

Research Article

An Expert System Based on Fisher Score and LS-SVM for Cardiac Arrhythmia Diagnosis

Ersen Yilmaz

Electrical-Electronic Engineering Department, Uludag University, Gorukle, 16059 Bursa, Turkey

Correspondence should be addressed to Ersen Yilmaz; ersen@uludag.edu.tr

Received 13 April 2013; Accepted 4 June 2013

Academic Editor: Michele Migliore

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

An expert system having two stages is proposed for cardiac arrhythmia diagnosis. In the first stage, Fisher score is used for feature selection to reduce the feature space dimension of a data set. The second stage is classification stage in which least squares support vector machines classifier is performed by using the feature subset selected in the first stage to diagnose cardiac arrhythmia. Performance of the proposed expert system is evaluated by using an arrhythmia data set which is taken from UCI machine learning repository.

1. Introduction

Developing expert systems for medical diagnosis has received increasing attention in the literature for the last few decades. These systems are designed by using knowledge discovery in patients' data and machine learning algorithms. They have potentials to optimize medical decisions, improve medical treatments, and reduce financial costs [1].

Electrocardiogram (ECG) is graphical representation of heart's electrical activity recorded through electrodes positioned at strategic points on a body. Although it is the oldest cardiologic test, it continues as the most commonly used cardiologic test [2]. Cardiac arrhythmias are any alterations of cardiac rhythm, and they cause disruption in normal synchronized contraction sequence of heart and influence pumping efficiency. Their type and occurrence frequency make characteristic changes on ECG [3]. They are important causes of morbidity and mortality [4]. Since they can be suppressed by drugs used in treatment of arrhythmias, early recognition is important [3].

An automated system for arrhythmia analysis was first used in the early 1960s [5]. Since then, many methods have been proposed for arrhythmia diagnosis. Performances of the important group of these methods [6–10] have been evaluated on the arrhythmia data set taken from UCI machine learning repository, which is donated by Guvenir et al. [6].

In [6], an algorithm, referred to as the VFI5 (Voting Feature Intervals), was proposed for arrhythmia diagnosis. Its performance was evaluated on the arrhythmia data set by running 10-fold-cross-validation. In the same study, feature weights were learned by Genetic Algorithm (GA).

Decision Trees (DTI), Feed-Forward Neural Networks (NN), and K-Nearest Neighbors (KNN) classifiers with a variable selection algorithm based on Markov Blanket (MB), called as HITON, were applied to diagnose cardiac arrhythmia [7]. Such classifiers' performances were evaluated on the arrhythmia data set by using 10-fold-cross-validation procedure for the feature subsets determined by HITON.

Three different machine learning algorithms, namely, OneR, J48, and Naive Bayes, were used for cardiac arrhythmia diagnosis [8]. Their performances were evaluated on the arrhythmia data set by splitting it into two mutually disjoint sets as training and testing. Three different percentages of splitting were used as 50% train-50% test, 70% train-30% test, and 80% train-20% test.

In [9], an Artificial Immune Recognition System (AIRS) with Fuzzy Weighted Preprocessing (FWP) was proposed for cardiac arrhythmia diagnosis. Performance of the method was evaluated on the arrhythmia data set by using both 10-fold-cross-validation and data-splitting procedures. Like in [8], the same methodology was used for splitting the data set into training and testing sets.

A Correlation-Based Feature Selection (CBFS) algorithm and Random Forests (RF) algorithm were used together as a diagnosing strategy for cardiac arrhythmia [10]. Strategy's performance was evaluated using the arrhythmia data set with and without random sampling by running 10-fold-cross-validation.

In this work, we propose an expert system based on Fisher Score (FS) and Least Squares Support Vector Machines (LS-SVM) for cardiac arrhythmia diagnosis. Its robustness is examined by running 10-fold-cross-validation using the arrhythmia data set taken from UCI machine learning repository. Performance of the method is evaluated in terms of classification accuracy. Additionally, confusion matrix, sensitivity, and specificity rates are presented in order to analyze the system's performance in detail.

2. Fisher Score (FS)

FS [11] is one of the most widely used supervised feature selection algorithm for determining the most discriminative subset of features. It computes a score for each feature and then selects the desired number of features according to their scores. Given a data set of N records $(x_i, y_i)_{i=1}^N$ with $x_i \in \mathfrak{R}^p$ and $y_i \in \{1, 2, \dots, c\}$, where x_i is the input vector which has p features and y_i is the corresponding class label, the most discriminative subset of m features is determined in two sequential steps.

In the first step, FSs for all features are computed by using [12]

$$F(j) = \frac{\sum_{k=1}^c n_k (\mu_k^j - \mu_j)^2}{(\sigma^j)^2}, \quad j = 1, 2, \dots, p, \quad (1)$$

where n_k represents the number of records in class k , μ^j and $\sigma^j = \sqrt{\sum_{k=1}^c n_k (\sigma_k^j)^2}$ are the mean and the standard deviations of the entire data set corresponding to feature j , respectively, and μ_k^j and σ_k^j denote the mean and the standard deviations of class k corresponding to feature j , respectively. Then, in the second step, top m ranked features with high scores are selected as the most discriminative features.

3. Least Squares Support Vector Machines (LS-SVM)

Support Vector Machines (SVM) is a supervised learning algorithm based on the structural risk minimization principle of statistical learning theory [13]. SVM was first introduced to machine learning community by Boser et al. [14], and since then it has been successfully used for both regression and classification problems.

In classification problems, the objective of SVM is to separate data into two different classes with a maximum margin while minimizing empirical classification error. Detailed information about SVM can be found in [13–15].

The major drawback of SVM is its higher computational load arising from the need to solve the constrained quadratic programming problem. This drawback is overcome by LS-SVM proposed by Suykens and Vandewalle [16], which solves

a set of linear equations instead of the quadratic programming problem.

Given a training data set of N records $(x_i, y_i)_{i=1}^N$ with $x_i \in \mathfrak{R}^p$ and $y_i \in \pm 1$, where x_i is p dimensional input vector and y_i is the corresponding class label, LS-SVM requires minimization of the following optimization problem:

$$\min_{w, b, e} J_{LS}(w, e) = \frac{1}{2} \|w\|^2 + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \quad (2)$$

subject to $y_i(w^T \varphi(x_i) + b) = 1 - e_i$, $i = 1 \dots N$, where w is a parameter vector, b is a bias term, e_i is prediction error for the record i , γ is a regularization parameter, and $\varphi(x_i)$ is a nonlinear mapping function of the records from input space to higher dimensional feature space. The corresponding Lagrangian function for (2) is defined as follows:

$$L_{LS}(w, b, e; \alpha) = J_{LS}(w, e) - \sum_{i=1}^N \alpha_i \{y_i [w^T \varphi(x_i) + b] - 1 + e_i\}, \quad (3)$$

where α_i 's are Lagrange multipliers.

According to optimality conditions $\partial L_{LS} / \partial \alpha_i = 0$, $\partial L_{LS} / \partial w = 0$, $\partial L_{LS} / \partial b = 0$, and, $\partial L_{LS} / \partial e_i = 0$, we can get $y_i [w^T \varphi(x_i) + b] - 1 + e_i = 0$, $w = \sum_{i=1}^N \alpha_i y_i \varphi(x_i)$, $\sum_{i=1}^N \alpha_i y_i = 0$, and $\alpha_i = \gamma e_i$, for $i = 1 \dots N$.

Defining $Y = [y_1; \dots; y_N]$, $\vec{1} = [1; \dots; 1]$, $\alpha = [\alpha_1; \dots; \alpha_N]$, $e = [e_1; \dots; e_N]$, and $Z = [\varphi(x_1)^T y_1; \dots; \varphi(x_N)^T y_N]$ and after eliminating w and e , the following set of linear equations is obtained [16]:

$$\left[\begin{array}{c|c} 0 & -Y^T \\ \hline Y & ZZ^T + \gamma^{-1} I \end{array} \right] \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{1} \end{bmatrix}. \quad (4)$$

Mercer's condition can be applied to the matrix $\Omega = ZZ^T$

$$\Omega_{i,j} = y_i y_j \varphi(x_i)^T \varphi(x_j) = y_i y_j K(x_i, x_j), \quad i, j = 1, \dots, N, \quad (5)$$

where $K(x_i, x_j)$ is a kernel function representing product of two vectors in feature space, that is, $\varphi^T(x_i) \varphi(x_j)$.

LS-SVM classifier is expressed as in (6) and found by solving the set of linear equations given in (4) as follows:

$$f(x) = \text{sign} \left(\sum_{i=1}^N y_i \alpha_i K(x, x_i) + b \right). \quad (6)$$

4. Arrhythmia Data

The arrhythmia data set used in this study was taken from UCI machine learning repository [6]. This data set has 452 ECG records described by 279 features, 206 features are linear, and the remaining 73 features are nominal. The first four features are age, sex, height, and weight, and the rest are features extracted from patients' ECG signals. Each record

TABLE 1: Confusion matrix.

| Predicted | Actual | |
|-----------|---------------------|---------------------|
| | Positive | Negative |
| Positive | TP (true positive) | FP (false positive) |
| Negative | FN (false negative) | TN (true negative) |

TP represents an instance, which is actually positive and predicted by the model as positive.

FN represents an instance, which is actually positive but predicted by the model as negative.

TN represents an instance, which is actually negative and predicted by the model as negative.

FP represents an instance, which is actually negative but predicted by the model as positive.

belongs to one of the 16 classes. The class 01 refers to normal ECG, the classes 02–15 refer to different arrhythmia types, and the class 16 refers to unclassified records. There are 22 unclassified records in the arrhythmia data set. Additionally, about 0.33% of the feature values in the data set are missing. Detailed description of the data set can be found on <http://archive.ics.uci.edu> (last accessed: April, 2013).

5. Performance Evaluation

The proposed expert system's performance is examined by running 10-fold-cross-validation. Four different measures, which are classification accuracy, confusion matrix, sensitivity, and specificity, are used for performance evaluation. Cross-validation procedure and these four measures are explained in the following subsections.

5.1. Cross-Validation (CV). CV is a widely used statistical method to evaluate classifiers' performances by splitting a data set into two sets as training and testing. In CV, the training and the testing sets must cross over in successive rounds, and in this way each record has a chance of being validated against [17]. For 10-fold-cross-validation, the data set is divided into 10 equal sized folds, and 10 iterations are performed. In each iteration step, one of the 10-fold is used for testing, and the remaining ninefold are used for training. In this way, at the end of the ten iteration steps, each record in the data set is used once for testing purpose.

5.2. Classification Accuracy. Classification accuracy is the most commonly used measure for determining performance of classifiers. It is the rate of number of correct predictions made by a model over a data set [18].

5.3. Confusion Matrix. Confusion matrix shows predicted and actual classifications. A confusion matrix for a classification problem with two classes is of size 2×2 , and it is given in Table 1 [18].

5.4. Sensitivity and Specificity. Sensitivity is the true positive rate, and specificity is the true negative rate [18]. They are defined as in (7) and (8), respectively,

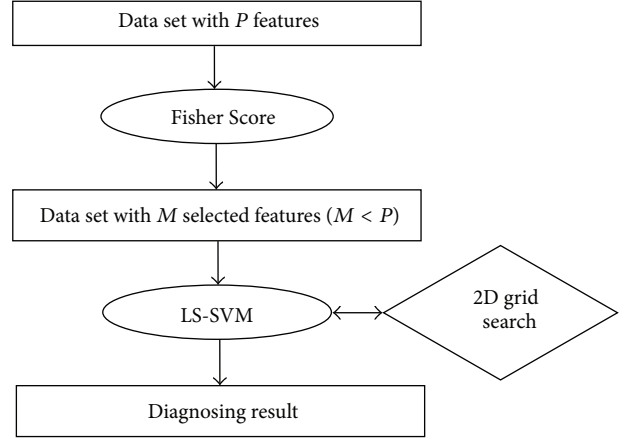


FIGURE 1: The proposed expert system's architecture.

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \quad (7)$$

$$\text{Specificity} = \frac{TN}{FP + TN}. \quad (8)$$

6. The Proposed Expert System

The proposed expert system for arrhythmia diagnosis is described in this section. The system has two stages. Its architecture is given in Figure 1.

In the first stage, the feature selection algorithm FS is used to reduce the feature space dimension of the arrhythmia data set, and different sets of features are obtained. Then, in the second stage, LS-SVM classifier is performed on these feature subsets, meanwhile parameters of the classifier are optimized by using two-dimensional (2D) grid search. According to the performance results of different feature subsets, the most discriminative feature subset with the best classifier parameters are chosen, and the optimal model for expert system is created.

LS-SVM has a Gaussian kernel function given by

$$K(x, x_i) = \exp\left(-\frac{1}{2\sigma^2}(x - x_i)^2\right). \quad (9)$$

Parameters of LS-SVM, which are penalty factor γ and kernel width σ^2 , are optimized by using 2D grid search.

7. Experiments

In our experiments, unclassified records (class 16) are excluded from the data set, and the rest of the records (430 records) are grouped into two categories as presence (class 02–15) or absence (class 01) of arrhythmia. The nearest neighbour method is used to impute missing values.

The nearest neighbour is one of the most popular nonparametric missing value estimation methods. Its main advantage is its simplicity. The method uses different distance metrics to determine the similarity between the target and the reference records [19]. In this study, distances are computed

TABLE 2: Classification accuracies for tenfold.

| Fold-1 | Fold-2 | Fold-3 | Fold-4 | Fold-5 | Fold-6 | Fold-7 | Fold-8 | Fold-9 | Fold-10 |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|
| %90.69 | %76.74 | %72.09 | %81.39 | %86.05 | %90.70 | %81.40 | %79.07 | %88.37 | %74.42 |

TABLE 3: Confusion matrix of the proposed expert system.

| Predicted | Actual | |
|------------|------------|--------|
| | Arrhythmia | Normal |
| Arrhythmia | 157 | 49 |
| Normal | 28 | 196 |
| Total | 185 | 245 |

by using well-known Euclidean distance. The distance from record i to record j is given by

$$d(x_i, x_j) = \sqrt{(x_i - x_j)^T (x_i - x_j)}, \quad i, j = 1, 2, \dots, N. \quad (10)$$

Missing feature values in the records are filled with the corresponding feature values of the nearest neighbour record which is selected by using (10) as the most similar record from the same class.

In the proposed expert system, LS-SVM classifier is trained nonincrementally. Therefore, the system is order independent [20, 21]. Performance evaluation of the system is made by running 10-fold-cross-validation. LS-SVM parameters (γ, σ^2) are selected by using 2D grid search on the intervals [0.001, 1000]. In the experiments, 55 different feature subsets and 400 different LS-SVM parameters sets for each feature subset are tested.

The experiments are performed in the following sequential steps.

Step 1. Different subsets of the features are obtained by the feature selection method FS.

Step 2. The arrhythmia data set is randomly split into 10-fold of almost equal size while maintaining the class distributions in each fold roughly the same as those in the data set.

Step 3. One of the feature subsets obtained in Step 1 is fed into LS-SVM.

Step 4. LS-SVM parameters are set to initial values, which are the first values of the selected intervals for the parameters.

Step 5. LS-SVM with the determined parameter values is performed by running 10-fold-cross-validation.

Step 6. Classification accuracies of tenfold and overall classification accuracy of these tenfold are obtained.

Step 7. If all values in the intervals are fed into LS-SVM, then LS-SVM parameter values with the highest overall classification accuracy are recorded for the relevant feature subset; go to Step 9, otherwise go to Step 8.

Step 8. New values of LS-SVM parameters are determined by 2D grid search on the intervals; go back to Step 5.

Step 9. If all feature subsets are fed into LS-SVM, then go to Step 11, otherwise go to Step 10.

Step 10. A new feature subset is fed into LS-SVM, and go back to Step 4.

Step 11. The feature subset with the highest overall classification accuracy is chosen as the best discriminative subset, and the relevant parameter values are used for optimum values for LS-SVM classifier.

8. Discussions

In our experiments, the highest overall classification accuracy is achieved when size of the feature subset is 65, and LS-SVM parameters are as follows: $\gamma = 0.1$ and $\sigma^2 = 5$.

10-fold-cross-validation results of the proposed expert system with this feature subset and LS-SVM parameters are summarized in Tables 2, 3, and 4. Classification accuracies for tenfold are obtained as in Table 2.

Overall classification accuracy of the proposed system is computed by averaging the classification accuracies of tenfold, which is 82.09%.

In order to analyze the proposed expert system's performance in detail, a confusion matrix is built and sensitivity and specificity measures are computed. Table 3 shows the confusion matrix of the proposed expert system.

Sensitivity and specificity rates of the proposed expert system are obtained as 84.86% and 80.00%, respectively.

In order to make a comparison, classification accuracies of the studies in the literature and our proposed expert system are given in Table 4. Performances of all methods given in Table 4 were evaluated on the same cardiac arrhythmia data set taken from UCI machine learning repository.

It can be seen from the comparison table that the proposed expert system achieves a remarkable classification accuracy rate of 82.09% and it is superior to other methods except RF-CBFS with random sampling strategy. Note that the classification accuracy rate of 90% for the RF-CBFS method was achieved by randomly sampling the data set so that the class distributions were changed in the training stage.

9. Conclusions

In this work, an expert system based on FS and LS-SVM is proposed for cardiac arrhythmia diagnosis. A Gaussian radial basis function is used as a kernel of LS-SVM, and the parameters of LS-SVM are optimized by using 2D grid search. The proposed system's performance is evaluated using a real data set with respect to classification accuracy with 10-fold-cross-validation. Additionally, confusion matrix, sensitivity,

TABLE 4: Comparison of proposed expert system with the studies in the literature.

| Method | Performance criteria | Maximum classification accuracy |
|------------------------------------|--|---------------------------------|
| VFI5-GA [6] | 10-fold-cross-validation | 68% |
| DTI-HITON | | 71.87% |
| NN-HITON [7] | 10-fold-cross-validation | 60.38% |
| KNN-HITON | | 65.30% |
| OneR | | 58.09% |
| J48 [8] | 70% train-%30 test | 74.26% |
| Naïve Bayes | | 75.00% |
| AIRS-FWP [9] | 10-fold-cross-validation 80% train-%20 test | 76.20% 80.71% |
| RF-CBFS [10] | 10-fold-cross-validation | 76.30% |
| RF-CBFS with random sampling | | 90.00% |
| Proposed expert system (FS-LS-SVM) | 10-fold-cross-validation | 82.09% |

and specificity rates are presented for further analysis of the system.

The experiments on the arrhythmia data set show that 65 features are sufficient for the proposed expert system to perform significantly well in distinguishing among normal and arrhythmia ones, and the system achieves a remarkable classification accuracy rate of 82.09%. The sensitivity and the specificity rates are obtained as 84.86% and % 80.00, respectively. According to empirical results, it is concluded that the proposed expert systems can help clinicians make better diagnosis of cardiac arrhythmia.

Acknowledgments

The authors would like to thank UCI Repository of Machine Learning Databases for being a valuable resource: Bache, K. & Lichman, M. (2013); UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/>); Irvine, CA, USA: University of California, School of Information and Computer Science.

References

- [1] A. G. Floares, "Using computational intelligence to develop intelligent clinical decision support systems," in *Computational Intelligence Methods for Bioinformatics and Biostatistics*, vol. 6160 of *Lecture Notes in Computer Science*, pp. 266–275, 2010.
- [2] M. G. Khan, *Rapid ECG Interpretation*, Humana Press, Totowa, NJ, USA, 2003.
- [3] D. A. Coast, R. M. Stern, G. G. Cano, and S. A. Briller, "An approach to cardiac arrhythmia analysis using hidden Markov models," *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 9, pp. 826–836, 1990.
- [4] M. T. Keating and M. C. Sanguinetti, "Molecular and cellular mechanisms of cardiac arrhythmias," *Cell*, vol. 104, no. 4, pp. 569–580, 2001.
- [5] C. L. Feldman, "Computer detection of cardiac arrhythmias: historical review," *American Review of Diagnostics*, vol. 2, no. 5, pp. 138–145, 1983.
- [6] H. A. Guvenir, B. Acar, G. Demiroz, and A. Cekin, "Supervised machine learning algorithm for arrhythmia analysis," in *Proceedings of the 24th Annual Meeting on Computers in Cardiology*, pp. 433–436, Lund, Sweden, September 1997.
- [7] C. F. Aliferis, I. Tsamardinos, and A. Statnikov, "HITON: a novel Markov blanket algorithm for optimal variable selection," in *Proceedings of the AMIA Annual Symposium*, pp. 21–25, Washington, DC, USA, 2003.
- [8] T. Soman and P. O. Bobbie, "Classification of arrhythmia using machine learning techniques," *WSEAS Transactions on Computers*, vol. 4, no. 6, pp. 548–552, 2005.
- [9] K. Polat, S. Şahan, and S. Güneş, "A new method to medical diagnosis: artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia," *Expert Systems with Applications*, vol. 31, no. 2, pp. 264–269, 2006.
- [10] A. Özçift, "Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis," *Computers in Biology and Medicine*, vol. 41, no. 5, pp. 265–271, 2011.
- [11] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [12] Q. Gu, Z. Li, and J. Han, "Generalized fisher score for feature selection," in *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence (UAI '11)*, pp. 266–273, Barcelona, Spain, July 2011.
- [13] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
- [14] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "Training algorithm for optimal margin classifiers," in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pp. 144–152, Pittsburgh, Pa, USA, July 1992.
- [15] Y. Lin, "Support vector machines and the Bayes rule in classification," *Data Mining and Knowledge Discovery*, vol. 6, no. 3, pp. 259–275, 2002.
- [16] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [17] P. Refaailzadeh, L. Tang, and H. Liu, "Cross-validation," in *Encyclopedia of Data Base Systems*, L. Liu and M. T. Özsu, Eds., pp. 532–538, Springer, 2009.

- [18] R. Kohavi and F. Provost, "Glossary of terms," *Machine Learning*, vol. 30, pp. 271–274, 1998.
- [19] B. N. I. Eskelson, H. Temesgen, V. Lemay, T. M. Barrett, N. L. Crookston, and A. T. Hudak, "The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases," *Scandinavian Journal of Forest Research*, vol. 24, no. 3, pp. 235–246, 2009.
- [20] S. T. Velert, *Incremental learning approaches to biomedical decision problems [Ph.D. thesis]*, Polytechnic University of Valencia, 2012.
- [21] S. Rüping, "Incremental learning with support vector machines," in *Proceedings of the 1st IEEE International Conference on Data Mining (ICDM '01)*, pp. 641–642, San Jose, Calif, USA, December 2001.



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

