USING SUPPORT VECTOR MACHINES (SVMS) WITH REJECT OPTION FOR HEARTBEAT CLASSIFICATION

Zahia Zidelmal, Ahmed Amirou
Mouloud Mammeri University (UMMTO), Tizi-Ouzou, Algeria

Adel Belouchrani
Electrical Engineering Department, Ecole Nationale Polytechnique, Algiers, Algeria

Keywords: Premature Ventricular Contraction (PVC), Beat Classification, Support Vector Machines, Reject option.

Abstract: In this paper, we introduce a new system for ECG beat classification using Support Vector Machines (SVMs) classifier with a double hinge loss. This classifier has the option to reject samples that cannot be classified with enough confidence. Specifically in medical diagnoses, the risk of a wrong classification is so high that it is convenient to reject the sample. After ECG preprocessing, feature selection and extraction, our decision rule uses dynamic reject thresholds following the cost of rejecting a sample and the cost of misclassifying a sample. Significant performance enhancement is observed when the proposed approach was tested with the MIT/BIH arrythmia database. The achieved results are represented by the error reject tradeoff and a sensitivity higher than 99%, being competitive to other published studies.

1 INTRODUCTION

Premature Ventricular Contraction (PVC) is an ectopic contraction caused by ventricular cells erroneously acting as a pacemaker. PVCs are characterized by the premature occurrence of bizarre shaped QRS complexes (see figure 1). Counting the occurrence of ectopic beats is of particular interest to support the detection of ventricular tachycardia and to evaluate the regularity of the depolarization of the ventricles.

As a large amounts of data are often analysed and stored when examining cardiac signals, computers can be used to automate signal processing. Accordingly, several algorithms have been proposed for the detection and classification of heartbeats together with signal processing techniques.

Classical techniques extract heuristic ECG descriptors, such as the QRS morphology (Chazal et al., 2004) and interbeat R-R intervals (Chazal et al., 2004), (Jecova et al., 2004). Other ECG descriptors rely on QRS frequency components calculated either by Fourier transform (Minami et al., 1999) or by computationally efficient algorithms with filter banks (Afonso and Tompkins, 1999). Some methods apply QRS template matching procedures, based on different transforms, e.g., Karhunen-Loeve transform (Gomez-Herrero et al., 2006), Hermite functions (Lagerholm et al., 2000) to approximate the variety of temporal and frequency characteristics of the QRS complex waveforms. Other techniques for computerized arrhythmia detection employ cross-correlation with predefined ECG templates (Krasteva and Jecova, 2007)

Several discriminative techniques such as artificial neural networks have been developed and used to exploit their natural ability in pattern-recognition tasks for successful classification of ECG beat (Yeap, 1990). The latter include linear back-propagation network discriminants, Self-organizing maps with learning vector quantization (Hu et al., 1997) where Hu et al customized a heartbeat classifier to a specific patient (local classifier) and then combined it with a global classifier using a mixture of experts approach (MOE). Among all these methods, Support Vector Machines (SVMs) have enjoyed a strong success in this application field (Osowski et al., 2004) where Osowski et al. combined multiple classifiers by the weighted voting principle. but leads to a computationally highly expensive approach.

Even though the performance of all these techniques, misclassifications cannot be completely elim-

Zidemal Z., Amirou A. and Belouchrani A. (2009).
USING SUPPORT VECTOR MACHINES (SVMS) WITH REJECT OPTION FOR HEARTBEAT CLASSIFICATION.
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, pages 204-210
DOI: 10.5220/0001431602040210

inated and, thus, can produce severe penalties. This motivates the introduction of a reject option in the classifiers. Although of its practical interest, this option has not received enough attention since the publications of Chow on the error reject tradeoff (Chow, 1957)-(Chow, 1970). A notable proposal of a reject rule explicitly designed for SVMs has been presented in (Fumera and Roli, 2002), (Kwok, 1999), (Tortorella, 2004) and recently in (Herbei and Wegkamp, 2006) and (Bartlett and Wegkamp, 2007). To our Knowledge, this option was not used in heartbeat classifiers although of its particular interest in medical field where the risk of a wrong classification is so high that it is convenient to reject the sample.

In this paper, SVMs working in binary classification mode and different preprocessing techniques of ECG waveform are used for ectopic heartbeat recognition. We introduce a cost-sensitive reject rule for SVMs using a double hinge loss. This classifier is based on the one provided by Bartlett and Wegkamp (Bartlett and Wegkamp, 2007). This heuristic approach is coherent with the theoretical foundations of SVMs, which are based on the structural risk minimization (SRM) principle (Vapnik, 1995), (Cristianini and Shawe-Taylor, 2000). Under this framework, the rejection region must be determined during the training phase of the classifier.

The paper is organized as follows. Section 2 presents the preprocessing techniques and how diagnostic features were selected and computed. Section 3 recalls Bayes'rule for binary classification with rejection. This section also presents the SVMs in this framework and the learning criterion dedicated for the problem at hand. This proposed method is tested empirically in Section 4. Finally, Section 5 briefly concludes the paper.

2 DATA PREPARATION

In this work, we concentrate on the classification of normal beats and abnormal beats (PVC beats) see Figure 1.

Some records from the MIT-BIH arrhythmia database (http://physionet.org/ physiobank/database/mitdb) with 360 Hz sampling frequency are used. Each record is accompanied by an annotation file in which each ECG beat has been identified by expert cardiologists. These labels, referred to as 'truth' annotation and are used to develop the SVMs classifier and to evaluate its performance. Since this study is to evaluate the performance of a binary classifier that can identify a premature ventricular contraction, some records presenting a high occurrence of PVC beats were selected.

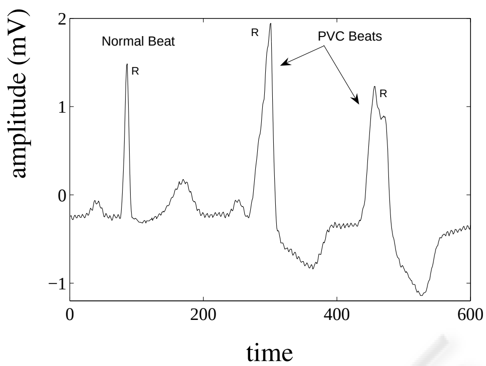

Figure 1: Segment of ECG (record 208) showing the morphology of PVC heartbeats (second and third ones).

2.1 Pre-processing

The objective of this paper is to classify the QRS beats as normal or abnormal ones. Before performing this task, several pre-processing steps were performed on the raw data to study their effects upon the performance of the classifier. In fact the electrocardiogram (ECG) from body electrodes are corrupted by noise. Usually, two principal sources of ECG noise can be distinguished: first one caused by the physical parameters of the recording equipment and the second one representing the bioelectrical activity also called background activity or baseline wander. Several noise removal techniques were recently developed. In this work, we reduce the high frequency noise by thresholding the ECG wavelet coefficients (Donoho, 1995). The baseline wander is removed by setting to zero the approximation coefficients vector of the sixth level of wavelet decomposition.

2.2 Feature Extraction and Selection

The QRS complex in ECG signal varies with origination and conduction path of the activation pulse in the heart. When the activation pulse originates in the atrium and travels through the normal conduction path, the normal QRS complex has a sharp and narrow deflexion and the spectrum contains high frequency components. When the activation pulse originates in the ventricle and does not travel through the normal path, the QRS becomes wide and the high frequency components of the spectrum are attenuated.

A set of algorithms from signal conditioning to measurements of average wave amplitudes, durations, morphology, and areas is usually adopted to perform a quantitative description of a heartbeat and a parameter extraction. In this study, we used some parameters such as the instantaneous R-R interval, average R-R interval, width, morphology and mobility of the QS segment.

• The R peaks were detected using a robust method based on wavelet coefficients (Lepage, 2003) witch was compared to the well known Pan and Tompkins algorithm (Pan and Tompkins, 1985). Once the R peaks are detected, the instantaneous R-R interval is calculated as the difference between the QRS peak of the present beat and the previous one.

• The average R-R interval is calculated as the average R-R interval over the previous ten intervals. The peaks Q and S are detected using simple peak detection method leading to the width of the QS segment.

• The morphology of the beat is captured by four Linear Predictive Coding (LPC) coefficients. The basic idea of this technique is that future values of a discrete signal are estimated as a linear function of previous samples. The most common representation is

$$\widehat{y}_n = \sum_{k=1}^{p} a_k y_{n-k} \tag{1}$$

where a_k is the k^{th} linear prediction coefficient, p is the order of the predictor and $\widehat{y}_n$ the present predicted sample.

• A frequency feature was extracted by computing mobility factor (MB) as defined in (Ramaswamy et al., 2004)

$$MB(x) = \sqrt{\frac{var(x')}{var(x)}} \tag{2}$$

where x is the original ECG signal from point Q to S, $var(x)$, the variance of x and x' the first derivative of x. MB is basically a ratio of energy of higher frequency signal over the energy of the signal. Since the ectopic beats have longer QS segments, the higher frequency energy will be lower. The information of each beat is stored as a 7-element vector, with the first three elements representing the temporal parameters, the next three elements representing the morphological information and the last one is the mobility of the QS segment.

3 CLASSIFICATION WITH REJECTION

Classification aims at predicting a class label $y \in \mathcal{Y}$ from an observed pattern $\mathbf{x} \in \mathcal{X}$. For this purpose, we construct a decision rule d that typically assigns a label to any $\mathbf{x} \in \mathcal{X}$. In binary problems, where the class is tagged $+1$ or -1, the two types of possible errors are: false positive (FP), where examples labeled -1 are categorized in the positive class, incurring a loss c_-; false negative (FN), where examples labeled $+1$ are categorized in the negative class, incurring a loss c_+. We consider here problems where some samples may be not categorized. The classifier d based on the one provided by (Bartlett and Wegkamp, 2007) has the option to reject samples that cannot be classified with enough confidence. This decision to abstain, will be denoted ® and incurs a loss r. The losses pertaining to each possible decision are recapped in table 1 and illustrated on figure 2.

Table 1: Losses for each possible pair of label and decision.

$d(\mathbf{x})$ \ y	$+1$	-1
$+1$	0	c_-
0	r	r
-1	c_+	0

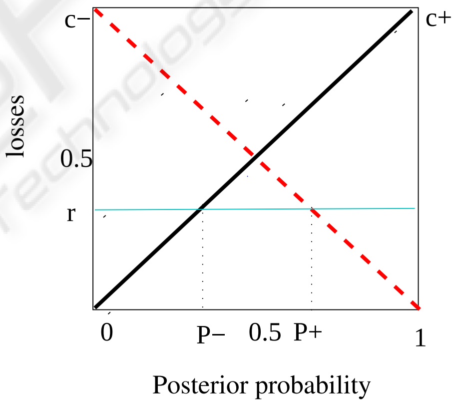

Figure 2: Illustration of different risks vs. posterior probabilities.

3.1 Bayes' Rule with Reject Option

Bayes´s decision theory is the paramount framework in statistical decision theory, where decisions are taken to minimize the expected loss

$$\begin{aligned} L(d) &= c_+ \mathrm{P}(Y=1, d(X)=-1) + \\ &\quad c_- \mathrm{P}(Y=-1, d(X)=1) + \\ &\quad r\,\mathrm{P}(d(X)=®) \ . \end{aligned} \tag{3}$$

From figure 2, one gets that rejection is a viable option if and only if

$$0 \leq r \leq \frac{c_- c_+}{c_- + c_+} \ . \tag{4}$$

If we assume that $c_- = c_+ = 1$, the condition (4) becomes simply $0 \leq r \leq \frac{1}{2}$. In particular, this implies that the loss of rejecting a pattern should be lower than the loss of making an error. Bayes rule can then be expressed simply, using two thresholds $p_- = r$ and $p_+ = 1 - r$ (see Figure 3).

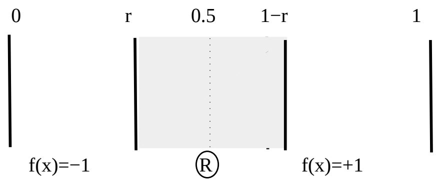

Figure 3: Bayes' rule with reject option.

Where, assuming that (4) holds, Bayes' rule with reject option can then be stated as

$$d(\mathbf{x}) = \begin{cases} -1 & \text{if } \mathrm{P}(Y=1|X=\mathbf{x}) < p_- \ , \\ +1 & \text{if } \mathrm{P}(Y=1|X=\mathbf{x}) > p_+ \ , \\ \circledR & \text{otherwise} \ . \end{cases} \tag{5}$$

Since the paper of Chow (Chow, 1970), this type of rule is sometimes referred to as Chow's rule.

3.1.1 Bayes Rule with Weighted Errors

Often, in practice, FN errors are more costly than FP errors or vice-versa. It is the case in medical field where misclassifying a sick patient as healthy is, in general, far worse than the reverse. In order to accommodate for that, we consider the risk function (3) with $c_+ > c_-$. Bayes rule with weighted errors can then be expressed using new thresholds

$$\begin{aligned} p_- &= r \\ p_+ &= 1 - \theta r \end{aligned} \tag{6}$$

where θ represents the ratio of the costs of the two types of error.

One of the major inductive principle is the empirical risk minimization, where one minimizes the empirical counterpart of the expected loss (3). In classification, this principle is usually NP-hard to implement. Hence, as Bayes decision rule is defined by conditional probabilities, many classifiers first estimate the conditional probability $\widehat{\mathrm{P}}(Y=1|X=\mathbf{x})$, and then plug this estimate in (5) to build the decision rule.

$$d(\mathbf{x}) = \begin{cases} -1 & \text{if } \widehat{\mathrm{P}}(Y=1|X=\mathbf{x}) < p_- \ , \\ +1 & \text{if } \widehat{\mathrm{P}}(Y=1|X=\mathbf{x}) > p_+ \ , \\ \circledR & \text{otherwise} \ . \end{cases} \tag{7}$$

3.2 SVMs Classifier with a Double Hinge Loss

In this section, we show how the standard SVM optimization problem is modified when the hinge loss is replaced by a double hinge loss. The optimization problem is first written using a compact notation, and the dual problem is then derived.

3.2.1 Double Hinge

The double hinge loss function $\phi_r(yf(\mathbf{x}))$, displayed in Figure 4 was proposed by (Bartlett and Wegkamp, 2007) in the context of binary classification with rejection. It is a positive, convex and piecewise linear loss function.

$$\phi_r(yf(\mathbf{x})) = \begin{cases} 1 - \frac{1-r}{r} yf(\mathbf{x}) & \text{if } yf(\mathbf{x}) < 0 \ , \\ 1 - yf(\mathbf{x}) & \text{if } 0 \leq yf(\mathbf{x}) < 1 \ , \\ 0 & \text{otherwise} \ , \end{cases} \tag{8}$$

The function f is estimated by the minimization of a regularized empirical risk on the training samples $\mathcal{T} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$

$$\sum_{i=1}^{n} \phi_r(y_i f(\mathbf{x}_i)) + \lambda \Omega(f) \ , \tag{9}$$

where ϕ_r is the loss function and $\Omega(\cdot)$ is a regularization functional, such as the (squared) norm of f in a Reproducing Kernel Hilbert Space (RKHS)$\mathcal{H}$, $\Omega(f) = \|f\|^2_{\mathcal{H}}$.

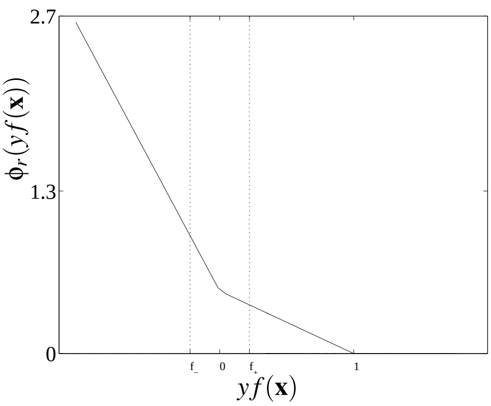

Figure 4: Loss function ϕ_r versus margin $yf(\mathbf{x})$ for $r = 0.1$ f_+ and f_- are the reject thresholds to be computed after the training step.

3.2.2 Optimization Problem

As in standard SVMs, we consider the regularized empirical risk on the training sample. Introducing the double hinge loss (8) results in an optimization problem that is similar to the standard SVMs problem.

Let C a constant to be tuned by cross-validation, we define $D = C(\frac{1-2r}{r})$; The optimization problem reads

$$\min_{f,b} \frac{1}{2}\|f\|^2_{\mathcal{H}} + C\sum_{i=1}^{n} |1 - y_i(f(\mathbf{x}_i + b))|_+ + \mathcal{D}\sum_{i=1}^{n} |-y_i(f(\mathbf{x}_i + b))|_+ \quad , \qquad (10)$$

where $|\cdot|_+ = \max(\cdot, 0)$. Minimizing (10) is a quadratic problem. This is best seen with the introduction of slack variables ξ and γ

$$\begin{cases} \min\limits_{f,b,\xi,\gamma} & \frac{1}{2}\|f\|^2_{\mathcal{H}} + C\sum_{i=1}^{n}\xi_i + \mathcal{D}\sum_{i=1}^{n}\gamma_i \ , \\ \text{s.t.} & \xi_i \geq 1 - y_i(f(\mathbf{x}_i) + b), \\ & \gamma_i \geq -y_i(f(\mathbf{x}_i) + b), \\ & \xi_i \geq 0 \ , \quad \gamma_i \geq 0 \quad i = 1, \dots, n \ . \end{cases} \qquad (11)$$

As for standard SVMs, the dual formulation of (11) leads to efficient optimization algorithms. To compute the solution, we use an active set algorithm following a strategy that proved to be efficient for standard SVMs. The SimpleSVM algorithm (Vishwanathan et al., 2003; Loosli et al., 2005) solves the SVM training problem by a greedy approach in which one solves a series of small problems. First, the training examples are assumed to be either support vectors or not, and the training criterion is optimized considering that the partition of examples is fixed. This optimization results in a new partition of examples in support and non-support vectors. These two steps are iterated until some level of accuracy is reached. Note that this algorithm compute the bias b in the same step as computing the lagrange multipliers α_i

After training, we can represent f as the finite sum

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x) + b \qquad (12)$$

where $\alpha_1, ..., \alpha_n, b$ is the solution of the dual of problem (11)

3.3 Estimation of the Posterior Probability

The SVM does not provide a probability measure. Given a raw score value, the estimation of the probability is a post processing step. One method of producing probabilistic outputs was proposed by (Platt, 2000). This method approximates the posterior probability by a two-parameter logistic function of the form

$$P_{A,B}(x) = \frac{1}{1 + \exp(Af(\mathbf{x}) + B)} \quad , \qquad (13)$$

Where $P_{A,B}(x) \approx \widehat{\mathrm{P}}(Y = y|X = \mathbf{x})$. The best parameters (A, B) are then estimated by minimizing the negative log likelihood of a validation set of labelled samples $\{(x_i, y_i)\}_{i=1}^{l}$, which is a cross-entropy error function:

$$\min_{A,B} - \sum_{i=1}^{l} t_i log(p_i) + (1 - t_i) log(1 - p_i) \quad , \qquad (14)$$

where $p_i = P_{A,B}(x_i)$ and $t_i = \frac{y_i+1}{2}$. A pseudo code for resolving (14) can be found in (Platt, 2000) and (Lin et al., 2003).

After mapping the SVM outputs to posterior probabilities, the decision rule (7) can be applied.

4 EXPERIMENT

From the clinical observations, we obtained 7 features. Six temporal features and one spectral feature. A support Vector Machine (SVMs) with reject option was used to classify these features. The classifier was constructed separately for each record selected in the MIT/BIH database. The data set has been uniformly divided into a training set, a validation set and a test set.

To learn the classifier, we considered a Gaussian kernel (characterized by a width σ) of the following form $K_\sigma(x,y) = \exp(-\frac{\|x-y\|^2}{\sigma^2})$. The kernel parameter σ and the penalization parameter C are firstly optimized by 5 fold cross-validation using SVMs with double hinge loss. The training set is partitioned in 5 subsets where the proportion of positives examples is identical. Each subset is iteratively used as a training set while the remaining ones are used as test sets. Note that the features are normalized before each training session. After learning, the best parameters (A, B) of the logistic function are estimated by fitting on the validation set.
We can see on figure 5 the reject region produced by the SVM classifier for symmetric misclassification losses with $r = 0.45$ and for asymmetric misclassification losses with $r = 0.30$.

As advocated in (Chow, 1970), a complete description of the performance of a recognition system with reject option is given by the error-reject tradeoff. Since the error rate E and the reject rate R are monotonic functions of r, we can compute the tradeoff E versus R from $E(r)$ and $R(r)$.

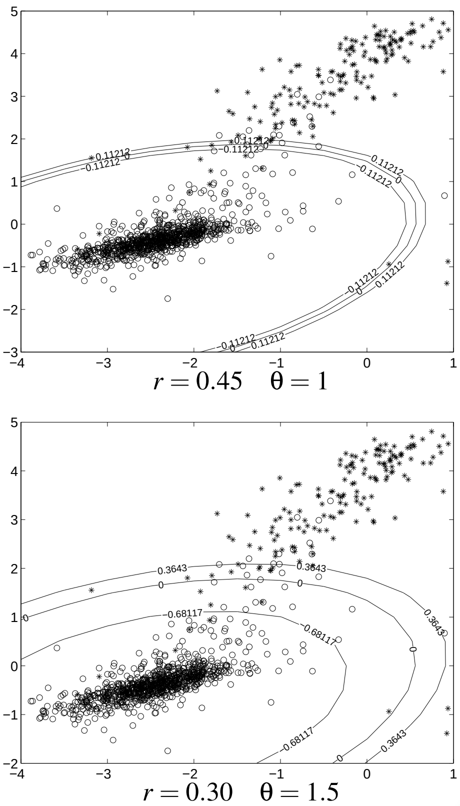

Figure 5: The reject region induced by the reject thresholds in correspondence to the cost of rejecting samples. $r = 0.45$ and $\theta = 1$ (Top); $r = 0.3$ and $\theta = 1.5$ (bottom). The circles($\circ$) indicate the negative class N, wile the asts ($*$) indicate the positive class P.

Varying r between 0.5 and 0.1 with $\theta = 1$, the mean results obtained using the selected records are reported on Figure 6.

The error rate $E = E(R)$, as a function of the reject rate Figure 6 (Top) decreases at a nearly constant rate of roughly 5%.

For example, if we set our rejection thresholds to exclude 5% of the cases, this means that the decision rule can be specified to classify 95% of the cases with a very low misclassification rate and identify the remaining 5% as hard cases that need special considerations.

Another interesting performance criterion is the sensitivity representing the fraction of real events that are correctly detected. $SE = \frac{TP}{TP+FN}$ where True Positive (TP) are the samples labelled $+1$ categorized in the positive class , and False Negative (FN), are the samples labelled $+1$ categorized in the negative class. We show on Figure 6 (bottom) that we obtained more than 98% of sensitivity with no rejection and more than 99% of sensitivity after rejecting less than 5% of instances. Note that the reject threshold follows

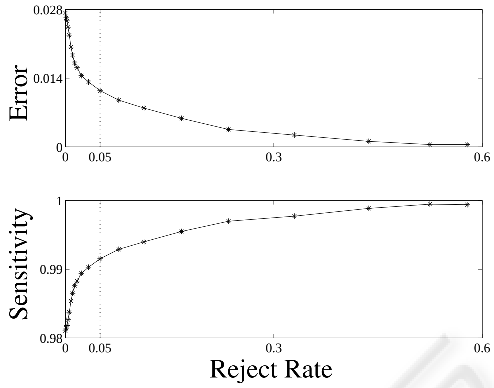

Figure 6: Error Reject curve obtained using the proposed method (Top); Sensitivity vs. reject rate (bottom). While varying r between 0.5 and 0.1 with $\theta = 1$.

the real cost of rejecting a sample and the real cost of misclassifying a sample to optimize the classification cost. This, is the goal of this cost sensitive reject rule.

These results are very competitive to other published studies e.g., (Krasteva and Jecova, 2007) getting 98,4% of sensitivity, (Osowski et al., 2004) obtaining 95,9% of accuracy using a computationally highly expensive approach and (Chazal et al., 2004) obtaining 77,7% of sensitivity for distinguishing Ventricular Ectopic Beats (VEB) from non-VEBs.

5 CONCLUSIONS

This paper presents a new heartbeat classifier using Support Vector Machines with an embedded reject option. The proposed system accomplishes preprocessing, feature extraction /selection and recognition tasks for recognition of Premature Ventricular Contraction (PVC) beats.

For this purpose a cost-sensitive reject rule for SVMs is used together with a double hinge loss for asymmetric classification. For each class, the loss of rejecting a pattern is assumed to be lower than the loss of making an error. A training criterion based on a convex and piecewise linear loss function is proposed. Under this framework, the rejection region is determined during the training phase of the classifier. Our decision rule uses dynamic reject thresholds following the cost of rejecting a sample and the cost of misclassifying a sample to optimize the classification cost.

Our results shown above illustrate a good error reject tradeoff and indicate that if we set our rejection thresholds to exclude less than 5% of the cases, the sensitivity of the classifier becomes higher than 99%, being competitive to other published studies.

This paper has focused on binary classification problems since only ECG records containing normal beats and PVC beats were selected. Extension to multi-category classification is also possible.

REFERENCES

Afonso, O. and Tompkins, W. (1999). Classification of premature ventricular complexes using filter bank features, induction of decision trees and a fuzzy rule-based system. *Med. Biol. Eng*, 37:560–565.

Bartlett, P. L. and Wegkamp, M. H. (2007). Classification with a reject option using a hinge loss. Technical Report M980, Department of Statistics, Florida State University.

Chazal, P., O'Dwyer, M., and Reilly, R. (2004). Automatic classification of hearthbeats using ecg morphology and hearthbeat interval features. *Trans.Biom.Eng*, 51(7):1196–1206.

Chow, C. K. (1957). An optimum character recognition system using decision function. *IRE Trans. Electronic Computers*, EC-6(4):247–254.

Chow, C. K. (1970). On optimum recognition error and reject tradeoff. *IEEE Trans. on Information Theory*, 16(1):41–46.

Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines*. Cambridge University Press.

Donoho, D. (1995). Denoising by soft thresholding. *IEEE, Trans on Info Theory*, 41(3):613–627.

Fumera, G. and Roli, F. (2002). Support vector machines with embedded reject option. In Lee, S.-W. and Verri, A., editors, *Pattern Recognition with Support Vector Machines: First International Workshop*, Lecture Notes in Computer Science, pages 68–82. Springer.

Gomez-Herrero, G., Jecova, I., Krasteva, V., and Egiazarian, K. (2006). Relative estimation of the karhunen loeve transform basis functions for detection of ventricular ectopic beats. *IEEE Comput.Cardiol*, 33:569–572.

Herbei, R. and Wegkamp, M. H. (2006). Classification with reject option. *The Canadian Journal of Statistics*, 34(4):709–721.

Hu, Y. H., Palready, S. H., and Tompkins, W. J. (1997). A patient-adaptable ecg beat classifier using a mixture of experts approach. *Trans.Biom.Eng*, 44(4):891–900.

Jecova, I., Bortolan, G., and Christov, I. (2004). Pattern recognition and optimal parameter selection in premature ventricular contraction calssification. *IEEE, Computer in cardiology*, 31:357–360.

Krasteva, V. and Jecova, I. (2007). Qrs template matching for recognition of ventricular ectopic beats. *Annals of Biomedical Engineering*, 35(12):2065–2076.

Kwok, J. T. (1999). Moderating the outputs of support vector machine classifiers. *IEEE Trans. on Neural Networks*, 10(5):1018–1031.

Lagerholm, M., Peterson, G., Braccini, G., Edenbrandt, L., and Soërnmo, L. (2000). Clustering ecg complex using hermite functions and self-organizing maps. *IEEE.Trans.Biom.Eng*, 47:838–848.

Lepage, R. (2003). Detection et analyse de l'onde p d'un electrocardiogramme: Application au dépistage de la fibrillation auriculaire. *Thse de Doctorat de l'université de Bretagne Occidentale*.

Lin, H., Lin, C., and Weng, R. C. (2003). A note on platt's probabilistic outputs for support vector machines. *Technical report, National Taiwan University, Taipei 116, Taiwan*.

Loosli, G., Canu, S., Vishwanathan, S., and Chattopadhay, M. (2005). Boite à outils SVM simple et rapide. *RIA – Revue d'Intelligence Artificielle*, 19(4/5):741–767.

Minami, K., NAkajima, H., and Toyoshima, T. (1999). Real-time discrimination of ventricular tachyarrhythmia with fourier-transform neural network. *Trans.Biom.Eng*, 46:179–185.

Osowski, S., Hoai, L, T., and Markiewicz, T. (2004). Support vector machine-based expert system for reliable heartbeat recognition. *Trans.Biom.Eng*, 51(4):582–589.

Pan, J. and Tompkins, W. J. (1985). A real-time qrs dtection algorithm. *Trans.Biom.Eng*, 32(3):230–236.

Platt, J. C. (2000). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*.

Ramaswamy, P., Cota, N. G., Chan, K. L., and Shankar, M. K. (2004). Multi-parameter detection of ectopic heartbeats. *IEEE, Int.Workshop.BioCAS*.

Tortorella, F. (2004). Reducing the classification cost of support vector classifiers through an ROC-based reject rule. *Pattern Analysis & Applications*, 7(2):128–143.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer Series in Statistics. Springer.

Vishwanathan, S. V. N., Smola, A., and Murty, N. (2003). SimpleSVM. In Fawcett, T. and Mishra, N., editors, *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, pages 68–82. AAAI.

Yeap, T. H. (1990). Ecg beat classification by a neural network. *Annual international conf of IEEE Eng in Med and Biol*, 12:1457–1458.