KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Alexander Alexeis Suarez-Leon, Carolina Varon, Rik Willems, Sabine Van Huffel, Carlos Roman Vazquez-Seisdedos (2018) **T-Wave End Detection Using Neural Networks and Support Vector machines,** Volume 96, 1 May 2018, Pages 116–127 Computers in Biology and Medicine
Archived version	Final version / pdf
Published version	https://www.sciencedirect.com/science/article/pii/S0010482518300507
Journal homepage	https://www.sciencedirect.com/journal/computers-in-biology-and-medicine
Author contact	carolina.varon@esat.kuleuven.be Phone number *+ 32 (0) 16 32 64 17*
IR	url in Lirias https://lirias.kuleuven.be/handle/123456789/xxxxxx

(article begins on next page)

T-Wave End Detection Using Neural Networks and Support Vector Machines

Alexander Alexeis Suárez-León[a,b,*], Carolina Varon[b,c], Rik Willems[b,d], Sabine Van Huffel[b,c], Carlos Román Vázquez-Seisdedos[a]

[a] *Universidad de Oriente, Faculty of Telecommunications, Informatics and Biomedical Engineering, Santiago de Cuba, Cuba*
[b] *KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Leuven, Belgium*
[c] *Imec, Leuven, Belgium*
[d] *UZ Leuven, Leuven, Belgium*

Abstract

Background and Objective: This paper studies a new approach for detecting the end of the T-wave in the electrocardiogram (ECG) using Neural Networks and Support Vector Machines. *Methods:* Both, Multilayer Perceptron (MLP) neural networks and Fixed-Size Least-Squares Support Vector Machines (FS-LSSVM) were used as regression algorithms to determine the end of the T-wave. Different strategies for selecting the training set such as random selection, k-means, robust clustering and maximum quadratic (Rényi) entropy were evaluated. Individual parameters were tuned for each method during training and the results are given for the evaluation set. A comparison between MLP and FS-LSSVM approaches was performed. Finally, a fair comparison of the FS-LSSVM method with other state-of-the-art algorithms for detecting the end of the T-wave was included. *Results:* The experimental results show that FS-LSSVM approaches are more suitable as regression algorithms than MLP neural networks. Despite the small training sets used, the FS-LSSVM methods outperformed the state-of-the-art techniques. *Conclusion:* FS-LSSVM can be successfully used as a T-wave end detection algorithm in ECG even with small training set sizes.

Keywords: ECG, T-Wave end, Neural Networks, FS-LSSVM

1. Introduction

Torsades de pointes (TdP) is a specific electrocardiographic form of polymorphic ventricular tachycardia. Sustained or prolonged TdP can lead to ventricular fibrillation and Sudden Cardiac Death (SCD) [1]. TdP has been associated with either, congenital or acquired Long QT Syndrome (LQTS). Congenital LQTS diseases are due to mutations of the genes that encode the

*Corresponding author
Email addresses: aasl@uo.edu.cu (Alexander Alexeis Suárez-León), carolina.varon@esat.kuleuven.be (Carolina Varon), rik.willems@kuleuven.be (Rik Willems), sabine.vanhuffel@esat.kuleuven.be (Sabine Van Huffel), cvazquez@uo.edu.cu (Carlos Román Vázquez-Seisdedos)

functions of sodium (Na+) and potassium (K+) channels. On the other hand, the blockade of the inward potassium rectifier (IKr) channel due to the effects of pro-arrhythmic drugs intake is the main cause of acquired LQTS [2] [3].
The QT interval is associated to the ventricular repolarization of the heart. It starts at the beginning of the QRS complex (first deflection) and lasts till the end of the T-wave. It is also known that QT interval depends on the previous RR interval. Several formulas have been defined to correct the QT interval, leading to a family of corrected QT intervals (QTc) which includes Bazett, Fridericia, Frammighan among others. It has been demonstrated that the main factors associated with increased risk of QT prolongation and TdP include sex (females are more susceptible than males), advanced age, bradycardia, congestive heart failure, long QTc and others [1]. Long QTc (Bazett) has been defined as QTc$\leq$450 ms for men and QTc $\leq$ 460 ms for women. A severely prolonged QTc with an increased risk of TdP is defined as QTc$\leq$500 ms [4]. Normally, the T-wave end point (Te) is at baseline level of the ECG signal, which is usually contaminated by noise and interference in ambulatory ECG. Another issue is that there is no consensus on the leads and the number of them that should be used for assessing Tend. Thus, a gold-standard for the T wave end is difficult to define [5]. Nevertheless, proper T-wave end detection remains as an open problem. Furthermore, in the case of drug induced QT prolongation detection it finds its main application nowadays.

There are multiple approaches proposed in the literature for estimating Te. For instance, Vila et al. [6] presented a Te detection algorithm based on a mathematical model. An ECG delineation method was proposed in [7] using a quadratic spline wavelet and four dyadic scales. Other studies are based on area computation [8] [9]. Discrete wavelets, area-curve length index and thresholds were used for delineating the ECG in [10]. A mathematical model of skewed Gaussian functions combined with the trapezium's area method was described in [11]. Other methods include the partially collapsed Gibbs sampler [12], the phasor transform [13] and piecewise derivative Dynamic Time Warping [14].

One possible approach is to consider that Te location depends on the samples of the current beat. In other words, detecting the Te point can be formulated as the problem of finding a certain function which estimates the position of Te from the samples contained in a given interval, i.e., a regression problem. The problem is to find a $\mathbf{\Phi}(\mathbf{x})$ function which maps $\mathbf{x} \in \Re^d$ to $y \in \Re$, where $\mathbf{x}$ is a segment of the signal, d is the length of the segment and y is the location of the Te. From the point of view of the multivariate function regression analysis this is a heuristic approach. In [15] a similar approach was taken for modeling the Te using neural networks. However, the latter study focuses on the properties of the fitted model rather than on the detection problem itself. Up to our knowledge there are no previous studies that address and deeply analyze machine learning based Tend detection algorithms. This paper proposes a new approach for detecting the end of the T-wave using Multi-layer Perceptron (MLP) and Fixed-Size Least-Squares Support Vector Machines (FS-LSSVM) as regression algorithms. Besides, both methods are evaluated using different approaches and the results are throughly discussed. As a result an state-of-the-art detector is obtained. The paper is organized as follows: in Section 2 we present the methodology of the detection algorithms based on both MLP neural networks and Fixed-Size LSSVM. The results of the test phase on QTDB are given in Section 3 and discussed in Section 4. Finally, the conclusions are presented in Section 5.

2. Materials and Methods

MLP and LSSVM are both supervised approaches that have been used in pattern recognition problems [16] [17] However, in this study we used both approaches for solving a regression problem. These types of algorithms require the availability of annotated data beforehand in order to support the training process. Then, during the learning process the algorithm learns through input-output pairs examples. The training step consists of selecting a subset of heartbeats in order to construct and tune the regression algorithm. The test step uses the trained algorithm to predict the end of the T-wave for beats which were not previously included in the training set, see Fig. 1a.
Below, the dataset used in this study is described, and then, the next sections explain in detail both, the training and testing phases and its stages.

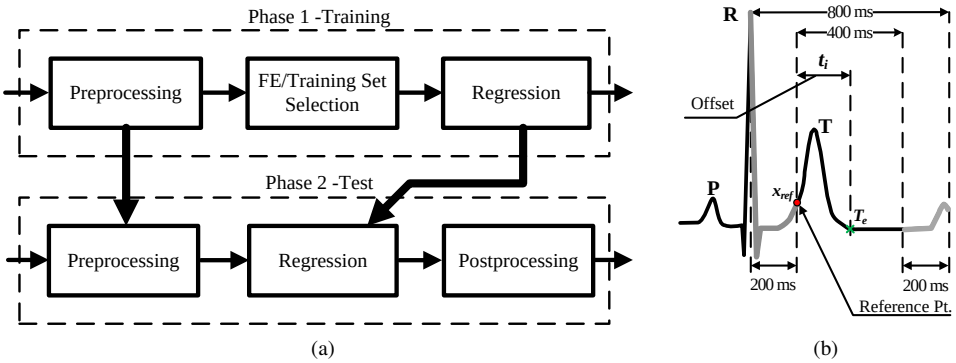


Figure 1: General approach, (a) training and test work-flow for Te detection algorithm and (b) segmentation method.

2.1. Dataset

This study uses the QT Database (QTDB) [18] which is publicly available at Physionet [19]. QTDB was designed for evaluating the performance of algorithms for event detection in ECG. It consists of short segments (15 min) extracted from 105 Holter recordings, each with two channels. All records have a sampling frequency of 250 Hz. Two cardiologists annotated the database: Cardiologist 1 (C1) annotated 3542 T-wave ends in 103 records, and Cardiologist 2 (C2) annotated 402 Te in 11 records. For the purpose of this study, only the annotations provided by C1 are considered, so the total number of available patterns is N_A = 3542. Cardiologist 1 dataset is used because of two reasons. First, there is a big difference in the amount of annotations from both cardiologists, and second, most of the annotations of C2 overlap with the annotations of C1. In practice, C1 and C2 are commonly used together for inter-observer variability studies. From the two channels available in each one of the 103 records annotated by C1, only the first one is used for evaluating the methods.

2.2. Preprocessing

The pre-processing step includes three stages: filtering, R-peak detection, and heartbeat segmentation. The filter stage uses a fourth order zero-phase band-pass Butterworth filter to deal

with both baseline wandering and high frequency noises. The cut-off frequencies are 0.5 Hz and 50 Hz for high-pass and low-pass respectively.
The next step detects the R peaks using an algorithm based on parabolic fitting [20]. The heartbeat segmentation is as follows: for each heartbeat, a 100 samples vector (400 ms) is extracted from a reference point (x_{ref}) at R + 50 samples (200 ms), where R is the R-peak location of the current heartbeat. This segmentation has two goals, (1) to select a relatively small interval which includes the Te and (2) to bypass the high energy and frequency content of the QRS complex, see Fig. 1b.
The Te location for the current heartbeat is given by the position of the x_{ref} point for the current heartbeat and an offset. The latter is the desired output of the regression function to estimate. Here we have used both, MLP and FS-LSSVM as regression algorithms. When MLP is used, the target vector is normalized dividing by 100 samples (400 ms) as follows,

$$t_i = \frac{Te_i - R_i - 50}{100}, \tag{1}$$

where t_i is the i-th component of the target vector, i.e. the offset of the i-th heartbeat selected for training, Te_i and R_i are the annotated Te and the R point for the current heartbeat respectively. On the other hand, the division by 100 samples is not needed in FS-LSSVM, so only the numerator of (1) is used.

2.3. Feature extraction and training set selection

The number of components of each vector is 100. This value is still large to be used as input to a regression algorithm. In order to reduce the size of the input vector for both, MLP and SVM approaches, a feature extraction algorithm must be used. Here, the Discrete Cosine Transform (DCT) is used for reducing the dimension of the input data. The DCT $y(k), k = 0, 1, \ldots, L-1$ of a data sequence $x(n), n = 0, 1, \ldots, L-1$ is defined as,

$$\begin{aligned} y(0) &= \frac{\sqrt{2}}{L} \sum_{n=0}^{L-1} x(n) \\ y(k) &= \frac{2}{L} \sum_{n=0}^{L-1} \left(x(n) \cos\left(\frac{(2n+1)k\pi}{2L} \right) \right), \end{aligned} \tag{2}$$

where L is the length of the data sequence (100-samples vector) $x(n)$, L is the number of components, i.e. $L = 100$.
The DCT is data independent and has high energy-packing efficiency, i.e, the DCT has the property of compacting the energy into few coefficients. It has been widely used in compression algorithms [21], [22] and also as feature extraction method in ECG morphological recognition [23]. Nevertheless, it is not only desirable to decrease the dimensionality of the observations but also the size of the training set. Since a small size of the training set can lead to a strong negative impact on the performance of the regression algorithms, a selection procedure for training vectors should be implemented. The method could be based on multiple criteria. Here we evaluate random selection, clustering methods and the quadratic Rényi entropy criterion.
The simplest method for clustering is the well-known k-means algorithm. However k-means is not robust in the presence of outliers. In fact, one extreme value can lead to the failure of the algorithm. Hence, robust clustering methods should be used. Here, we use two trimmed k-means

based approaches, the Trimmed k-means (TKMEANS) algorithm [24] and TCLUST [25]. The TKMEANS algorithm allows that in a given dataset there could be vectors which cannot be assigned to any cluster (outliers). The algorithm requires two parameters: the number of clusters (k) and the percentage of trimmed observations (α), see algorithm 1.

Algorithm 1 Trimmed k-means (TKMEANS)

Require $k \geq 2, \alpha \leq 0.5$

1: Initialize k random centers, $\mathbf{m}_j^0$, where $j = 1, \ldots, k$

2: Build the set $\mathbf{H}$ which includes the $n(1-\alpha)$ vectors closest to the centers, where (n) is the number of observations.

3: Divide $\mathbf{H}$ into k subsets where, $\mathbf{H}_j$ contains the vectors in $\mathbf{H}$ closer to the center $\mathbf{m}_j$ than to the other centers.

4: Update centers $\mathbf{m}_j^{l+1}$ such that each $\mathbf{m}_j^{l+1}$ is the sample mean of the vectors in $\mathbf{H}_j$.

5: Repeat steps 1-4 $Q = 15$ times (default value), keeping the best solution in the sense of minimizing the function,

$$\arg\min_{\mathbf{H}} \min_{\mathbf{m}_1,\ldots,\mathbf{m}_k} \sum_{\mathbf{x}_i \in \mathbf{H}} \left\| \mathbf{x}_i - \mathbf{m}_j \right\|^2. \tag{3}$$

6: stop

Since Trimmed k-means is defined using the Euclidean distance, it assumes spherical distribution for each cluster. Notwithstanding, in certain problems this assumption might not be valid, i.e., one or more clusters might have different scatter structures. For these types of problems, heterogeneous clustering would yield better results. TCLUST belongs to the class of heterogeneous clustering algorithms, and it is based on the probabilistic framework proposed in [26] and [27]. It requires three parameters: the number of clusters (k), the proportion of trimmed observations (α) and the upper bound of the ratio between the largest (M_n) and the smallest (m_n) eigenvalue of all covariance matrices (c). The algorithm is described in detail below.
It is noticeable that the value of Q can be tuned in both algorithms. Here, the default value ($Q = 15$) is used. Although the probability of an efficient clustering increases with the value of Q, the execution time also grows, especially in TCLUST. Several trials with values 10, 15 and 20 were performed for both algorithms, however, no evidence of better results was found for the values 10 and 20. In fact, a noticeable increased execution time for TCLUST was observed for ($Q = 20$). So, the default value ($Q = 15$) is kept.

2.4. Multilayer Perceptron

The MLP neural networks used in this study have three layers: input, hidden and output. While the number of output units is given by the nature of the problem, choosing the number of input and hidden units is a critical step when configuring a neural network. This is because training the network is a problem with a large number of free parameters (weights and biases). The output for a general three layer network with N_I input units, N_H hidden neurons, and N_O output units is given by,

Algorithm 2 Heterogeneous clustering (TCLUST)

Require $k \geq 2, \alpha \leq 0.5, c \geq M_n/m_n$

1: Initialize k random centers ($\mathbf{m}_j^0$), covariance matrices (Σ_j^0) and weights, (π_j^0), where $j = 1, \ldots, k$.

2: Build the set $\mathbf{H}$ which includes the $n(1-\alpha)$ vectors with the largest values for the function,

$$\max_{j=1,\ldots,k} \pi_j^l f\left(\mathbf{x}; \mathbf{m}_j^l, \Sigma_j^l\right). \tag{4}$$

where $f(\cdot, \mathbf{m}, \mathbf{\Sigma})$ is the probability density function of the multivariate normal distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{\Sigma}$.

3: Divide $\mathbf{H}$ into k subsets where $\mathbf{H}_j$ contains the vectors in $\mathbf{H}$ such that the maximum value of the function (4) is attained to this j.

4: Update centers $\mathbf{m}_j^{l+1}$, covariance matrices Σ_j^{l+1}, and weights π_j^{l+1} with the sample mean, the sample covariance matrices, and the proportion of vectors in $\mathbf{H}_j$ respectively.

5: Compute

$$\begin{cases} M_n = \max\limits_{j=1,\ldots,k} \max\limits_{l=1,\ldots,p} \lambda_l(\Sigma_j) \\ m_n = \min\limits_{j=1,\ldots,k} \min\limits_{l=1,\ldots,p} \lambda_l(\Sigma_j), \end{cases} \tag{5}$$

where $\lambda_l(\Sigma_j)$ are the eigenvalues of the covariance matrices.

6: **if** $M_n/m_n > c$

7: Solve a quadratic programming problem in order to force the constraint $c \geq M_n/m_n$

8: **end if**

9: Repeat steps 1-8 $Q = 15$ times (default value), keeping the best solution in the sense of minimizing the function,

$$\sum_{j=1}^{k} \sum_{\mathbf{x}_i \in \mathbf{H}_j} \log\left(\pi_j^l f\left(\mathbf{x}_i; \mathbf{m}_j^l, \Sigma_j^l\right)\right), \tag{6}$$

with $\# \cup_{j=1}^{k} \mathbf{H}_j = [n(1-\alpha)]$

10: **stop**

$$y_k = f_o\left[\sum_{j=1}^{N_H} w_{jk} f_h\left(\sum_{i=1}^{N_I} w_{ij}x_i + \theta_j\right) + \theta_k\right], \tag{7}$$

where f_o and f_h are, respectively, the hidden and output activation functions, w_{ij} is the weight from the input neuron i to the hidden unit j, w_{jk} is the weight from the hidden neuron j to the output unit k, x_i is the input i, θ_j and θ_k are the biases for the hidden and output units respectively, and the indexes $i = 1, \ldots, N_I,\ j = 1, \ldots, N_H, k = 1, \ldots, N_O$.
Normally $f_o(x) = x$, in this problem $N_O = 1$ and $f_h = \tanh(\cdot)$. Thus, the expression (7) can be rewritten as,

$$y = \sum_{j=1}^{N_H} w_j \tanh\left(\sum_{i=1}^{N_I} w_{ij}x_i + \theta_j\right) + \theta. \tag{8}$$

The number of parameters to adjust (provided that $N_O = 1$) is given by,

$$N_p = N_H(N_I + 2) + 1. \tag{9}$$

Generally, neural networks need a high number of examples for training. There are several criteria for selecting the number of training patterns. Here, we examined training set sizes lower than the 30% of the dataset size. Given the number of parameters, a common criterion is to select the training set size q times the number of free parameters,

$$N_H(N_I + 2) + 1 \leq 0.3q^{-1}N_A, \tag{10}$$

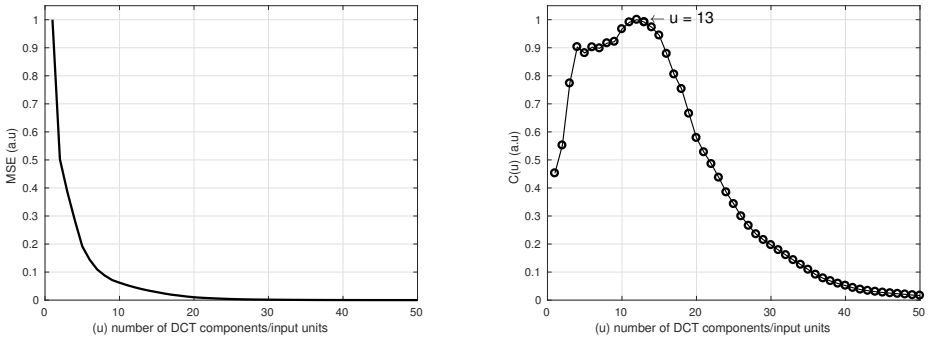
where $q \geq 2$. The expression (10) has three degree of freedom $(N_I; N_H; q)$. Although a well-known practical criterion suggests $q = 10$, given the amount of available patterns ($N_A = 3542$) the latter value for this criterion is impractical. So, we explore first the possible values of N_I and N_H, and then the value of q can be computed from (10).
From both values N_I and N_H, the number of input units (N_I) is the most relevant because it is equal to the number of the DCT components to keep (u) which is an expression of the grade of compression that we apply to the signal. On the other hand, N_H can be always upper-bounded as a multiple of N_I e.g. $N_H \leq 2N_I$. So, the problem of selecting the number of input units (N_I) is also the problem of selecting the number of DCT components to keep (u).

In order to satisfy (10), it is necessary to choose the lowest possible value for u since N_I is a multiplier term on the left side of (10). Appropriate values for u can be given by the total mean squared error (MSE) in the reconstruction of the 30% of the dataset when u components are kept, see Fig. 2a. At first glance, appropriate values for u seem to be on the interval [30, 40], however, such values are not small enough, this is due to the fact that this criterion does not take into account (10). In Fig. 2b we have explored a criterion which is based on both, MSE and the number of parameters (N_p) as follows,

$$C(u) = \text{mse}(\mathbf{X}^u)N_p(u),\ u = 1, \ldots, L, \tag{11}$$

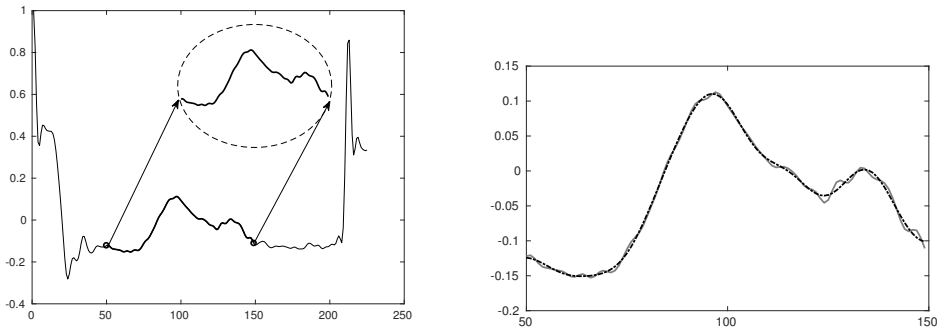
where $\mathbf{X}^u$ is the truncated $N_A \times u$ matrix of the dataset when the first u components of the DCT are kept, $N_p(u)$ is the number of parameters to adjust as defined in (9) considering $N_I = u$ and bounding N_H as $N_H \leq 2N_I = 2u$. The value $u = 13$ is selected since it is the first value for which the error decreases faster than the complexity grows, where the complexity is seen as the

number of free parameters to adjust (N_p). Figures 3a-3b illustrate the DCT reconstruction of an annotated heartbeat from QTDB using the first 13 components. Using (10), $N_I = u = 13$ and $q = 3$ is clear that $N_H \leq 23$. In order to provide a larger margin, i.e., to increase q, the value $N_H = 19$ is selected.



(a) (b)

Figure 2: Criteria for selecting the number of input units, (a) total mean squared error (MSE) in the reconstruction of the data using u components and (b) and trade-off Complexity-MSE ($C(u)$). In order to clarify the interval of interest only the first 50 values of both, MSE(u) and $C(u)$ criteria are drawn.



(a) (b)

Figure 3: DCT reconstruction of an annotated beat from QTDB (record sel102, first heartbeat) using 13 components (a) segment of interest for detecting Te (b) the original segment (gray continuous line) and the 13-components DCT reconstructed segment (black dash-dotted line).

2.5. Fixed-Size LSSVM

One advantage of Support Vector Machines (SVM) algorithms is that the training set size does not explicitly depend on the dimension of the input vectors. Thus, a higher number of DCT components than in the MLP case can be used. The latter is exploited here by increasing the number of Least-Squares Support Vector Machines (LSSVM) is a special formulation of SVM which replaces the inequality constraints on the optimization problem by equality constraints.

Consequently, the optimization process can be formulated and solved by means of a system of linear equations instead of using a quadratic programming algorithm [17].

Fixed-Size Least-Squares Support Vector Machines (FS-LSSVM) is a method for solving large-scale classification/regression problems. It takes advantages of the primal-dual LS-SVM formulations. So, in problems with large datasets could be of benefit to solve the SVM in the primal space [28]. Such approach would require an explicit expression of the feature map or an approximation to this map. In order to efficiently obtain the approximation to the map, the Nyström method has been proposed [29]. Nyström method uses a low rank approximation of the kernel matrix by selecting m rows or columns of this matrix. Fig. 4 shows a general diagram for a FS-LSSVM construction.

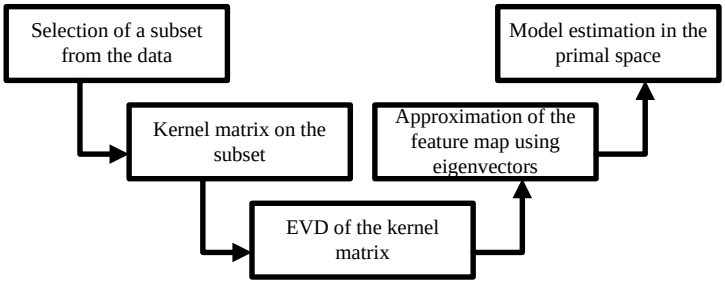


Figure 4: Fixed-Size LSSVM construction stages.

First, the vectors for training are selected by using the active prototype vector selection method [28] [30]. This procedure searches for prototype vectors (PV) which maximize the quadratic Rényi entropy criterion ($H^m_{R2}(x)$). One formal definition of the quadratic Rényi entropy [31] is as follows,

$$H^m_{R2}(x) = -\log \int f(x)^2 dx, \tag{12}$$

where $f(\cdot)$ is the probability density function. In practice, the information potential ($\hat{H}^m_{R2}(\mathbf{X})$) is used. The latter is an estimator of $H^m_{R2}(\mathbf{X})$ determined by kernel density estimation methods. Here the link between Kernel Principal Components Analysis (KPCA) and nonparametric density estimation is exploited [32],

$$\hat{H}^m_{R2}(\mathbf{X}) = -\log\left[\frac{1}{m^2}\mathbf{1}^{\mathbf{T}} \cdot \mathbf{K} \cdot \mathbf{1}\right], \tag{13}$$

where $\mathbf{K}$ is the kernel matrix given by

$$\mathbf{K} = \kappa(h^m_{R2}; \mathbf{x}_i, \mathbf{x}_j), \tag{14}$$

here $\kappa(\cdot)$ is the Gaussian kernel and m is the number of vectors. Using the eigenvalue decomposition of this matrix, the information potential can be estimated as [33],

$$\hat{H}^m_{R2}(\mathbf{X}) = -\log\left[\frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}^2 \cdot \lambda_i\right], \tag{15}$$

where v_{ij} are the components of the i-th eigenvector and λ_i is the i-th eigenvalue of the following eigenvalue problem,

$$\mathbf{K} \cdot \mathbf{v}_i = \lambda_i \cdot \mathbf{v}_i, \quad i = 1, \ldots, m. \tag{16}$$

Here we employed a slightly modified version of active PV selection which is detailed in algorithm 3 [28] [30]. The main difference with respect to the algorithm proposed in [30] is that the algorithm below runs for a predefined number of iterations (m), which is equal to the number of support vectors.

Algorithm 3 Active PV selection for training

Require $m \ll n$

1: Let the training set $\mathcal{T}_n = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)\}$.
Choose a subset of prototype vectors $\mathcal{S}_m \subset \mathcal{T}_n$ of size m at random and define the complement subset $\mathcal{S}_{n-m} = (\mathcal{S}_m)^c \in \mathcal{T}_n$.

2: **for** $i = 1$ **to** m

3: Randomly select one sample point from both subsets $\mathbf{x}^* \in \mathcal{S}_m$ and $\mathbf{x}^+ \in \mathcal{S}_{n-m}$, then
swap($\mathbf{x}^*, \mathbf{x}^+$)

4: Compute

$$\begin{cases} \mathcal{E}_1 = \hat{H}^m_{R2}(\mathbf{x}_1, \ldots, \mathbf{x}^+, \ldots, \mathbf{x}_m) \\ \mathcal{E}_2 = \hat{H}^m_{R2}(\mathbf{x}_1, \ldots, \mathbf{x}^*, \ldots, \mathbf{x}_m) \end{cases} \tag{17}$$

5: **if** $\mathcal{E}_1 > \mathcal{E}_2$

6: $\mathbf{x}^+ \in \mathcal{S}_m$ and $\mathbf{x}^* \notin \mathcal{S}_m$, $\mathbf{x}^* \in \mathcal{S}_{n-m}$,

7: **else**

8: $\mathbf{x}^* \in \mathcal{S}_m$ and $\mathbf{x}^+ \notin \mathcal{S}_m$, $\mathbf{x}^+ \in \mathcal{S}_{n-m}$,

9: **end if**

10: **end for**

11: **stop**

The computation of $\hat{H}^m_{R2}$ has a free parameter (h^m_{R2}) which is the bandwidth for the kernel density estimation. There is a wide variety of methods for selecting the bandwidth value such as the Silverman's Rule of Thumb [34], the Sheather and Jones method [35], [30], among others.

After selecting the prototype vectors, the kernel matrix for this subset is computed. Next, the eigenvalue decomposition (EVD) of the kernel matrix is performed. An approximation of the feature map is obtained from the eigenvectors of the previous decomposition. Given this approximation of the feature map, the model is constructed on the primal space. In this case a linear ridge regression is performed as follows,

$$\hat{\mathbf{w}} = \min \|\mathbf{F}\mathbf{w} - \mathbf{t}\|^2 + \|\mathbf{\Gamma}\mathbf{w}\|^2, \tag{18}$$

where $\mathbf{F}$ is the matrix of features, $\hat{\mathbf{w}}$ is the vector which contains the optimal hyperplane coefficients including the bias term, $\mathbf{t}$ is the vector of outputs (target vector), and $\mathbf{\Gamma}$ is the regularization matrix. For an L_2-regularization, $\mathbf{\Gamma}$ is given by,

$$\mathbf{\Gamma} = \gamma \mathbf{I}, \tag{19}$$

where γ is the regularization parameter and $\mathbf{I}$ is the identity matrix. The solution of (18) is as follows,

$$\hat{\mathbf{w}} = \left(\mathbf{F}^T\mathbf{F} + \mathbf{\Gamma}^T\mathbf{\Gamma}\right)^{-1}\mathbf{F}^T\mathbf{t}. \tag{20}$$

The γ is determined using the second and the third levels of inference provided by the Bayesian framework for LSSVM [33], [36].

2.6. Post-processing

Due to the fact that the output of both, the neural networks and the FS-LSSVM, are offsets from the reference point, a post-processing is necessary in order to estimate the actual Te. In the case of MLP the output is also in the range [0 1], so the estimated Te is determined by,

$$\hat{t}_i = \left[R_i + 50 + A\tilde{t}_i\right], \tag{21}$$

where $[\cdot]$ means rounding towards the nearest integer, A is a constant which depends on the algorithm, i.e., $A = 100$ for MLP and $A = 1$ for FS-LSSVM and $\tilde{t}_i$ is the output of the regression algorithm.

2.7. Experiments

2.7.1. Experiment I: Clustering + MLP

In this experiment we compare the performance of the neural network for different methods of training set selection. Four approaches, random selection, simple k-means, Trimmed k-means and TCLUST are evaluated. The neural network used in this experiment has the structure 13:19:1 (286 parameters). The number of clusters is fixed to 6. For each method, 50 iterations are performed, and a fixed percentage of 25% is used (one of four heartbeats are selected for training). For the latter training set size (25%), it is possible to select the α value in the overall interval [0-25]. However, extreme values for α i.e. near to zero or 25 should be avoided because of the loss of diversity of the training set. Thus, a proportion of 3:2 is kept, i.e. the α value is set to 0.15 and the other 10% is available for common (clustered) heartbeats. For heterogeneous clustering the upper bound eigenvalue ratio is $c = 1000$. This value was selected after several simulations using the values $1, 10, 100, 1000, 10000$. The value $c = 1$ corresponds to a weighted version of TKMEANS and $c = 10000$ is a highly heterogeneous algorithm. The training set selection strategy is as follows: all trimmed vectors are used as part of the training set (15%), 10% of the vectors at each cluster selected at random complete the training set.

2.7.2. Experiment II: Best Clustering Methods + MLP

This experiment studies the influence of the training set size (p) and the number of clusters for the best clustering methods. The same (13:19:1) network structure is used. The number of cluster (k) is variable from 2 to 12. For each pair (p, k), 10 iterations are performed. The training set size is sorted from 16% to 30%. The trimmed percentage parameter is selected equal to 0.15 for both TKMEANS and TCLUST methods. The latter uses the same upper bound eigenvalue ratio $c = 1000$. The training set selection strategy is similar to the experiment I. All trimmed vectors are used as part of the training set (15%). The [1-15]% of the beats at each cluster are selected at random in order to complete the training set.

2.7.3. Experiment III: Fixed-Size LSSVM

In this experiment we examine the performance of Fixed-Size LSSVM as a regression method. Here we use random selection and active prototype vector selection (based on Rènyi entropy criterion, see algorithm 3). The training/validation set size varies from 16% to 30%. Both methods are described in detail below. In the random selection approach, a number of observations corresponding to the percentage are chosen at random. After this first selection, a grid search procedure on this set is implemented in order to determine the parameter capacity (C_p) and the kernel parameter (σ). The search sets are $C_p = \{10, 50, 100, 150, 200\}$ and $\sigma = \{0.1, 0.2, 0.5, 1, 2, 4, 10\}$. After tuning C_p and σ, Rényi's entropy criterion, i.e., active PV algorithm is applied in order to select the final training set according to the capacity determined in the previous step. The bandwidth parameter is chosen as $h^m_{R2} = \sigma$.

In the active PV selection approach, the bandwidth is preselected equal to one ($h^m_{R2} = 1$). The best set according to the Rényi entropy criterion is selected after 10 iterations of the algorithm 3. Once the vectors are chosen, a grid search procedure is performed in order to determine both, the capacity (C_p) and the kernel parameter (σ). This grid search is carried out on the dataset selected in the previous step. The search set for the kernel parameter is the same as above, i.e., $\sigma = \{0.1, 0.2, 0.5, 1, 2, 4, 10\}$. On the other hand, the capacity set is $C_p = \{\frac{2}{15}, \frac{1}{5}, \frac{4}{15}, \frac{1}{3}, \frac{2}{5}\}N_T$, where N_T is the size of the training/validation set. For each pair (C_p, σ) 10 iterations are performed. The best pair (C_p, σ) is selected for constructing the regression algorithm together with the training set.

The last part of the experiment compares MLP and FS-LSSVM as regression algorithms. A fixed number of clusters (6) is selected in the case of both MLP approaches (TKMEANS+MLP and TCLUST+MLP). Then, the best algorithm in terms of precision and accuracy (with priority on the former) is selected and compared with the FS-LSSVM approach.

2.7.4. Experiment IV: Comparison with other T-wave end detection algorithms

In this experiment we compare the performance of the algorithms described above with some of the state-of-the-art methods on detecting Te. Such comparison is not straightforward due to fact that the proposed methods require supervised training. Thus, there will be observations with almost zero error beforehand (those which belong to the training set). Consequently, the real performance will be biased when it is computed using the whole dataset. This limitation can be eliminated if just the test set is considered. However, the test set is a subset of the global set, which implies that the results will be given for different datasets.

Here, we compare the methods by Zhang [8], wavelet-based ECG delineation algorithm [7], [19] and trapezium areas [9] with a Fixed-Size LSSVM using Rényi entropy as selection criterion and 15% of the training set size. The former methods are selected because they are state-of-the-art techniques that are also publicly available on-line, which allows to evaluate the results with the same test sets used by the algorithms studied here. Both, Zhang's method and wavelet-based ECG delineation algorithm are used without tunning up their parameters. On the other hand, the trapezium method requires two parameters, the length of the search window (LEN) in milliseconds and the slope (SLP) of the heartbeat (-1 for positive T-waves and +1 for negative ones). The best pair LEN-SLP is used for each record evaluated. The trapezium method also requires the detection of the T-wave peak (Tm). The Tm points are detected using the wavelet-based ECG

delineation algorithm.

In order to avoid an unfair comparison, only the test set is considered. Moreover, the results corresponds to the worst case in 10 iterations of the algorithm. Besides, since the machine learning algorithms were trained with the first leads of each record, it is necessary to extend the procedure to more than one lead. The approach followed here is straightforward and consists of training another FS-LSSVM using the second lead versions of the heartbeats selected in the first lead. Thus, the advantage is that no new selection process is required. It is noticeable that maximum Rényi entropy criterion is in general not longer valid for the second lead. Other approaches that take into account this fact can be evaluated as well. For instance, the training set size can be divided among both leads and the selection procedure can be performed independently. However, the evaluation of the latter approach as well as others is beyond the scope of this study.

2.8. Performance assessment and testing

Here two performance measures are used. On one hand, the accuracy is quantitatively assessed by the mean value of the error in the location of Te in milliseconds,

$$\mu_e = \frac{1}{N_S}\sum_{i=1}^{N_S} e_i = \frac{1}{N_S}\sum_{i=1}^{N_S} (\hat{t}_i - Te_i), \tag{22}$$

where N_S is the number of patterns in the test set and e_i is the error for the current heartbeat. On the other hand, the precision is quantitatively assessed by the corrected sample standard deviation of the location error in milliseconds,

$$\sigma_e = \sqrt{\frac{1}{N_S - 1}\sum_{i=1}^{N_S} (e_i - \mu_e)^2}. \tag{23}$$

All the results, except those ones on Tables 2-4, are given for the test set. Thus, the previous measures are considered as unique set measures.

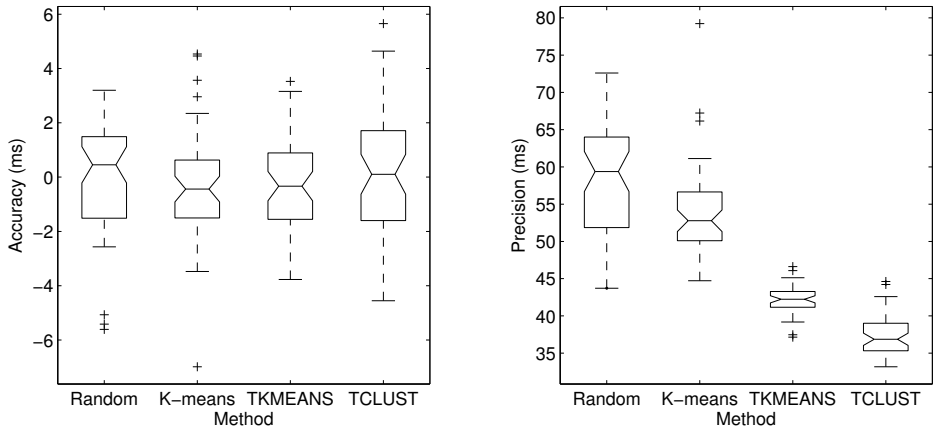
3. Results

The figures 5a and 5b show the results on accuracy and precision respectively for the experiment I. From the data in Fig. 5b, it is apparent that the best results were obtained for both trimmed approaches, i.e., TKMEANS and TCLUST.

In the second experiment only the precision parameter is presented for both robust clustering methods. Fig. 6 shows the general behavior of the precision with respect to the number of clusters (k) and the training set size (p).

Turning now to the study on the individual influence of each parameter, on one hand the Fig. 6a shows the behavior of precision for TKMEANS+MLP and TCLUST+MLP with respect to (k) and independently of (p). On the other hand, the Fig. 6b shows the behavior of the precision for both clustering methods with respect to (p) independently of (k).

Fig. 7 shows the results for the experiment III using FS-LSSVM. Both parameters, accuracy and precision are considered here. A comparison in terms of accuracy and precision between



(a) (b)

Figure 5: Accuracy (a) and precision (b) for MLP based Te detection algorithms with random, k-means, trimmed k-means (TKMEANS) and TCLUST training set selection.

both MLP-based methods and FS-LSSVM is presented in Figures 8a and 8b. The graph was obtained by averaging the accuracy and precision of the three methods for each training set size. In the case of MLP-based methods, the k parameters (the number of clusters) were selected using the best performances in term of precision shown in Fig. 6a. Thus, $k = 5$ for TKMEANS+MLP and $k = 10$ for TCLUST+MLP.

The results of experiment IV are given in the tables 1 and 2. The performance measures (μ_e, σ_e) were used. QTDB recording stratification according to Te accuracy and precision are given for the four algorithms. In Table 2, the results for each record are classified in one of four groups using the criterion of the CSE Working Party [37]. Group I, records where $\sigma_e < 30.6$ ms and $\mu_e < 15$ ms; Group II, $\sigma_e < 30.6$ ms and $\mu_e > 15$ ms; Group III, $\sigma_e > 30.6$ ms and $\mu_e < 15$ ms and Group IV, $\sigma_e > 30.6$ ms and $\mu_e > 15$ ms.

Table 1: Performance comparison using unique set measures and the testing set

Method	**Lead 1**	**Both leads**
	$\mu_e \pm \sigma_e$ (ms)	$\mu_e \pm \sigma_e$ (ms)
Zhang [8]	7.20 ± 60.32	0.68 ± 38.66
Trapezium [9]	-0.36 ± 72.65	-0.97 ± 58.24
Wavelet-ECG [7] [19]	-6.01 ± 65.27	-1.28 ± 59.10
RE + FS-LSSVM 15% **(this study)**	-2.10 ± 42.70	-3.19 ± 29.27

4. Discussion

Although there were no significant differences between algorithms in the accuracy parameter, experiment I clearly pointed out that trimmed k-means based methods outperform the random

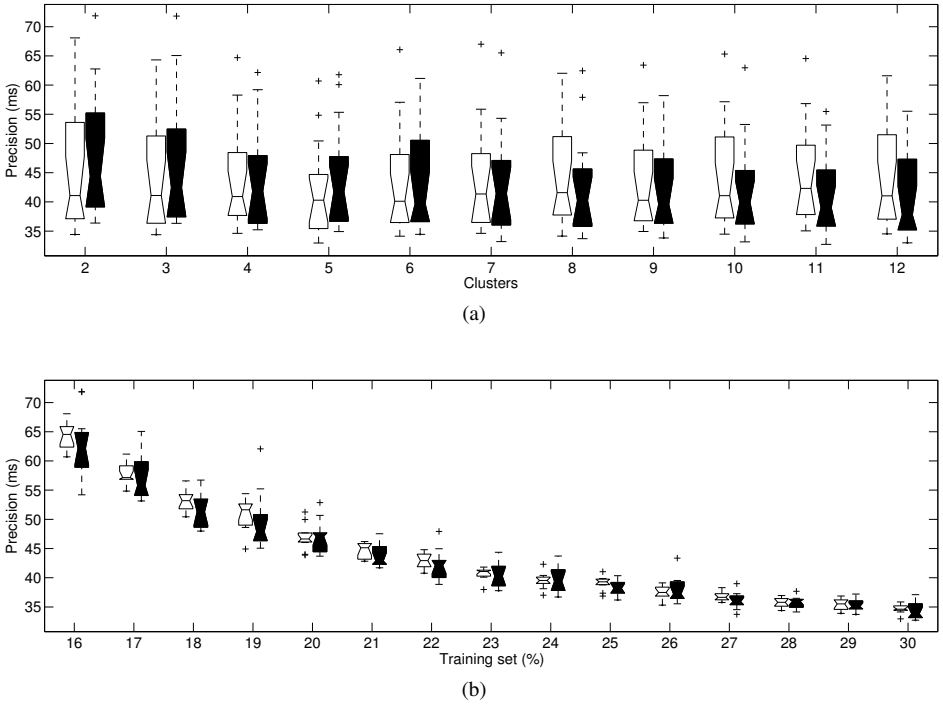
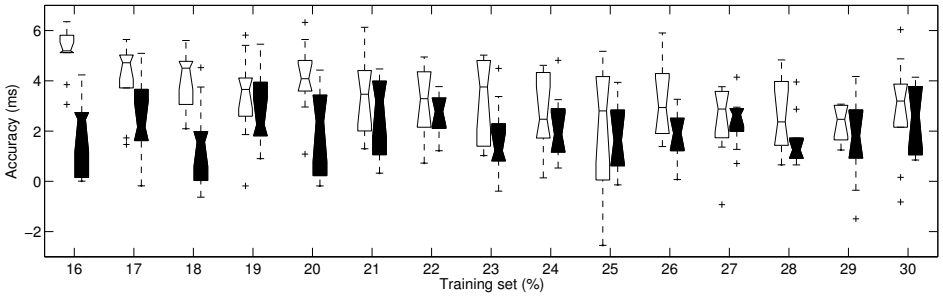


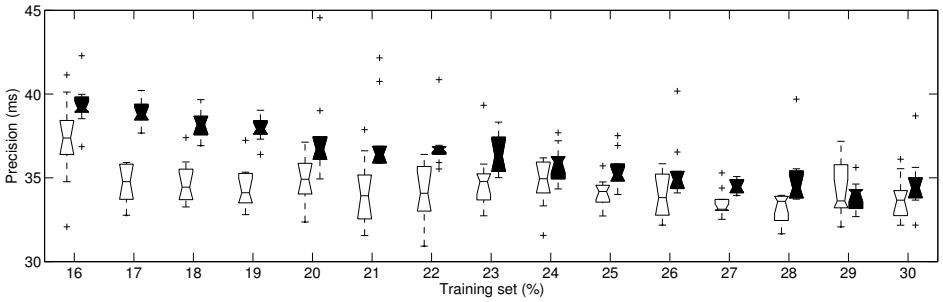
Figure 6: Results for TKMEANS and TCLUST methods. Precision of TKMEANS (white) and TCLUST (black) with respect to the number of clusters (a). Precision of TKMEANS (white) and TCLUST (black) with respect to the training set size (b).

and k-means based selection strategies. The median of the accuracy was close to 0 ms for all methods which suggests that this behavior is inherent to the MLP network and it does not depend on the selection of the training set. Both, TKMEANS and simple k-means had almost the same accuracy (TKMEANS is slightly better) while TCLUST was the most disperse approach in terms of accuracy. The precision parameter clearly pointed out the advantages of robust clustering selection over random and simple k-means, see Fig. 5b. On the other hand, again TKMEANS shown the lowest Inter Quartile Range (IQR) while TCLUST based selection shows the best precision (median).

Experiment II compared both trimmed k-means approaches. The findings suggested that generally TCLUST yields better performances in terms of the median of the precision. Notwithstanding, also noticeable results arose with respect to the number of clusters and the training set sizes. TKMEANS based selection seems to have lower dispersion when the number of clusters is in a central range, i.e., from four to seven. On the other hand, TCLUST improves its performance when the number of clusters increases, Fig.6a. The median value of precision for TKMEANS is between 40.1 ms and 42.3 ms while TCLUST varies in the range 37.8 ms to 44.4 ms. It is also noticeable that both methods pointed out that small values for the number of clusters (2 and 3) are unsuitable. It is not strange that the performance improves as the training set size increases for both methods, Fig. 6b. Nevertheless, for lower percentages (16% - 19%) TCLUST shows



(a)



(b)

Figure 7: Performance indexes for random (white) and Rényi entropy (black) selection strategies. Random and Rényi entropy selection accuracy (a) and Random and Rényi entropy selection precision (b).

a lower median than TKMEANS. This could mean that the scatter structure of the clusters tend to be less spherical for small training sets. Thus, the use of the heterogeneous clustering algorithm offers more advantages than TKMEANS. However, TKMEANS shows lower IQR than TCLUST. The former is also more independent from the number of clusters than TCLUST. Although the median precision of TCLUST is slightly better than TKMEANS, these differences are not statistically significant. In order to point out whether a method is overall the best, the authors would recommend to use TKMEANS rather than TCLUST due to the computational cost increase that implies the use of TCLUST instead of TKMEANS. Finally, from Fig. 6b it is clear that the training set size parameter is of more relevance in the performance of the detection algorithm than the number of clusters.

Experiment III evaluated FS-LSSVM as regression functions. In terms of accuracy, the Rényi entropy selection outperforms the random selection, Fig. 7a. It is noticeable that for the random selection approach, the precision barely depends on the training set size. This is a direct consequence of the initial randomness of the selection process. The a-posteriori refinement of the training set using Rényi entropy criterion does not allow the possibility to include vectors outside the selected training set which may increase the information potential, i.e., the information potential is upper-bounded beforehand. This also explains the larger values in the accuracy parameter of the random selection approach. The precision parameter analysis pointed out that the random selection had better precision than Rényi entropy based selection. Thus, the random

Table 2: QTDB Recording stratification according to Te accuracy and precision for Lead 1, (T): amount of records in each group, (%): percentage with respect to the total amount of records (103). Boldfaced values represent the best results for each group.

Method	**Group I**		**Group II**		**Group III**		**Group IV**	
	(T)	**(%)**	**(T)**	**(%)**	**(T)**	**(%)**	**(T)**	**(%)**
Zhang [8]	**69**	**67.0**	8	7.8	**15**	**14.6**	11	10.6
Trapezium [9]	59	57.3	5	4.9	24	23.3	15	14.5
Wavelet-ECG [7] [19]	63	61.2	**2**	**2.0**	27	26.2	11	10.6
RE+FS-LSSVM 15% **(this study)**	66	64.1	19	18.4	**15**	**14.6**	**3**	**2.9**

Table 3: QTDB recording stratification according to Te accuracy and precision for both leads, (T): amount of records in each group, (%): percentage with respect to the total amount of records (103).

Method	**Group I**		**Group II**		**Group III**		**Group IV**	
	(T)	**(%)**	**(T)**	**(%)**	**(T)**	**(%)**	**(T)**	**(%)**
Zhang [8]	98	96.1	5	3.9	0	0.0	0	0.0
Trapezium [9]	88	85.4	1	1.0	12	11.7	2	1.9
Wavelet-ECG [7] [19]	86	83.5	3	2.9	12	11.7	2	1.9
RE+FS-LSSVM 15% **(this study)**	103	100	0	0.0	0	0.0	0	0.0

selection strategy yielded also acceptable results, although worser accuracies than in the Rényi entropy strategy might be expected.

The comparison between MLP methods and FS-LSSVM is presented in Fig. 8. Except for small training set sizes (16%-18%) the mean accuracies rapidly tends to be in a band between -2 and 2 for both MLP methods. Conversely, RE+FS-LSSVM is always stable around 2 ms of mean error. This could be a consequence of the stronger and more efficient regularization mechanisms of FS-LSSVM approach. On the other hand, the MLP approaches are highly sensitive to the training set size. Thus, performances better than $\sigma_e \leq 40$ ms are achievable only when the training set sizes are larger than 25% of the whole dataset. The previous discussion points out the FS-LSSVM as the best global approach for detecting the T-wave end for small training set sizes.

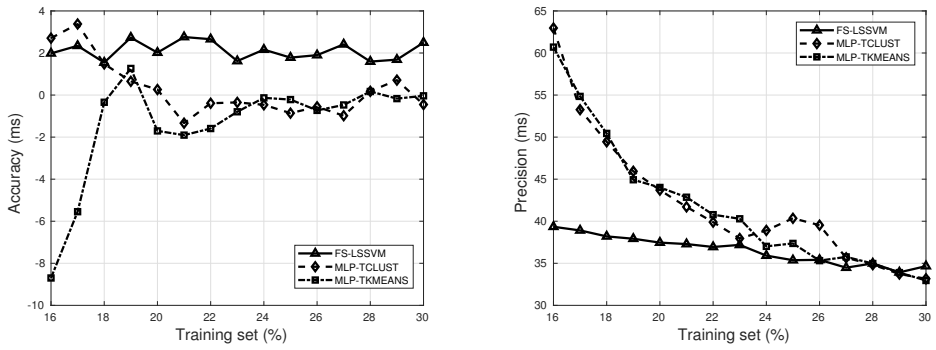
Table 1 points out that the algorithm based on RE+FS-LSSVM outperforms the other three methods in terms of precision for both cases, i.e. when the first lead in the record is used and when both leads are used. The noticeable differences among both data columns in Table 1 are explained by the differences in the method for evaluating the error. Regarding the evaluation using one lead, the error is defined as the difference between the output of the algorithm and the annotated position. However, when both leads are used, the reported error for each beat is given by the smallest difference between the outputs of the algorithm for each lead and the corresponding annotation. In other words, the best Te estimation is selected for each beat. Thus, with this criterion, it is expected that the respective performances of the algorithms increase. This criterion is also known as best-beat-per-cardiac-cycle and it is a standard approach for reporting results of T wave end detection algorithms. This table is given only for the testing set of the algorithm, so there's no influence of the training set in the results and the generalization capabilities of the FS-LSSVM can be clearly appreciated. The mean error value is overall the best only when the first

Table 4: Te detection algorithms performance comparison.

Study	($\mu_e \pm \sigma_e$ (ms))
Madeiro et al., 2013 [11]	2.8 ± 15.3
A. Martínez et al., 2010 [13]	5.8 ± 22.7
Lin et al., 2010 [12]	4.3 ± 20.8
Ghaffari et al., 2009 [10]	0.8 ± 10.7
Zhang et al., 2006 [8]	0.3 ± 17.4
Martínez et al., 2004 [7]	-1.6 ± 18.1
Vila et al., 2000 [6]	0.8 ± 30.3
This study	-3.0 ± 16.9

lead is considered. However, it remains below 1 sample (4 ms @250 Hz) which is an expression of the robustness of the method.

Data shown in tables 2-4 should be interpreted with caution because the training set output was included for computing those indexes. The main issue is that it is too difficult to compute those indexes without including the training set. Hence, such tables are given as reference because they are two recommended approaches for evaluating Te detection performance algorithms. Thus, they provide a general idea on the performance of the proposed approach with respect to other studies. Besides, it is worth to note that despite the indexes on Tables 2-4 were computed including the output of the algorithm for the training set, due to the implicit regularization mechanism of FS-LSSVM, no zero output error it is expected for the training set. Considering these facts a brief discussion on these data is presented below.



(a) (b)

Figure 8: Comparison between methods, TKMEANS+MLP, TCLUST+MLP and RE + FS-LSSVM. Accuracy (a) and precision (b).

Tables 2 and 3 show the performance of the algorithms using the criterion recommended by the CSE Working Party. Regarding this criterion, for the records in Group I the algorithm performs well i.e. the output of the algorithm is comparable to the output produced by a cardiologist. The output of the algorithm for the records in Group II has a large bias and acceptable precision.

This is interpreted as the method has a significant systematic error (offset). Conversely, Group III shows a low bias output but high standard deviation, so the algorithm has a significant random error. Finally, Group IV corresponds to records where the output is considered unpredictable (both, the bias and the standard deviation of the error are large). Using only the first lead (see Table 2), Zhang's algorithm is the one which has good performance in the largest number of records (69) followed by the FS-LSSVM approach (66). On the other hand, the algorithm using FS-LSSVM shows the smallest number of records in Group IV (2). This means that the proposed method is very robust since unpredictable output was observed for only 3 records in the database. Table 3 confirms the latter showing that for all records FS-LSSVM has a Group I type output if both leads are used by the algorithm. Finally, Table 4 include several studies on Te detection. As one can see, the proposed method lead to state-of-the-art Te detection algorithm.

5. Conclusions

This study has shown a novel approach for detecting the end of the T-wave in ECG using neural networks and Support Vector Machines. It was also shown that training set selection strategies can reduce the amount of patterns required to obtain acceptable performances. The findings suggest that for MLP-based methods the required training set size is around 25% while for FS-LSSVM even a 15% of the global dataset for training is enough to obtain more than acceptable performances. FS-LSSVM shows the best performance on precision parameter (σ_e). It also shows high stability, even for small training set sizes. Another major finding was that robust clustering-based and Active PV selections of the training set are more effective than others, e.g., random, simple k-means or other *ad-hoc* selections.

Although the algorithms proposed here requires training, the results show an increase in the precision of the output which leads to an increment on the reliability of the algorithm from the point of view of the user. This is specially relevant in the case of a test where a large number of Te points have to be determined. Another interesting property of the proposed approach is that the training set size is the only one parameter which is left to the user choice. The rest of the parameters are internally determined by the algorithm. Thus, the user does not have to deal with a number of difficult to understand parameters. Conversely, she/he can decide beforehand the number of heartbeats that she/he is willing to manually annotate and the rest of the process can be done automatically. Although 15% may be seen as a too high percentage, actually this is a relatively small training set size and the proposed method have acceptable performances. Besides, it can be argued that the method has been evaluated considering that all heartbeats, coming from 103 different records (also coming from different original databases) belong to a single record. In other words, it is unlikely to find the diversity that QTDB presents in a long-term record. Hence, in practice smaller training set sizes can be used with good results. Additionally, the method may be combined with unsupervised approaches such as those ones discussed above. In this hybrid approach, the unsupervised method could be added to the preprocessing stage in order to provide an initial guess of the Te for the heartbeats selected for training. Thus, the expert would have only to check and correct the output of the unsupervised method for a small number of heartbeats. This would lead to a decrease in the time needed for manually annotating the training set.

Despite unsupervised approaches may be more suitable for short-term analysis, this is not the case for long-term recordings where both, signal quality and morphology change. In fact, in

long-term analysis, the latter is the common scenario and usually, a considerable amount of time is spent in correcting the T-wave end detection errors of unsupervised approaches. Tunning the parameters of these methods is not always possible neither intuitive. Moreover, when tunning is available, it frequently leads to poorer results. On the other hand, the proposed approach depends on both the cardiologist expertise and the number of heartbeats the user is willing to annotate. So, the main question is... What would the user prefer to do? To correct an unknown number of erroneous detections or to be proactive and to annotate a small number of heartbeats in order to gain in robustness and confidence. The results of this work open a new alternative for cardiologists that prefer to annotate instead of correcting.

Acknowledgment

This work has been supported by the Belgian Development Cooperation through VLIR-UOS (Flemish Interuniversity Council-University Cooperation for Development) in the context of the Institutional University Cooperation programme with Universidad de Oriente.
Carolina Varon is a postdoctoral fellow of the Research Foundation-Flanders (FWO).
The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (nº339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

References

[1] W. Zareba, Drug induced QT prolongation, Cardiology Journal 14 (6) (2007) 523–533.
[2] S. Nachimuthu, M. D. Assar, J. M. Schussler, Drug-induced QT interval prolongation: mechanisms and clinical management, Therapeutic Advances in Drug Safety 3 (5) (2012) 241–253.
[3] J. Nielsen, F. Wang, C. Graff, J. K. Kanters, QT dynamics during treatment with sertindole, Therapeutic Advances in Psychopharmacology 5 (1) (2015) 26–31.
[4] F. S. Riad, A. M. Davis, M. P. Moranville, J. F. Beshai, Drug-induced QTc prolongation, Am J Cardiol 119 (2017) 280–283.
[5] F. Monitillo, M. Leone, C. Rizzo, A. Passantino, M. Iacoviello, Ventricular repolarization measures for arrhythmic risk stratification, World Journal of Cardiology 8 (1) (2016) 57–73.
[6] J. A. Vila, Y. Gang, J. M. R. Presedo, M. Fernández-Delgado, S. Barro, M. Malik, A new approach for TU complex characterization, IEEE Transactions on Biomedical Engineering 47 (6) (2000) 764–772.
[7] J. P. Martínez, R. Almeida, S. Olmos, A. P. Rocha, P. Laguna, A wavelet-based ECG delineator: evaluation on standard databases, IEEE Transactions on biomedical engineering 51 (4) (2004) 570–581.
[8] Q. Zhang, A. I. Manriquez, C. Médigue, Y. Papelier, M. Sorine, An algorithm for robust and efficient location of T-wave ends in electrocardiograms, IEEE Transactions on Biomedical Engineering 53 (12) (2006) 2544–2552.
[9] C. R. Vázquez-Seisdedos, J. E. Neto, E. J. Marañón-Reyes, A. Klautau, R. C. Limão de Oliveira, New approach for T-wave end detection on electrocardiogram: Performance in noisy conditions, Biomedical engineering online 10 (1) (2011) 1.
[10] A. Ghaffari, M. Homaeinezhad, M. Akraminia, M. Atarod, M. Daevaeiha, A robust wavelet-based multi-lead electrocardiogram delineation algorithm, Medical engineering & physics 31 (10) (2009) 1219–1227.
[11] J. P. Madeiro, W. B. Nicolson, P. C. Cortez, J. A. Marques, C. R. Vázquez-Seisdedos, N. Elangovan, G. A. Ng, F. S. Schlindwein, New approach for T-wave peak detection and T-wave end location in 12-lead paced ECG signals based on a mathematical model, Medical engineering & physics 35 (8) (2013) 1105–1115.
[12] C. Lin, C. Mailhes, J.-Y. Tourneret, P-and T-wave delineation in ECG signals using a Bayesian approach and a partially collapsed Gibbs sampler, IEEE transactions on biomedical engineering 57 (12) (2010) 2840–2849.
[13] A. Martínez, R. Alcaraz, J. J. Rieta, Application of the phasor transform for automatic delineation of single-lead ECG fiducial points, Physiological measurement 31 (11) (2010) 1467.

[14] A. Zifan, S. Saberi, M. H. Moradi, F. Towhidkhah, Automated ECG Segmentation Using Piecewise Derivative Dynamic Time Warping, International Journal of Biological and Life Sciences 1 (3) (2005) 181–185.
[15] W. Bystricky, A. Safer, Modelling T-end in Holter ECGs by 2-layer perceptrons, in: Computers in Cardiology, 2002, IEEE, 2002, pp. 105–108.
[16] C. Rahime, Ö. Yüksel, Comparison of fcm, pca and wt techniques for classification ecg arrhythmias using artificial neural network, Expert Systems with Applications 33 (2) (2007) 286–295.
[17] J. A. Suykens, J. Vandewalle, Least Squares Support Vector Machine classifiers, Neural processing letters 9 (3) (1999) 293–300.
[18] P. Laguna, R. G. Mark, A. Goldberg, G. B. Moody, A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG, in: Computers in Cardiology 1997, IEEE, 1997, pp. 673–676.
[19] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, H. E. Stanley, Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220.
[20] A. I. Manriquez, Q. Zhang, An algorithm for QRS onset and offset detection in single lead electrocardiogram records, in: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2007, pp. 541–544.
[21] N. Ahmed, T. Natarajan, K. R. Rao, Discrete cosine transform, IEEE transactions on Computers 100 (1) (1974) 90–93.
[22] K. R. Rao, P. Yip, Discrete cosine transform: algorithms, advantages, applications, Academic press, 2014.
[23] C. R. Vázquez-Seisdedos, A. A. Suárez-León, J. E. Neto, A comparison of different classifiers architectures for electrocardiogram artefacts recognition, in: J. Ruiz-Shulcloper, G. S. di Baja (Eds.), Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Vol. 8259 of Lecture Notes in Computer Science, Springer, 2013, pp. 254–261.
[24] J. A. Cuesta-Albertos, A. Gordaliza, C. Matrán, Trimmed k-means: An attempt to robustify quantizers, The Annals of Statistics 25 (2) (1997) 553–576.
[25] L. A. García-Escudero, A. Gordaliza, C. Matrán, A. Mayo-Iscar, A general trimming approach to robust cluster analysis, The Annals of Statistics (2008) 1324–1345.
[26] M. T. Gallegos, Maximum likelihood clustering with outliers, in: Classification, Clustering, and Data Analysis, Springer, 2002, pp. 247–255.
[27] M. T. Gallegos, G. Ritter, A robust method for cluster analysis, Annals of Statistics (2005) 347–380.
[28] J. A. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, Least Squares Support Vector Machines, World Scientific Pub. Co., 2002.
[29] C. Williams, M. Seeger, Using the Nyström method to speed up kernel machines, in: Proceedings of the 14th annual conference on neural information processing systems, no. EPFL-CONF-161322, 2001, pp. 682–688.
[30] K. De Brabanter, J. De Brabanter, J. A. Suykens, B. De Moor, Optimized fixed-size kernel models for large data sets, Computational Statistics & Data Analysis 54 (6) (2010) 1484–1504.
[31] D. Erdogmus, J. C. Principe, Generalized information potential criterion for adaptive system training, IEEE Transactions on Neural Networks 13 (5) (2002) 1035–1044.
[32] M. Girolami, Orthogonal series density estimation and the kernel eigenvalue problem, Neural computation 14 (3) (2002) 669–688.
[33] T. Van Gestel, J. A. Suykens, G. Lanckriet, A. Lambrechts, B. De Moor, J. Vandewalle, Bayesian framework for Least-Squares Support Vector Machine classifiers, Gaussian processes, and kernel Fisher Discriminant Analysis, Neural computation 14 (5) (2002) 1115–1147.
[34] B. W. Silverman, Density estimation for statistics and data analysis, Vol. 26, CRC press, 1986.
[35] S. J. Sheather, M. C. Jones, A reliable data-based bandwidth selection method for kernel density estimation, Journal of the Royal Statistical Society. Series B (Methodological) (1991) 683–690.
[36] K. Pelckmans, J. A. Suykens, T. Van Gestel, J. De Brabanter, L. Lukas, B. Hamers, B. De Moor, J. Vandewalle, LS-SVMLab: a MATLAB/C toolbox for least squares support vector machines, Tutorial. KULeuven-ESAT. Leuven, Belgium.
[37] The CSE Working Party, Recommendations for measurement standards in quantitative electrocardiography, European Heart Journal 6 (10) (1985) 815–825.