Hindawi Publishing Corporation
The Scientific World Journal
Volume 2013, Article ID 419187, 10 pages
http://dx.doi.org/10.1155/2013/419187

Research Article

Feature Selection Method Based on Artificial Bee Colony Algorithm and Support Vector Machines for Medical Datasets Classification

Mustafa Serter Uzer,[1] Nihat Yilmaz,[1] and Onur Inan[2]

[1] *Electrical-Electronics Engineering, Faculty of Engineering, Selcuk University, Konya, Turkey*
[2] *Computer Engineering, Faculty of Engineering, Selcuk University, Konya, Turkey*

Correspondence should be addressed to Mustafa Serter Uzer; msuzer@selcuk.edu.tr

Received 25 May 2013; Accepted 6 July 2013

Academic Editors: J. Yan and Y. Zhang

This paper offers a hybrid approach that uses the artificial bee colony (ABC) algorithm for feature selection and support vector machines for classification. The purpose of this paper is to test the effect of elimination of the unimportant and obsolete features of the datasets on the success of the classification, using the SVM classifier. The developed approach conventionally used in liver diseases and diabetes diagnostics, which are commonly observed and reduce the quality of life, is developed. For the diagnosis of these diseases, hepatitis, liver disorders and diabetes datasets from the UCI database were used, and the proposed system reached a classification accuracies of 94.92%, 74.81%, and 79.29%, respectively. For these datasets, the classification accuracies were obtained by the help of the 10-fold cross-validation method. The results show that the performance of the method is highly successful compared to other results attained and seems very promising for pattern recognition applications.

1. Introduction

Pattern recognition and data mining are the techniques that allow for the acquirement of meaningful information from large-scale data using a computer program. Nowadays, these techniques are extensively used, particularly in the military, medical, and industrial application fields, since there is a continuously increasing amount and type of data in these areas, due to advanced data acquisition systems. For this reason, for the obtained data set, data reduction algorithms are needed for filtering, priority sorting, and providing redundant measurements to detect the feature selection. By using these algorithms, quality data is obtained, which in turn raises the quality of the analyzing systems or the success of the recognition systems. In particular, medical applications with ever-increasing popularity and use of advanced technology are the most important field in which these algorithms are used. Many new algorithms developed in the field of medicine are tested on the disease data presented for the common use of all the scientists, and their performances are compared. The datasets from UCI database are very popular for this purpose. The algorithm developed and tested on hepatitis, liver disorders, and diabetes data from UCI was compared with studies in the literature that use the same datasets. These data sets consist of diseases that are commonly encountered in society and significantly reduce the quality of life of patients. The selected data sets are comprised of a variety of test and analysis device data and personal information about the patients. The main objective our work is the integration of the developed systems to these test and analysis devices and to provide a fully automatic assistance to the physician in the creation of diagnosis systems for the diseases. The diagnosis systems, which can be easily used during routine controls, will make the timely information and the early treatment of patients possible.

For the dataset recognition aiming diagnosis of the diseases, we propose a two-stage approach. The first stage has used the clustering with ABC algorithm as selection criteria for feature selection, and, thus, more effective feature selection methods have been constituted. Hence, it has been

made possible both to select the related features faster and to reduce the feature vector dimensions. In the second stage, the reduced data was given to the SVM classifier and the accuracy rates were determined. The k-fold cross-validation method was used for improving the classifier reliability. The datasets we have worked on have been described in the Background section. As it is seen from the results, the performance of the proposed method is highly successful compared to other results attained and seems very promising for pattern recognition applications.

1.1. Background. The developed approach has been tested for the diagnosis of liver diseases and diabetes, which are commonly seen in the society and both reduce the quality of life. In the developed system, the hepatitis and liver disorders datasets were used for the diagnosis of liver disease, and the Diabetes dataset was used for the diagnosis of diabetes.

The liver disease diagnostics studies using the Hepatitis dataset were as follows: Polat and Güneş [1] proposed a new diagnostic method of hepatitis disease based on a hybrid feature selection (FS) method and artificial immune recognition system (AIRS) using fuzzy resource allocation mechanism. The obtained classification accuracy of the proposed system was 92.59%. A machine learning system studied by Polat and Güneş [2] was conducted to identify hepatitis disease. At first, the feature number of dataset on hepatitis disease was reduced from 19 to 10 by using in the feature selection (FS) subprogram and C4.5 decision tree algorithm. Then, fuzzy weighted preprocessing was used for weighting the dataset after normalizing between 0 and 1. AIRS classifier system was used while classifying the weighted input values. The classification accuracy of their system was 94.12%. Principal component analysis (PCA) and artificial immune recognition system (AIRS) were conducted for hepatitis disease prediction in the study by Polat and Güneş [3]. Classification accuracy, 94.12%, was obtained with the proposed system using 10-fold cross-validation. A method which had an accuracy value of 96.8% for hepatitis dataset was proposed by Kahramanli and Allahverdi [4], and in this method extracting rules from trained hybrid neural network was presented by using artificial immune systems (AISs) algorithm. An automatic diagnosis system using linear discriminant analysis (LDA) and adaptive network based on fuzzy inference system (ANFIS) was proposed by Dogantekin et al. [5] for hepatitis diseases. This automatic diagnosis system of hepatitis disease diagnostics was obtained with a classification accuracy of about 94.16%. Bascil and Temurtas [6] realized a hepatitis disease diagnosis based on a multilayer neural network structure that used the Levenberg- Marquardt algorithm as training algorithm for the weights update with a classification accuracy of 91.87% from 10-fold cross-validation.

The studies for the diagnosis of liver disease in using the liver disorders dataset were as follows: by Lee and Mangasarian [7], smoothing methods were applied to generate and solve an unconstrained smooth reformulation of the support vector machine for pattern classification using a completely arbitrary kernel. They termed such reformulation a smooth support vector machine (SSVM). Correct classification rate of the proposed system with CV-10 was 70.33% for liver disorders dataset. In Van Gestel et al.'s [8] article, the Bayesian evidence framework was combined with the LS-SVM classifier formulation. Correct classification rate of proposed system with CV-10 was 69.7% for liver disorders dataset. Gonçalves et al. [9] a new neuro-fuzzy model, especially created for record classification and rule extraction of databases, named as inverted hierarchical neuro-fuzzy BSP System (HNFB). Correct classification rate of this system was 73.33% for liver disorders dataset. Özşen and Güneş [10] aimed to contribute to an artificial immune system AIS by attaching this aspect and used the Euclidean distance, Manhattan distance, and hybrid similarity measure with simple AIS. Correct classification rate of the proposed system with AWAIS was 70.17%, with hybrid similarity measure 60.57%, with the Manhattan distance 60.21%, with the Euclidean distance 60.21% for liver disorders. Li et al. [11] proposed a nonlinear transformation method based on fuzzy to find classification information in the original data attribute values for a small dataset and used a support vector machine (SVM) as a classifier. Correct classification rate of the proposed system was 70.85% for liver disorders. Chen et al. [12] proposed an analytical approach by taking an integration of particle swarm optimization (PSO) and the 1-NN method. Correct classification rate of proposed system with 5-fold cross-validation was 68.99% for liver disorders dataset. A hybrid model based on integrating a case-based reasoning approach and a particle swarm optimization model were proposed by Chang et al. [13] for medical data classification.

Another disease that we selected is diabetes. Some of the most important studies conducted on this dataset are as follows: Şahan et al. [14] proposed attribute weighted artificial immune system (AWAIS) with weighting attributes due to their important degrees in class discrimination and using them for the Euclidean distances calculation. AWAIS had a classification accuracy of 75.87 using 10-fold cross-validation method for diabetes dataset. Polat and Güneş [15] worked on diabetes disease using principal component analysis (PCA) and adaptive neuro-fuzzy inference system (ANFIS). The obtained test classification accuracy was 89.47% by using the 10-fold cross-validation. Polat et al. [16] proposed a new learning system which is cascade and used generalized discriminant analysis and least square support vector machine. The classification accuracy was obtained as 82.05%. Kahramanli and Allahverdi [17] presented a hybrid neural network that achieves accuracy value of 84.24% using artificial neural network (ANN) and fuzzy neural network (FNN) together. Patil et al. [18] proposed hybrid prediction model (HPM) which uses Simple k-means clustering algorithm for verifying the chosen class labels and then using the classification algorithm on the result set. Accuracy value of HPM was 92.38%. Isa and Mamat [19] presented a modified hybrid multilayer perceptron (HMLP) network for improving the conventional one, and the average correct classification rate of the proposed system was 80.59%. Aibinu et al. [20] proposed a new biomedical signal classification method using complex-valued pseudo autoregressive (CAR) modeling approach. The presented technique obtained a classification accuracy of 81.28%.

(1) Start with the empty set $Y_0 = \{\varnothing\}$
(2) Select the next best feature
$x^+ = \arg\max[J(Y_k + x)]$; $x \notin Y_k$
(3) Update $Y_{k+1} = Y_k + x^+$; $k = k + 1$
(4) Goto 2

Pseudocode 1: Pseudo code for SFS [22].

2. Preliminaries

2.1. Feature Selection. Feature selection provides a smaller but more distinguishing subset compared to the starting data, selecting the distinguishing features from a set of features and eliminating the irrelevant ones. Reducing the dimension of the data is aimed by finding a small important features set. This results in both reduced processing time and increased classification accuracy.

The algorithm developed in this study was based on the sequential forward selection (SFS) algorithm, which is popular in these algorithms. SFS is a method of feature selection offered by Whitney [21]. Sequential forward selection is the simplest greedy search algorithm which starts from the empty set and sequentially adds the feature x^+ for obtaining results in the highest objective function $J(Y_k + x^+)$ when combined with the features Y_k that have already been selected. Pseudo code is given Pseudocode 1 for SFS [22].

In summary, SFS begins with zero attributes and then evaluates the whole feature subsets with only one feature, and the best performing one adds this subset to the best performing feature for subsets of the next larger size. This cycle repeats until there is no improvement in the current subset [23].

The objection function is critical for this algorithm. Finding the highest value of this function is an optimization problem. Clustering is an ideal method for the detection of feature differentiation. The developed method can be summarized using the ABC algorithm for feature selection aiming clustering problem adaptation.

2.1.1. Clustering with Optimization. Clustering is a grouping process running on the multi-dimentional data by using similarities. Distance criteria are used to evaluate similarities in samples set. Clustering problems can be expressed as the placement of every object into one K cluster for a given N number of objects and minimizing the sum of squares of the Euclidean distances between the centers of these objects in the cluster to which they belong. The function that uses the clustering algorithm is given in (1) [24] for minimizing:

$$J(w, z) = \sum_{i=1}^{N} \sum_{j=1}^{K} w_{ij} \left\| x_i - z_j \right\|^2. \quad (1)$$

Here, N is the number of samples, K is the number of clusters, x_i $(i = 1, \ldots, N)$ is the place of the ith sample, and the center of the jth sample z_j $(j = 1, \ldots, N)$ can be obtained by (2):

$$z_j = \frac{1}{N_j} \sum_{i=1}^{N} w_{ij} x_i. \quad (2)$$

Here, N_j is the number of samples in the jth cluster, and w_{ij} is the relationship of j cluster and x_i sample with a value of 1 or 0. If the sample i (x_i) belongs to the j cluster, w_{ij} is 1, otherwise that it is 0.

The clustering process that separates objects into groups can be performed by supervised or unsupervised learning. Training data in unsupervised clustering (also known as automatic clustering) does not need to set class tags. In supervised clustering, however, it should be specified so that the classes can learn the tags. In this study, the datasets used should contain class information since supervised clustering was used. Therefore, the optimization aims to find the centers of clusters by making the objective function minimize, which is the total of the samples distances to centers [24]. In this study, the sum of distances between all training cluster samples and the cluster center $(p_i^{\mathrm{CL}_{\mathrm{known}}(x_j)})$ that samples belong to in the n-dimensional Euclidean space are minimized for adaptation [24]. Consider the following:

$$f_i = \frac{1}{D_{\mathrm{Train}}} \sum_{j=1}^{D_{\mathrm{Train}}} d\left(x_j, p_i^{\mathrm{CL}_{\mathrm{known}}(x_j)}\right). \quad (3)$$

Here, D_{Train} is the number of training samples, and the total expression in the cost function is for normalizing the number to a value between 0.0 and 1.0. The $p_i^{\mathrm{CL}_{\mathrm{known}}(x_j)}$ value indicates the center of the class that belongs to the sample that is used according to training data. Here, the ABC algorithm was chosen as the optimization method for clustering. Thus, ABC, as a new clustering method, can also be used in the feature selection algorithms.

2.2. Artificial Bee Colony (ABC) Algorithm. Artificial bee colony (ABC) algorithm, as a population-based stochastic optimization proposed by Karaboga in [24–26], realize the intelligent foraging behavior of honey bee swarms. It can be used for classification, clustering and optimization studies. Pseudocode of the ABC algorithm is given as Pseudocode 2.

An artificial group of bees in the ABC algorithm consists of three different groups: employed bees, onlooker bees, and scout bees. In this algorithm, the number of bees employed in the colony also equals the number of onlooker bees. Additionally, the number of employed bees or onlooker bees equals the number of solutions in the population. An onlooker bee is the bee that waits in the dance area to make the food source selection decision. An onlooker bee is named employed bee once it goes to a food source. An employed bee that has consumed the food source turns into a scout bee, and its duty is to perform a random search to discover new resources. Food supply position—which represents the solution to the optimization problem—and the amount of nectar

(1) Load training samples
(2) Generate the initial population z_i, $i = 1, \ldots, \mathrm{SN}$
(3) Evaluate the fitness (f_i) of the population
(4) set cycle to 1
(5) **repeat**
(6) **FOR** each employed bee {
Produce new solution v_i by using (6)
Calculate the value f_i
Apply greedy selection process}
(7) Calculate the probability values p_i for the solutions (z_i) by (5)
(8) **FOR** each onlooker bee {
Select a solution z_i depending on p_i
Produce new solution v_i
Calculate the value f_i
Apply greedy selection process}
(9) **If** there is an abandoned solution for he scout
then replace it with a new solution which will
be randomly produced by (7)
(10) Memorize the best solution so far
(11) cycle = cycle + 1
(12) until cycle = MCN

Pseudocode 2: Pseudo-code of the ABC algorithm [24].

in the food source depends on the quality of the associated solution. This value is calculated in (4).

$$\mathrm{fit}_i = \frac{1}{1 + f_i} \tag{4}$$

SN in the algorithm indicates the size of the population. At first, the ABC algorithm produces a distributed initial population $P(C = 0)$ of SN solutions (food source positions) randomly, where SN means the size of population. Each z_i solution is a D-dimensional vector for $i = 1, 2, 3, \ldots, \mathrm{SN}$. Here, D is the numbers of cluster products and input size for each dataset. After startup, an investigation is repeated on employed bees, onlooker bees, and scout bees processes until the number of population of positions ($C = 1, 2, \ldots, \mathrm{MCN}$) is completed. Here, MCN is the maximum cycle number.

An employed bee makes a small change in position due to the local knowledge in its memory, and a new source is generated. This bee makes a comparison of the nectar amount (fitness amount) of a new source with the nectar amount of previous source and decides which one is higher. If the new position is higher than the old one then it is assimilated into its memory and the old one is forgotten. Otherwise, the position of the previous one stays in its memory. All employed bees that complete the task of research share the position and nectar food source information with the onlooker bees that are in the dance area.

An onlooker bee evaluates the nectar information of all employed bees and chooses a food source depending on the probability of the nectar amount. This probability value (p_i) is calculated in (5). Just like the employed bees, the onlooker bee modifies the situation from memory and it checks the nectar amount of the candidate source. If its nectar amount is higher than the previous one and the new position is assimilated into memory and the old one is forgotten, then

$$p_i = \frac{\mathrm{fit}_i}{\sum_{n=1}^{\mathrm{SN}} \mathrm{fit}_n}, \tag{5}$$

where SN is the number of food sources which is equal to the number of employed bees and the fitness of the fit_i solution given in (4). The f_i given in (3) is the cost function of the cluster problem. ABC uses (6) for producing a candidate food position:

$$v_{ij} = z_{ij} + \phi_{ij}\left(z_{ij} - z_{kj}\right). \tag{6}$$

Here, $k \in \{1, 2, \ldots, \mathrm{SN}\}$ and $j \in \{1, 2, \ldots, D\}$ are randomly selected indexes. k is a random value different from i. ϕ_{ij} is a random number between $[-1, 1]$ which controls the production of neighboring food sources around z_{ij} and represents comparison of two food sources to a bee.

While onlooker and employed bees perform exploitation in the search area, scout bees control the discovery process and replace the consumed nectar food source with a new food source in the ABC algorithm. If the position cannot be improved as a previously determined cycle number, this food source is accepted as abandoned. The previously determined cycle number is defined as the "*limit*" for abandonment. In this case, there are three control parameters in ABC: the number of food sources (SN) which is equal to the number of employed and onlooker bees, the maximum cycle number (MCN), and the limit value.

If an abandoned source is assumed to be z_i and $j \in \{1, 2, \ldots, D\}$, the scout looks for a new source to replace z_i. This process is described by (7):

$$z_i^j = z_{\min}^j + \mathrm{rand}(0, 1)\left(z_{\max}^j - z_{\min}^j\right). \tag{7}$$

After (v_{ij}) which is each candidate position is produced, the position is evaluated by ABC and its performance is compared with previous one. The performance is compared with the previous one. If the new food source has an equal amount or more nectar than the old one, the new one takes place instead of the old food source in memory. Otherwise, the old one stays in its place in memory. So a greedy selection mechanism is used to make selections among the old source and one of the candidates.

2.3. Support Vector Machines (SVMs). SVM is an effective supervised learning algorithm used in classification and regression analyses for applications like pattern recognition, data mining, and machine learning application. SVM was developed in 1995 by Cortes and Vapnik [27]. Many studies have been conducted on SVM: a flexible support vector machine for regression, an evaluation of flyrock phenomenon based on blasting operation by using support vector machine [28, 29].

In this algorithm, there are two different categories separated by a linear plane. The training of the algorithm is determining the process for the parameters of this linear plane. In multiclass applications, the problem is categorized into groups as belonging either to one class or to others. SVM's use in pattern recognition is described below.

An n-dimensional pattern (object) x has n coordinates, $x = (x_1, x_2, \ldots, x_n)$, where each x is a real number, $x_i \in R$ for $i = 1, 2, \ldots, n$. Each pattern x_j belongs to a class $y_j \in \{-1, +1\}$. Consider a training set T of m patterns together with their classes, $T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$. Consider a dot product space S, in which the patterns x are embedded, $x_1, x_2, \ldots, x_m \in S$. Any hyperplane in the space S can be written as

$$\{x \in S \mid w \cdot x + b = 0\}, \quad w \in S,\ b \in R. \tag{8}$$

The dot product $w \cdot x$ is defined by

$$w \cdot x = \sum_{i=1}^{n} w_i x_i. \tag{9}$$

A training set of patterns can be separated as linear if there exists at least one linear classifier expressed by the pair (w, b) which correctly classifies all training patterns as can be seen in Figure 1. This linear classifier is represented by the hyperplane $H(w \cdot x + b = 0)$ and defines a region for class +1 patterns ($w \cdot x + b > 0$) and another region for class −1 patterns ($w \cdot x + b < 0$).

After the training process, the classifier becomes ready for prediction of the class membership on new patterns, different from training. The class of a pattern x_k is found from the following equation:

$$\text{class}(x_k) = \begin{cases} +1 & \text{if } w \cdot x_k + b > 0 \\ -1 & \text{if } w \cdot x_k + b < 0. \end{cases} \tag{10}$$

Thus, the classification of new patterns relies on only the sign of the expression $w \cdot x + b$ [30].

Sequential Minimal optimization is used in the training stage of SVM. SMO algorithm is a popular optimization method used to train the support vector machine (SVM). The dual presentation of an SVM primal optimization problem is indicated in (11):

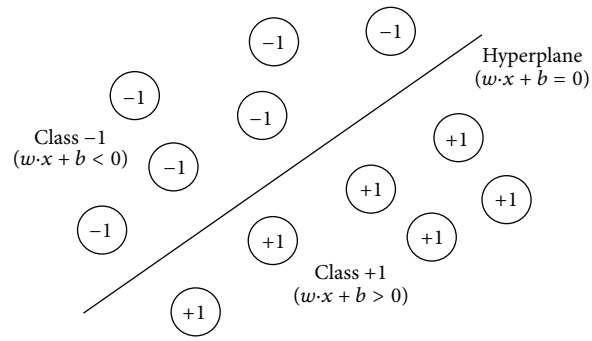

Figure 1: Linear classifier defined by the hyperplane H ($w \cdot x + b = 0$).

$$\max_{\alpha} \quad \Psi(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j k\left(x_i, x_j\right) \alpha_i \alpha_j$$
$$\text{subject to} \quad \sum_{i=1}^{N} y_i \alpha_i = 0, \quad 0 \le \alpha_i \le C,\ i = 1, \ldots, n, \tag{11}$$

where x_i is a training sample, $y_i \in \{-1, +1\}$ is the corresponding target value, α_i is the Lagrange multiplier, and C is a real value cost parameter [31].

2.4. Performance Evaluation. Four criteria for performance evaluation of hepatitis, liver disorders and diabetes datasets were used. These criteria are classification accuracy, confusion matrix, analysis of sensitivity and specificity, and k-fold cross-validation.

2.4.1. Classification Accuracy. In this study, the classification accuracies for the datasets are measured with the following the equation:

$$\text{accuracy}(T) = \frac{\sum_{i=1}^{N} \text{assess}(t_i)}{N}, \quad t_i \in T,$$
$$\text{assess}(t_i) = \begin{cases} 1, & \text{if classify}(t_i) \equiv \text{correctclassification}, \\ 0, & \text{otherwise}, \end{cases} \tag{12}$$

where T is the classified set of data items (the test set) and N is the number of testing samples of the dataset. We will also show the accuracy of our performed k-fold cross-validation (CV) experiment.

2.4.2. Confusion Matrix. The confusion matrix includes four classification performance indices: true positive, false positive, false negative, and true negative as given in Table 1. They are also usually used in the two-class classification problem to evaluate the performance.

2.4.3. Analysis of Sensitivity and Specificity. The following expressions were used for calculating sensitivity, specificity,

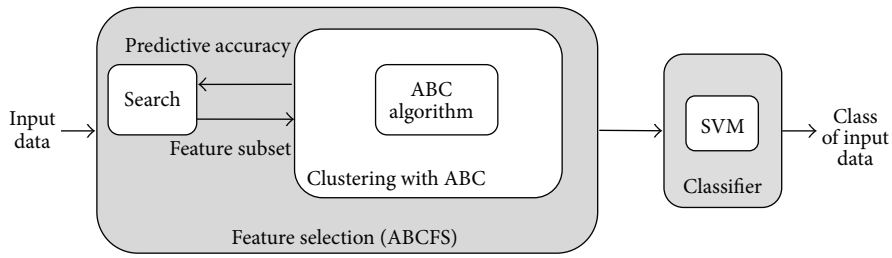

Figure 2: Block diagram of the proposed system.

Table 1: The four classification performance indices included in the confusion matrix.

Actual class	Predicted class	
	Positive	Negative
Positive	True positive (TP)	False negative (FN)
Negative	False positive (FP)	True negative (TN)

positive predictive value, and negative predictive value; we use [32]:

$$\begin{aligned} \text{Sensitivity } (\%) &= \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100, \\ \text{Specificity } (\%) &= \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100, \\ \text{Positive predctive value } (\%) &= \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100, \\ \text{Negative predicitive value } (\%) &= \frac{\text{TN}}{\text{TN} + \text{FN}} \times 100. \end{aligned} \tag{13}$$

2.4.4. k-Fold Cross-Validation. k-fold cross-validation is used for the test result to be more valuable [33]. In k-fold cross-validation, the original sample is divided into random k subsamples, one of which is retained as the validation data for model testing and the remaining k-1 sub-samples are used for training The cross-validation process is then repeated k times (the folds), with each of the k sub-samples used exactly once as the validation data. The process is repeated k times (the folds), with each of the k sub-samples used only once as the validation data. The average of k results from the folds gives the test accuracy of the algorithm [34].

3. Experimental Work

Less distinctive features of the data set affect the classification negatively. Such data especially decrease the speed and the system performance significantly. With the proposed system, using the feature selection algorithm, the features with less discriminant data were eliminated. The reduced data set increased the testing success of the classifier and the rate of the system. From Figure 2, the proposed system has two phases. At the first phase, as selection criteria, clustering with ABC algorithm was used for feature selection, and, thus, a more effective feature selection method was constituted. Hence, it has been made possible both to select the related features in a shorter period of time and to reduce the dimension of the feature vector. At second stage, the obtained reduced data is supplied to the SVM classifier to determine the accuracy rates. The k-fold cross-validation was used for the classifier reliability improvement.

In this study, ABCFS + SVM system is suggested in order to solve the three classification problem named as Hepatitis dataset, Liver Disorders dataset, Diabetes dataset, respectively.

3.1. Datasets. We used the dataset from the UCI machine learning database [35], which is commonly used among researchers for classification, that gives us a chance to compare the performance of our method with others. The datasets of this work can be defined shortly as follows.

3.1.1. Hepatitis Dataset. This dataset was donated by Jozef Stefan Institute, Yugoslavia. The purpose of the dataset is to predict the presence or absence of hepatitis disease from the different medical tests results of a patient. This database contains 19 attributes. There are 13 binary and 6 discrete values. Hepatitis dataset includes 155 samples from two different classes (32 "die" cases, 123 "live" cases). This dataset contains missing attribute values. We substituted the missing data by frequently encountered values of own class. Attributes of symptoms that are obtained from patient are given in Table 2 [3, 35].

3.1.2. Liver Disorders Dataset. The liver disorders dataset is named as BUPA liver disorders. The liver disorders database includes 6 features, that is, MCV, alkphos, SGPT, SGOT, gammaGT and drinks. There are 345 data in total and each sample is taken from an unmarried man. Two hundred of them are chosen for one class with the remaining 145 are in the other. The first 5 features are all blood tests which are sensitive to liver disorders that arise from excessive alcohol consumption. This dataset is donated by Richard S. Forsyth et al. in 1990. The attributes are given in Table 3 [13].

3.1.3. Diabetes Dataset. This dataset contains 768 samples, where each sample has 8 features which are eight clinical findings. All patients of the dataset are Pima Indian women in which the youngest one is 21 years old and living near Phoenix, Arizona, USA. The binary target variable can take "0" or "1." If it takes "1," it means a positive test for Diabetes, or if it takes "0," it means a negative test. There are 268 different cases in class "1" and 500 different cases in class "0." The features and parameters are given in Table 4 [16].

Table 2: Range values and attribute names for hepatitis dataset [35].

The number of attribute	The name of attribute	Interval of attribute
1	Age	7–78
2	Sex	Male, Female
3	Steroid	No, Yes
4	Antivirals	No, Yes
5	Fatigue	No, Yes
6	Malaise	No, Yes
7	Anorexia	No, Yes
8	Liver big	No, Yes
9	Liver firm	No, Yes
10	Spleen palpable	No, Yes
11	Spiders	No, Yes
12	Ascites	No, Yes
13	Varices	No, Yes
14	Bilirubin	0.3–8
15	Alk phosphate	26–295
16	SGOT	14–648
17	Albumin	2.1–6.4
18	Protime	0–100
19	Histology	No, Yes

Table 3: Range values and attribute names for liver disorders dataset [35].

The number of attribute	The name of attribute	Description of the attribute	Interval of attribute
1	MCV	Mean corpuscular volume	65–103
2	Alkphos	Alkaline phosphatase	23–138
3	SGPT	Alamine aminotransferase	4–155
4	SGOT	Aspartate aminotransferase	5–82
5	gammaGT	Gamma-glutamyl transpeptidase	5–297
6	Drinks	Number of half-pint equivalents of alcoholic beverages drunk per day	0–20

3.2. Feature Selection with ABC. In the system, a searching process runs to find the best feature subset same like sequential forward selection algorithm. Prediction accuracy for feature selection is found by ABC clustering. Pseudocode of the developed feature selection algorithm based on ABC is given in Pseudocode 3.

In Pseudocode 3, n is sample count and p is desired feature count which is selected as providing the highest performance criteria. While *data* represents the entire dataset, *Data_all* includes the features that are considered chosen. *Train_data_all* is generated by taking 75% of the data found in all classes of *Data_all*. *Test_data_all* is generated by taking 25% of the data that are found in all classes of *Data_all*.

Table 4: Features and parameters of the diabetes dataset.

Features	Mean	Standard deviation	Min	Max
Number of times pregnant	3.8	3.4	0	17
Plasma glucose concentration, 2 h in an oral glucose tolerance test	120.9	32.0	0	199
Diastolic blood pressure (mm Hg)	69.1	19.4	0	122
Triceps skinfold thickness (mm)	20.5	16.0	0	99
2-hour serum insulin (mu U/mL)	79.8	115.2	0	846
Body mass index (kg/m^2)	32.0	7.9	0	67.1
Diabetes pedigree function	0.5	0.3	0.078	2.42
Age (years)	33.2	11.8	21	81

Table 5: List of datasets.

Databases	Number of classes	Samples	Number of features	Number of selected features	Selected features
Hepatitis	2	155	19	11	12, 14, 13, 15, 18, 1, 17, 5, 16, 2, 4
Liver disorders	2	345	6	5	5, 3, 2, 4, 1
Diabetes	2	768	8	6	2, 8, 6, 7, 4, 5

Table 6: List of classification parameters.

Parameters	Value
Method	SVM
Optimization algorithm	SMO
Validation method	k-fold cross-validation (10-fold CV)
Kernel_Function	Linear
TolKKT	$1.0000e-003$
MaxIter	15000
KernelCacheLimit	5000
The initial value	Random

Table 7: Performance of classification for the hepatitis, liver disorders, and diabetes datasets.

Performance criteria	Hepatitis dataset	Liver disorders dataset	Diabetes dataset
Classification accuracy (%)	94.92	74.81	79.29
Sensitivity (%)	97.13	88.22	89.84
Specificity (%)	88.33	56.68	59.61
Positive predictive value (%)	96.91	73.99	80.63
Negative predictive value (%)	88.33	78.57	75.65

Thus, a uniform dispersion was obtained according to the classes. Train data that belongs to each class (*train_data_class*) is trained by the ABC algorithm, which has been modified to cluster. At the end of training, 10 feature vectors named food and representing each class are obtained. Goodness of the chosen feature cluster is described by the food values

```
n = sample_count;
Selected_features = {∅}
For p = 1 to desired_feature_count
  For c = 1 to feature_count
    Data_all = data(Selected_features + feature(c));
    For i = 1 to class_number
      Train_data_class(i) = partition(rand(Data_all(class == i)), 0.75)
      Test_data_class(i) = partition(rand(Data_all(class = i)), others);
        [foods(i)] = Modified_ABC_algortihm (train_data_class(i), performance_function);
    End for i
   Test_data_all = merge(test_data_class);
   For i = 1: size(Test_data_all)
     For k = 1: class_count
       For j = 1: count(foods(k))
         distance(k, j) = oklid_distance(foods(j, k)-test_data(i));
       End for j
       min_dist(k) = min(distance(k));
     End for k
     [result_of_test_data(i), class_of_test_data(i)] = min(min_dist);
   End for i
   Performance_criteria(feature(c))
= sum(class_of_test_data(i) == class_of_test_data_expected);
  End for c
 Best_feature(c) = arg max(performance_criteria(feature(c))
 Selected_features = Selected_fetures + best_feature(c)
 End for p
```

Pseudocode 3: Pseudo-code of developed feature selection algorithm based on ABC.

Table 8: Classification accuracies obtained by our method and other classifiers for the hepatitis dataset.

Author (year)	Method	Classification accuracy (%)
Polat and Güneş (2006) [1]	FS-AIRS with fuzzy res. (10-fold CV)	92.59
Polat and Güneş (2007) [2]	FS-Fuzzy-AIRS (10-fold CV)	94.12
Polat and Güneş (2007) [3]	AIRS (10-fold CV)	76.00
	PCA-AIRS (10-fold CV)	94.12
Kahramanli and Allahverdi (2009) [4]	Hybrid system (ANN and AIS) (without k-fold CV)	96.8
Dogantekin et al. (2009) [5]	LDA-ANFIS	94.16
Bascil and Temurtas (2011) [6]	MLNN (MLP) + LM (10-fold CV)	91.87
Our study	ABCFS + SVM (10-fold CV)	**94.92**

Table 9: Classification accuracies obtained by our method and other classifiers for the liver disorders dataset.

Author (year)	Method	Classification accuracy (%)
Lee and Mangasarian (2001) [7]	SSVM (10-fold CV)	70.33
van Gestel et al. (2002) [8]	SVM with GP (10-fold CV)	69.7
Gonçalves et al. (2006) [9]	HNFB-1 method	73.33
Özşen and Güneş (2008) [10]	AWAIS (10-fold CV)	70.17
	AIS with hybrid similarity measure (10-fold CV)	60.57
	AIS with Manhattan distance (10-fold CV)	60.21
	AIS with Euclidean distance (10-fold CV)	60.00
Li et al. (2011) [11]	A fuzzy-based nonlinear transformation method + SVM	70.85
Chen et al. (2012) [12]	(PSO) + 1-NN method (5-fold CV)	68.99
Chang et al. (2012) [13]	CBR + PSO (train: 75%-test: 25%)	76.81
Our study	ABCFS + SVM (train: 75%-test: 25%)	**82.55**
	ABCFS + SVM (10-fold CV)	**74.81**

Table 10: Classification accuracies obtained by our method and other classifiers for diabetes dataset.

Author (year)	Method	Classification accuracy (%)
Şahan et al. (2005) [14]	AWAIS (10-fold CV)	75.87
Polat and Güneş (2007) [15]	Combining PCA and ANFIS	89.47
Polat et al. (2008) [16]	LS-SVM (10-fold CV)	78.21
	GDA-LS-SVM (10-fold CV)	82.05
Kahramanli andAllahverdi (2008) [17]	Hybrid system (ANN and FNN)	84.2
Patil et al. (2010) [18]	Hybrid prediction model (HPM) with reduced dataset	92.38
Isa and Mamat (2011) [19]	Clustered-HMLP	80.59
Aibinu et al. (2011) [20]	AR1 + NN (3-fold CV)	81.28
Our study	ABCFS + SVM (train: 75%-test: 25%)	**86.97**
	ABCFS + SVM (10-fold CV)	**79.29**

accuracy representing the test dataset. The error value is found by taking the difference between the test data class and the food value class having a minimum Euclidean distance to the test data class.

The performance value shows the suitability of the added property. The most appropriate property value does not belong to the chosen properties cluster. This process is repeated by starting from an empty cluster up until the desired feature number. The decline in the value of rising performance trend is for determining the maximum number of features. In summary, in ABCFS, it starts from selecting the feature set as empty, then adds the feature(c) that results in the highest objective function.

We selected colony size 20, maximum cycle/generation number (MCN) 300, and limit value 200. The algorithm was run 100 times. Performance value is found by taking the average of these algorithm results.

The datasets used for evaluating ABCFS performance and their features are as follows: the number of classes, the number of samples, the number of features and the number of selected features, which are given in Table 5.

3.3. SVM Classification Parameters. The reliability of the classifier was provided by the k-fold cross-validation method. While this classifier was used, the training was performed according to the parameters in Table 6.

4. Experimental Results and Discussion

ABCFS + SVM method test results developed for the hepatitis dataset, liver disorders dataset and diabetes datasets are given in Pseudocode 3. These test results contain the classification performance values achieved by the developed methodology by the help of 10-fold cross-validation. The performance values include average classification accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of the proposed system which are given in Table 7. The results of the study show that the average correctness rate of the studies performed so far on all used datasets by employing the method of k-fold cross-validation is a very promising result.

For the hepatitis dataset, the comparisons with the other systems are given in Table 8.

For the liver disorders dataset, the comparisons with the other systems are given in Table 9.

For the diabetes dataset, the comparisons with the other systems are given in Table 10.

5. Conclusions

This study was designed for use in the diagnosis of liver and diabetes. In these databases that were used, there are some redundant and low-distinctive features. These features are very important factors affecting the success of the classifier and the system processing time. In the system we have developed, the elimination of these redundant features increased the system speed and success. The artificial bee Colony (ABC) algorithm, which is a very popular optimization method, was used for the feature selection process in the study. The ABC-based feature selection algorithm that was developed in this study is the first example of the ABC algorithm used in the field of feature selection. The databases that are subjected to feature selection are classified using SVM. In order to achieve a reliable performance of the classifier, the 10-fold cross-validation method was used. The system results were compared with the literature articles that use the same databases. Classification accuracy of the proposed system reached 94.92%, 74.81%, and 79.29% for hepatitis dataset, liver disorders dataset and diabetes dataset, respectively. Obtained results show that the performance of the proposed method is highly successful compared to other results attained and seems very promising for pattern recognition applications.

Acknowledgment

The authors would like to thank Selcuk University Scientific Research Projects Coordinatorship for the support of this paper.

References

[1] K. Polat and S. Güneş, "Hepatitis disease diagnosis using a new hybrid system based on feature selection (FS) and artificial immune recognition system with fuzzy resource allocation," *Digital Signal Processing*, vol. 16, no. 6, pp. 889–901, 2006.

[2] K. Polat and S. Güneş, "Medical decision support system based on artificial immune recognition immune system (AIRS), fuzzy

weighted pre-processing and feature selection," *Expert Systems with Applications*, vol. 33, no. 2, pp. 484–490, 2007.

[3] K. Polat and S. Güneş, "Prediction of hepatitis disease based on principal component analysis and artificial immune recognition system," *Applied Mathematics and Computation*, vol. 189, no. 2, pp. 1282–1291, 2007.

[4] H. Kahramanli and N. Allahverdi, "Extracting rules for classification problems: AIS based approach," *Expert Systems with Applications*, vol. 36, no. 7, pp. 10494–10502, 2009.

[5] E. Dogantekin, A. Dogantekin, and D. Avci, "Automatic hepatitis diagnosis system based on linear discriminant analysis and adaptive network based on fuzzy inference system," *Expert Systems with Applications*, vol. 36, no. 8, pp. 11282–11286, 2009.

[6] M. S. Bascil and F. Temurtas, "A study on hepatitis disease diagnosis using multilayer neural network with Levenberg Marquardt training algorithm," *Journal of Medical Systems*, vol. 35, no. 3, pp. 433–436, 2011.

[7] Y. J. Lee and O. L. Mangasarian, "SSVM: a smooth support vector machine for classification," *Computational Optimization and Applications*, vol. 20, no. 1, pp. 5–22, 2001.

[8] T. van Gestel, J. A. K. Suykens, G. Lanckriet, A. Lambrechts, B. de Moor, and J. Vandewalle, "Bayesian framework for least-squares support vector machine classifiers, Gaussian processes, and Kernel Fisher discriminant analysis," *Neural Computation*, vol. 14, no. 5, pp. 1115–1147, 2002.

[9] L. B. Gonçalves, M. M. B. R. Vellasco, M. A. C. Pacheco, and F. J. de Souza, "Inverted Hierarchical Neuro-Fuzzy BSP system: a novel neuro-fuzzy model for pattern classification and rule extraction in databases," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 36, no. 2, pp. 236–248, 2006.

[10] S. Özşen and S. Güneş, "Effect of feature-type in selecting distance measure for an artificial immune system as a pattern recognizer," *Digital Signal Processing*, vol. 18, no. 4, pp. 635–645, 2008.

[11] D. C. Li, C. W. Liu, and S. C. Hu, "A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets," *Artificial Intelligence in Medicine*, vol. 52, no. 1, pp. 45–52, 2011.

[12] L. F. Chen, C. T. Su, K. H. Chen, and P. C. Wang, "Particle swarm optimization for feature selection with application in obstructive sleep apnea diagnosis," *Neural Computing and Applications*, vol. 21, no. 8, pp. 2087–2096, 2012.

[13] P. C. Chang, J. J. Lin, and C. H. Liu, "An attribute weight assignment and particle swarm optimization algorithm for medical database classifications," *Computer Methods and Programs in Biomedicine*, vol. 107, no. 3, pp. 382–392, 2012.

[14] S. Şahan, K. Polat, H. Kodaz, and S. Güneş, "The medical applications of attribute weighted artificial immune system (AWAIS): diagnosis of heart and diabetes diseases," in *Artificial Immune Systems*, vol. 3627 of *Lecture Notes in Computer Science*, pp. 456–468, 2005.

[15] K. Polat and S. Güneş, "An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease," *Digital Signal Processing*, vol. 17, no. 4, pp. 702–710, 2007.

[16] K. Polat, S. Güneş, and A. Arslan, "A cascade learning system for classification of diabetes disease: generalized discriminant analysis and least square support vector machine," *Expert Systems with Applications*, vol. 34, no. 1, pp. 482–487, 2008.

[17] H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," *Expert Systems with Applications*, vol. 35, no. 1-2, pp. 82–89, 2008.

[18] B. M. Patil, R. C. Joshi, and D. Toshniwal, "Hybrid prediction model for t ype-2 diabetic patients," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8102–8108, 2010.

[19] N. A. M. Isa and W. M. F. W. Mamat, "Clustered-hybrid multilayer perceptron network for pattern recognition application," *Applied Soft Computing Journal*, vol. 11, no. 1, pp. 1457–1466, 2011.

[20] A. M. Aibinu, M. J. E. Salami, and A. A. Shafie, "A novel signal diagnosis technique using pseudo complex-valued autoregressive technique," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9063–9069, 2011.

[21] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1119–1125, 1994.

[22] L. Ladha and T. Deepa, "Feature selection methods and algorithms," *International Journal on Computer Science and Engineering*, vol. 3, no. 5, pp. 1787–1797, 2011.

[23] M. Sasikala and N. Kumaravel, "Comparison of feature selection techniques for detection of malignant tumor in brain images," in *Proceedings of the Annual IEEE INDICON '05*, pp. 212–215, December 2005.

[24] D. Karaboga and C. Ozturk, "A novel clustering approach: artificial bee colony (ABC) algorithm," *Applied Soft Computing Journal*, vol. 11, no. 1, pp. 652–657, 2011.

[25] D. Karaboga and B. Akay, "A comparative study of artificial bee colony algorithm," *Applied Mathematics and Computation*, vol. 214, no. 1, pp. 108–132, 2009.

[26] B. Akay and D. Karaboga, "A modified artificial bee colony algorithm for real-parameter optimization," *Information Sciences*, vol. 192, pp. 120–142, 2012.

[27] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[28] H. Amini, R. Gholami, M. Monjezi, S. R. Torabi, and J. Zadhesh, "Evaluation of flyrock phenomenon due to blasting operation by support vector machine," *Neural Computing and Applications*, vol. 21, no. 8, pp. 2077–2085, 2012.

[29] X. B. Chen, J. Yang, and J. Liang, "A flexible support vector machine for regression," *Neural Computing and Applications*, vol. 21, no. 8, pp. 2005–2013, 2012.

[30] O. Ivanciuc, *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and T. R. Cundari, 2007.

[31] T. W. Kuan, J. F. Wang, J. C. Wang, P. C. Lin, and G. H. Gu, "VLSI design of an SVM learning core on sequential minimal optimization algorithm," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 4, pp. 673–683, 2012.

[32] K. Polat and S. Güneş, "Breast cancer diagnosis using least square support vector machine," *Digital Signal Processing*, vol. 17, no. 4, pp. 694–701, 2007.

[33] D. François, F. Rossi, V. Wertz, and M. Verleysen, "Resampling methods for parameter-free and robust feature selection with mutual information," *Neurocomputing*, vol. 70, no. 7–9, pp. 1276–1288, 2007.

[34] N. A. Diamantidis, D. Karlis, and E. A. Giakoumakis, "Unsupervised stratification of cross-validation for accuracy estimation," *Artificial Intelligence*, vol. 116, no. 1-2, pp. 1–16, 2000.

[35] C. L. Blake and C. J. Merz, *University of California at Irvine Repository of Machine Learning Databases*, 1998, http://www.ics.uci.edu/~mlearn/MLRepository.html.