the distance between  $X_i$  and  $X_j$  is less than a prescribed value  $\epsilon$ .

$$R_{i,j} = \begin{cases} 0 & \text{if } \|X_i - X_j\| < \epsilon \\ b & \text{if } \|X_i - X_j\| \geq \epsilon \end{cases} \quad (1)$$

Since graphically evaluation would be a difficult task, recurrence quantification analysis (RQA) was evolved in order to provide a quantification measure. In comparison to other methodologies, RQA requires no assumption about the linearity or stationary about the data for analyzing nonlinear signals. In the present study, recurrence rate (%REC), determinism (%DET) and Shannon entropy (ENTR) are considered [15] and they are defined as follows.

- Recurrence rate (%REC): It refers to the density of recurrence plots. It is the probability of a particular state of recurrence. Corresponds to the correlation sum.

$$\%REC = \frac{1}{N^2} \sum_{i,j=1}^N R(i, j). \quad (2)$$

- Determinism (%DET): It relates to determinism/passivity of the system. It is the ratio of recurrence spots placed on the diagonals to all recurrence spots.

$$\%DET = \frac{\sum_{l=l_{\min}}^N P(l)}{\sum_{i,j} R_{i,j}} \quad (3)$$

$P(l)$  is the histogram of the lengths  $l$  of the diagonal lines.

- ENTR: In a recurrence plot, it refers to measure of complexity. It is also known as Shannon entropy. The Shannon entropy of the probability distribution of the diagonal line lengths  $p(l)$ :

$$ENTR = - \sum_{l=l_{\min}}^N p(l) \ln p(l) \quad (4)$$

where

$$p(l) = \frac{P(l)}{\sum_{l=l_{\min}}^N P(l)} \quad (5)$$

Grassberger et al. mention that the delay is a non-critical parameter and hence recommended not to be concerned with finding the optimal value of  $\tau$  [23]. However, a methodology was adopted in selecting this parameter. The changes in the value of %REC with changes in  $\tau$  were estimated. If these changes were smooth, there was no artifact due to parameter selection. Based on this procedure, the value of  $\tau$  was selected. A similar procedure is followed in the selection of the value of radius. An appropriate radius is chosen such that the value of %REC is relatively low (i.e., 1–5%) [24]. Based on this analysis, values of  $m=1$ , a delay of  $\tau=1$  and a

radius of  $\epsilon=15$  were chosen. The distance norm was set to Euclidean.

### 2.3.2. Approximate entropy

To quantify levels of complexity within a time series, Pincus introduced the concept of ApEn. ApEn measures the “likelihood that runs of patterns that are close remain close on next incremental comparisons” [25]. ApEn was initially developed to analyze physiological data, such as heart rate. The details about the algorithm are introduced in Pincus [17] and Al-Angari and Sahakian [26]. For a given  $N$  data points from a time series  $x(n)=x(1), x(2), \dots, x(N)$  where  $N$  is the length, Two parameters,  $m$  (length of the data segment being compared) and  $r$  (similarity criterion) are specified. For a finite length of data points  $N$ , the approximate entropy is estimated as

$$\text{ApEn}(m, r, N) = [\phi^m(r) - \phi^{m+1}(r)] \quad (6)$$

### 2.3.3. Sample entropy

SampEn was developed to reduce the bias caused by self matching. SampEn demonstrates comparatively consistent and is an independent of the data length. Recently, SampEn has been applied for the analysis of various other biomedical signals. Especially, in the analysis of the epileptic activities using EEG signals. Song et al. [27] used SampEn to reveal the characteristics of epileptic EEG signals and identify the EEG signals regarding the existence of seizure or not. Sample Entropy is given as:

$$\text{SampEn}(m, r, N) = \ln[\phi^m(r) - \phi^{m+1}(r)] \quad (7)$$

Although  $m$  and  $r$  are critical in determining the outcome of ApEn and SampEn, no guidelines exist for optimizing their values [17]. For most of the research studies the values of  $m$  and  $r$  are usually  $m=1$  or  $m=2$  and  $r$  between 0.1 and 0.25 times the standard deviation of the original time series, as suggested by Pincus [25]. In present study, we have computed ApEn and SampEn based on the work concluded by Jennifer et al. [28]. In our studies we have taken  $m=2$  and  $r$  is equal to 20% of the standard deviation of the amplitude of time series. And  $N$  is set to 8000 (data length of the VAG signal).

### 2.3.4. Wavelet energy

With the three nonlinear methods introduced above, wavelet based energy of the sub-band signals (D2, D3, D4, D5, D6, D7) are computed and results are compared. The sub-band signals (D2, D3, D4, D5, D6, D7) are not used directly as entries of the feature vector [22]. Instead, the energy in each of the sub-band

signals (D2, D3, D4, D5, D6, D7) are used. The wavelet based energy for each sub-band can be represented as:

$$\begin{bmatrix} X2 \\ X3 \\ X4 \\ X5 \\ X6 \\ X7 \end{bmatrix} = \begin{bmatrix} \log \left( \sum_n |D2(n)| \right) \\ \log \left( \sum_n |D3(n)| \right) \\ \log \left( \sum_n |D4(n)| \right) \\ \log \left( \sum_n |D5(n)| \right) \\ \log \left( \sum_n |D6(n)| \right) \\ \log \left( \sum_n |D7(n)| \right) \end{bmatrix} \quad (8)$$

## 2.4. Feature selection

Many factors contribute in degrading the performance of classification in machine learning algorithms. Among them are the irrelevant and redundant data which hampers the over classification model, thereby do contributing to discrimination among the classes. This challenge escalates when the data is of high dimensionality.

Thus, to overcome these factors, feature selection techniques play an important role for a successful building of the classification model [29]. The number of features is reduced as feature selection removes the irrelevant and redundant features and retains the most significant, relevant and stable features. The performance of the classifier's accuracy increases while using the feature selection algorithms.

### 2.4.1. Apriori algorithm

Apriori algorithm is an influential algorithm for data mining and is especially known for learning association [20]. In data mining, association is usually referred as, for a given element set, the algorithm pursuits to find subset which are common to at least a minimum number. It follows a bottom up approach where predominant subsets are extended one element at a time and groups of candidates are tested against the data. When there is not any further successful extension, the algorithm terminates [30].

### 2.4.2. Genetic algorithm

Genetic algorithms (GA) are one of the examples of evolutionary algorithms (EAs), which is derived from Darwin's theory of evolution associated with natural selection [21]. For an optimization problem in GA the population of strings also termed as chromosomes, are encoded as candidate solutions called individuals. Usually binary strings of 0s and 1s or bit string represent the genetic information that is chromosome and the solution is encoded by such sets of bits. For a new population of individuals or termed as next generation, genetic operators are then applied to the individuals. Crossover and mutation are two main genetic operators applied over here.

In crossover, two offspring strings from two parent strings are created by copying selected bits from each parent. However, in mutation, the value of a single bit (with a small probability) changes randomly. Also, in order to increase the probability that the single bit can survive throughout the evolutionary process, a fitness function is used to measure the quality of an individual. Large search spaces are dealt efficiently by a GA, and other algorithms have less chance to arrive at a local optimal solution [31].

## 2.5. Classification

### 2.5.1. Least squares support vector machines (LS-SVM)

Support Vector Machine (SVM) is used for nonlinear estimation, function estimation, density estimation and classification proposed by Vapnik [32]. SVM is a machine learning algorithm based on statistical learning of Vapnik-Chervonenkis dimension [33]. One of the major drawback of SVM was its high computation load. Suykens suggested LS-SVM to single out this liability [34]. LS-SVM computes linear equation instead of solving the convex for quadratic programming problem in case of SVM. Due to the ease of its implementation, it has found lots of uses in the area of adaptive signal processing. In the present study, we focus on LS-SVM.

### 2.5.2. Random forest

Random forest is based on the decision tree classification techniques. It is an ensemble modeling which combines the results from different models [35]. Random forest is defined by Breiman as "combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest" [36]. The problem of over-fitting found in decision tree is overcome by random forest.

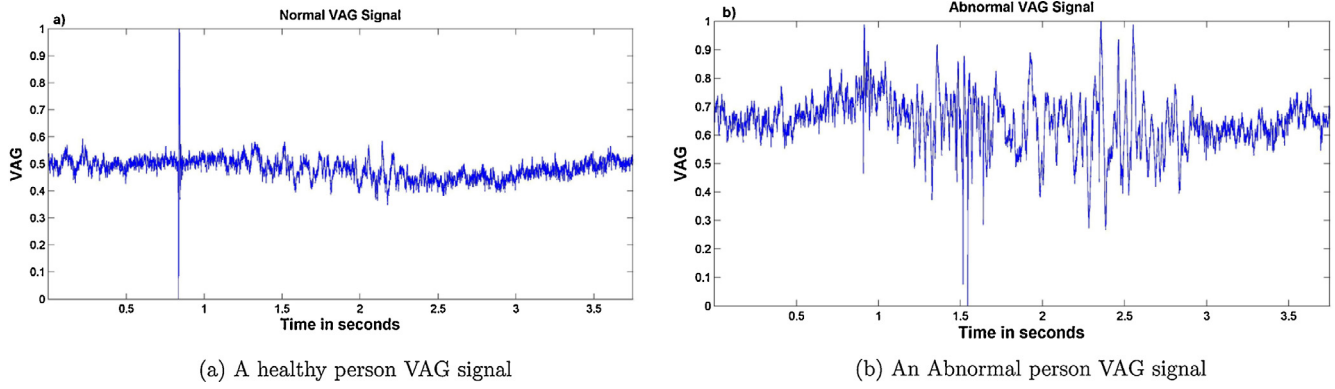
The effectiveness of the classifier techniques can be evaluated using different performance parameters [37]. And they are as sensitivity (SEN), specificity (SPF), accuracy (ACC), positive predictive value (PPV), negative predictive value (NPV), Matthews correlation coefficient (MCC) and false discovery rate (FDR).

## 3. Experiments

### 3.1. Data set

The experiment data set for data acquisition of VAG signal was performed at the university of Calgary, Canada. This setup was authorized by Conjoint Health Research Ethics Board of the University of Calgary. The setup required the subject to sit on an adamant surface table with leg dropping freely in the air. At the mid patella of the knee joint, an accelerometer was mounted. The subject was asked to perform flexion and extension of the knee joint. The active movements represented by VAG signal were captured in the form of electric voltage generated due to deceleration and acceleration. The signal acquired was pre-amplified to avoid anti-aliasing. Data acquisition board and National Instruments software LabView were used for amplifying and digitizing the signal. The VAG signal was





**Fig. 2 – VAG signal of both normal and abnormal person.**

per-filter (10Hz to 1kHz) and amplified. The sampling frequency of VAG signal was 2000Hz and was digitized with a resolution of 12 bits [38].

In the current investigation, we have used data set as reported in the earlier literature. The data set contains a total of 89 VAG samples. 51 subjects having no physiological disorder condition were considered as normal while 38 subjects suffering from various kinds of knee joint disorders such as meniscal tears, anterior cruciate ligament and tibial chondromalacia injuries were considered as abnormal. A sample VAG signal of a normal and abnormal subject is shown in Fig. 2(a) and (b) respectively. We can observe that the abnormal signal exhibits a high level of irregularity in the time domain, as illustrated in Fig. 2(b).

### 3.2. Results

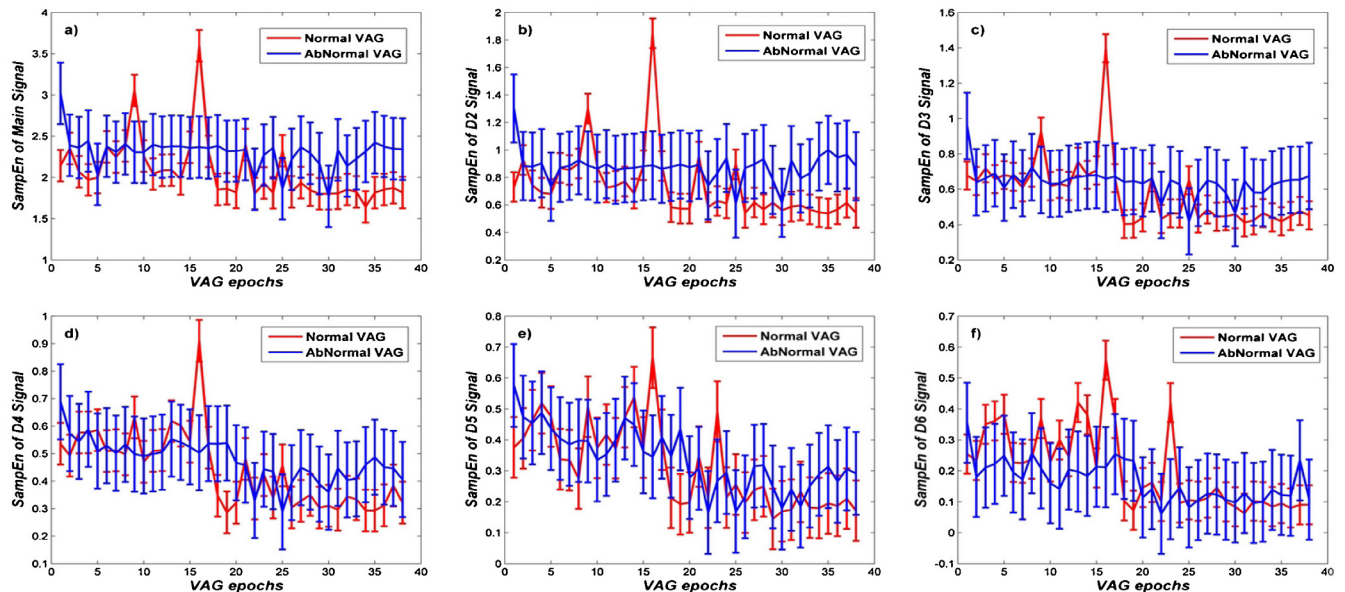
#### 3.2.1. Feature extraction

In this study, we have implemented the proposed method of feature extraction, feature selection and classification

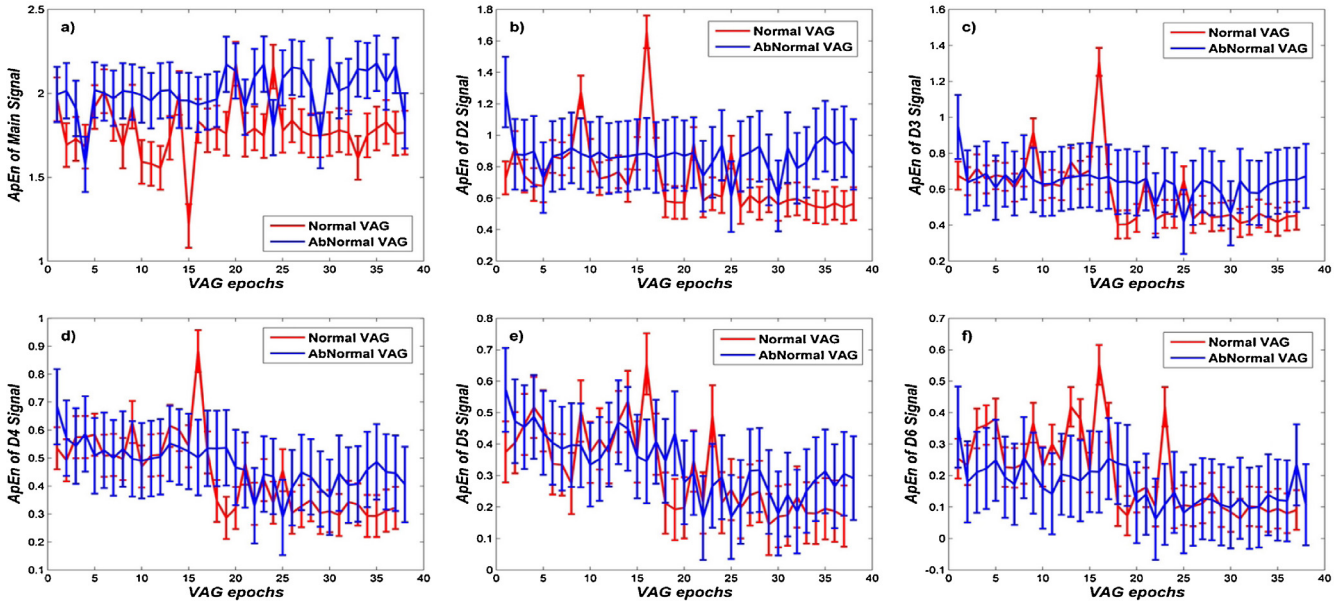
techniques in MATLAB 2014A. Initially, the feature set was consisted of twenty-four features and they were extracted as follows:

- (1) RQA features: percent recurrence, determinism and Shannon Entropy of the frequency distribution of the diagonal line lengths.
- (2) ApEn of main signal and sub-band D2–D7 signals.
- (3) SampEn of main signal and sub-bands D2–D7 signals.
- (4) Energy of the main signal and sub-band D2–D7 signals.

Fig. 3 represents the error plot for SampEn. Error bar represents the range and standard deviation (SD). They are called as descriptive error bars because they illustrate the data spread. Range error bars encompass the lowest and highest values. The figure represents error plot of sample entropy extracted from original VAG signal and its sub-band signals D2, D3, D4, D5, D6, D7. From the error plot, we can notice a substantial overlap of normal and abnormal VAG signal states computed from the original VAG signal. Thereby, these distinctions are



**Fig. 3 – Error plot of SampEn for normal and abnormal VAG. (a) Evaluated from the original VAG signals. (b)–(f) Evaluated from sub-bands D3, D4, D5, D6, D7 respectively (for 38 VAG epochs).**



**Fig. 4** – Error plot of ApEn for normal and abnormal VAG. (a) Evaluated from the original VAG signals. (b)–(f) Evaluated from sub-bands D3, D4, D5, D6, D7 respectively (for 38 VAG epochs).

apparently more visible in sub-band signals D2, D3, D4, D5, D6, D7. Similarly, error plots for ApEn and RQA's parameters have been plotted in Figs. 4 and 5 respectively.

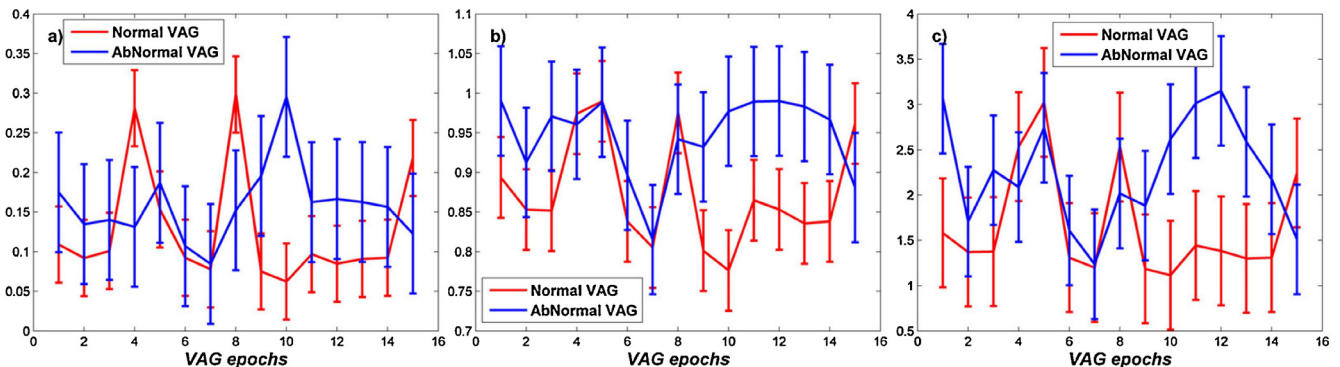
### 3.2.2. Feature selection

**3.2.2.1. Apriori algorithm.** The twenty-four features extracted characterize a VAG signal and most of these features might be irrelevant or redundant to build an effective classification model. Thus, in order to develop a highly accurate classification model, the most significant, relevant and stable features should be selected. A K-means clustering followed by apriori algorithm analysis is performed over the twenty-four features to select most significant, relevant and stable features and thus, reduces the dimensionality of the feature set. The proposed method involves dividing each of the features of the data into two classes using K-means clustering followed by simple one bit encoding of each group. Feature data belonging to one group were indicated with zeros and those belonging to the other were denoted as ones. The encoded feature vectors for all the recorded signals were further analyzed for

recurrences. Features that acted as primary key were chosen as most significant, stable and relevant features. The feature set consisted of six features, which showed the maximum recurrence. The compressed feature set constituted the following extracted features:

- (1) ApEn of the main signal.
- (2) ApEn of sub-band D6.
- (3) ApEn of sub-band D7.
- (4) Determinism (Recurrence Quantification Analysis).
- (5) Shannon Entropy (Recurrence Quantification Analysis).
- (6) SampEn Entropy of sub-band D6.

**3.2.2.2. Genetic algorithm.** Similarly, to improve the performance of the classification model, a genetic algorithm as feature selection algorithm is also considered in the present study [39]. The genetic algorithm is able to choose most significant, relevant and stable features. The genetic algorithm chooses four features, which showed the



**Fig. 5** – Error plot for RQA parameters: (a) %REC; (b) %DET; (c) ENTR (for 15 VAG epochs).



**Table 2 – Kruskal–Wallis test for the features.**

Feature	ApEn Main	ApEn D6	ApEn D7
p-Value	7.78e–08	0.4992	0.1123
Feature	DET	ENTR	SampEn D6
p-Value	0.1371	0.1559	0.0165
Feature	SampEn D2	Energy D2	Energy D5
p-Value	2.75e–04	7.02e–05	0.4110

maximum recurrence. The feature set constituted the following:

- (1) ApEn of the main signal.
- (2) SampEn of sub-band D2.
- (3) Wavelet Energy of sub-band D2.
- (4) Wavelet Energy of sub-band D5.

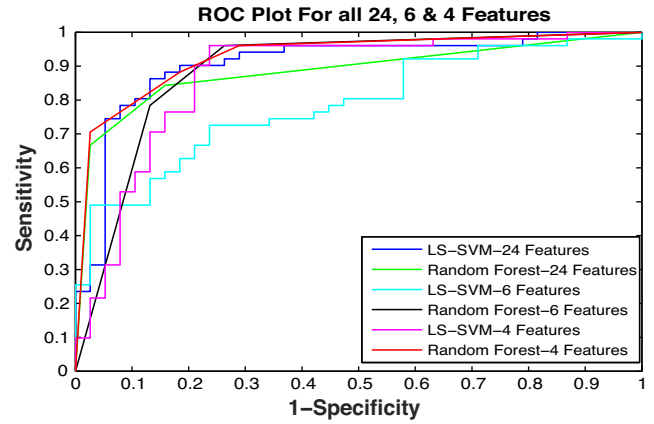
To further verify the discriminability between the features, the Kruskal–Wallis test is performed on the features extracted. Kruskal–Wallis is a test of variance using population variance among groups. The *p*-value returned from the test indicates the variation between the two data. Smaller *p*-value of 0.05 indicates a very strong presumption against the null hypothesis, i.e., it hints that two feature data belongs to sets and hence could be a useful feature in distinguishing the two classes. The results obtained from the Kruskal–Wallis test for the selected features are presented in Table 2.

### 3.2.3. Classification

The extracted features are given as inputs to classifiers, LS-SVM and random Forest. Cross validation is used for selecting optimal kernel/model parameters for the classifiers. Radial basis function (RBF) is used as kernel function for LS-SVM classifier. The regularization constant is chosen appropriately to avoid over-fitting. Sensitivity, specificity, accuracy and area under receiver operating curve (ROC-AUC) are computed to compare the performance of the classifier. Targets are set as ‘1’ for abnormal signals and ‘0’ for normal signals. In present work, a 10-fold cross-validation [40] procedure has been used in order to assess the classification performance of the classifier. All classifiers considered in this study underwent a 10-fold cross-validation, with the same folds being used by each classifier.

In the random forest, a validation procedure is followed for finding the number of decision trees in the classifier. A part of the data-set is set aside for validation and the number of trees is alternated, till an ideal validation accuracy is obtained. The number of decision trees chosen for the ensemble is taken as five.

A receiver operating characteristic (ROC) curve was illustrated to compare the performance of the classifier. ROC curve is a graphical representation of the true positive rate (TPR, also named sensitivity) against the false positive rate (FPR,  $FPR = 1 - \text{specificity}$ ) [41]. The area under the ROC curve (ROC-AUC) is an effective way of comparing the performance of different features (all 24 features, 6 features from apriori

**Fig. 6 – ROC plot of classifiers with 24, 6 and 4 features.**

algorithm and 4 features from a genetic algorithm) using LS-SVM and random forest. The area under ROC curve (ROC-AUC) illustrates how considerably a parameter would be able to distinguish between two diagnostic groups. Larger ROC area implies a better classification accuracy [42]. The ROC plot for two classifiers using feature selection algorithms is observed in Fig. 6.

Tables 3 and 4 show the comparison between all twenty-four features and features selected by feature selection algorithms using LS-SVM and random forest respectively. The number of features selected vary with respect to the feature selection algorithm. Six features are selected by using the apriori algorithm, while four features are selected by genetic algorithm. For all the above cases, the classification is carried out by LS-SVM and random forest and are shown in Tables 3 and 4 respectively.

From Table 3, it is evident that using feature selection algorithms, the performance of the LS-SVM classifier increases. The accuracy of LS-SVM using all twenty-four features is 93.18% and by using apriori algorithm (six features) as feature selection algorithm, its accuracy increases to 94.31%. While using genetic algorithm (four features) its accuracy is 91.01%. Features obtained by using the apriori algorithm also provided a low FDR of 0.0892 than compared to all twenty-four features (0.1071) and genetic algorithm (0.1052). The high accuracy obtained from the apriori algorithm along with low FDR, provided a good classification tool to classify VAG signals more efficiently. Table 4 provided a similar case study of classifying the VAG signal using random forest. From the Table 4, it is evident that features obtained from feature selection algorithms perform better than the complete twenty-four features. The classification accuracy, using all twenty-four features is 86.52% while using the apriori algorithm, is 89.77% whereas classification accuracy is highest using genetic algorithm of 91.01%.

From Table 4, the ROC-AUC is also highest in genetic algorithm with 0.92 and standard error (S.E) of 0.0306. Using apriori algorithm, the ROC-AUC is 0.8863 S.E: 0.0388 whereas it was lowest for all twenty-four features. The MCC was also higher by using genetic algorithm (0.8159) while it was lowest for twenty-four features (0.7231). The features selected using genetic algorithm yielded a lower FDR in comparison to the features selected by the apriori algorithm and all twenty-four

**Table 3 – LS-SVM.**

Parameters	LS-SVM					
Feature selection	All 24 Features		Apriori		Genetic	
Confusion matrix	50 6	1 32	51 5	1 32	51 6	2 30
ROC-AUC	0.8916 ± 0.0375		0.8161 ± 0.0481		0.8715 ± 0.0406	
Sensitivity	0.9803		0.9807		0.9622	
Specificity	0.8421		0.8648		0.8333	
Accuracy	0.9318		0.9431		0.9101	
PPV	0.8929		0.9107		0.8947	
NPV	0.9697		0.96		0.9375	
MCC	0.8423		0.8628		0.8137	
FDR	0.1071		0.0892		0.1052	

**Table 4 – Random forest.**

Parameters	Random forest					
Feature selection	All 24 features		Apriori		Genetic	
Confusion matrix	49 9	3 28	50 8	2 29	48 5	3 33
ROC-AUC	0.8842 ± 0.0387		0.8863 ± 0.0388		0.92 ± 0.0306	
Sensitivity	0.9423		0.9615		0.9411	
Specificity	0.7568		0.7837		0.8684	
Accuracy	0.8652		0.8977		0.9101	
PPV	0.8448		0.862		0.9056	
NPV	0.9032		0.9354		0.9167	
MCC	0.7231		0.7704		0.8159	
FDR	0.1552		0.1379		0.0943	

features. Considering the results obtained from [Tables 3 and 4](#), it is evident that, genetic algorithm performs extremely well with random forest as classifier. Thus, four features obtained from the genetic algorithm have successfully classified more number of VAG signals into normal and abnormal classes using random forest.

From [Tables 3 and 4](#) it is clear that LS-SVM performed much better with respect to random forest in terms of accuracy. The accuracy of LS-SVM using six features selected from the apriori algorithm with 94.31% was highest among other results. The proposed methodology would enable the instrument designer to design a highly accurate instrument. This feature selection technique can be applied to biomedical system as well.

#### 4. Discussion

VAG signals are characterized by multi-component, nonstationary and nonlinearity. The conventional signal processing techniques are not suitable for the analysis of VAG signal. The main objective of this paper is to enhance the classification accuracy by distinguishing abnormal signals from normal ones for diagnostic purposes. In the present work, we have proposed a wavelet decomposition based feature extraction and classification techniques for classifying the normal and abnormal VAG signals. Initially the original VAG signals were subjected to a wavelet decomposition into sub-band signals. Feature extraction based on nonlinear methods, i.e., RQA, ApEn and SampEn were computed for the original and sub-band signals distinctly. Additionally, a wavelet-based energy was also extracted. The extracted features from the sub-band

signals D2, D3, D4, D5, D6, D7 provided a significant discrimination between normal and abnormal VAG signal than the original VAG signal. These distinction were quite observable from the error plot of SampEn and ApEn, as shown in [Figs. 3 and 4](#) respectively. The proposed methodology characterized a VAG signal with twenty-four features.

Another important task for designing a low computing system is to develop a highly sensitive classifier with a low false detection rate. Since wavelet based feature extraction had extracted twenty-four features. Most of these extracted features are irrelevant and redundant. Thereby, these features will not contribute in improving the classification model. Thus, feature selection algorithm contributes in selecting the most significant, relevant and stable features. And thus, the feature selection algorithm has improvised the classification accuracy.

Apriori algorithm and genetic algorithm were used as feature selection techniques. Six features were selected using the apriori algorithm, while four features were selected by genetic algorithm. The reluctant features were discarded by genetic and apriori algorithm. This distinction was also evident from Kurskal–Wallies test as shown in [Table 2](#). The selected features were given as input to the proposed classifier algorithms (LS-SVM and random forest) and their performances were compared. The performance of LS-SVM using a feature selection algorithm using the apriori algorithm was highest. When compared to random forest, it gave an accuracy of 94.31% which can be observed from [Tables 3 and 4](#). Results conveyed that, the performance of proposed methodology with apriori algorithm as feature selection algorithm was highest with an accuracy of 94.31% using the apriori algorithm.

**Table 5 – Comparison of the proposed methodology with the existing methodologies.**

Parameter Classifiers	Proposed method LS-SVM	Fluctuation features [9] Bayes classification	FF,VMS [6] K-nearest neighbor	Matching pursuit decomposition [5] Dynamic weight classifier
Sensitivity	1.0	0.714	0.6429	NA
Specificity	0.8648	0.978	0.9362	NA
Accuracy	94.31%	88.00%	80.00%	88.76%

The results obtained were also compared with the previous literature studies using the same set of data-set as shown in the Table 5. The performance of the proposed LS-SVM method with feature selection algorithm using apriori algorithm is superior to the fluctuation feature which uses kernel density modeling with input parameters fractional scaling index and average envelope amplitude (Accuracy=88%, SEN%71.43%, SPE: 85.11%) [9]. The results of the proposed method are also compared with the  $k$  nearest neighbor classifier ( $k=5$ ) with features as form factor and variance of mean squared (VMS) (Accuracy=80%, SEN%71.43%, SPE: 85.11%) [6]. It also performed better when compared to the radial-basis function neural network using form factors, skewness, kurtosis and entropy features (ROC-AUC: 0.8172) [38]. From the table it is evident that the proposed wavelet based feature extraction along with feature selection algorithm using apriori provided the better classification accuracy and ROC-AUC.

This work was quite restricted for classification of normal and abnormal VAG signals only. The higher accuracy obtained from our methodology could provide an aid to establish a standard protocol for the discrimination and early stage detection of knee pathologies using the VAG signals. Thereby, developing a low cost based computer aided diagnosis (CADx) tool for knee joint disorders. This work could be further directed in classifying the various extent of knee joint disorders. This diagnosis could provide physician a benefit in clinical application.

## 5. Conclusion

In this paper, a new feature extraction technique based on wavelet decomposition has been proposed. The proposed methodology could provide a highly accurate diagnostic system for the non-invasive diagnosis and monitoring of knee joint disorders. Nonlinear features were able to differentiate much more effectively between normal and abnormal VAG signal. The differences were more pronounced in the sub-band signals (D2, D3, D4, D5, D6, D7). The proposed methodology characterizes twenty-four features for a VAG signal. Hence, to discard the irrelevant and redundant features, two feature selection algorithms were considered to select the most significant, relevant and stable features. The performance of the feature selection algorithm was evaluated by two classifiers. The results verified that the performance of the feature selection algorithms were superior to all twenty-four features in terms of accuracy and other parameters. Classifying parameters such as sensitivity, specificity, MCC, etc. also validated a higher performance of LS-SVM in comparison to random forest. The results obtained also provided much better performance in comparison with previous literature. The proposed methodology in our works could find a potential application in the knee-joint pathology for early detection and monitoring.

## 6. Conflict of interest

None declared.

## Acknowledgment

The authors are extremely grateful to Prof. Rangaraj M. Rangayyan, Dr. Cyril B. Frank, Dr. Gordon D. Bell, Prof. Yuan-Ting Zhang and Prof. Sridhar Krishnan of University of Calgary, Canada for the work of data acquisition.

## REFERENCES

- [1] V. Vigorita, B. Ghelman, D. Mintz, *Orthopaedic Pathology, M – Medicine Series*, Lippincott Williams & Wilkins, 2008.
- [2] B. Manaster, J. Crim, Z. Rosenberg, *Diagnostic and Surgical Imaging Anatomy: Knee, Ankle, Foot*, Amirsys, 2007.
- [3] G. McCoy, J. McCrea, D. Beverland, W. Kernohan, R. Mollan, *Vibration arthrography as a diagnostic aid in diseases of the knee. A preliminary report*, J. Bone Jt. Surg. Br. 69-B (2) (1987) 288–293.
- [4] Y. Wu, *Knee Joint Vibroarthrographic Signal Processing and Analysis*, Springer-Verlag, 2014.
- [5] S. Cai, S. Yang, F. Zheng, M. Lu, Y. Wu, S. Krishnan, *Knee joint vibration signal analysis with matching pursuit decomposition and dynamic weighted classifier fusion*, Comput. Math. Methods Med. (2013), <http://dx.doi.org/10.1155/2013/904267>.
- [6] K. Liu, X. Luo, S. Yang, S. Cai, F. Zheng, Y. Wu, *Classification of knee joint vibroarthrographic signals using k-nearest neighbor algorithm*, in: 2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE), 2014, pp. 1–4, <http://dx.doi.org/10.1109/CCECE.2014.6900933>.
- [7] K.S. Kim, J.H. Seo, J.U. Kang, C.G. Song, *An enhanced algorithm for knee joint sound classification using feature extraction based on time-frequency analysis*, Comput. Methods Progr. Biomed. 94 (2) (2009) 198–206, <http://dx.doi.org/10.1016/j.cmpb.2008.12.012>.
- [8] R.M. Rangayyan, F. Oloumi, Y. Wu, S. Cai, *Fractal analysis of knee-joint vibroarthrographic signals via power spectral analysis*, Biomed. Signal Process. Control 8 (2013) 23–29, <http://dx.doi.org/10.1016/j.bspc.2012.05.004>.
- [9] S. Yang, S. Cai, F. Zheng, Y. Wu, K. Liu, M. Wu, Q. Zou, J. Chen, *Representation of fluctuation features in pathological knee joint vibroarthrographic signals using kernel density modeling method*, Med. Eng. Phys. 36 (10) (2014) 1305–1311.
- [10] S. Xie, S. Krishnan, *Wavelet-based sparse functional linear model with applications to EEGs seizure detection and epilepsy diagnosis*, Med. Biol. Eng. Comput. 51 (1–2) (2013) 49–60, <http://dx.doi.org/10.1007/s11517-012-0967-8>.
- [11] J. Xie, C. Wang, *Using support vector machines with a novel hybrid feature selection method for diagnosis of erythematous-squamous diseases*, Expert Syst. Appl. 38 (2011) 5809–5815, <http://dx.doi.org/10.1016/j.eswa.2010.10.050>.

- [12] S. Maldonado, R. Weber, A wrapper method for feature selection using Support Vector Machines, *Inf. Sci.* 179 (13) (2009) 2208–2217, <http://dx.doi.org/10.1016/j.ins.2009.02.014>.
- [13] S. Gunal, O.N. Gerek, D.G. Ece, R. Edizkan, The search for optimal feature set in power quality event classification, *Expert Syst. Appl.* 36 (2009) 10266–10273, <http://dx.doi.org/10.1016/j.eswa.2009.01.051>.
- [14] J.-P. Eckmann, S.O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems, *EPL (Europhys. Lett.)* 4 (9) (1987) 973.
- [15] C.L. Webber, J.P. Zbilut, Dynamical assessment of physiological systems and states using recurrence plot strategies, *J. Appl. Physiol. (Bethesda, Md.: 1985)* 76 (1994) 965–973.
- [17] S. Pincus, Approximate entropy as a measure of system complexity, *Proc. Natl. Acad. Sci.* 88 (6) (1991) 2297–2301, <http://dx.doi.org/10.1073/pnas.88.6.2297>.
- [18] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, *Am. J. Physiol. Heart Circ. Physiol.* 278 (2000) H2039–H2049.
- [19] D.E. Lake, J.S. Richman, M.P. Griffin, J.R. Moorman, Sample entropy analysis of neonatal heart rate variability, *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 283 (2002) R789–R797, <http://dx.doi.org/10.1152/ajpregu.00069.2002>.
- [20] R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases, *J. Comput. Sci. Technol.* 15 (1994) 487–499, <http://dx.doi.org/10.1007/BF02948845>.
- [21] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989, <http://dx.doi.org/10.1007/s10589-009-9261-6>.
- [22] A. Shueb, H. Edwards, J. Connolly, B. Bourgeois, S. Ted Treves, J. Guttag, Patient-specific seizure onset detection, *Epilepsy Behav.* 5 (4) (2004) 483–498.
- [23] P. Grassberger, T. Schreiber, C. Schaffrath, Nonlinear time sequence analysis, *Int. J. Bifurc. Chaos* 01 (03) (1991) 521–547, <http://dx.doi.org/10.1142/S0218127491000403>.
- [24] G.L. Pallecchia, K. Shockley, Application of recurrence quantification analysis: influence of cognitive activity on postural fluctuations, in: M.A. Riley, G.C. Van Orden (Eds.), *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*, 2005, pp. 95–141.
- [25] S.M. Pincus, R.R. Viscarello, Approximate entropy: a regularity measure for fetal heart rate analysis, *Obstet. Gynecol.* 79 (1992) 249–255.
- [26] H.M. Al-Angari, A.V. Sahakian, Use of sample entropy approach to study heart rate variability in obstructive sleep apnea syndrome, *IEEE Trans. Biomed. Eng.* 54 (2007) 1900–1904, <http://dx.doi.org/10.1109/TBME.2006.889772>.
- [27] Y. Song, J. Crowcroft, J. Zhang, Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine, *J. Neurosci. Methods* 210 (2012) 132–146, <http://dx.doi.org/10.1016/j.jneumeth.2012.07.003>.
- [28] J.M. Yentes, N. Hunt, K.K. Schmid, J.P. Kaipust, D. McGrath, N. Stergiou, The appropriate use of approximate entropy and sample entropy with short data sets, *Ann. Biomed. Eng.* 41 (2) (2013) 349–365, <http://dx.doi.org/10.1007/s10439-012-0668-3>.
- [29] D. Koller, M. Sahami, Toward optimal feature selection, in: *International Conference on Machine Learning*, 1996, pp. 284–292, Article ID: 393144.
- [30] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, A.I. Verkamo, Fast discovery of association rules, *Adv. Knowl. Discov. Data Min.* 12 (1996) 307–328.
- [31] O. Ludwig, U. Nunes, Novel maximum-margin training algorithms for supervised neural networks, *IEEE Trans. Neural Netw.* 21 (6) (2010) 972–984, <http://dx.doi.org/10.1109/TNN.2010.2046423>.
- [32] V.N. Vapnik, A.Y. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, *Theory Probab. Appl.* (1971), <http://dx.doi.org/10.1137/1116025>.
- [33] V.N. Vapnik, The nature of statistical learning theory, *IEEE Trans. Neural Netw.* 8 (1995), <http://dx.doi.org/10.1109/TNN.1997.641482>.
- [34] J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, *Neural Process. Lett.* 9 (1999) 293–300, <http://dx.doi.org/10.1023/A:1018628609742>.
- [35] T.K. Ho, Random decision forests, in: *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, 1995, pp. 278–282, <http://dx.doi.org/10.1109/ICDAR.1995.598994>.
- [36] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32, <http://dx.doi.org/10.1023/A:1010933404324>.
- [37] A.T. Azar, S.A. El-Said, Performance analysis of support vector machines classifiers in breast cancer mammography recognition, *Neural Comput. Appl.* 24 (5) (2014) 1163–1177, <http://dx.doi.org/10.1007/s00521-012-1324-4>.
- [38] R.M. Rangayyan, Y.F. Wu, Screening of knee-joint vibroarthrographic signals using statistical parameters and radial basis functions, *Med. Biol. Eng. Comput.* 46 (2008) 223–232, <http://dx.doi.org/10.1007/s11517-007-0278-7>.
- [39] A.T. Azar, H.I. Elshazly, A.E. Hassanien, A.M. Elkorany, A random forest classifier for lymph diseases, *Comput. Methods Progr. Biomed.* 113 (2) (2014) 465–473, <http://dx.doi.org/10.1016/j.cmpb.2013.11.004>.
- [40] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *International Joint Conference on Artificial Intelligence*, vol. 14, 1995, pp. 1137–1143, <http://dx.doi.org/10.1067/mod.2000.109031>.
- [41] M.H. Zweig, G. Campbell, Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine [published erratum appears in *Clin Chem* 1993 Aug;39(8):1589], *Clin. Chem.* 39 (1993) 561–577.
- [42] M. Greiner, D. Pfeiffer, R.D. Smith, Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests, *Prev. Vet. Med.* 45 (2000) 23–41, [http://dx.doi.org/10.1016/S0167-5877\(00\)00115-X](http://dx.doi.org/10.1016/S0167-5877(00)00115-X).