

Research Article

Health Monitoring for Elderly: An Application Using Case-Based Reasoning and Cluster Analysis

Mobyen Uddin Ahmed, Hadi Banaee, and Amy Loutfi

Center for Applied Autonomous Sensor Systems, Örebro University, 701 82 Örebro, Sweden

Correspondence should be addressed to Mobyen Uddin Ahmed; mobyen.ahmed@oru.se

Received 24 March 2013; Accepted 18 April 2013

Academic Editors: T.-C. Chen, G. L. Foresti, Z. Liu, and R. Rada

Copyright © 2013 Mobyen Uddin Ahmed et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a framework to process and analyze data from a pulse oximeter which remotely measures pulse rate and blood oxygen saturation from a set of individuals. Using case-based reasoning (CBR) as the backbone to the framework, records are analyzed and categorized according to their similarity. Record collection has been performed using a personalized health profiling approach in which participants wore a pulse oximeter sensor for a fixed period of time and performed specific activities for pre-determined intervals. Using a variety of feature extraction methods in time, frequency, and time-frequency domains, as well as data processing techniques, the data is fed into a CBR system which retrieves most similar cases and generates an alarm according to the case outcomes. The system has been compared with an expert's classification, and a 90% match is achieved between the expert's and CBR classification. Again, considering the clustered measurements, the CBR approach classifies 93% correctly both for the pulse rate and oxygen saturation. Along with the proposed methodology, this paper provides a basis for which the system can be used in the analysis of continuous health monitoring and can be used as a suitable method in home/remote monitoring systems.

1. Introduction

Today, the possibility to remotely monitor physiological health parameters provides a new approach for disease prevention and early detection [1, 2]. Furthermore, such health monitoring systems could be useful for the elderly in independent and assisted living [3]. In developing health monitoring systems, several intelligent data processing methods have been proposed in the literature, for instance, neural network (NN) [4] and support vector machine (SVM) [5]. These methods are often black box methods and make it difficult for experts to gain further insight into the structure presented in the data.

In this paper, a clinical decision support system (CDSS) has been proposed where case-based reasoning (CBR) approach [6] is applied to analyze and process the data coming from a pulse oximeter that contain measurements of both pulse rate and blood oxygen saturation. A case-based reasoning (CBR) [6–17] approach can work in a way close to human reasoning, for example, it solves a new problem applying previous experiences, which is more common for doctors,

clinicians, or engineers. In the proposed system, CBR is the part of a large framework where first the data is preprocessed and features are extracted to find significant parameters of interest using time, frequency, and time-frequency domain features. In addition, to use the features directly, the CBR system can also produce similar matches based on input from a clustering approach that increases the retrieval process by looking only within the most similar group of cases. The method presented here is verified using data collected from a group of individuals who followed a controlled sequence of activities called a health profile. In the health profile, users are requested to perform specific activities for a predefined time intervals for example, deep breathing, walking, and so forth. Using the health profiling, ground truth data could be obtained, and a further verification has been done by three human experts who provided general classification of whether the data from a specific user seemed normal.

Previous work related to the proposed approach ranges from processing of data from physiological sensors that have typically relied on time domain features [2, 18, 19] to classifiers which may utilize known information from

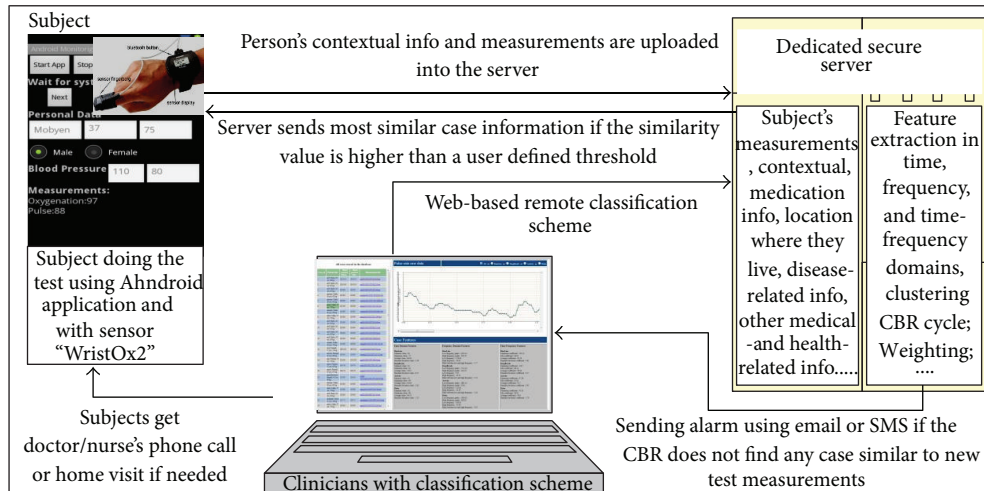


FIGURE 1: The steps and the data flow of the CDSS.

disease databases or consider only the physiological parameters [1, 20]. A notable example [21] in which the authors presented a health condition monitoring system using electrocardiogram (ECG), heart rate (HR), oxygen saturation, impedance pneumography, and activity patterns is included. The sensors were mounted in the patient's garments, and the service is applied in rehabilitation of cardiac ambulatory stable cardiorespiratory patients. Studies that consider only continuous biomedical data (i.e. pulse rate) and respective signal processing methods can be found in [22, 23]. All these works have demonstrated that the recent advancement in sensor technology could provide continuous detection of, for example, pulse, blood oximetry, and level of physical activity. Furthermore, through appropriate sensor data processing, it is also possible to detect changes and provide alerts and warnings to users [23]. However, in most of these studies, the main goal is to classify data into different labeled classes, or find similar cases by determining features. In this work, classification of data in a meaningful way, that is, retrieving the effective solution for similar measurements, and enabling diagnosis of anomaly cases are currently missing in this area. A further potential development which has been addressed in the current paper is to take into account personalization of the data processing which could be achieved by considering user conditions and tailoring the system towards specific groups of individuals, for example, elderly.

2. Overview of the Proposed Clinical Decision Support System (CDSS)

The case-based remote classification scheme, that is, the CDSS proposed here, consists of three modules, and the steps of the approach are illustrated in Figure 1. The *client side module* contains an Android application which is deployed in a smart phone using 2.3.3 programming environment. The Android application consists of a personalized health profiling approach and a Bluetooth communication protocol.

The personalized health profiling approach is a four-session approach in 9-minute duration, and the sessions are *baseline*, *deep breath*, *activity*, and *relax*. The details of the personalized health profiling approach can be found in [3]. The "WristOx2" sensor is connected as shown in Figure 1 and sends data through the Bluetooth communication. The details of Bluetooth communication and information about "WristOx2" sensor can be found in [24].

After each test, the measurements are stored into the smart phone in ".txt" format and finally uploaded to the server. When the *server side application module* receives the data file, the file is parsed, the features are extracted and a new problem case is formulated, and the CBR cycle is initiated as presented in Figure 2. Here, *k*-means clustering is used to group the cases offline. The new case is then entered into CBR cycle and in the retrieval step, the similarity value is calculated among the stored clustered cases.

The server retrieves the most similar cases depending on user defined similarity threshold and finally sends the most similar case information to the subject. At the same time, the server also sends an SMS and email notification to clinicians and generates an alert if the stored cases are not similar to the new case. Here, the system also considers user defined similarity threshold. Also, a web-enabled case-based CDSS is proposed here. Here, clinicians can see details about contextual information and measurements about a subject enabling remote monitoring of the subject. In the web-enabled CDSS, the system provides a number of functionalities, such as (1) *features level analysis*, (2) *CBR classification with clustered measurements*, and (3) *CBR classification with expert opinion*. In *features level analysis*, user can consider extracted features in different individual domains or their combination, that is, time, frequency, and time-frequency domains. In *CBR classification with clustered measurements*, users have the opportunity to see similar measurements individually, that is, similar pulse rate or similar oxygen saturation signals. Here, the signals are clustered separately using unsupervised clustering algorithm (i.e. *k*-means). Moreover, the CDSS can

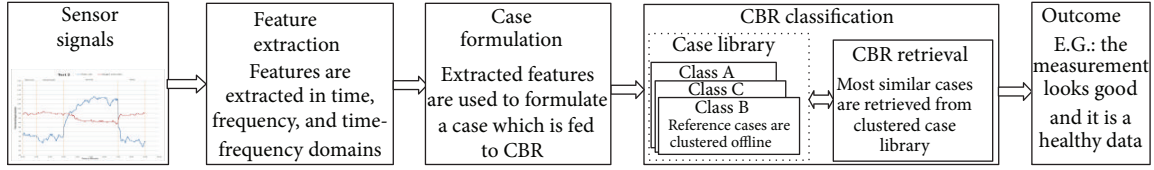


FIGURE 2: Different steps in the server system.

TABLE 1: A summary table with several criteria about the data collection.

Criteria	Values
Number of the subjects	15
Number of the measurements	29
Age range	65 to 83 years
Weight range	60 to 97 Kg
Height range	160 to 181 cm
Number of the male subjects	7
Number of the female Subjects	8
Speed on treadmill	3.0 to 5.0 km/h
Blood pressure range before	Systolic 132 to 201 diastolic 74 to 116
Blood pressure range after	Systolic 117 to 234 diastolic 63 to 119
Have diseases	6 subjects
Pulse rate	55 to 146 beats per minute
Oxygen saturation	82 to 100%

be also used while there is no expert classification available, that is, if the CBR system failed to retrieve any similar cases with a higher similarity value as a threshold.

3. Study Design and Data Analysis

Data were collected from 15 elderly persons using a 9-minute personalized health profiling approach with 4 sessions (i.e., baseline, deep breath, activity, and relax) discussed in [3]. The summary of the data observation is illustrated in Table 1. As it can be seen from Table 1, 29 measurements were collected from 15 subjects (7 were male and 8 were female) between the age of 65 and 83 years. Among the 15 subjects, 6 of them were diagnosed previously of having 1 or 2 kinds of different diseases and of which they were at the time of the experiment undergoing treatment. The others were comparatively healthy. However, some of them were having high blood pressure during the data collection.

During the test, the subjects used a treadmill in which the minimum speed was 3 km per hour and the maximum speed was 5 km per hour. The speed was considered according to the subject's capability of walking. The pulse rate among the subjects was 55 beats per minute as the minimum and 146 beats per minute as the maximum and oxygen saturation between 82% and 100% was observed. The sensor reading was closely monitored during data collection to ensure that the connection was not loose. The preprocessing done directly after data collection removed any erroneous data values caused by a loose collection, and one measurement was found

with a huge amount of sensor error in data reading and therefore it was not considered.

The collective 29 measurements of (a) pulse rate and (b) oxygen saturation are plotted in Figure 3, where the baseline, deep breath, activity, and relax are the 4 sessions. As it can be seen from Figure 3(a), the pulse rate varies between 65 and 95 in baseline and deep breath; however, the pulse rate increases when subjects are doing the activity task and decreases while they are resting (relax). Similarly, the oxygen saturation during the activity drops for some of the subjects as shown in Figure 3(b). Here, the changes of pulse rate and oxygen saturation are highly individual due to health factors, metabolic activity, and so forth. Thus, interpreting/analysing the measurements and understanding the large variations in the measurements from diverse subjects require knowledge and experience which is often very difficult to model even for an expert of the domain.

4. Features Extraction and Case Formulation

The collected measurements from the pulse oximeter are processed to extract features in time domain, frequency domain, and time-frequency domain. The numbers of features are considered based on previous study [4, 5, 25–27] and expert opinion. In the *time domain*, statistical features like maximum, minimum, arithmetic mean, and standard deviation of data are considered.

Frequency domain is also considered for the feature extraction as it has been observed in the previous works [25, 26] that the distinction between healthy and diseased people's data is visible in the frequency domain. In order to obtain the frequency domain features, a power spectral density (PSD) was calculated as shown in Figure 4. The PSD is the squared amplitude of the discrete Fourier transform (DFT) which is achieved by using fast Fourier transform (FFT) algorithm on the process measurements. In DFT, an input function is required which should be discrete, and this is done by sampling the pulse rate or oxygen saturation over a period of time. N length of time-series sequence is transformed into another sequence of N complex numbers by the function presented in (1), where T is the sampling frequency

$$Y \left[\frac{n}{NT} \right] = \frac{1}{N} \sum_{k=0}^{N-1} y[kT] e^{-i2\pi nkT/NT}. \quad (1)$$

Here, the measurements are processed through a scaling procedure to sample the measurements in 1 Hz frequency range. From the power spectral density low-frequency power, high-frequency power, low-frequency power to high-frequency power ratio, low-frequency peak power spectral

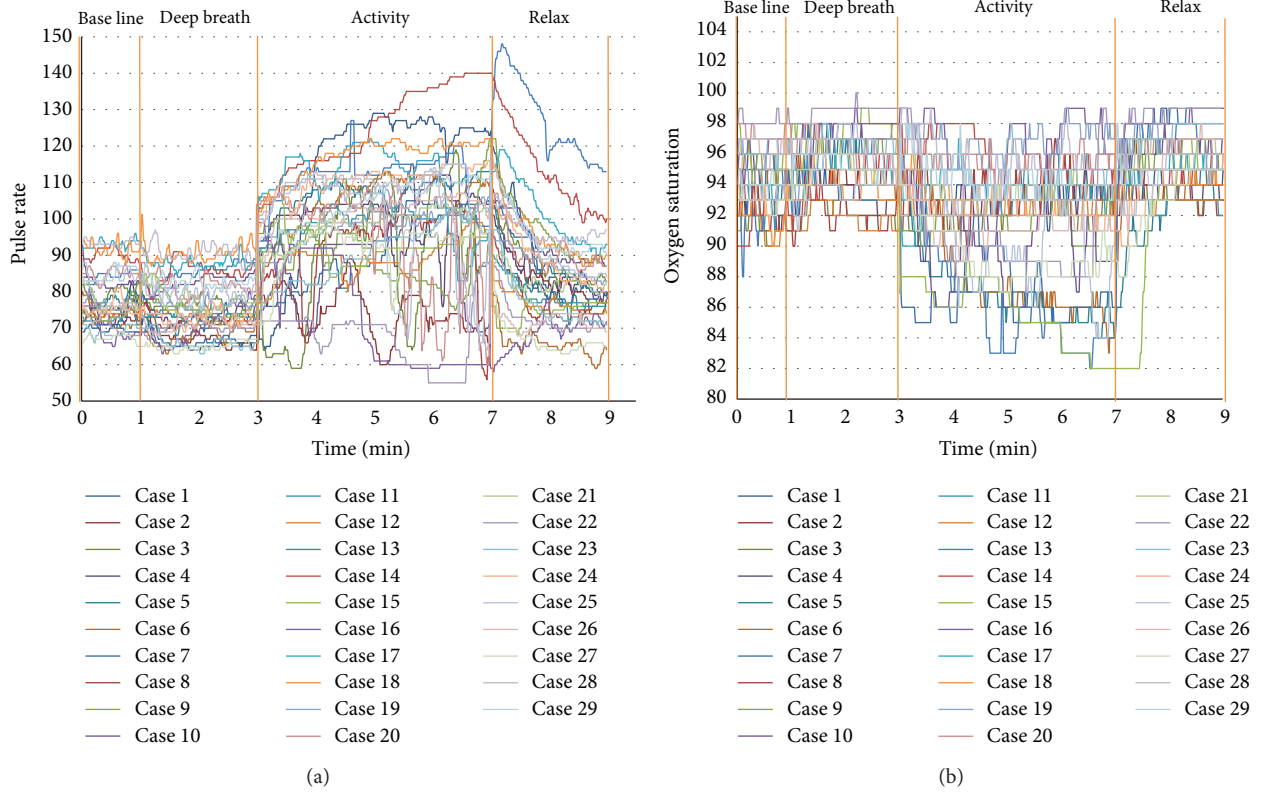


FIGURE 3: 29 measurements to illustrate the large variations during the 4 steps. y -axis: (a) pulse rate in beats per minute, (b) oxygen saturation in percentage and x -axis: time in minutes (total 9 minutes).

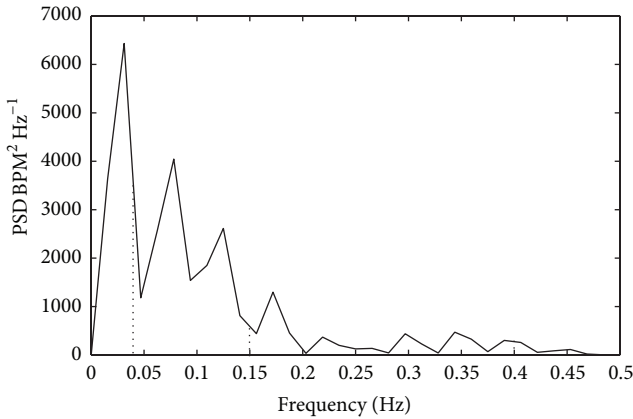


FIGURE 4: PSD plot of pulse rate, LF and HF region is marked by dotted line.

density, and high-frequency peak power spectral density were calculated [25]. Frequencies between 0.04 Hz and 0.15 were considered as low-frequency, and frequencies between 0.15 and 0.4 were considered as high-frequency as shown in Figure 4 [25]. The power in high- and low-frequency regions was calculated by numerical integration of the power spectral density of the corresponding frequency range. The unit of the power spectral density and power for the pulse rate was BPM² (beats per minute) Hz⁻¹ and BPM²,

respectively. Similarly, the frequency domain features for the oxygen saturation were calculated, with the unit of the power spectrum density and the power of (%)² Hz⁻¹ and (%)², respectively.

In *time-frequency domain* features, a discrete wavelet transform (DWT) is performed since it can keep the information of both time and frequency. Statistical features maximum, minimum, arithmetic mean, and standard deviations were calculated from the approximation coefficient of wavelet decomposition of level 1 [27]. The function “Daubechies 2” was used as the mother wavelet. The continuous wavelet transform linked to mother wavelet $\psi(t)$ can be defined by

$$W(a, b) = \int_{-\infty}^{\infty} y(t) \psi_{ab}(t) dt, \quad (2)$$

where $y(t)$ is any square integral function and a, b are scaling and translation parameters, respectively. Evaluating the continuous wavelet at dyadic interval of the signal can be expressed by

$$y(t) = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} d_j(k) 2^{j/2} \psi(2^j t - k), \quad (3)$$

where d_j is the discrete wavelet coefficient of the signal $y(t)$. Symmetric padding was used to make the data samples power of two to implement discrete wavelet transform [27].

In CBR, generally a case can be comprised using problem and solution part, in which the problem contains a vector of

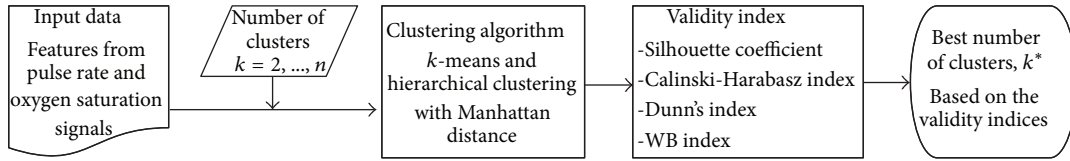


FIGURE 5: Schema diagram of clustering validity process.

features to describe a problem, and the solution part contains classification. To formulate a case, all features from the three domains in four sessions plus the subjects' contextual information (age, weight, gender, and blood pressure) were taken. Blood pressure was measured twice, once before and once after the end of taking measurements from pulse oximeter. As a result, each case contains total 59 ($4 \times 13 + 7$) features in problem description part of the cases. The list of the features with detail information is presented in our previous article [3]. As the solution of a case, this part of the cases contains expert classification where the classification is done considering three different classes. The three classes are (a) Class A, the measurement *looks good and it is a healthy data*, (b) Class B, the measurement *looks ok and it is a normal data*, and (c) Class C, the measurement *looks different and needs to look more in detail*.

5. Cluster Analysis in the System

In order to group the measurements, clustering methods have been applied. This group information helps CBR to do a fast retrieval; that is, it looks only most similar group of cases. Moreover, clinicians could get benefit in situations when CBR fails to retrieve any similar case for a given high threshold similarity value. To find a set of unlabeled similar groups of time-series, several clustering methods have been reported in the literature. Depending on problem and application, different kinds of clustering algorithms are applied [28]. The two most common algorithms of time-series clustering are hierarchical clustering and k -means in which the similarity function is effective on result [29].

Figure 5 illustrates the schema diagram for the process of finding the best number of clusters and the best algorithm for clustering data. In this process, two clustering algorithms, hierarchical and k -means, are applied on the feature vector of data. In *hierarchical algorithm*, in each step, pairwise distance of vectors (clusters) is computed and then the similar vector (clusters) are merged into a new cluster until all data goes to a single cluster [30]. To find the similarity between clusters, single linkage distance is used [29]. In *k-means algorithm*, k vectors are selected as initial centers of clusters. Then other vectors are assigned to these k centers based on the defined similarity function. (In this work we used Manhattan distance.) Then the center of each cluster is updated to the average of its members. These steps are repeated until the center of clusters is fixed [29].

In the step of Figure 5, to find the best clustering algorithm and also the best number of clusters, for each result of clustering step, some general methods and indices are applied

to measure the quality of clustering algorithms and validation of obtained clusters. Here, we used four validity indices. Silhouette coefficient (SC) [31] and Dunn's index (DI) [32] are the popular validity indices that show the compactness of data within the clusters and separation between clusters. Calinski-Harabasz index (CH) [32] is a sum of square-based index that evaluates the quality of partitioning [31]. In these three methods, the maximum of the index value determines the optimum clustering. WB index [33] is a new validity index that emphasizes the effect of sum of squares within cluster with multiplying the number of clusters. This method determines the optimum number of clusters by minimal value of index. By comparing the results of validity indices, this process will be able to recognize the best clustering algorithm and cluster numbers (k^*).

6. Case-Based Classification

The function of the case-based classification is to retrieve the most similar cases. Weighting of features is an important task for retrieving similar cases. To determine the degree of importance of each feature, expert knowledge is needed. As another alternative, feature weighting could be done automatically using some artificial intelligence techniques. In our previous research [7], the automatic weighting is investigated by learning from the case base; that is, it distinguishes individual features in terms of discriminating powers on the discretized universes of features. However, it does not show a better result than the expert's weighting. Moreover, in order to perform automatic weighting of features, the algorithm needs a high volume of cases with their corresponding classification. Therefore, in some medical domains, automatic weighting is not preferable. In this work, weights of the features are defined by a domain expert. These weights are then used to retrieve similar cases; the detail of features weighting is presented in our previous paper [3]. Similarity of a feature value between two cases (i.e., a target case and one case from library) is measured using the normalized Manhattan distance between the feature values of the two cases. Nonnumeric features such as gender are converted to a numeric value by substituting the contextual value with a numeric one (1 for male, 0 for female). The Manhattan distance function to calculate the similarity of a feature between two cases is shown in (4), where T_i and S_i are the i th feature values of target and source case, respectively,

$$\text{sim}(T_i, S_i) = 1 - \frac{|T_i - S_i|}{\max\{T_i, \text{Max}(i)\} - \min\{T_i, \text{Min}(i)\}} \quad (4)$$

Here, $\text{Max}(i)$ and $\text{Min}(i)$ represent the *maximum* and *minimum* values of the feature i obtained from the whole case library. Then “max” and “min” functions compare the values between the new case feature T_i and *maximum* and *minimum* values obtained from the case library. The function returns 1 if the values are the same and returns 0 if the values are dissimilar. This is known as a local similarity function

$$\text{sim}(T, S) = \frac{\sum_i^n w_i \times \text{sim}(T_i, S_i)}{\sum_i^n w_i}. \quad (5)$$

The similarity between the two cases is then measured using the weighted average of all the features that are to be considered. The function for calculating similarity between two cases T and S with n features is presented in (5), where w_i is the weight of the feature i defined by an expert of the domain. Note that, in the weight vector w_i is also considered the weight of three domains (i.e., time, frequency, and time-frequency features) and the weight of four sessions (i.e., *baseline*, *deep breath*, *activity* and *relax*). In CDSS, the CBR system retrieves both of the cases of pulse rate and oxygen saturation in parallel since there is no domain knowledge defined to combine them. However, a weighting approach is added in the CDSS to find an overall classification. Here, according to expert, the measurement of the pulse rate is more important than the oxygen saturation. The weighted average function presented in (6) to calculate a complete similarity value, that is, to find an overall classification by the CBR retrieval, is used:

$$\text{Sim}_{\text{case}}(\text{Pulse}, \text{Oxy}) = \frac{W_P \times \text{Sim}_{\text{pulse}} + W_O \times \text{Sim}_{\text{oxy}}}{W_P + W_O}, \quad (6)$$

where $\text{Sim}_{\text{case}}(\text{Pulse}, \text{Oxy})$ calculated the complete similarity value considering the individual similarity value of pulse rate as $\text{Sim}_{\text{pulse}}$ and the similarity value of oxygen saturation as Sim_{oxy} . W_P and W_O are the weight given by the expert.

7. Experimental Works

The experimental work is conducted as: (1) *proof-of-concept prototype*, (2) *classification accuracy of CBR with controlled measurements*, (3) *classification accuracy of CBR with clustered measurements*, and (4) *classification accuracy of CBR considering expert classification*.

7.1. The Proof-of-Concept Prototype. The web-based CDSS has been built using PHP and JavaScript programming language, and case libraries are developed in MySQL database. The main goal of this experiment is to see whether the several functions implemented in the web-based system are working properly or not. The DSS using case-based retrieval is verified by implementation as a prototype in which all the implemented methods are compared according to their outcome, that is, in terms of the technical point of view. According to Watson [17], these trials have been conducted through the following 4 tests: (1) *retrieval accuracy*, (2) *retrieval consistency*, (3) *case duplication*, and (4) *global test*. For the test *retrieval accuracy*, a “leave-one-out” retrieval technique is

used; that is, one case is taken from the case library as a query case, and then the system retrieves the most similar cases. Among the retrieved cases, the query case is also retrieved as the top similar case with the similarity value 1.0; that is, the similarity value of two same cases is computed as 100%. To test the *retrieval consistency*, the same query is used to perform more than one similar search, and if it has been found that the same stored cases have been retrieved with the same similarity, then the implemented retrieval function is considered to have consistency. It is also observed that no cases are identical during retrieval except the query case, when it matches itself; thus, *case duplication* is also checked. Regarding the global test, the classification accuracy of the CBR system is performed and discussed in the next section.

7.2. Classification Accuracy of CBR with Controlled Measurements. For these experiments, the proposed system uses the sessions from health profiling. That is, each case is then further divided into 4 categories according to the sessions that are, *baseline*, *deep breath*, *activity*, and *relax*. The reason behind this is that the measurements are very much controlled; subjects are following the tasks according to the condition of health profiling. Moreover, there is always a question about the gold standard of expert. The main goal of this experiment is to see how accurately the CBR approach can classify the signals. Two case libraries are used with 15 subjects; one is for only pulse rate (29 signals) and another one is for oxygen saturation (29 signals). Each case library contains 116 cases (29 measurements \times 4 sessions) with 4 classes in total, and each case contains 13 features extracted (both from time domain, frequency domain, and time-frequency domain) from the signals. For CBR classification, a “leave-one-out” retrieval technique (i.e., 1 out of 29 cases) is applied and the top most retrieved similar case ($k = 1$) is considered. If the considered case is retrieved from the same class, then we count 1 as correct classification, and percentage of each class is calculated after querying all 29 cases. The results for percentage of correct classification considering pulse rate case library are shown in Table 2.

Table 2 illustrates the results for the pulse rate cases; the first three columns present the classification accuracy considering only each of individual domain features. That is, when it is only time domain, all the other features in frequency and time-frequency domains are weighted as “0.” It can be observed, CBR approach has achieved the highest percentage of classification accuracy while considering the frequency domain features. The average value of the classification accuracy for the frequency domain features is 85.3%, for the time domain features it is 44%, and for the time-frequency domain features it is 48.3%. However, when the features are combined, the highest accuracy has been increased on average to $\approx 5\%$ which is presented in the last column of Table 2. The combination has been performed using a weight vector in which both the time and time-frequency domains features get 10% each, and frequency domain features get 80%. Average classification accuracy is around 90% while it combines all the domain features.

TABLE 2: Percentage of correct classifications of each class considering pulse rate case library.

Sessions	Using only time domain features (TmF)	Using only frequency domain features (FrqF)	Using only time-frequency domain features (TmFrqF)	Using a combination of features TmF = 10%, TmFrqF = 10%, and FrqF = 80%
Baseline	51.7%	96.5%	62.0%	96.5%
Deep breath	27.6%	79.3%	34.5%	83.0%
Activity	45.0%	96.5%	45.0%	100.0%
Relax	51.7%	69.0%	51.7%	80.0%
Average	44.0%	85.3%	48.3%	89.8%

TmF: time domain features, FrqF: frequency domain features, and TmFrqF: time-frequency domain features.

TABLE 3: Percentage of correct classifications of each class considering oxygen saturation case library.

Sessions	Using only time domain features (TmF)	Using only frequency domain features (FrqF)	Using only time-frequency domain features (TmFrqF)	Using a combination of features TmF = 10%, TmFrqF = 10%, and FrqF = 80%
Baseline	31.0%	100.0%	27.5%	100.0%
Deep breath	24.1%	65.5%	34.5%	72.0%
Activity	62.0%	100%	51.1%	100.0%
Relax	27.6%	38.0%	24.7%	48.0%
Average	36.2%	76.0%	34.5%	80.0%

TmF: time domain features, FrqF: frequency domain features, and TmFrqF: time-frequency domain features.

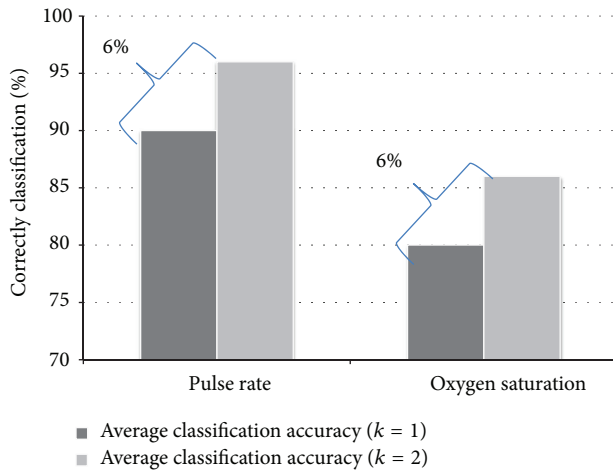


FIGURE 6: The classification accuracy increased 6% while considering top 2 retrieved similar cases, that is, $k = 2$.

Similarly, Table 3 presents the classification results based on the oxygen saturation case library.

Table 3 presents results for the oxygen saturation; it shows the highest percentage of correct classification using the frequency domain features. The result has increased 4% on an average while it combines the three domain features. The average value of the correct classification is 80% with combination of features and 76% without combination, that is, only considering the frequency domain.

Figure 6 presents the average value of correctly classified cases both for the pulse rate and oxygen saturation data. It

TABLE 4: Confusion matrix for pulse rate (a) and oxygen saturation (b) considering $k = 2$, using CBR with extracted features.

(a)				
	Baseline	Deep breath	Activity	Relax
Baseline	29	0	0	0
Deep breath	0	28	0	1
Activity	0	0	29	0
Relax	1	3	0	25
Accuracy in percentage	100%	96.5%	100%	86.2%

(b)				
	Baseline	Deep breath	Activity	Relax
Baseline	29	0	0	0
Deep breath	0	25	0	4
Activity	0	0	29	0
Relax	0	12	0	17
Accuracy in percentage	100%	86.2%	100%	59%

shows the comparison result in the classification accuracy values while considering the top most similar retrieved case, that is, $k = 1$ and top 2 most similar retrieved cases, that is, $k = 2$. It can be seen from Figure 6 that the overall accuracy is increased by 6% while considering $k = 2$ and the classification accuracy for the pulse rate is 96% and for the oxygen saturation is 86%. The confusion matrices both for (Table 4(a)) the pulse rate and (Table 4(b)) oxygen saturation while considering top 2 cases are presented in Table 4.

The distribution of the classification using CBR can be seen both from the (a) pulse rate and (b) oxygen saturation in Table 4; the *baseline* and *activity* classes are classified 100%

correctly. However, the classification in *deep breath* and *relax* are missing. For example, in *deep breath* for the pulse data (Table 4(a)), one case is misclassified as *relax*, and in *relax* 4 cases are misclassified, 3 as *deep breath* and 1 as *baseline*. Similarly, Table 4(b), 4 *deep breath* cases are misclassified as *relax*, and 12 of *relax* class cases are misclassified as *deep breath*.

7.3. Classification Accuracy of CBR with Clustered Measurements. This experimental work has been conducted in twofold: (a) *evaluating the clustering algorithms* and (b) *classification accuracy* using CBR. Here, the entire measurement is considered and clustered using well-known clustering algorithm; that is, the measurements include all the sessions, baseline, deep breath, activity, and relax. However, the measurements of pulse rate and oxygen saturation are still separated and clustered separately.

(a) *Evaluating the Clustering Algorithms.* The main aim for this experimental work is to identify the clustering method and number of clusters. Here, all the extracted features (i.e., features in time, frequency, and time-frequency domains) form 29 measurements are considered.

Two well-known clustering algorithms, *k*-means and single linkage hierarchical clustering, are applied [29, 30], and up to 5 clusters are generated by each algorithm. In order to make meaningful clusters and also find the best clustering method, four validation indices, Silhouette coefficient (SC) [31], Dunn’s index (DI) [34], Calinski-Harabasz index (CH) [32], and WB index (WB) [33], are compared. As it can be seen from Figure 7, for pulse rate (a) and oxygen saturation (b), *k*-means clustering is better than single linkage hierarchical clustering. Also the best number of clusters is 3 for both. In SC, DI, and CH the maximum value shows the best clustering and the best number of clusters (star points), and in WB the minimum value is the best. According to the above evaluation, *k*-means clustering algorithm with 3 clusters achieved the highest evaluation value. Therefore, here the clustered cases with *k*-means are considered. The distributions of cases in 3 clusters (i.e., Classes A, B, and C) for pulse rate are presented in Table 5(a), and Classes A’, B’, and C’ for the oxygen saturation are presented in Table 5(b).

(b) *Classification Accuracy of CBR.* The main goal is to see how close the CBR system can classify compared to the clustering method. The class information is added into the case solution part, and the “leave-one-out” retrieval technique is applied. Here, the top 2 most similar cases ($k = 2$) are considered; that is, if both the query case and one of the retrieved cases belong to a similar class, then the number of correctly classified cases is counted as 1.

Finally, a percentage of the correctly classified cases are calculated on a total of 29 cases, and the results are presented in Table 6. From Table 6, when $k = 1$, CBR performs $\approx 80\%$ correctly for the pulse rate and $\approx 90\%$ for the oxygen saturation as close to the clustering algorithm (i.e., *k*-means). However, the percentage of correctly classified cases is improved while $k = 2$ and the achieved value is $\approx 93\%$ for both the pulse rate and oxygen saturation. For the pulse rate (a) and oxygen saturation (b) the confusion matrices of

TABLE 5: Case distribution in 3 clusters both for pulse rate (a) and oxygen saturation (b).

(a)			
Criteria	Class A	Class B	Class C
Case.ids	1, 3, 5, 7, 14, 15, 16, 17, 19, 25, 26	2, 4, 11, 12, 18, 20, 22, 28, 30	8, 9, 10, 13, 21, 23, 24, 27, 29
Total number of cases	11	9	9
(b)			
Criteria	Class A’	Class B’	Class C’
Case.ids	1, 3, 7, 8, 10, 12, 17, 22, 24, 28, 30	2, 5, 13, 14, 15	4, 9, 11, 16, 18, 19, 20, 21, 23, 25, 26, 27, 29
Total number of cases	11	5	13

TABLE 6: Classification accuracy in CBR considering clustering.

29 measurements	Percentage of correctly classified cases ($K = 1$)	Percentage of correctly classified cases ($K = 2$)
Pulse rate	79.30%	93.10%
Oxygen saturation	89.65%	93.10%
Average	84.5%	93.1%

TABLE 7: Confusion matrix for pulse rate (a) and oxygen saturation (b) considering $k = 2$.

(a)				
	Class A	Class B	Class C	Total
Class A	11	0	0	11
Class B	0	9	0	9
Class C	1	1	7	9
Accuracy in percentage	100%	100%	78%	29
(b)				
	Class A’	Class B’	Class C’	Total
Class A’	11	0	0	11
Class B’	0	3	2	5
Class C’	0	0	13	13
Accuracy in percentage	100%	60.0%	100%	29

correctly classified cases using CBR ($k = 2$) compare to the clustered cases are presented in Table 7.

7.4. Classification Accuracy of CBR Considering Expert Classification. The main goal of this experimental work is to investigate the classification performance of the CBR approach compared to an expert of the domain. Here, the word “expert” means a panel which consists of three researchers working mostly with biomedical sensor signals. The panel has examined and analyzed the cases and performed an overall classification; that is, all the 29 test data (both pulse and oxygen saturation measurements) from 15 subjects were classified by the panel into three classes. The three classes are

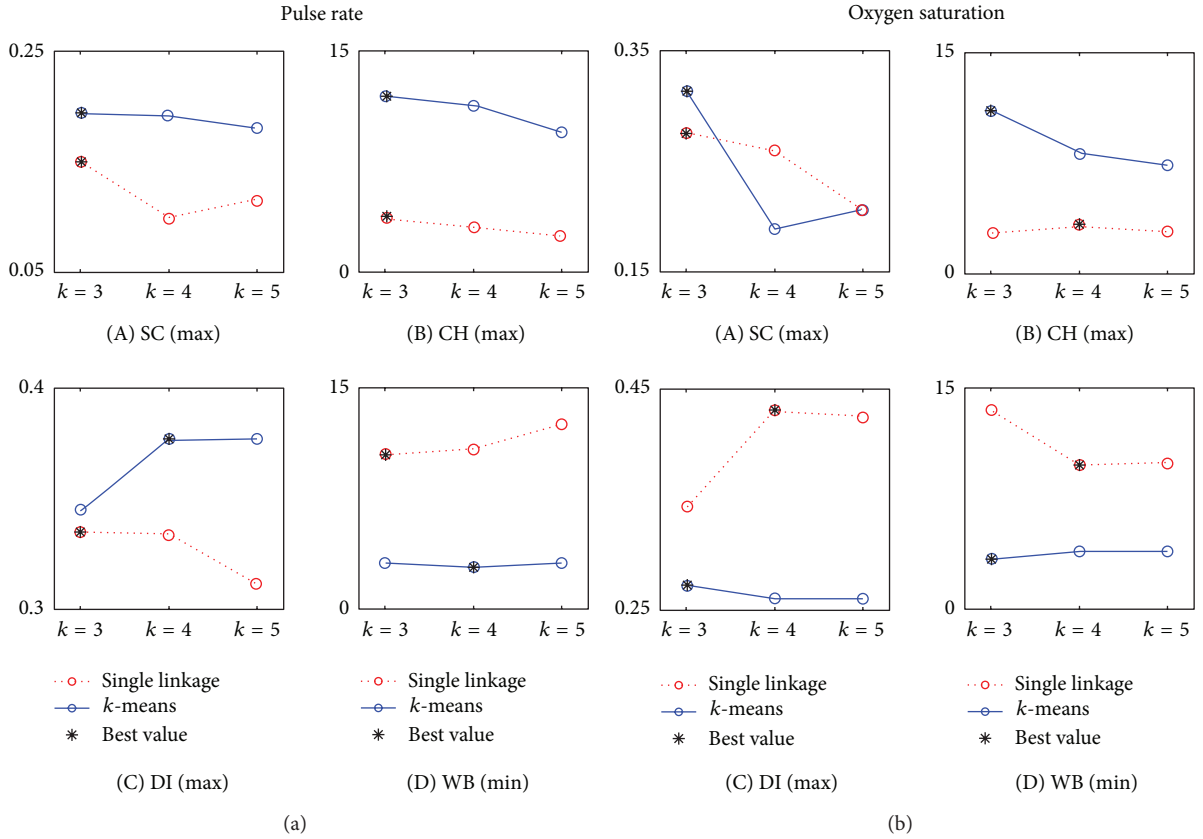


FIGURE 7: The results of different validity indices on k -means and hierarchical single linkage clustering for $k = 3$, $k = 4$, and $k = 5$ in pulse rate (a) and oxygen saturation (b). (A) Silhouette coefficient (SC), (B) Calinski-Harabasz index (CH), (C) Dunn's index (DI), and (D) WB index (WB).

(i) Class A, the measurement *looks good and it is a healthy data*, (ii) Class B, the measurement *looks ok and it is a normal data*, and (iii) Class C, the measurement *looks different and needs to look more in detail*. A majority voting approach is considered to perform the final classification on cases. For example, for a case “Case_id_7.1,” a set of classification is achieved from the panel; that is, Case_id_7.1 = (Class B, Class B, and Class A) and Class B is taken for final classification of the case. Note that the panels agree with each other in 72.4% of the cases (21 cases out of 29).

Similar to the previous experiments, the “leave-one-out” retrieval technique is applied for CBR, and a percentage of correctly classification cases are calculated. Note that the final similarity value is calculated by the weighted average of individual similarity value of pulse rate and oxygen saturation where pulse rate receives a higher importance than the oxygen. Moreover, the experiment has been conducted considering both $k = 1$ and $K = 2$ as discussed earlier. The performance of the CBR system in terms of classification accuracy is illustrated in Table 8.

It can be observed from Table 8 that the CBR approach can classify with 92% accuracy for Class A, 37.5% for Class B, and 75% for Class C while considering the singular top most similar retrieved case (i.e. $k = 1$). The performance of the CBR classification increased while considering two

TABLE 8: Classification accuracy in CBR considering expert overall classification.

	Total number of cases	Percentage of correctly classified cases ($K = 1$)	Percentage of correctly classified cases ($K = 2$)
Class A	13	92.3%	100.0%
Class B	8	37.5%	62.5%
Class C	8	75.0%	75.0%
Total	29	72.3%	83.0%

top similar classes; that is, Class A achieved 100%, Class B achieved 62.5%, and Class C remains the same. So, in total, the accuracy of the CBR system was 72.3% for $k = 1$ and 83% for $k = 2$. The confusion matrix of correctly classified cases using CBR and $k = 2$ while comparing to the expert classes is presented in Table 9.

According to Table 9, 3 classes out of 8 cases in Class B were misclassified as Class A, and 2 classes out of 8 cases in Class C were misclassified as Class A, and Class B respectively. According to the class definition defined by the expert, Class A and Class B, represent as healthy or normal conditions of the subjects, and Class C is the opposite. So, in order to perform a sensitivity and specificity analysis, 29 cases were

TABLE 9: Confusion matrix considering $k = 2$ in CBR.

	Class A	Class B	Class C	Total
Class A	13	0	0	13
Class B	3	5	0	8
Class C	1	1	6	8
Accuracy in percentage	100%	62.5%	75%	29

TABLE 10: Statistical analysis of the CBR classification considering $k = 1$.

Criteria/indices	Values
Total cases	29
Not healthy (Class C) group (P)	8
Healthy (Class A + Class B) group (N)	21
True positive (TP)	6
False positive (FP)	1
True negative (TN)	20
False negative (FN)	2
Sensitivity = $TP/(TP + FN)$	≈ 0.75
Specificity = $TN/(FP + TN)$	≈ 0.95
Accuracy = $(TP + TN)/(P + N)$	≈ 0.90

further divided into 2 groups: *healthy* = (Class A + Class B) and *not healthy* = Class C. Therefore, the *healthy* group consists of 21 cases, and *not healthy* group consists of 8 cases. Note that this time we only consider $K = 1$ in CBR retrieval, and the results are presented in Table 10.

As can be observed from Table 10, in *healthy* group, 20 cases are correctly classified that represent true negative (TN), and 1 is misclassified that represent false positive (FP). Similarly, in *not healthy* group, 6 cases are correctly classified that represent true positive (TP), and 2 are misclassified that represent false negative (FN). So, the sensitivity, specificity, and overall accuracy achieved by the CBR approach are 75%, 95%, and 90%, respectively, while considering only one top similar retrieved case (i.e. $k = 1$).

8. Summary and Discussion

This paper proposed a case-based clinical decision support system to monitor individuals' health condition remotely. Here, a pulse oximeter is used together with a personalized health profile protocol. The pulse oximeter is low-cost and easy to use sensor in any environment. The protocol helps to establish individual health profile.

The system has been designed in three-tier client server architecture. The measurements are collected through an Android application, and a case library has been implemented in a remote server. The development of the server side consists of the feature extraction part to find key parameters of interest using time, frequency, and time-frequency domains. Besides, this paper introduces a time-series similarity measure and a clustering approach to group the cases offline which makes the retrieval process faster by looking only within the most similar group of cases. This will be helpful for large datasets, and our future target

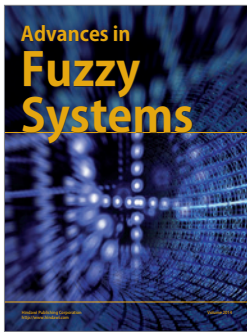
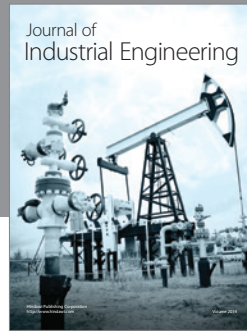
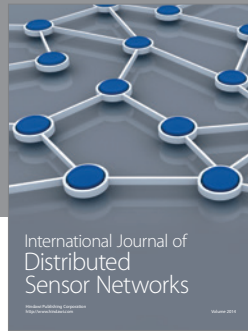
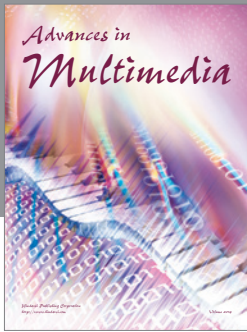
is to test the system in a large-scale environment. A web-enabled CDSS with advanced signal analysis capability could be accessed via specific clients, for example, physician at a health clinic in order to monitor health. In the web-enabled CDSS, the system provides a number of functionalities, such as (1) features level analysis, (2) CBR classification with clustered measurements and (3) CBR classification with expert opinion.

The proposed CDSS has been evaluated considering *controlled measurements*, using *clustered measurements*, and considering *expert classification*. As can be observed from Figure 6, for the pulse rate, the classification accuracy of the CBR approach is 96%, and for the oxygen saturation it is 86% considering the controlled measurements. Here, the evaluation shows better performance when the three domains features are combined rather than single domain features. According to the confusion matrices presented in Table 4, the cases in *baseline* and *activity* conditions are well classified than the other two conditions, that is, *deep breath* and *relax*. The cases belong to these conditions are classified as each other and could be explained by the fact that the two conditions *deep breath* and *relax* are rather similar [35]. Note that the classification performance for CBR has been achieved by considering the 2 top most similar retrieved cases (i.e., $K = 2$) since the similarity value of these two cases is often very close to each other (e.g., 96.34 and 96.10). From these experiments, it can be observed that using different test data sets, methods and considering expert classification, the system can perform its classification task and generate alarms in an anomalous situation. The results show that the approach could be applicable to use in a clinical remote health monitoring. Nevertheless, the performance of the system could be experimented considering a large case library with reference cases. To show the superiority of the proposed approach, a comparison with the other approaches such as neural networks or support vector machine could be obtained, and this is now ongoing. Moreover, the system should be verified in real clinical environment for day-to-day usage.

References

- [1] S. Youm, G. Lee, S. Park, and W. Zhu, "Development of remote healthcare system for measuring and promoting healthy lifestyle," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2828–2834, 2011.
- [2] J. Gong, S. Lu, R. Wang, and L. Cui, "PDhms: pulse diagnosis via wearable healthcare sensor network," in *Proceedings of the IEEE International Conference on Communications (ICC '11)*, pp. 1–5, 2011.
- [3] M. U. Ahmed, A. M. Islam, and A. Loutfi, "A case-based patient identification system using pulse oximeter and a personalized health profile," in *Proceedings of the Health Sciences at 20th International Conference on Case-Based Reasoning (ICCB '12)*, B. Isabelle, M. Stefania, and M. Cindy, Eds., Springer, Lyon, France, September 2012.
- [4] Z. Jin, J. Oresko, S. Huang, and A. C. Cheng, "HeartToGo: a personalized medicine technology for cardiovascular disease prevention and detection," in *Proceedings of the IEEE/NIH Life*

- Science Systems and Applications Workshop (LiSSA '09)*, pp. 80–83, April 2009.
- [5] N. Krupa, M. A. MA, E. Zahedi, S. Ahmed, and F. M. Hassan, "Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine," *BioMedical Engineering Online*, vol. 10, article 6, 2011.
 - [6] A. Aamodt and E. Plaza, "Case-based reasoning: foundational issues, methodological variations, and system approaches," *AI Communications*, vol. 7, no. 1, pp. 39–59, 1994.
 - [7] S. Begum, M. U. Ahmed, P. Funk, N. Xiong, and B. Von Schéele, "A case-based decision support system for individual stress diagnosis using fuzzy similarity matching," *Computational Intelligence*, vol. 25, no. 3, pp. 180–195, 2009.
 - [8] J. M. Corchado, J. Bajo, and A. Abraham, "GerAmi: improving healthcare delivery in geriatric residences," *IEEE Intelligent Systems*, vol. 23, no. 2, pp. 19–25, 2008.
 - [9] M. U. Ahmed, S. Begum, P. Funk, N. Xiong, and B. von Schéele, "A three phase computer assisted biofeedback training system using case-based reasoning," in *Proceedings of the 9th European Conference on Case-based Reasoning*, Trier, Germany, 2008.
 - [10] S. Montani, L. Portinale, G. Leonardi, R. Bellazzi, and R. Bellazzi, "Case-based retrieval to support the treatment of end stage renal failure patients," *Artificial Intelligence in Medicine*, vol. 37, no. 1, pp. 31–42, 2006.
 - [11] I. Bichindaritz, "Prototypical case mining from biomedical literature for bootstrapping a case base," *Applied Intelligence*, vol. 28, no. 3, pp. 222–237, 2008.
 - [12] M. U. Ahmed, S. Begum, P. Funk, N. Xiong, and B. von Schéele, "Case-based reasoning for diagnosis of stress using enhanced cosine and fuzzy similarity," *International Journal of Transactions on Case-Based Reasoning on Multimedia Data*, vol. 1, no. 1, pp. 3–19, 2008.
 - [13] M. U. Ahmed, S. Begum, P. Funk, N. Xiong, and B. von Schéele, "A multi-module case based biofeedback system for stress treatment," *International Journal of Artificial Intelligence in Medicine*, vol. 51, no. 2, pp. 107–115, 2010.
 - [14] M. U. Ahmed and P. Funk, "A computer aided system for post-operative pain treatment combining knowledge discovery and case-based reasoning," in *Case-Based Reasoning Research and Development*, vol. 7466 of *Lecture Notes in Computer Science*, pp. 3–16, 2012.
 - [15] S. Begum, M. U. Ahmed, P. Funk, N. Xiong, and M. Folke, "Case-based reasoning systems in the health sciences: a survey of recent trends and developments," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 41, no. 4, pp. 421–434, 2011.
 - [16] S. Begum, M. U. Ahmed, and P. Funk, "Case-based systems in health sciences: a case study in the field of stress management," *WSEAS Transactions on Systems*, vol. 8, no. 3, pp. 344–354, 2009.
 - [17] I. Watson, *Applying Case-Based Reasoning: Techniques For Enterprise Systems*, Morgan Kaufmann Publishers, San Francisco, Calif, USA, 1997.
 - [18] M. Stacey and C. McGregor, "Temporal abstraction in intelligent clinical data analysis: a survey," *Artificial Intelligence in Medicine*, vol. 39, no. 1, pp. 1–24, 2007.
 - [19] I. Yoo, P. Alafaireet, M. Marinov et al., "Data mining in healthcare and biomedicine: a survey of the literature," *Journal of Medical Systems*, vol. 36, no. 4, pp. 2431–2448, 2011.
 - [20] S. Jeong, C. H. Youn, E. B. Shim, M. Kim, Y. M. Cho, and L. Peng, "An integrated healthcare system for personalized chronic disease care in home-hospital environments," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 4, pp. 572–585, 2012.
 - [21] G. K. Pang, "Health monitoring of elderly in independent and assisted living," in *Proceedings of the International Conference on Biomedical Engineering (ICoBE '12)*, pp. 553–556, February 2012.
 - [22] R. M. Rahman and F. R. M. Hasan, "Using and comparing different decision tree classification techniques for mining ICDDR,B Hospital Surveillance data," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11421–11436, 2011.
 - [23] C. M. Chen, "Web-based remote human pulse monitoring system with intelligent data analysis for home health care," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2011–2019, 2011.
 - [24] G. Koshmak, *An Android Based Monitoring and Alarm System for Patients with Chronic Obtrusive Disease [M.S. thesis]*, Department of Technology at Örebro University.
 - [25] K. Dingli, T. Assimakopoulos, P. K. Wraith, I. Fietze, C. Witt, and N. J. Douglas, "Spectral oscillations of RR intervals in sleep apnoea/hypopnoea syndrome patients," *European Respiratory Journal*, vol. 22, no. 6, pp. 943–950, 2003.
 - [26] C. Zamarrón, F. Gude, J. Barcala, J. R. Rodriguez, and P. V. Romero, "Utility of oxygen saturation and heart rate spectral analysis obtained from pulse oximetric recordings in the diagnosis of sleep apnea syndrome," *Chest*, vol. 123, no. 5, pp. 1567–1576, 2003.
 - [27] D. Cvetkovic, E. D. Übeyli, and I. Cosic, "Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: a pilot study," *Digital Signal Processing*, vol. 18, no. 5, pp. 861–874, 2008.
 - [28] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.
 - [29] T. W. Liao, "Clustering of time series data: survey," *Pattern Recognition*, vol. 38, pp. 1857–1874, 2005.
 - [30] C. A. Ratanamahatana, J. Lin, D. Gunopulos, E. J. Keogh, M. Vlachos, and G. Das, "Mining time series data," in *Data Mining and Knowledge Discovery Handbook*, pp. 1049–1077, 2010.
 - [31] L. Vendramin, R. J. G. B. Campello, and E. R. Hruschka, "On the comparison of relative clustering validity criteria," in *Proceedings of the 9th SIAM International Conference on Data Mining (SDM '09)*, pp. 729–740, May 2009.
 - [32] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, "Understanding of internal clustering validation measures," in *Proceedings of the 10th IEEE International Conference on Data Mining (ICDM '10)*, pp. 911–916, December 2010.
 - [33] Q. Zhao, M. Xu, and P. Fränti, "Sum-of-squares based cluster validity index and significance analysis," in *Adaptive and Natural Computing Algorithms*, vol. 5495 of *Lecture Notes in Computer Science*, pp. 313–322, 2009.
 - [34] K. R. Žalik and B. Žalik, "Validity index for clusters of different sizes and densities," *Pattern Recognition Letters*, vol. 32, no. 2, pp. 221–234, 2011.
 - [35] M. U. Ahmed and A. Loutfi, "Physical activity classification for elderly based on pulse rate," in *Proceedings of the 10th International Conference on Wearable Micro and Nano Technologies for Personalized Health Tallinn*, June 2013.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

