Multi-Class SVM Classification Combined with Kernel PCA Feature Extraction of ECG Signals

Maya Kallas*†, Clovis Francis*, Lara Kanaan‡, Dalia Merheb‡, Paul Honeine†, Hassan Amoud§

*Laboratoire d'Analyse et de Surveillance des Systèmes (LASYS), Faculty of Engineering, Lebanese University, Lebanon
†Institut Charles Delaunay (UMR CNRS 6279), Université de Technologie de Troyes, France
‡Department of Telecommunications, Faculty of Engineering, Holy Spirit University of Kaslik, Lebanon
‡Azm Center for Research in Biotechnology and its Applications, Lebanese University, Lebanon

***Abstract*—The cardiovascular diseases are one of the main causes of death around the world. Automatic detection and classification of electrocardiogram (ECG) signals are important for diagnosis of cardiac irregularities. This paper proposes to apply the Support Vector Machines (SVM) classification, to diagnose heartbeat abnormalities, after performing feature extraction on the ECG signals. The experiments were conducted on the ECG signals from the MIT-BIH arrhythmia database [1] to classify two different abnormalities and normal beats. Kernel Principal Component Analysis (KPCA) is used for feature extraction since it performes better than PCA on ECG signals due to their nonlinear structures. This is demonstrated in a previous work [2]. Two multi-SVM classification schemes are used, One-Against-One (OAO) and One-Against-All (OAA), to classify the ECG signals into different disease categories. The experiments conducted show that SVM combined with KPCA performs better than that without feature extraction. Moreover, our results show a better performance in Gaussian KPCA feature extraction with respect to other kernels. Furthermore,the performance of Gaussian OAA-SVM combined with KPCA has higher average accuracy than Gaussian OAA-SVM in ECG classification.**

***Index Terms*—kernel machines, ECG signals, Kernel Principal Component Analysis, Support Vector Machines, Multi-class classification**

I. INTRODUCTION

With the recent developments in technology, physicians have powerful tools to observe the operation of the heart muscle and thus to establish their diagnosis. Among cardiovascular examinations, electrocardiogram (ECG) analysis is the most commonly practiced. This is due to the fact that ECG presents useful information about the rhythm and the electrical activity of the heart. Thus, it is widely used for the diagnosis of cardiac arrhythmias. For effective diagnostics, the study of the ECG signal must be carried out for several hours. For this reason, numerous researchers have been interested in enabling computers to classify the abnormal ECG signals.

Several algorithms have been proposed in order to detect and classify the cardiac diseases observed in the ECG signals. Initially introduced for two-class classification based on the support vector machines (SVM) with Vapnik [3], some improvements are later introduced such as robustness and low computational cost [4]. Nevertheless, real-life problems require more than two-class. An extension can be done at different levels: on the one hand during the learning process [5] for feature space and hyperplane computation, and on the other hand during the decision process [6] based on classical results. More precisely for our case, the ECG signals, in [7], a neuro-fuzzy approach of the ECG-based classification of heart rhythms is described. In [8], two classification systems based on the support vector machines (SVM) approach are implemented. The first exploits features based on high-order statistics, while the second uses the coefficients of Hermite polynomials. In [9], a new approach for feature selection and classification of cardiac arrhythmias based on Adaptive Feature Selection and SVM is proposed.

A technique used to extract feature is the Principal Component Analysis (PCA) [10]. The main idea of this method is to obtain some uncorrelated variables from a large number of correlated variables. However, this technique reveals only linear structures in a given dataset. An extension of this method to the nonlinear systems is the kernel PCA. The concept of this technique is to map the data into a high dimensional feature space, where PCA is applied on the mapped data. In the case of KPCA, the dimension of the feature space is much larger than the input space leading to a pre-image problem [11]. Solving this issue is out of the scope of this paper and treated in details in other papers [12], [13]. See [14] for a recent review.

In this paper, we propose to classify some ECG signals into different classes: the normal case, and two abnormal cases, Premature Ventricular Contraction (PVC) and Left Bundle Branch Block (LBBB). To this end, two different multi-class SVM classification schemes are used: the One-Against-One and the One-Against-All with a Gaussian kernel. On one hand, they are applied directly on real ECG signals, and on the other hand, they are used with kernel principal component analysis (KPCA) which is used for feature extraction in some high dimensional feature space.

The remainder of this paper is organized as follows: Section II outlines the basic concepts involving KPCA, which is used to extract the structures that are non-linearly related to the input space by solving an eigenvalue problem in a high dimensional feature space. Section III describes the principle of Support Vector Machine (SVM) for solving multi-class arrhythmia classification problems. Section IV presents the experimental results that evaluates the adopted techniques performance. By the end, a final conclusion and recommendations

978-1-4673-0747-5/12/$31.00 ©2012 IEEE

for future work are presented.

II. Kernel PCA

Principal component analysis (PCA) is a mathematical technique whose purpose is to transform a number of correlated variables into a number of uncorrelated variables called "*principal components*" (PC). These PC accounts for the maximum variance of the data set. The redundancy of the original variables means that they are measuring the same concept [10]. The principal components are computed as a linear transformation of the data set and their weights are chosen so they will be commonly uncorrelated. Each PC contains new information of the data set and is ordered so that the first few components account for most of the variability [10].

The conventional PCA detects only linear structures in a given dataset. A more generalized technique has been introduced to learn the nonlinearities using kernels, the so-called kernel PCA (KPCA). The KPCA can reveal nonlinear kernel principal components that are more appropriate to complex and nonlinear data such as face images, handwritten digits and natural signals. The basic idea of KPCA is to map the original data into a high dimensional space via a specific function and then to apply the standard PCA algorithm on it. The linear PCA in the high dimensional feature space corresponds to a nonlinear PCA in the original input space and can find the most interesting direction [10]. Suppose we have L observations of dimension N each. The PCA can find up to N non-zero eigenvalues contrary to KPCA that can extract up to L non-zero eigenvalues. However, for both of them, the first few principal components account for the maximum of the variance, and the mean squared error for the reconstruction patterns calculated by these first few PCs is minimal. A main advantage is that the principal components are uncorrelated. Then, one of the main advantages of KPCA is that by choosing a specific kernel function we can have a previous idea about the type of the nonlinear components before extracting them [15]. As said before one of the main advantages of standard PCA towards KPCA is that we can reconstruct easily the original data by simply using the principal components.

III. Multi-class SVM

Support Vector Machines were introduced with the combination of kernel-based feature space and hyperplane-classifier. Initially introduced, they were used to solve two-class learning problems. Many SVM algorithms were introduced. The main tasks are the selection of suitable kernel to represent data in the feature space and the computation of separating hyperplanes in order to minimize the empirical risk of miss-classification. Nonetheless, real-world classification tasks require more than two-class. While, large margin separators have been developed to binary problems, they can also be adapted to treat multi-class problems.

The idea is simply to transform the problem into k classes of k binary classifiers. There are two kinds of multi-class SVM system, One-Against-All (OAA) and One-Against-One (OAO). In our previous work [2] we emphasized the case of Binary SVM in conjunction with PCA and KPCA feature extraction. Our results show a better classification accuracy when using feature extraction by means of the Gaussian KPCA. For this reason we adopt in this paper this kernel for feature extraction before Multi-Class SVM classification.

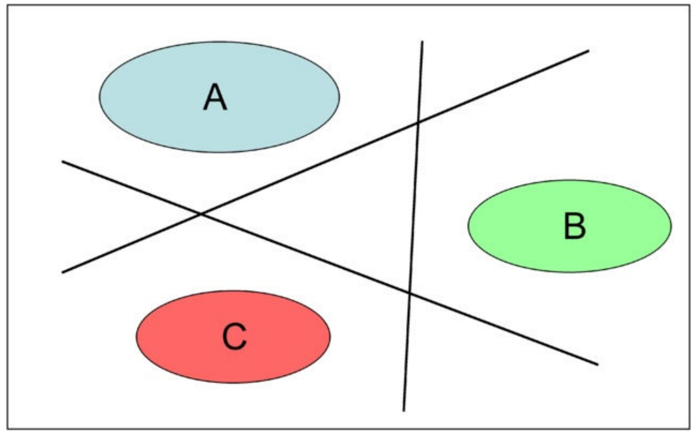

Fig. 1. Diagram of binary OAA region boundaries on a basic problem.

A. One-Against-All (OAA) SVM

The approach "One-Against-All" is the simplest and oldest method of decomposition, it was introduced by Vapnik in 1995[16]. Consider a c-class problem, where we have L training samples, the input is a set $T_{\boldsymbol{XY}} = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_L, y_L)\}$ of training vectors $\boldsymbol{x}_i \in \boldsymbol{X} \subseteq \mathbb{R}^N$ and corresponding hidden states $y_i \in \boldsymbol{Y} = \{1, 2, \ldots, c\}$ which are the label classes.

The initial formulation of the one-against-all method states that a data point would be classified under a certain class if and only if that class's SVM accepted it and all other classes' SVMs rejected it. While accurate for tightly clustered classes, this method leaves regions of the feature space undecided where more than one class accepts or all classes reject. Figure 1 illustrates this formulation.

An improvement of the performance of OAA SVMs was suggested by Vapnik in 1998[17]. The simplest solution to solve a multi-class SVM is to break it down into a set of binary sub-problems and build SVM independently for each of them. This strategy, called "One-Against-All" consists on building SVMs equal to the number of classes. Each SVM is then trained to separate the data of a class that will be labeled 1, from those of all other classes that are labeled -1. Thus, each SVM is associated with a class and its output before thresholding belongs to the class. The decision rule is therefore generally used to allocate the unknown data to the class corresponding to the SVM with the largest output value [18], [19], [20]. This is demonstrated in figure 2.

B. One-Against-One (OAO) SVM:

Another decomposition method is "One-Against-One", also known as "pairwise coupling", "all pairs" or "round robin"

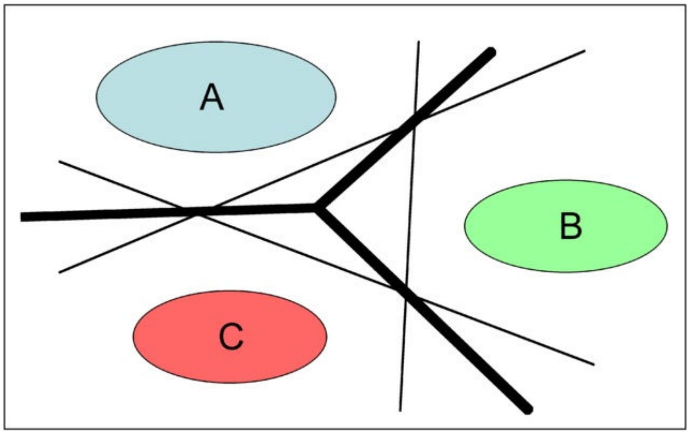

Fig. 2. Diagram of continuous OAA region boundaries on a basic problem.

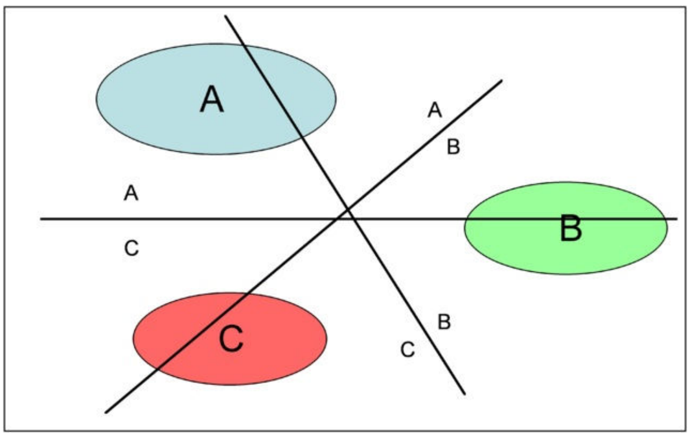

Fig. 3. Diagram of OAO-SVM decision boundaries on a basic problem.

[20]. This method consists on building an SVM for each pair of classes, or $c(c-1)/2$ SVM for a problem with c classes. Each classifier is trained to separate the data from one class to those of another class. The decision rule used is generally the majority voting called "max-wins voting". Each SVM votes for a class and the unknown data is finally associated with the class receiving most of the votes [18], [19]. More advanced methods include using decision graphs to determine the class selected in a similar manner to knockout tournaments[21]. This concept is showed in figure 3.

IV. Results

We applied the aforementioned multi-class SVM classifiers on ECG signals, therefore, the signal itself is divided into k classes of k binary classifiers. In our study, we considered three different classes: a class for the normal case and two classes for two different abnormal cases, Premature Ventricular Contraction (PVC) and Left Bundle Branch Block (LBBB). A PVC is an extrasystole involving the ventricles of the heart, sometimes producing accompanying palpitations, while a LBBB is the failure of the cardiac impulse to propagate down the left bundle branch, resulting in early activation of the right side of the septum and the right ventricular myocardium.

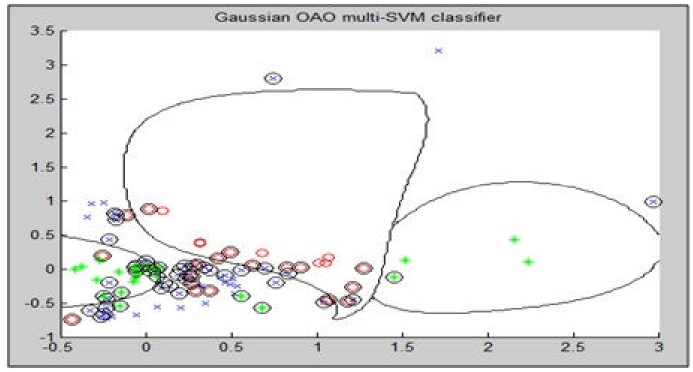

Fig. 4. Gaussian OAO multi-SVM classifier.

Our target is to optimize the classifier performance and to reach a higher percentage in the accuracy. The inputs of the cross validation algorithm are the number of folds, the training data, the kernel parameters σ and the regularization constant $\boldsymbol{C}$. Each pair of kernel arguments and constant $\boldsymbol{C}$ forms a set called model. In order to attempt this result, we tested various parameters values for the SVM kernel and we evaluated them to search for the best one by applying 5-fold cross validation. In [2], when using a two-class SVM, we had proven that the SVM classifier based on the Gaussian kernel presents better results than with other kernels when dealing with original data mapped into a high dimensional feature space, taken into consideration the accuracy defined as the ratio of corrected classified beats over the total number of beats, the sensitivity identified by the ratio of people that the test considered sick among all that are truly sick, the specificity described by the ratio of people that the test considered healthy among all that are truly healthy and the positive predictivity characterized by the percent of people with a positive test who have the disease. Thus, in the multi-SVM classification, we considered also the Gaussian kernel. After applying the 5-fold cross validation method in the Gaussian kernel function, we obtained that for $\boldsymbol{C}$ equals 100 and σ equals 1 the accuracy is maximum. For testing purposes, we used 10 normal signals, 10 PVC signals and 14 LBBB signals. We used Multi SVM classification of ECG signals without feature extraction and after feature extraction.

Figures 4 and 5 show the Gaussian OAO and OAA multi-SVM classifiers with the models above. Figure 6 represents the percentage of correctly classified ECG signals in each class with and without feature extraction. The normal heartbeats can be entirely separated from other abnormal heartbeats with a 100% accuracy, the PVC heartbeats have a 70% accuracy in both cases OAO-SVM or OAA-SVM, and the LBBB heartbeats have an 85.71% accuracy in OAO-SVM and 92.85% in OAA-SVM case.

A. Multi-SVM Classification after feature extraction

The inputs of the multi-SVM classifier are the Kernel Principal Components. After applying the 5-fold cross validation

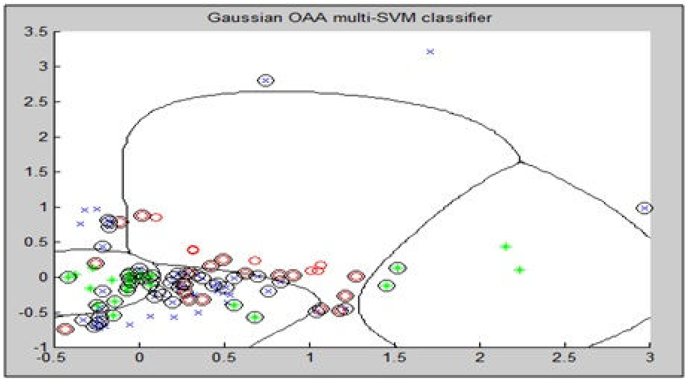

Fig. 5. Gaussian OAA multi-SVM classifier.

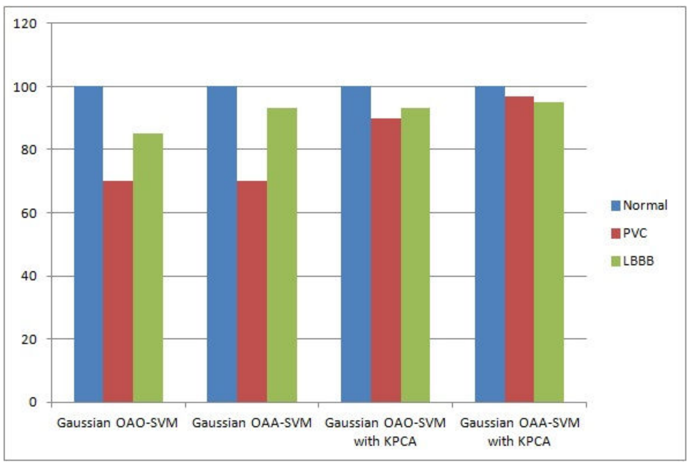

Fig. 6. Performance of the Gaussian OAO-SVM, Gaussian OAA-SVM, Gaussian OAO-SVM combined with KPCA, and Gaussian OAA-SVM combined with KPCA for the cases of normal, PVC and LBBB classes.

method in the Gaussian kernel function, we obtained that for $\boldsymbol{C}$ equals 100 and σ equals 0.1 the accuracy is maximum.

The NORMAL heartbeats have 100% accuracy in both cases OAO-SVM or OAA-SVM while combined with KPCA, the PVC heartbeats have a 90% accuracy in OAO-SVM and 97.34% in OAA-SVM, and the LBBB heartbeats have a 92.65% accuracy in OAO-SVM and 94.85% in OAA-SVM case.

These results illustrate the superiority of KPCA OAA-SVM when dealing with Multi-SVM classification.

V. Conclusion

Electrocardiogram (ECG) supervising is the most important and efficient way for preventing heart attacks. Our work presented an integrated classification method, which combines the Support Vector Machine (SVM) with the Kernel Principal Component Analysis (KPCA) for classification of different types of cardiac abnormalities. The classification experiments of two types of arrhythmias and normal beats were performed on the MIT-BIH Arrhythmia Database. The Multi SVM classifiers were trained, cross-validated and tested on the original beats before and after feature extraction.

Our results showed that Multi-SVM is applied to classify the signals into three classes: Normal, Premature Ventricular Contraction (PVC) and Left Bundle Branch Block (LBBB). Multi-SVM parameters were also chosen according to a 5-fold-cross validation method. Experimentation confirmed a very high average classification accuracy of 97.39% for the Gaussian multiple-classifier (OAA) combined with KPCA in the classification of three types of beats which is higher than the accuracy obtained with the OAO classifier.

In our future work, we intend to enlarge the number of classes to cover other cardiac arrhythmias.

Acknowledgment

This work is partly supported by the Lebanese University and the French-Lebanese research program CEDRE No. 10 SCI F15/L5.

References

[1] R. Mark and G. Moody, “MIT-BIH arrythmia database,” 2007. [Online]. Available: http://ecg.mit.edu/dbinfo.html

[2] L. Kanaan, D. Merheb, M. Kallas, C. Francis, H. Amoud, and P. Honeine, “PCA and KPCA of ECG signals with binary SVM classification,” in *IEEE Workshop on Signal Processing Systems SiPS’2011*, Beirut, Lebanon, 4-7 Octobre 2011.

[3] B. E. Boser, L. M. Guyon, and V. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the 5th annual ACM workshop on computational learning theory*. ACM Press, 1992, pp. 144–152.

[4] P. Bartlett and J. Shawe-Taylor, *Generalization performance of support vector machines and other pattern classifiers*. Cambridge, MA, USA: MIT Press, 1999, pp. 43–54.

[5] G. Rätsch, A. J. Smola, and S. Mika, “Adapting codes and embeddings for polychotomies,” in *NIPS*, 2002, pp. 513–520.

[6] E. L. Allwein, R. E. Schapire, and Y. Singer, “Reducing multiclass to binary: a unifying approach for margin classifiers,” *J. Mach. Learn. Res.*, vol. 1, pp. 113–141, September 2001.

[7] T. Linh, S. Osowski, and M. Stodolski, “On-line heart beat recognition using hermite polynomials and neuro-fuzzy network,” *IEEE Transactions On Instrumentation And Measurement*, vol. 52, no. 4, pp. 1224–1231, August 2003.

[8] S. Osowski, T. Linh, and T. Markiewicz, “Support vector machine-based expert system for reliable heartbeat recognition,” *IEEE Transactions On Biomedical Engineering*, vol. 51, no. 4, pp. 582–589, Avril 2004.

[9] W. Kao, C. Yu, C. Shen, W. Chen, and P. Hsiao, “Electrocardiogram analysis with adaptive feature selection and support vector machines,” in *IEEE Asia Pacific Conference on Circuits and Systems, APCCAS*, Singapore, 4-7 December 2006, pp. 1783–1786.

[10] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, Oct. 2002.

[11] F. Castells, P. Laguna, L. Sörnmo, A. Bollmann, and J. Roig, “Principal component analysis in ECG signal processing,” *EURASIP J. Appl. Signal Process.*, vol. 2007, pp. 98–98, January 2007.

[12] H. Chou, Y. Chen, Y. Shiau, and T. Kuo, “A high performance compression algorithm for ECG with irregular periods,” in *Proc. IEEE International Workshop on Biomedical Circuits and Systems*, 2004.

[13] L. I. Smith, “A tutorial on principal components analysis introduction,” *Statistics*, vol. 51, no. 1, p. 52, 2002.

[14] P. Honeine and C. Richard, “Pre-image problem in kernel-based machine learning,” *IEEE Signal Processing Magazine, Special Issue On Dimensionality Reduction Via Subspace and Manifold Learning*, vol. 28 (2), March 2011.

[15] D. Patra, M. K. Das, and S. Pradhan, “Integration of FCM, PCA and neural networks for classification of ECG arrhythmias,” *IAENG International Journal of Computer Science*, February 2010.

[16] V. N. Vapnik, *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995.

[17] ——, *Statistical Learning Theory*. Wiley-Interscience, September 1998.

[18] A. Pronobis and B. Caputo, “Confidence-based cue integration for visual place recognition,” in *in Proc. IROS07 Centre for Autonomous Systems Royal Institute of Technology*, SE-100 44 Stockholm, Sweden, 2007.

[19] J. Fu, C. Huang, and S. Lee, “A multi-class svm classification system based on methods of self-learning and error filtering,” Taiwan, Republic of China, 2008.

[20] J. Milgram, M. Cheriet, and R. Sabourin, “One Against One or One Against All: Which One is Better for Handwriting Recognition with SVMs?” in *Tenth International Workshop on Frontiers in Handwriting Recognition*, G. Lorette, Ed., Université de Rennes 1. La Baule (France): Suvisoft, October 2006.

[21] C. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, June 1998.