CrossMark

SYSTEMS-LEVEL QUALITY IMPROVEMENT

A decision support system to improve medical diagnosis using a combination of k-medoids clustering based attribute weighting and SVM

Musa Peker[1] ⓘ

Abstract The use of machine learning tools has become widespread in medical diagnosis. The main reason for this is the effective results obtained from classification and diagnosis systems developed to help medical professionals in the diagnosis phase of diseases. The primary objective of this study is to improve the accuracy of classification in medical diagnosis problems. To this end, studies were carried out on 3 different datasets. These datasets are heart disease, Parkinson's disease (PD) and BUPA liver disorders. Key feature of these datasets is that they have a linearly non-separable distribution. A new method entitled k-medoids clustering-based attribute weighting (kmAW) has been proposed as a data preprocessing method. The support vector machine (SVM) was preferred in the classification phase. In the performance evaluation stage, classification accuracy, specificity, sensitivity analysis, f-measure, kappa statistics value and ROC analysis were used. Experimental results showed that the developed hybrid system entitled kmAW + SVM gave better results compared to other methods described in the literature. Consequently, this hybrid intelligent system can be used as a useful medical decision support tool.

This article is part of the Topical Collection on *Systems-Level Quality Improvement*.

✉ Musa Peker
musa@mu.edu.tr

[1] Department of Information Systems Engineering, Faculty of Technology, Mugla Sitki Kocman University, 48000 Mugla, Turkey

Keywords Medical diagnosis · k-medoids clustering based attribute weighting · Support vector machine · Hybrid classification method · Decision support system

Introduction

Medical diagnosis refers to the process of identifying a particular disease by analyzing the symptoms. From a biomedical informatics aspect, a medical diagnosis is a classification operation incorporating a decision-making process that is based on available medical data. From this aspect, automatic medical diagnostic systems provide the advantages of structural computing power when large amounts of data are used. For example, with these systems, it is possible to learn from similar cases from a large database of patient records. Using this information, it can be possible to reach a decision quickly in terms of the current patient. This can be useful in helping the specialist. Moreover, these systems aim to minimize the possibility of physician error. The benefits of using such intelligent systems include increased diagnostic accuracy and a reduction in the time and cost associated with treatment [1–3].

Many researchers have been working on new computer-aided systems and technologies in order to help doctors diagnose particular diseases. Most of the newly-developed systems are tested on data with regard to diseases that has been gathered in the medical field and are open for use by all scientists. In this context, the performance of these systems is compared.

One of the most popular databases used for this purpose is the UCI machine learning repository [4]. The method proposed in this study was tested on heart disease, Parkinson's disease and BUPA liver disorders datasets obtained from this database and the results obtained were compared with studies in the literature. When selecting these datasets, diseases with

high mortality rates that affect the majority of society were selected. The summary of the information in the literature regarding studies that have been implemented on these datasets is presented below.

In the literature, there are studies carried out on Statlog heart disease dataset with the purpose of the diagnosis of heart disease. Duch et al. [5] have made comparative analyses using k-nearest neighbour (kNN), Manhattan with kNN, feature space mapping (FSM), and separability split value (SSV) algorithms. They obtained highest accuracy rate as 85.1 % with kNN algorithm. Sahan et al. [6] have presented a novel classification algorithm named as feature weighted artificial immune system (AWAIS). With the proposed method, they obtained 82.59 % classification accuracy. Polat and Gunes [7] have proposed a novel system based on a combination of attribute selection, artificial immune recognition system (AIRS) classifier and fuzzy weighted pre-processing. As a result, they obtained a classification accuracy of 92.59 %. Polat et al. [8] have developed a new system based on kNN based weighting pre-processing and AIRS with fuzzy resource allocation mechanism. They achieved 87 % classification accuracy. Ozsen and Gunes [9] have achieved 83.95 % classification accuracy with a new classifier called artificial immune system (AIS) with hybrid feature vectors. Kahramanli and Allahverdi [10] have used fuzzy neural network (FNN) algorithm for this problem. As a result, they obtained 86.8 % classification accuracy. Polat and Gunes [11] have proposed an attribute selection method named as kernel f-score feature selection (KFFS). In the study in which LS-SVM algorithm was used as a classification algorithm, 83.70 % classification accuracy was obtained. Das et al. [12] have proposed an ensemble method with three neural networks for diagnosis of heart disease. 89.01 % accuracy rate was achieved with the method. Subbulakshmi et al. [13] have proposed a novel learning algorithm for training of single layer feed-forward neural networks. They achieved classification accuracy of 87.50 % in diagnosis of heart disease with the method called as Extreme Learning Machine (ELM). Mantas and Abellán [14] have proposed a decision tree algorithm related to imprecise probabilities. They applied the algorithm called as Credal-C4.5 on different data sets. Researchers obtained 64.53 % classification accuracy for Statlog heart disease data set.

In the literature, numerous studies conducted on PD data set for diagnosis of PD. Shahbaba and Neal [15] have proposed a non-linear system based on Dirichlet mixtures. 87.7 % classification accuracy was obtained with the proposed method. Das [16] has applied 4 different methods for the diagnosis of PD. These methods are respectively, ANN, DMneural regression and decision trees. The highest accuracy rate was achieved with the ANN method. With this method, the accuracy rate of 92.9 % was achieved. Guo et al. [17] have developed a method based on genetic programming (GP) and

expectation maximization (EM). 93.1 % classification accuracy was obtained with the proposed method. Sakar and Kursun [18] have proposed a mutual information based attribute selection and a SVM based method and the accuracy rate of 92.75 % was obtained. Ozcift and Gulten [19] have proposed a method that combines 30 machine learning algorithms with rotation forest (RF) ensemble classifier. 87.13 % classification accuracy was obtained in the study in which the correlation based feature selection (CFS) algorithm was used as a feature selection algorithm. Aström and Koker [20] have obtained 91.2 % classification accuracy using a parallel neural network model for the diagnosis of Parkinson's disease. Luukka [21] has proposed a novel method based on fuzzy entropy measures and similarity classifier. 85.03 % classification accuracy was achieved with the proposed method. Li et al. [22] have used a non-linear fuzzy-based conversion method with SVM and have achieved 93.47 % classification accuracy. Ozcift [23] has used SVM attribute selection based rotation forest ensemble classifiers. With this method, the classification accuracy of 87.13 % was obtained. Polat [24] has applied k-nearest neighbor algorithm and fuzzy c-means based feature weighting method (FCMCBAW) and has achieved 97.93 % classification accuracy. Daliri [25] has proposed a method called as chi-square distance kernel-based SVM and obtained 91.2 % classification accuracy with this method. Zuo et al. [26] have presented a new method based on particle swarm optimization (PSO) which is one of heuristic optimization algorithms, and have obtained 97.47 % accuracy rate with the proposed method. Chen et al. [27] have proposed a method based on fuzzy kNN and principal component analysis (PCA). They obtained 96.07 % accuracy rate with the proposed method. Ma et al. [28] have proposed a method for this problem using a subtractive clustering based attribute weighting (SCBAW) and an extreme learning machine. High accuracy rates were obtained with the proposed method.

Some brief information about the studies in the literature on BUPA liver disorder data set for the detection of hepatic impairment is as follows. Pham et al. [29] have achieved 55.90 % classification accuracy by using the RULES-4 algorithm. Van Gestel et al. [30] have achieved 69.20 % accuracy rate with SVM algorithm. Goncalves et al. [31] have proposed a novel neuro-fuzzy model called as inverted hierarchical neuro-fuzzy BSP system (HNFB). The accuracy rate of 73.33 % was obtained for liver disorder data set. Polat et al. [32] have proposed AIRS with performance evaluation by fuzzy resource allocation mechanism for this problem. With the proposed method, they have achieved 83.36 % classification accuracy. Jin et al. [33] have developed a SVM with genetic-fuzzy feature transformation and achieved classification accuracy of 70.80 % with the proposed method. Ozsen and Gunes [34] have applied GA-AWAIS hybrid method and achieved 85.21 % classification accuracy. Lee and Mangarissan [35] have proposed two methods to classify this problem. These

methods are smooth SVM (SSVM) classifier and reduced SVM (RSVM) classifier. With these methods, they obtained 70.33 and 74.86 % classification accuracy, respectively. Li et al. [22] have proposed a non-linear fuzzy based conversion method with SVM. Classification accuracy of this method is 70.85 %. Chen et al. [36] have developed a method which uses 1-NN method and PSO algorithm together. The classification accuracy of this method is 68.99 % for this problem. Dehuri et al. [37] have proposed an improved PSO based evolutionary functional link ANN (ISO-FLANN) model for this problem. With the proposed model, they obtained 76.8 % classification accuracy. Shaoa and Deng [38] have proposed a coordinate descent margin based-twin SVM for classification. With the proposed model, they obtained 73.67 % accuracy rate. Savitha et al. [39] has proposed fully complex-valued RBF (FC-RBF) classifier for classification problems. They obtained 74.6 % classification accuracy with the proposed method. In order to classify noisy data, Mantas and Abellán [14] have proposed a decision tree algorithm which depends on imprecise probability. They applied the algorithm called as Credal-C4.5 on different data sets. Researchers have obtained classification accuracy of 64.53 % for BUPA liver disorder data set. López et al. [40] have carried out the training of SVM algorithm with multivariate normalization. With the proposed method, they obtained 72.17 % classification accuracy.

In this study a new data pre-processing method entitled k-medoids clustering-based attribute weighting has been proposed. Gunes et al.'s [41] work was an inspiration when this method was being developed. In their study, a k-means algorithm was preferred as the weighting method. K-means method is an effective method which also has some disadvantages. The major disadvantage is the sensitivity towards the objects referred to as outliers in the clustering phase [42]. An object with a huge value can significantly change the center point and the average of the cluster in which that subject is included. This change may disrupt the sensitivity of the cluster. To resolve this issue, instead of taking the average of the objects in the cluster, the closest object to the center point - called the medoid - can be used. This operation is performed by using the k-medoids method. In terms of this aspect, the proposed feature weighting method is expected to be effective. The purposes of kmAW are as follows: (i) to convert a non-linear separable dataset to a linear separable dataset and (ii) to gather similar or close data points. As a result of numerous trials using different algorithms, the SVM algorithm was preferred because it offers better performance with regard to kmAW.

The rest of this paper is organized as follows. In the "Methods" section, information is given about the methods used in this study. In the "Experimental design" section, information is presented about the datasets used and the experimental setup. In the "Experimental results and discussions" section, the experimental results and the discussion are

presented. Comparative analyses of the results obtained using the proposed method with the studies in the literature are also given in this section. Results and further targeted studies are shared in the "Conclusions" section.

Methods

Data preprocessing

Preprocessing methods are applied to input data in order classification algorithm to produce more effective results and to reduce the calculation load of algorithms used. Thanks to data preprocessing techniques, a data set with linearly non-separable distribution is converted into a data set with linearly separable data set [11]. In the literature, there have been several studies carried out on data preprocessing or transformation. Polat and Gunes [43] have proposed kNN based attribute weighting method to reduce changes in features in the data set and applied it to medical data sets. Tahir et al. [44] have proposed a hybrid method for feature weighting by utilizing kNN and tabu search algorithms. Sun [45] has proposed a novel feature weighting method based on RELIEF algorithm. Effective results were obtained with the study called as iterative RELIEF (I-RELIEF). Polat et al. [46] have proposed a new feature weighting method based on similarity measure between attributes and have implemented this method in classification of the Doppler signals to identify Atherosclerosis disease. As a data preprocessing method, Dua et al. [47] have presented an algorithm based on bonded component theory to extract the signs from the image and to reduce the image to appropriate size. Polat and Durduran [48] have proposed a new feature weighting called as feature weighting based on subtractive clustering to detect traffic accidents. Unal et al. [49] have presented pairwise fuzzy c-means based attribute weighting for improved classification.

When the literature is reviewed, it is seen that clustering methods are used for weighting method [43, 48, 49]. The most commonly used clustering methods are as follows, respectively: k-means clustering [50], k-medoids clustering, mountain clustering [51], subtractive clustering [52] and fuzzy c-means clustering [53]. In this study, feature weighting process is performed by using k-medoids clustering method which is one of effective clustering methods with low computational load.

K-medoids clustering

The k-medoids clustering algorithm has been proposed to remove the noise and extreme sensitivity of k-means algorithm to the exceptional data. The foundation of k-medoids algorithm is based on finding k representing objects representing various structural features of the data [54]. Representative object is the most central object of the cluster that minimizes

the average distance to other objects. Therefore, the division method is based on the logic of minimization of total of the uniqueness between each object and its reference point. The representative objects are mostly called as medoids in clustering literature [55]. Process steps of the k-medoids clustering algorithm are as follows.

Step 1. Determination of the k cluster number.
Step 2. The selection of k objects as initial medoids.
Step 3. Assigning the remaining objects to the nearest cluster with medoid x.
Step 4. Calculating the objective function. (Error squares criteria: the sum of the distances of whole objects for nearest medoids)
Step 5. Random selection of non medoid y point.
Step 6. Replacement of x and y point, If the replacement of x and y would minimize the objective function.
Step 7. Processes between Step 3 and Step 6 are repeated until there is no change. Objective function specified in step 4 is calculated using Eq. (1).

$$Cost(N, Y) = \sum_{i=1}^{x} \min_{j=1}^{y} \left(d\left(n_j, p_i\right) \right) \tag{1}$$

N denotes the set of medoids, Y denotes data set, x represents the number of patterns, y denotes the number of sets, n_j is j th medoid, p_i is i th pattern and d is a distance function.

K-medoids clustering based attribute weighting (kmAW)

Attributes weighting method is based on the principle of reducing the change in features that form data set. Similar data in the same feature is collected and differentiation ability of the classifier is increased via this weighting method. The name of the proposed feature weighting method in this study is the kmAW. The kmAW works as follows: Initially, cluster medoids are found by k-medoids clustering method. The average values of the features are then calculated according to clusters. Two ratios are obtained at the following stage. The first ratio is *medoid value / mean value*. The other ratio is *mean value / medoid value*. Each data in the dataset is multiplied by one of these ratios. If the data value is larger than the medoid value, then it is multiplied by the ratio with the small value. If the data value is smaller than the medoid value, it is then multiplied by the ratio with the large value. If the data value is equal to the medoid value, then it will be multiplied by 1. Consequently, we ensure that the weighted data will be closer to the medoid value. In Fig. 1, the flow chart of the kmAW method is presented.

In Fig. 2, processing steps of the kmAW method are explained on a simple data set which has m sample and n feature. The explanation is on the weighting of the values belong to f_1 feature. In the figure, $f_1, f_2, f_3, \ldots, f_n$ refers to the features. Features and their values are presented in the field number 1. In the field number 2, process of allocation to the classes is carried out as a result of the implementation of k-medoids algorithm. x_1 and x_2 classes are given as an example. Weighting coefficients ($wc1$ and $wc2$) are calculated with kmAW and obtained values are multiplied by feature values. New values of feature are given in the field number 3. $wc1$ and $wc2$ weighting coefficients are calculated as follows. The calculation of weighting coefficient ($wc1$) according to the x_1 class. The calculation of $wc1$ is based on x_1 class is performed according to Eqs. (2) and (3).
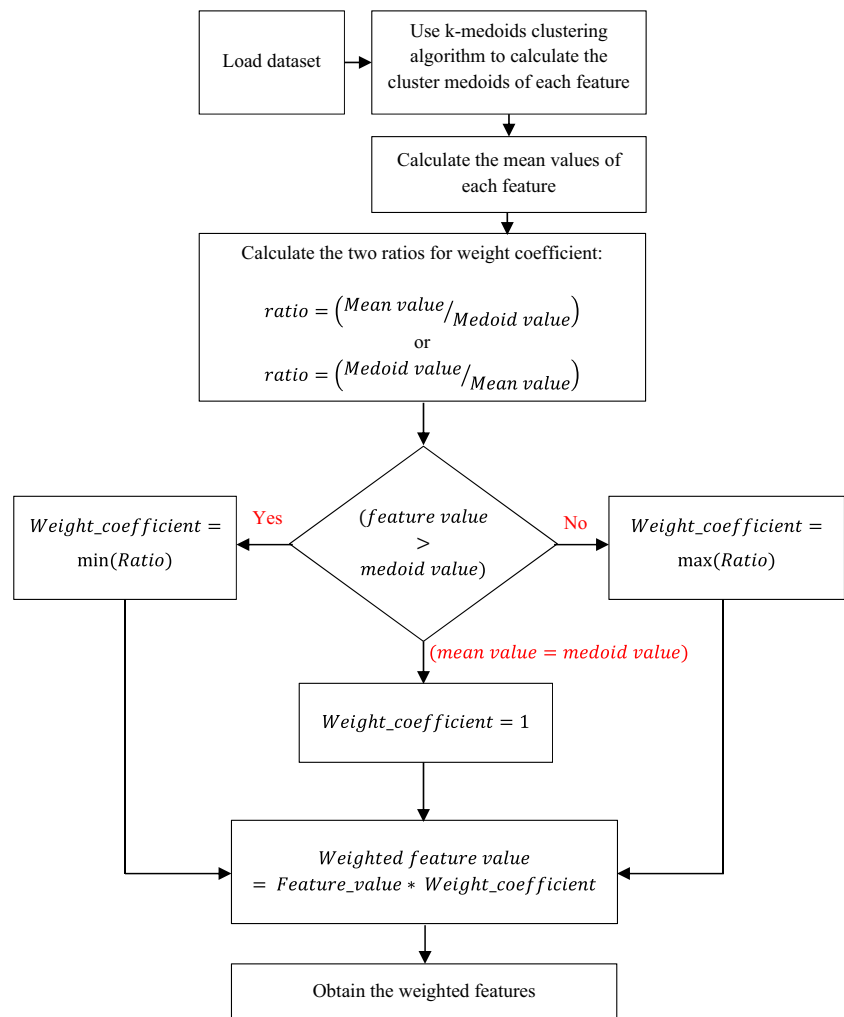
$$k_1 = \frac{\sum_{x=1}^{y} a_x}{y} \tag{2}$$

$$wc1 = \frac{k_1}{l_1} \ OR \ wc1 = \frac{l_1}{k_1} \ OR \ wc1 = 1 \tag{3}$$

where $x = 1, 2, \ldots, a_y$, are the values of f_1 feature. For example, the values of f_1 feature based on x_1 class in Fig. 2/area no 2 are a_1 and a_3. y shows how many values the related feature has in x_1 class, while k_1 is the average feature value. l_1 represents the medoid value of the related feature in x_1 class as a result of the implementation of k-medoids clustering method. $wc1$ is the weighting coefficient. If the feature value is less than the medoid value, then wc1 with the large value is used, but if the feature value is greater than medoid value, then wc1 with the small value is used. If the feature value is equal to the medoid value, then the $wc1 = 1$ equation is used. Calculation of $wc2$ based on x_2 class is performed according to Eqs. (4) and (5).

$$k_2 = \frac{\sum_{a=1}^{b} a_a}{b} \tag{4}$$

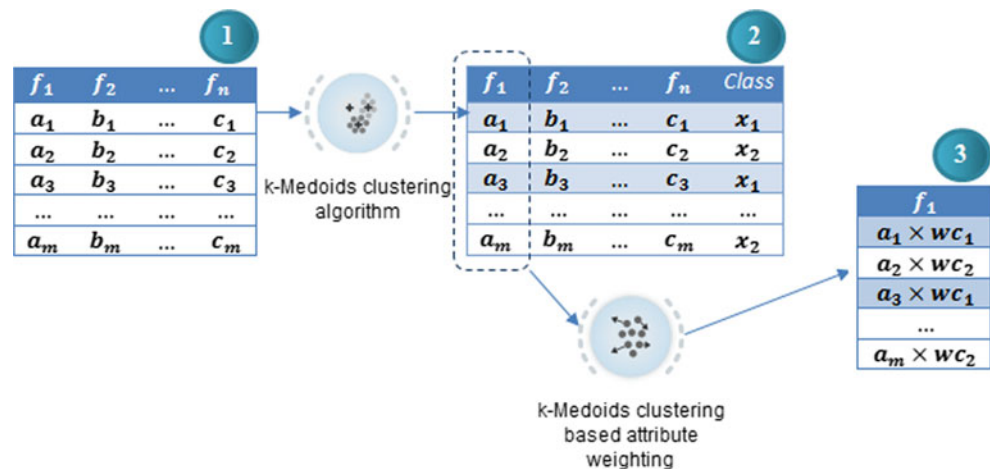$$wc2 = \frac{k_2}{l_2} \ OR \ wc2 = \frac{l_2}{k_2} \ OR \ wc2 = 1 \tag{5}$$

where $a = 1, 2, \ldots, a_b$, are the values of f_2 feature. For example, the values of f_1 feature are based on x_2 class in Fig. 2/area no 2 are a_2 and a_m. b shows how many values the related feature has in x_2 class, k_2 is the average feature value, while l_2 represents the medoid value of the related feature in x_2 class as a result of the implementation of the k-medoids clustering method. $wc2$ is the weighting coefficient. If the feature value is less than the medoid value, then $wc2$ with the large value is used, while if the feature value is greater than the medoid value, then $wc2$ with the small value is used. If the feature value is equal to the medoid value, then the $wc2 = 1$ equation is used.

Fig. 1 The flow chart of the kmAW

Load dataset → Use k-medoids clustering algorithm to calculate the cluster medoids of each feature

Calculate the mean values of each feature

Calculate the two ratios for weight coefficient:

$$ratio = \left(\frac{Mean\ value}{Medoid\ value}\right)$$

or

$$ratio = \left(\frac{Medoid\ value}{Mean\ value}\right)$$

$(feature\ value > medoid\ value)$

Yes → $Weight\_coefficient = min(Ratio)$

No → $Weight\_coefficient = max(Ratio)$

$(mean\ value = medoid\ value)$

$Weight\_coefficient = 1$

$Weighted\ feature\ value = Feature\_value * Weight\_coefficient$

Obtain the weighted features

Support vector machine (SVM)

SVM was developed first by Vapnik [56] for regression and classification studies. This algorithm is an efficient classification algorithm based on statistical learning theory.

Mathematical algorithms of SVM were originally designed for classification problem of two-class linear data, and then generalized for classification of multi-class non-linear data. The working principle of SVM is based on the fact that estimating the optimal decision function that can separate the two

Fig. 2 Weighting of values of a feature (f_1) with kmAW

classes from each other, in other words, identifying of the hyperplane that can most properly separate the two classes from each other [56, 57].

In the classification with SVM, separating of the samples of two classes which are generally shown as $\{-1, +1\}$ with class labels is aimed with the help of a decision function obtained by training data. Hyperplane which can separate training data most properly is determined using said decision function. As shown in Fig. 3a, many hyperplane that can separate two-class data from each other can be plotted. However, the aim of SVM is to find hyperplane that maximizes the distance between the nearest points to it. As shown in Fig. 3b, hyperplane which makes optimum differentiation with maximizing the limit is called as optimum hyperplane. The points limiting the width of the limit are referred as the support vectors.

In a two-class linearly separable classification problem, if training data consisting of k numbers of samples for SVM training is assumed as $\{x_i, y_i\}$, $i = 1, 2, \ldots, k$ then inequalities of optimum hyperplane are as follows:

$$\text{For } y_i = +1, w.x_i + b \geq +1$$
$$\text{For } y_i = -1, w.x_i + b \leq -1 \tag{6}$$

$x \epsilon R^N$ and indicates N-dimensional space, $y \epsilon \{-1, +1\}$ indicates class labels, w indicates weight vector (normal to hyperplane) and b indicates trend value. In order to determine the optimal hyperplane, it is necessary to determine two hyperplanes parallel to this plane (Fig. 3b). Points which form these hyperplanes called as support vectors and these planes are denoted as $w.x_i + b = \pm 1$

To maximize the limit of the optimum hyperplane, w should be minimized. In the case, determining of optimal hyperplane requires the solution of the following constrained optimization problem.

$$min \left[\frac{1}{2} w^2 \right] \tag{7}$$

Accordingly, the restrictions are denoted as follows;

$$y_i(w.x_i + b) \geq +1 \text{ and } y \epsilon \{-1, +1\} \tag{8}$$

This optimization problem can be solved using Lagrange equations. After this operation, Eq. (9) is obtained.
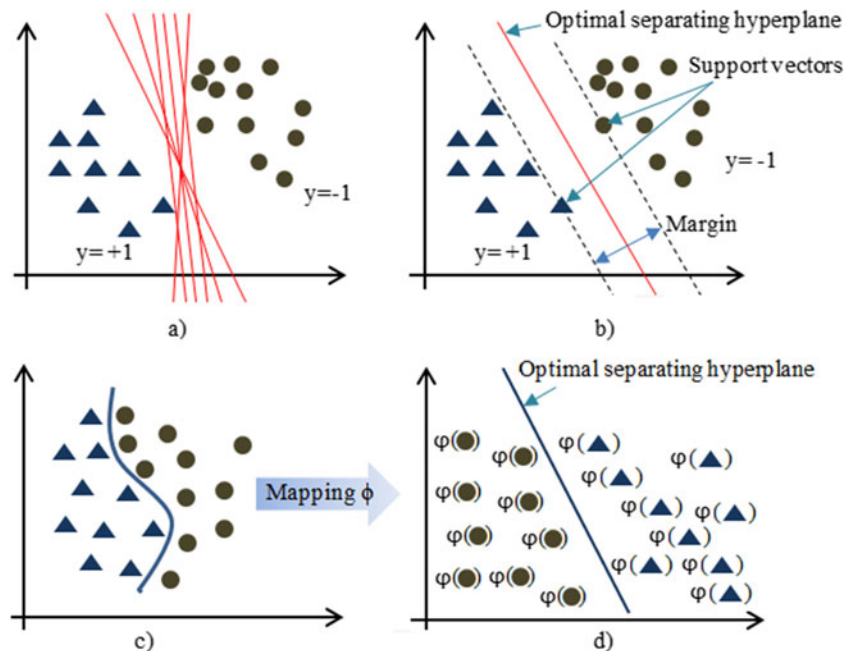
$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{k} \alpha_i y_i (w.x_i + b) + \sum_{i=1}^{k} \alpha_i \tag{9}$$

As a result, decision function for a linearly separable two-class problem can be calculated by using Eq. (10).

$$f(x) = sign \left(\sum_{i=1}^{k} \lambda_i y_i (x.x_i) + b \right) \tag{10}$$

In many problems, linearly separation of the data is not possible (Fig. 3c). In this case, the problem caused by being some of training data on the other side of hyperplane is solved by defining a positive artificial variable (ξ_i). The balance between maximizing the limit and minimizing the misclassification errors can be controlled by defining parameter ($0 < C < \infty$) to a plane (denoted with C) which has positive values [59]. Optimization problem for data that cannot be separated linearly using adjustment parameter and artificial variable:

Fig. 3 Geometric illustration of SVM [58]. a Hyperplanes for two-class problems (b) Optimal separating hyperplane and support vectors (c) Data that cannot be separated linearly (d) Determination of the hyper plane for data that cannot be separated linearly

$$min\left[\frac{\|w\|^2}{2} + C.\sum_{i=1}^{r}\xi_i\right] \qquad (11)$$

Accordingly, the limitations are expressed by Eq. (12);

$$y_i(w.\varphi(x_i) + b) - 1 \geq 1 - \xi_i$$
$$\xi_i \geq 0 \text{ and } i = 1, 2, \ldots, N \qquad (12)$$

As can be seen on Fig. 3d, for the solution of the optimization problem expressed in Eqs. (11) and (12), linearly inseparable data which is in the input space is displayed in a high dimensional space defined as feature space. Thus, data can be separated linearly and interclass hyperplane can be identified.

Nonlinear transformations can be done in SVM with the help of a kernel function mathematically expressed as $K(x_i, x_j) = \varphi(x_i).\varphi(x_j)$, in this way, linear separation of high dimension data is enabled. As a result, decision rule related to the solution of a two class problem which cannot be separated linearly using kernel function can be written as follows:

$$f(x) = sign\left(\sum_i \alpha_i y_i \varphi(x_i).\varphi(x_j) + b\right) \qquad (13)$$

Kernel function and determination of optimum parameters of kernel function is essential for a classification problem to be carried out with SVM. In this study, radial basis kernel function (RBF) is used as kernel function. RBF kernel function can be defined as Eq. (14)

$$K(x_i, x_j) = exp\left(-\gamma\|x_i - x_j\|^2\right), \ \gamma > 0 \qquad (14)$$

where γ is the kernel parameter.

The proposed kmAW + SVM hybrid method

Figure 4 gives the flow chart of the proposed method. In the first phase, data pre-processing step is performed. In this phase, attribute weighting was performed with the kmAW method. The obtained features were presented as input to SVM algorithm. Detailed information about these algorithms is presented in previous sections.

Experimental design

Data description

In this study, experiments were performed on three different data sets to determine the effectiveness of the proposed method. These data sets are: heart disease, PD and BUPA liver disorders. Brief information about these data sets is presented below.
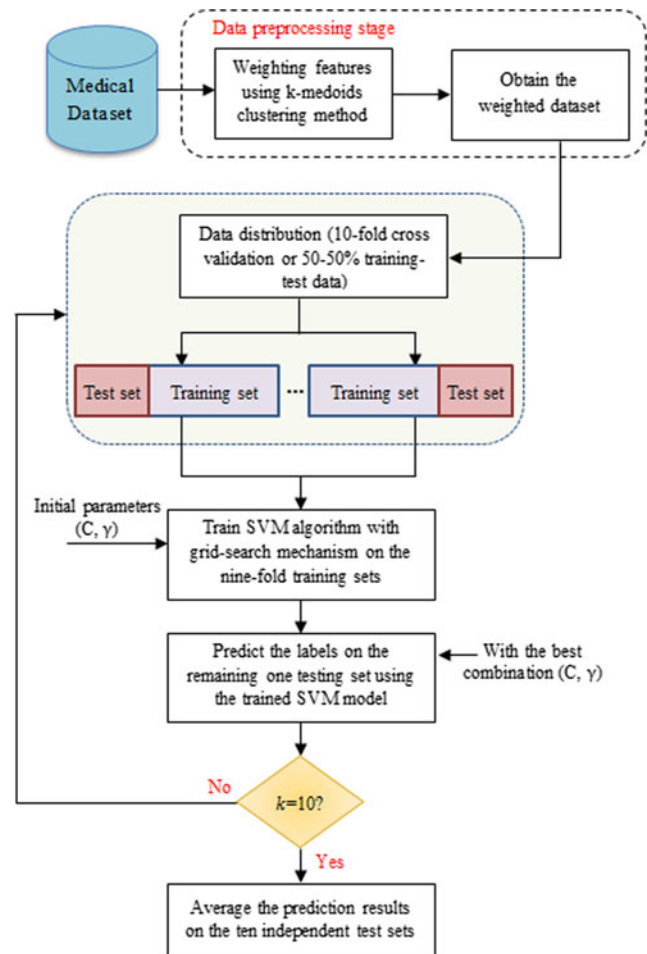

Fig. 4 Flow chart of the proposed medical diagnosis system according to 10-fold cross validation

Statlog heart disease data set

Statlog heart disease data set consists of a total of 270 data collected from patients with heart disease and healthy people [4]. 150 of these data belong to patients, while the remaining 120 belong to healthy individuals. Each data consists of 13 features listed in Table 1. Class information is recorded as 1 (no disease) and 2 (existence of disease) as the 14th feature of the data sequences.

Parkinson's disease dataset

PD data set is comprised of 195 biomedical sound measurements received from 8 healthy people and 23 Parkinson patients [4]. Properties of the PD data set are as follows; average, minimum and maximum sound fundamental frequency, disorder measured in fundamental frequency (Jitter (%), Jitter (absolute), Jitter: RAP, Jitter: PPQ and Jitter: DDP), amplitude irregularity measurements (Shimmer, Shimmer: APQ, Shimmer:APQ3, Shimmer:APQ5, Shimmer: DDA, Shimmer(dB)), the measurement of ratio between tone

Table 1 The features of the Statlog heart disease dataset

no	Feature	no	Feature
1	Age	8	Exercise induced angina
2	Sex	9	Maximum heart rate achieved
3	Chest pain type (four values)	10	Number of major vessels (0–3) colored by fluoroscopy
4	Serum cholesterol in mg/dl	11	The slope of the peak exercise ST segment
5	Resting blood pressure	12	Old peak = ST depression induced by exercise relative to rest
6	Resting electrocardiographic results (values 0, 1 and 2)	13	Thal: 3 = normal; 6 = fixed defect and 7 = reversible defect
7	Fasting blood sugar >120 mg/dl		

components in the audio and noise (HNR and NHR), two nonlinear dynamic complexity measurements (RPDE and D2), three measure of fundamental frequency variation (PPE, Spread1 and Spread2) and signal fractal scaling exponent (DFA). Table 2 shows the attributes of PD dataset [60, 61].

BUPA liver disorders data set

BUPA liver disorders dataset prepared by BUPA Medical Research Company contains 6 features and 345 samples consist of two classes [4]. Data were obtained from the patients with hepatic impairment and healthy subjects. 200 of this data were taken from healthy people with no hepatic impairment.

The remaining 145 data were taken from the patients with hepatic impairment. Each data consists of 6 properties. First 5 features of collected data samples are the blood test results and the last feature includes daily alcohol consumption. Table 3 shows statistical measurements of BUPA liver disorders data set.

Experimental setup

In all the experiments, the selection of training and test data was performed by the 10-fold cross-validation (CV) method and 50–50 % hold out methods. The reason for using two different data selection methods is for comparisons done by the studies presented in the literature to be fairer. Because, in the literature, in some studies 10-fold CV has been utilized, while in other studies 50–50 % hold out method has been used. For the determination of the stability and reliability of results, experiments were repeated 10 times. And the averages of obtained values were calculated.

In this study, determination of parameter of SVM algorithm is made as follows. RBF kernel function which often preferred on SVM applications was preferred as kernel function. Parameter values of this function which gave good results were found by using 10-fold CV on training data with grid search mechanism. Grid search mechanism is one the most commonly used methods for determining the regularization

Table 2 The features of the PD dataset

Feature	Description
MDVP: Fo (Hz)	Mean vocal fundamental frequency
MDVP: Flo (Hz)	Minimum vocal fundamental frequency
MDVP: Fhi (Hz)	Maximum vocal fundamental frequency
Shimmer: APQ 3 Shimmer: APQ 5 MDVP: Shimmer MDVP: Shimmer (dB) Shimmer: DDA MDVP: APQ	Several measures of variation in amplitude
MDVP: RAP MDVP: PPQ MDVP: Jitter (%) MDVP: Jitter (Abs) Jitter: DDP	Several measures of variation in fundamental frequency
NHR HNR	Two measures of ratio of noise to tonal components in the voice
DFA	Signal fractal scaling exponent
RPDE D2	Two non-linear dynamical complexity measures
Spread 1 Spread 2 PPE	Three non-linear measures of fundamental frequency variation

Table 3 The features of the BUPA liver disorder dataset

no	Features
1	MCV (mean corpuscular volume)
2	Alkphos (alkaline phosphatase)
3	SGPT (alanine aminotransferase)
4	Gamma GT (gamma-glutamyl transpeptidase)
5	SGOT (aspartate aminotransferase)
6	Drinks (number of half-pint equivalents of alcoholic beverages drunk per day)

parameter C and kernel parameter γ values [62]. Determination of effective parameter values by using 10-fold CV with grid search is preferred for the following reasons. First, cross-validation process may prevent overfitting problem. Second, required calculation time is not very much for determining the effective parameter values compared to other methods. Moreover, the grid-search can be readily parallelized because each (C, γ) is independent. In the grid search, the regularization parameter C was explored on $C = 2^{-10}, 2^{-4}, \ldots, 2^{10}$. The kernel parameter γ was explored on $\gamma = 2^{-10}, 2^{-9}, \ldots, 2^{5}$. We use LIBSVM software [63] to conduct SVM experiment.

Prediction performance of kmAW + SVM was measured with five evaluation method. These are, respectively; Accuracy, specificity, sensitivity, kappa statistic and f-measure value. Formulas for these parameters are shown in Eqs. (15)–(18).

$$Accuracy\ (CA) = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \qquad (15)$$

$$Specificity = \frac{TN}{FP + TN} \times 100\% \qquad (16)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \qquad (17)$$

$$f-measure = \frac{2 \times Precision x Recall}{Precision + Recall} \qquad (18)$$

where, true positive (TP) indicates the correct number of classification of disease data, False Negative (FN) indicates false number of classification of healthy data. True Negative (TN) indicates the correct number of classification of healthy data, False Positive (FP) indicates false number of classification of disease data. Precision is TP/(TP + FP) and Recall is TP/(TP + FN).

F-measure value is calculated depending on harmonic mean of precision and recall values of the classifier. This value is used as a performance evaluation metric to measure the classifier performance. F-measure takes values between 0 and 1 and it is expected for f-measure to take a value close to 1 in a high performance classification.

Kappa statistics were developed as an alternative to accuracy ratio measure for evaluation of the classifiers [64]. This value is used to calculate the compatibility between the evaluations made by two or more evaluators. Kappa statistic value can be calculated as shown in Eq. (19).

$$KS = \frac{P_0 - P_c}{1 - P_c} \qquad (19)$$

P_0 is the accuracy of the classifier, P_c represents the accuracy value obtained by random estimation on the same data set. Kappa statistic value is in the range of [−1, 1]. −1 represents an unsuccessful classification, 1 represents that a successful classification has been performed.

ROC curve is often used for self-identification of diagnostic test and to enable making a reliable comparison between tests [65]. In the coordinate system where ROC curve to be created, the actual positive value (sensitivity) of diagnostic test is located in the Y-axis, the false positive value (−1 specificity) is located in the X axis. ROC curve is plotted by combining points corresponding true positive and false positive at each intersection point. ROC curves show all the possible intersection points and allow for predictions about the frequency of different results (TP, TN, FP and FN) at each intersection point. The area under the ROC for a diagnostic test can take values between 0.50 and 1.00 depending on the activity level. When this area is greater the diagnostic test will have more differentiation ability [66].

Experimental results and discussions

The results obtained by applying the proposed method to 3 different datasets are presented below. First, attributes in the data set were weighted using the kmAW. Figures 5, 6 and 7 show the distribution of the original and weighted samples (in two classes) created by the best 3 principle components obtained by PCA for each database. As shown in the figures, the differentiation capability of the original dataset was significantly improved using the kmAW method. This is due to the gathering together of similar data after weighting. With the implementation of the kmAW method, it is observed that linearly inseparable datasets can be separated linearly.

The results obtained by applying the SVM algorithm are presented in Tables 4, 5 and 6. The results obtained for the Statlog heart disease dataset are given in Table 4. Accordingly, in terms of the 10-fold CV method, 90.82 % classification accuracy was obtained using the kmAW + SVM method. The accuracy rate obtained by the application of the SVM method to the original dataset is 81.48 %. In the 50–50 % range, an 89.29 % accuracy rate is obtained using the kmAW + SVM method, while 81.86 % classification accuracy is obtained with the original dataset + the SVM method. The kmAW + SVM method gave good results and this is also seen in the kappa statistic values. The highest kappa value obtained using the 10-fold CV and kmAW + SVM method is 0.8227. The effect of the weighted features is positive for the Statlog heart disease dataset.

The results obtained for the Parkinson disease dataset are presented in Table 5. When the table is examined it can be seen that a 98.95 % classification accuracy was obtained using the kmAW + SVM according to the 10-fold CV method. An 84.25 % classification accuracy was achieved with the use of the original dataset and the SVM method. In the 50–50 % range, the classification accuracy
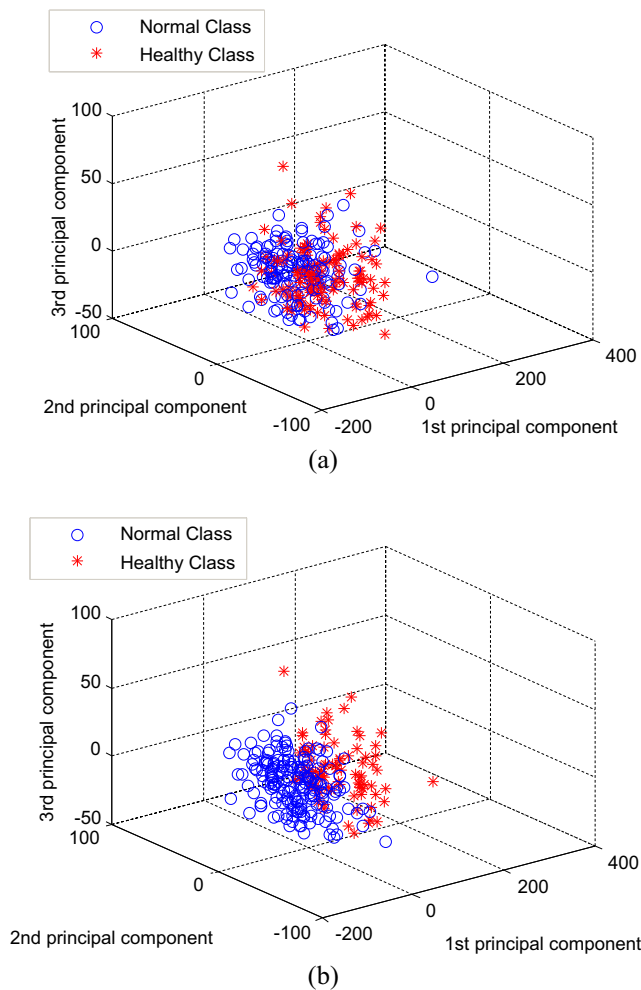
Fig. 5 Three-dimensional distribution (in two classes) of the original samples created by the best 3 principle components obtained after implementation of principal component analysis for the Statlog heart disease dataset, (**a**) for original features (**b**) for weighted features

Fig. 6 Three-dimensional distribution (in two classes) of the original samples created by the best 3 principle components obtained after implementation of principal component analysis for the PD dataset, a) for original features b) for weighted features

is 97.98 % using the kmAW + SVM method, and 81.75 % with the original dataset + SVM method. It can also be seen in the Kappa statistic values that the kmAW + SVM method gives good results. The highest kappa statistic value of 0.9735 was obtained using the 10-fold CV and kmAW + SVM method. The Kappa value was found to be 0.5870 with the original dataset + SVM method. Significant differences between the two methods is also observed with this aspect. Weighted features gave effective results for the PD dataset.

The results obtained for the BUPA liver disorder dataset are presented in Table 6. When the table is examined, according to the 10-fold CV method, it can be seen that 86.25 % classification accuracy was achieved using the kmAW + SVM method, and a 68.85 % classification accuracy was obtained with the application of the SVM algorithm to the original dataset. In the 50–50 % range, classification accuracy is 85.32 % using the kmAW + SVM
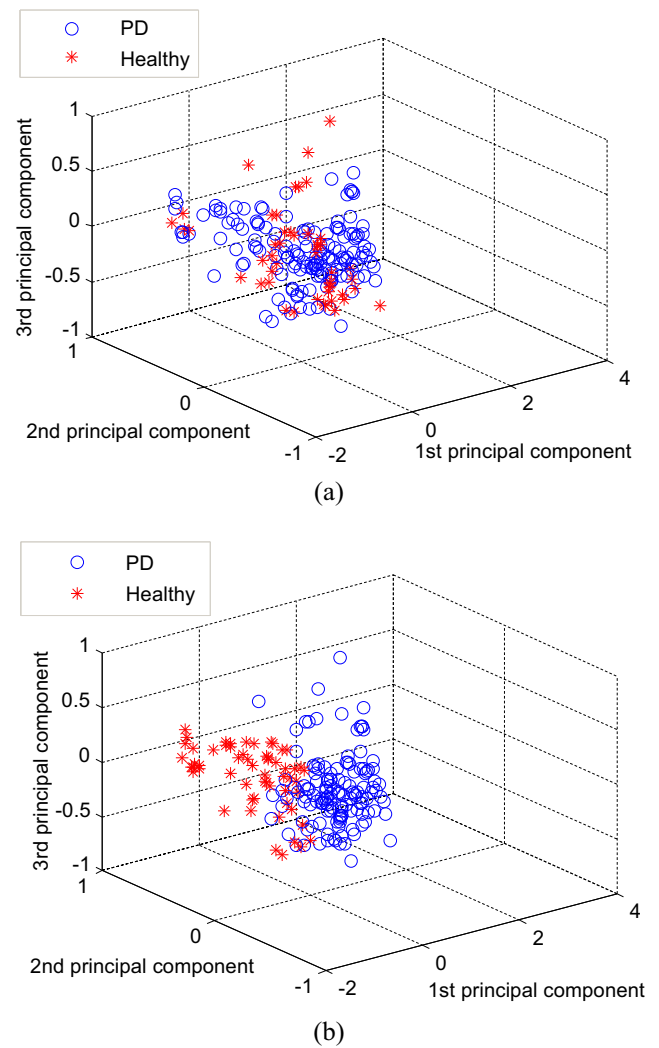
method, and 66.75 % with the original dataset + SVM method. In performance measurement metrics other than classification accuracy, it can also be seen that good results are obtained using the kmAW + SVM method. When kappa statistic values are analyzed, it is seen that the highest values are obtained with the kmAW + SVM method. Weighted features and the hybrid application of the SVM algorithm gave effective results for the BUPA liver disorder dataset.

When Tables 4, 5 and 6 are analyzed in general, it can be observed that the kmAW + SVM method gave better results than the original dataset + SVM method. The highest accuracy rate was obtained by implementing the 10-fold CV data distribution method.

ROC curves were also used for performance evaluation. A comparison of the 10-fold CV with the kmAW + SVM and the
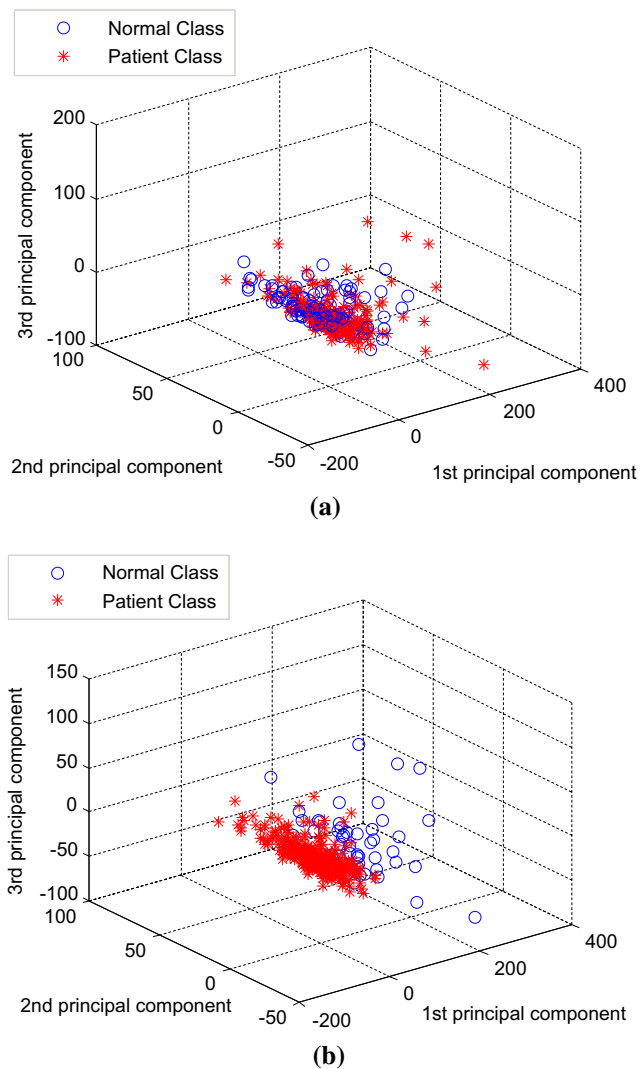
Fig. 7 Three-dimensional distribution (in two classes) of the original samples created by the best 3 principle components obtained after implementation of principal component analysis for the BUPA liver disorder dataset, (**a**) for original features (**b**) for weighted features

original dataset + SVM is presented using these curves. The ROC graph obtained for the heart disease dataset is presented in Fig. 8a. As seen on the ROC graph, there is a significant difference between the areas calculated for the two methods (AUC = 0.9195 for kmAW + SVM, AUC = 0.8125 for without feature weighting + SVM).

The ROC graph obtained for the PD dataset is presented in Fig. 8b. Accordingly, it is observed that there is a significant difference between the areas under the ROC curve (AUC). (AUC = 0.9808 for kmAW + SVM, AUC = 0.7955 for without feature weighting + SVM). It can be observed that the kmAW + SVM method gave better results.

The ROC graph obtained for the BUPA liver disorder dataset is given in Fig. 8c. A significant difference is observed between the two methods (AUC = 0.8685 for kmAW + SVM, AUC = 0.6832 for without feature weighting + SVM).

Table 4 The results obtained based on the performance evaluation criteria for the Statlog heart disease dataset

Features	Performance Metrics	10-fold CV	50–50 % training–testing
All original features	ACC	81.48 ± 5.99	81.86 ± 6.25
	Sensitivity	84.72 ± 6.38	84.67 ± 5.44
	Specificity	77.78 ± 7.05	74.44 ± 8.35
	f-measure	0.8291	0.8083
	Kappa	0.6269	0.5919
	AUC	0.8125	0.7960
After feature weighting	ACC	90.82 ± 3.25	89.29 ± 3.85
	Sensitivity	93.95 ± 2.98	90.54 ± 3.05
	Specificity	90.45 ± 4.42	87.88 ± 5.12
	f-measure	0.9268	0.8993
	Kappa	0.8227	0.7848
	AUC	0.9195	0.8921

Comparative analysis of the proposed method with the methods found in the literature is presented in Tables 7, 8 and 9. Comparative analysis for the Heart disease dataset is presented in Table 7. When Table 7 is examined, it can be seen that accuracy rates generally ranging between 80 and 88 % were obtained by other researchers. A 90.82 % classification accuracy was achieved with the proposed method for the same dataset.

Comparative analysis carried out in terms of previous studies for the PD dataset is presented in Table 8. As shown in the table, accuracy rates generally ranging between 83 and 98 % have been obtained by other researchers. With 98.95 % accuracy values, the proposed method gives better results compared to other studies.

Table 5 The results obtained based on the performance evaluation criteria for the PD dataset

Features	Performance Metrics	10-fold CV	50–50 % training–testing
All original features	ACC	84.25 ± 5.38	81.75 ± 5.75
	Sensitivity	67.05 ± 7.85	62.20 ± 8.15
	Specificity	91.05 ± 2.95	89.95 ± 3.76
	f-measure	0.6955	0.6475
	Kappa	0.5870	0.5230
	AUC	0.7955	0.7625
After feature weighting	ACC	98.95 ± 1.85	97.98 ± 2.35
	Sensitivity	96.12 ± 3.56	94.25 ± 3.67
	Specificity	100 ± 0	99.42 ± 1.15
	f-measure	0.9795	0.9599
	Kappa	0.9735	94.641
	AUC	0.9808	96.845

Table 6   The results obtained based on the performance evaluation criteria for the BUPA liver disorder dataset

Features	Performance Metrics	10-fold CV	50–50 % training–testing
All original features	ACC	68.85 ± 8.90	66.75 ± 9.10
	Sensitivity	62.75 ± 9.85	60.80 ± 9.98
	Specificity	73.88 ± 5.15	71.15 ± 6.23
	f-measure	0.6301	0.6021
	Kappa	0.3588	0.3153
	AUC	0.6832	0.6598
After feature weighting	ACC	86.25 ± 4.25	85.32 ± 5.35
	Sensitivity	82.72 ± 5.88	82.50 ± 6.59
	Specificity	88.90 ± 3.52	88.36 ± 3.66
	f-measure	0.8501	0.8327
	Kappa	0.7393	70.94
	AUC	0.8685	0.8545

A comparative analysis for the BUPA liver disease dataset is presented in Table 9. When the table is analyzed, it can be seen that an 86.25 % classification accuracy was achieved in this study, while accuracy values obtained by other researchers generally range between 60 and 86 %.

When we evaluate the situation in general, it can be observed that the developed method gave better results compared to the methods proposed in previous studies.

There are numerous algorithms available in the literature that are presented as feature weighting and classification algorithms. In terms of the 3 different datasets, a large number of experiments were performed with

different methods to achieve the best results. As a result of these experiments, it can be seen that the kmAW+ SVM hybrid method is more effective than the other methods. In the experiments, FCMCBAW, SCBAW and k-means clustering-based attribute weighting (KMCBAW) algorithms were used as the attribute weighting method. SVM, ANN, Random Forest, C4.5 Decision Tree and Naive Bayes algorithms which are often preferred in the literature as classification algorithms are selected. A comparative analysis based on hybrid methods obtained by different combinations is presented in Table 10. The results show that feature weighting methods give effective results.

When the results were analyzed, it can be seen that KMCBAW, FCMCBAW, SCBAW and kmAW feature weighting methods give promising results in terms of the classification of non-linear medical datasets. In general, it can be seen that SVM algorithms give better results with weighting methods. Similar results were obtained with regard to the kmAW and SCBAW methods. The nearest accuracy values to SVM were obtained using the ANN method.

Conclusions

This study proposes a hybrid system aimed at improving the classification accuracy of computer-aided medical diagnostic systems. Experiments were performed on data related to heart disease, PD and liver disorders. The main
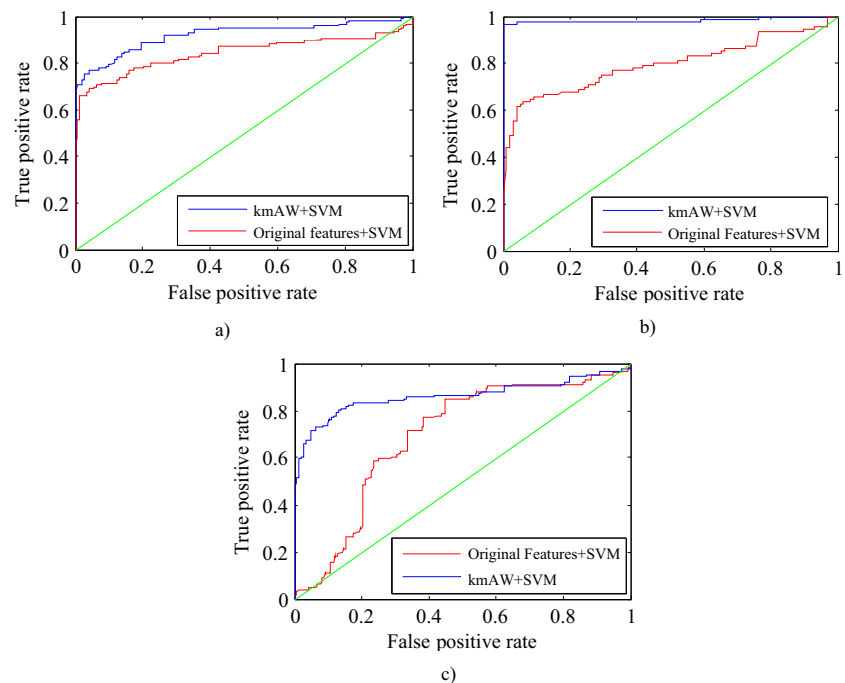

Fig. 8 ROC graphs (**a**) for the Statlog heart disease dataset (**b**) for the PD dataset c) for the BUPA liver disorder dataset

Table 7   Performance comparison of various methods in terms of accuracy (%) for the Statlog heart disease dataset

Authors	Method	Classification accuracy (%)
Duch et al. [5]	k-NN, k = 28, 7 features (10-fold CV)	84.6–85.6
	k-NN, k = 28, Manhattan (10-fold CV)	82.2–83.4
	FSM, 27 fuzzy rules	82
	SSV, 3 rules	80.2–83.4
Sahan et al. [6]	AWAIS (10-fold CV)	82.59
Ozsen and Gunes [9]	AIS algorithm with hybrid similarity measure (10-fold CV)	83.95
Kahramanli and Allahverdi [10]	Hybrid system using ANN and FNN (10-fold CV)	86.8
Polat and Gunes [11]	A hybrid of LS-SVM classifier and kernel f-score feature selection (50–50 % training–testing)	83.70
Subbulakshmi et al. [13]	Extreme learning machine (70–30 % training–testing)	87.50
Mantas and Abellán [14]	Decision tree based on imprecise probabilities (Credal C4.5) (10 fold CV)	80.33
Shao and Deng [38]	Coordinate descent margin based-twin SVM (10 fold CV)	84.44 ± 6.80
Ozsen et al. [67]	Kernel based AIS (5-fold CV)	85.93
Tian et al. [68]	Cooperative coevolutionary algorithm - elliptical basis function neural network (50–25–25 % training-validation-testing)	82.45
Torun and Tohumoglu [69]	Simulated annealing and fuzzy classifier (10 fold CV)	81.11 ± 5.91
Al-Obeidat et al. [70]	Particle swarm optimization for PROAFT (10 fold CV)	84.27
Jaganathan and Kuppuchamy [71]	Neural network threshold selection (10 fold CV)	85.19
Lim and Chan [72]	Bandler kohout-interval-valued fuzzy sets (BK-IVFS weighted) (5 fold CV)	85.56
Yang et al. [73]	Fuzzy class – label SVM (y_i - SVM) and Fuzzy SVM (F-SVM)	85.19
Ahmad et al. [74]	Improved hybrid genetic algorithm-multilayer perceptron network (75–25 % training–testing)	86.30
Ibrikci et al. [75]	Combined Kernels with Support Vector Machine (65–35 % training–testing)	88.89
Proposed Method	kmAW + SVM (50–50 % training–testing)	89.29
Proposed Method	kmAW + SVM (10 fold CV)	90.82

Table 8   Performance comparison of various methods in terms of accuracy (%) for the PD dataset

Authors	Method	Classification accuracy (%)
Shahbaba and Neal [15]	Dirichlet process mixtures (5 fold CV)	87.70
Das [16]	Variable selection + ANN (65–35 % training – testing)	92.90
Guo et al. [17]	GP-EM (10-fold CV)	93.10
Sakar and Kursun [18]	Mutual information + Support vector machine (booststrap with 50 replicates)	92.75
Ozcift and Gulten [19]	CFS-RF (10-fold CV)	87.10
Aström and Koker [20]	Parallel ANN	91.20
Spadoto et al. [21]	OPF gravitational search + PSO + OPF harmony search + OPF (20–30–50 % training-validation-testing)	84.01
Luukka [21]	Fuzzy entropy measures + Similarity classifier	85.03
Li et al. [22]	SVM + Fuzzy-based non-linear transformation	93.47
Polat [24]	kNN + FCMFW (50–50 % training–testing)	97.93
Daliri [25]	A chi-square distance kernel-based SVM (50–50 % training - testing)	91.20
Zuo et al. [26]	PSO-FKNN (10-fold CV)	97.47
Chen et al. [27]	PCA-FKNN (10-fold CV)	96.07
Psorakis et al. [76]	Multiclass multi-kernel relevance vector machines (10-fold CV)	89.47
Proposed Method	kmAW + SVM (50–50 % training - testing)	97.98
Proposed Method	kmAW + SVM (10 fold CV)	98.95

Table 9 Performance comparison of various methods in terms of accuracy (%) for the BUPA liver disorder dataset

Authors	Method	Classification accuracy (%)
Ozsen and Gunes [9]	AIS with hybrid similarity measure (10-fold CV)	60.57
Mantas and Abellán [14]	Decision tree based on imprecise probabilities (Credal C4.5)	64.53
Li et al. [22]	A fuzzy-based nonlinear transformation method + SVM	70.85
van Gestel et al. [30]	SVM with GP (10-fold CV)	69.7
Goncalves et al. [31]	Inverted hierarchical neuro-fuzzy binary space partitioning system	73.33
Polat et al. [32]	Fuzzy artificial immune recognition system (10-fold CV)	83.4
Lee and Mangasarian [35]	Reduced SVMs (10-fold CV)	74.9
Dehuri et al. [37]	Improved PSO and functional link artificial neural network (FLANN) (10-fold CV)	76.80
Shao and Deng [38]	Coordinate descent margin based-twin SVM (10-fold CV)	72.80 ± 5.31
Savitha et al. [39]	Fully complex valued RBF (10 fold CV)	74.6
López et al. [40]	Mahalanobis SVM	72.17
Torun and Tohumoglu [69]	Fuzzy classifier and Simulated annealing (10 fold CV)	74.13 ± 12.7
Al-Obeidat et al. [70]	Particle swarm optimization for PROAFT (10 fold CV)	69.31
Yang et al. [73]	Fuzzy class – label SVM (y_i - SVM) and fuzzy SVM (F-SVM)	74.78
Lin and Chang [77]	Case based reasoning + Particle swarm optimization (5 fold CV)	78.18
Wang et al. [78]	Spiking neural networks (SNNs)	56.6 ± 1.8
Ozsen and Yucelbas [79]	Ellipsoidal-AIS (5 fold CV)	85.59 ± 1.32
Proposed Method	kmAW + SVM (50–50 % training–testing)	85.32
Proposed Method	kmAW + SVM (10 fold CV)	86.25

innovation of this study lies in a hybrid system entitled kmAW + SVM which combines an efficient clustering attribute weighting method with a powerful classification algorithm. In this study, the kmAW method was used as a data preprocessing tool in order to improve the diagnosis accuracy of a SVM classifier, and to reduce the variance of the features in datasets. The classification accuracy of the proposed system for the Statlog heart disease dataset, the PD dataset and the BUPA liver disorder dataset reached 90.82, 98.95 and 86.25 %, respectively. Even a slight increase in accuracy rates is very significant in a key subject such as medical diagnosis. Hence, the method proposed here will contribute significantly to this field. Based on the results, the proposed method provides good results compared to the methods proposed in previous studies, and it appears promising for use with regard to medical diagnostic systems. As a result, an effective system which can be used as a computerized decision support system to help physicians in terms of medical diagnoses has been developed.

Table 10 Comparison of classification accuracies using different weighting methods and classification algorithms

Dataset	Hybrid Method (Feature Weighting Method + Classification Algorithm)				
		SVM	ANN	Random Forest	C4.5 Decision Tree	Naive Bayes
Heart Disease	SCBAW	90.25	89.75	83.35	84.12	82.05
	FCMCBAW	89.65	89.05	82.75	82.15	83.44
	KMCBAW	88.12	87.55	81.10	81.55	80.05
	kmAW	90.82	88.56	84.42	85.12	81.15
Parkinson's Disease	SCBAW	97.96	96.50	92.65	94.05	88.80
	FCMCBAW	97.55	95.45	91.35	93.45	90.50
	KMCBAW	96.50	95.05	90.88	92.26	87.75
	kmAW	98.95	96.75	92.46	93.82	89.95
BUPA liver disorder	SCBAW	86.05	85.65	82.25	83.55	80.05
	FCMCBAW	85.55	85.01	80.08	82.25	81.15
	KMCBAW	84.32	83.95	78.65	79.90	80.12
	kmAW	86.25	84.25	81.15	83.65	79.56

References

1. Das, R., Turkoglu, I., and Sengur, A., Diagnosis of valvular heart disease through neural networks ensembles. *Comput. Methods Programs Biomed.* 93(2):185–191, 2009.

2. Peker, M., A new approach for automatic sleep scoring: Combining Taguchi based complex-valued neural network and complex wavelet transform. *Comput. Methods Programs Biomed.* 2016. doi:10.1016/j.cmpb.2016.01.001.

3. Das, R., and Sengur, A., Evaluation of ensemble methods for diagnosing of valvular heart disease. *Expert Syst. Appl.* 37(7):5110–5115, 2010.

4. Bache, K., and Lichman, M., UCI machine learning repository. 2013, Available at http://archive.ics.uci.edu/ml.

5. Duch, W., Adamczak, R., and Grabczewski, K., A new methodology of extraction, optimization and application of crisp and fuzzy logical rules. *IEEE Trans. Neural Network* 12(2):277–306, 2001.

6. Sahan, S., Polat, K., Kodaz, H., and Gunes, S., The medical applications of attribute weighted artificial immune system (AWAIS): Diagnosis of heart and diabetes diseases. *Lect. Notes Comput. Sci.* 3627:456–468, 2005.

7. Polat, K., and Gunes, S., A hybrid approach to medical decision support systems: Combining feature selection, fuzzy weighted pre-processing and AIRS. *Comput. Methods Programs Biomed.* 88(2):164–174, 2007.

8. Polat, K., Sahan, S., and Gunes, S., Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-NN (nearest neighbour) based weighting preprocessing. *Expert Syst. Appl.* 32(2):625–631, 2007.

9. Ozsen, S., and Gunes, S., Effect of feature-type in selecting distance measure for an artificial immune system as a pattern recognizer. *Digit. Signal Process.* 18(4):635–645, 2008.

10. Kahramanli, H., and Allahverdi, N., Design of a hybrid system for the diabetes and heart diseases. *Expert Syst. Appl.* 35(1–2):82–89, 2008.

11. Polat, K., and Gunes, S., A new feature selection method on classification of medical datasets: Kernel F-score feature selection. *Expert Syst. Appl.* 36(7):10367–10373, 2009.

12. Das, R., Turkoglu, I., and Sengur, A., Effective diagnosis of heart disease through neural networks ensembles. *Expert Syst. Appl.* 36(4):7675–7680, 2009.

13. Subbulakshmi, C. V., Deepa, S. N., and Malathi, N., Extreme learning machine for two category data classification. *In 2012 I.E. International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, pp. 458–461, 2012.

14. Mantas, C. J., and Abellán, J., Credal-C4. 5: Decision tree based on imprecise probabilities to classify noisy data. *Expert Syst. Appl* 41(10):4625–4637, 2014.

15. Shahbaba, B., and Neal, R., Nonlinear models using Dirichlet process mixtures. *J. Mach. Learn. Res.* 10:1829–1850, 2009.

16. Das, R., A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Syst. Appl.* 37(2):1568–1572, 2010.

17. Guo, P. F., Bhattacharya, P., and Kharma, N., Advances in detecting Parkinson's disease. *in Medical Biometrics, vol. 6165 of Lect. Notes Comput. Sci*, pp. 306–314, 2010.

18. Sakar, C. O., and Kursun, O., Telediagnosis of Parkinson's disease using measurements of dysphonia. *J. Med. Syst.* 34(4):591–599, 2010.

19. Ozcift, A., and Gulten, A., Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. *Comput. Methods Programs Biomed.* 104(3):443–451, 2011.

20. Astrom, F., and Koker, R., A parallel neural network approach to prediction of Parkinson's disease. *Expert Syst. Appl.* 38(10):12470–12474, 2011.

21. Luukka, P., Feature selection using fuzzy entropy measures with similarity classifier. *Expert Syst. Appl.* 38(4):4600–4607, 2011.

22. Li, D. C., Liu, C. W., and Hu, S. C., A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets. *Artif. Intell. Med.* 52(1):45–52, 2011.

23. Ozcift, A., SVM feature selection based rotation forest ensemble classifiers to improve computer-aided diagnosis of Parkinson disease. *J. Med. Syst.* 36(4):2141–2147, 2012.

24. Polat, K., Classification of Parkinson's disease using feature weighting method on the basis of fuzzy c-means clustering. *Int. J. Syst. Sci.* 43(4):597–609, 2012.

25. Daliri, M. R., Chi-square distance kernel of the gaits for the diagnosis of Parkinson's disease. *Biomed. Signal Process. Contr.* 8(1):66–70, 2013.

26. Zuo, W. L., Wang, Z. Y., Liu, T., and Chen, H. L., Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach. *Biomed. Signal Process. Contr.* 8(4):364–373, 2013.

27. Chen, H. L., Huang, C. C., Yu, X. G., Xu, X., Sun, X., Wang, G., and Wang, S. J., An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. *Expert Syst. Appl.* 40(1):263–271, 2013.

28. Ma, C., Ouyang, J., Chen, H. L., and Zhao, X. H., An efficient diagnosis system for Parkinson's disease using kernel-based extreme learning machine with subtractive clustering features weighting approach. *Comput Math. Methods Med.* 2014. doi:10.1155/2014/985789.

29. Pham, D. T., Dimov, S. S., and Salem, Z., Technique for selecting examples in inductive learning. *In European Symposium on Intelligent Techniques (ESIT 2000)*, pp. 119–127, 2000.

30. Van Gestel, T., Suykens, J. A. K., Lanckriet, G., Lambrechts, A., De Moor, B., and Vandewalle, J., Bayesian framework for least squares support vector machine classifiers, Gaussian processes and kernel fisher discriminant analysis. *Neural. Comput.* 14(5):1115–1147, 2002.

31. Goncalves, L. B., Vellasco, M. B. R., Pacheco, M. A. C., and de Souza, F. J., Inverted hierarchical neuro-fuzzy BSP system: A novel neuro-fuzzy model for pattern classification and rule extraction in databases. *IEEE Trans. Syst. Man Cybern. C Appl. Rev.* 36(2):236–248, 2006.

32. Polat, K., Sahan, S., Kodaz, H., and Gunes, S., Breast cancer and liver disorders classification using artificial immune recognition system (AIRS) with performance evaluation by fuzzy resource allocation mechanism. *Expert Syst. Appl.* 32(1):172–183, 2007.

33. Jin, B., Tang, Y. C., and Zhang, Y. Q., Support vector machines with genetic fuzzy feature transformation for biomedical data classification. *Inform. Sci.* 177(2):476–489, 2007.

34. Ozsen, S., and Gunes, S., Attribute weighting via genetic algorithms for attribute weighted artificial immune system (AWAIS) and its application to heart disease and liver disorders problems. *Expert Syst. Appl.* 36(1):386–392, 2009.

35. Lee, Y. J., and Mangasarian, O. L., SSVM: A smooth support vector machine for classification. *Comput. Optim. Appl.* 20(1):5–22, 2001.

36. Chen, L. F., Su, C. T., Chen, K. H., and Wang, P. C., Particle swarm optimization for feature selection with application in obstructive sleep apnea diagnosis. *Neural Comput. Appl.* 21(8):2087–2096, 2012.

37. Dehuri, S., Roy, R., Cho, S. B., and Ghosh, A., An improved swarm optimized functional link artificial neural network (ISO-FLANN) for classification. *J. Syst. Software* 85(6):1333–1345, 2012.

38. Shao, Y. H., and Deng, N. Y., A coordinate descent margin based-twin support vector machine for classification. *Neural Network* 25:114–121, 2012.

39. Savitha, R., Suresh, S., Sundararajan, N., and Kim, H. J., A fully complex-valued radial basis function classifier for real-valued classification problems. *Neurocomputing* 78(1):104–110, 2012.

40. López, F. M., Puertas, S. M., and Arriaza, J. T., Training of support vector machine with the use of multivariate normalization. *Appl. Soft Comput.* 24:1105–1111, 2014.

41. Gunes, S., Polat, K., and Yosunkaya, S., Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting. *Expert Syst. Appl.* 37(12):7922–7928, 2010.

42. Han, J., Kamber, M., and Pei, J., *Data mining: Concepts and techniques*. Morgan Kaufmann, 2006.

43. Polat, K., and Gunes, S., A hybrid medical decision making system based on principles component analysis, k-NN based weighted preprocessing and adaptive neuro-fuzzy inference system. *Digit. Signal Process.* 16(6):913–921, 2006.

44. Tahir, M. A., Bouridane, A., and Kurugollu, F., Simultaneous feature selection and feature weighting using hybrid tabu search/k-nearest neighbor classifier. *Pattern Recogn. Lett.* 28(4):438–446, 2007.

45. Sun, Y., Iterative RELIEF for feature weighting: Algorithms, theories, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* 29(6):1035–1051, 2007.

46. Polat, K., Latifoglu, F., Kara, S., and Gunes, S., Usage of novel similarity based weighting method to diagnose the Atherosclerosis from carotid artery Doppler signals. *Med. Biol. Eng. Comput.* 46:353–362, 2008.

47. Dua, S., Singh, H., and Thompson, H. W., Associative classification of mammograms using weighted rules. *Expert Syst. Appl.* 36(5):9250–9259, 2009.

48. Polat, K., and Durduran, S. S., Subtractive clustering attribute weighting (SCAW) to discriminate the traffic accidents on Konya–Afyonkarahisar highway in Turkey with the help of GIS: A case study. *Adv. Eng. Software* 42(7):491–500, 2011.

49. Unal, Y., Polat, K., and Kocer, H. E., Pairwise FCM based feature weighting for improved classification of vertebral column disorders. *Comput. Biol. Med.* 46:61–70, 2014.

50. MacQueen, J. B., Some methods for classification and analysis of multivariate observations. *In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability,* pp. 281–297, 1967.

51. Bezdek, J. C., *Pattern recognition with fuzzy objective function algorithms*. Plenum Press, New York, 1981.

52. Yager, R. R., and Filev, D. P., Generation of fuzzy rules by mountain clustering. *J. Intell. Fuzzy Syst.* 24:209–219, 1994.

53. Chiu, S. L., Fuzzy model identification based on cluster estimation. *J. Intell. Fuzzy Syst.* 2:267–278, 1994.

54. Kaufman, L., and Rousseeuw, P., *Clustering by means of medoids*. North-Holland, 1987.

55. Kaufman, L., and Rousseeuw, P. J., *Finding groups in data: An introduction to cluster analysis*. Wiley, Hoboken, NJ, 1990.

56. Vapnik, V. N., *The nature of statistical learning theory*. Springer, NewYork, 1995.

57. Berikol, G. B., Yildiz, O., and Ozcan, I. T., Diagnosis of acute coronary syndrome with a support vector machine. *J. Med. Syst.* 40(4):1–8, 2016.

58. Su, L., Shi, T., Xu, Z., Lu, X., and Liao, G., Defect inspection of flip chip solder bumps using an ultrasonic transducer. *Sensors* 13(12):16281–16291, 2013.

59. Cortes, C., and Vapnik, V., Support vector network. *Mach. Learn.* 20(3):273–297, 1995.

60. Elbaz, A., Bower, J. H., Maraganore, D. M., McDonnell, S. K., Peterson, B. J., Ahlskog, J. E., Schaid, D. J., and Rocca, W. A., Risk tables for Parkinsonism and Parkinson's disease. *J. Clin. Epidemiol.* 55:25–31, 2002.

61. Little, M. A., McSharry, P. E., Hunter, E. J., and Ramig, L. O., Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans. Biomed. Eng.* 56:1015–1022, 2009.

62. Bergstra, J., and Bengio, Y., Random search for hyper-parameter optimization. *The J. Mach. Learn. Res.* 13(1):281–305, 2012.

63. Chang, C. C., and Lin, C. J., LIBSVM: A library for support vector machines. 2001, Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm.

64. Cohen, J., A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20(1):37–46, 1960.

65. Kocer, S., and Canal, M. R., Classifying epilepsy diseases using artificial neural networks and genetic algorithm. *J. Med. Syst.* 35(4):489–498, 2011.

66. Alickovic, E., and Subasi, A., Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier. *J. Med. Syst.* 40(4):1–12, 2016.

67. Ozsen, S., Gunes, S., Kara, S., and Latifoglu, F., Use of kernel functions in artificial immune systems for the nonlinear classification problems. *IEEE Trans. Inform. Tech. Biomed.* 13(4):621–628, 2009.

68. Tian, J., Li, M., and Chen, F., A hybrid classification algorithm based on coevolutionary EBFNN and domain covering method. *Neural Comput. Appl.* 18(3):293–308, 2009.

69. Torun, Y., and Tohumoglu, G., Designing simulated annealing and subtractive clustering based fuzzy classifier. *Appl. Soft Comput.* 11(2):2193–2201, 2011.

70. Al-Obeidat, F., Belacela, N., Carretero, J. A., and Mahanti, P., An evolutionary framework using particle swarm optimization for classification method PROAFTN. *Appl. Soft Comput.* 11(8):4971–4980, 2011.

71. Jaganathan, P., and Kuppuchamy, R., A threshold fuzzy entropy based featureselection for medical database classification. *Comput. Biol. Med.* 43:2222–2229, 2013.

72. Lim, C. K., and Chan, C. S., A weighted inference engine based on interval valued fuzzy relational theory. *Expert Syst. Appl.* 42:3410–3419, 2015.

73. Yang, C. Y., Chou, J. J., and Lian, F. L., Robust classifier learning with fuzzy class labels for large-margin support vector machines. *Neurocomputing* 99:1–14, 2013.

74. Ahmad, F., Isa, N. A. M., Hussain, Z., and Osman, M. K., Intelligent medical disease diagnosis using improved hybrid genetic algorithm-multilayer perceptron network. *J. Med. Syst.* 37(2):1–8, 2013.

75. Ibrikci, T., Ustun, D., and Kaya, I. E., Diagnosis of several diseases by using combined kernels with support vector machine. *J. Med. Syst.* 36(3):1831–1840, 2012.

76. Psorakis, I., Damoulas, T., and Girolami, M. A., Multiclass relevance vector machines: Sparsity and accuracy. *IEEE Trans. Neural Network* 21(10):1588–1598, 2010.

77. Lin, J. J., and Chang, P. C., A particle swarm optimization based classifier for liver disorders classification, *in: International Conference on Computational Problem-Solving (ICCP)*, pp. 3–5, 2010.

78. Wang, J., Belatreche, A., Maguire, L., and McGinnity, T. M., An online supervised learning method for spiking neural networks with adaptive structure. *Neurocomputing* 144:526–536, 2014.

79. Ozsen, S., and Yucelbas, C., On the evolution of ellipsoidal recognition regions in artificial immune systems. *Appl. Soft Comput.* 31:210–222, 2015.