

# Distinguishing Between Supply Ischaemic and Non-Supply Ischaemic ST Events using a Relevance Vector Machine

CB Vilakazi, L Tarassenko, GD Clifford

Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom

## Abstract

*In this paper, we apply a sparse Bayesian learning algorithm called the Relevance Vector Machine (RVM) which was used to classify the 1126 ischaemic ST events and 1126 non-supply ischaemic ST events in the Long Term ST Database as supply or non-supply ST episodes. A Genetic Algorithm (GA) method was used to identify which of the extracted features used as input to the RVM were the most important with respect to the model's performance. The GA indicated that 9 of the 35 extracted features were the most relevant. The 9 features that were selected are heart rate variability, slope of the ST segment, energy in the QRS complex and Mahalanobis distance of the first five Karhunen Loève Transform of the QRS complex and ST segment for differentiation between supply and non-supply ischaemic ST episodes. The classification accuracy achieved using the 35 features was 80.1% on the test set. When using the 9 most relevant features determined from the GA, the classification accuracy rose to 87.4%.*

## 1. Introduction

Myocardial ischaemia is the leading cause of death in the industrialized countries hence early diagnosis and treatment are very important [1]. Myocardial ischaemia stems from insufficient supply of blood to the myocardium due to blockages in the coronary artery. Myocardial ischaemia can be defined as an imbalance between oxygen/nutrient delivery with regard to myocardial requirements. Supply ischaemia results from a partial occlusion of a coronary artery, reducing the amount of oxygenated blood to the myocardium. The term demand ischaemia refers to a condition where an increased oxygen demand caused by exercise, tachycardia or emotion, leads to a transitory imbalance [1].

ST segment changes provide a sensitive marker for early diagnosis of myocardial ischaemia in ECG recording [2]. It is also known that changes in the ST segment can result from a wide variety of other causes such as changes

in heart rate, conduction pattern, position of the subject, and noise in the ECG. Heart rate-related and ischaemic ST events are characterized by length and extremum deviation. In contrast, body position changes and conduction changes events are characterized by a sudden shift in the ST level and the time at which they occur [2]. Heart-rate related, body position and conduction changes events can be grouped together as non-supply ischaemic or non-ischaemic ST events [3]. Several automatic transient ST segment deviation techniques using various approaches have been published using the Long-Term ST Database (LTST DB) [2–6]. In 2003, the Physionet/Computers in Cardiology Challenge [7] aim was to classify ST changes as ischaemic or non-supply ischaemic using available annotated records of the LTST DB.

In this paper, we introduce the use of the Relevance Vector Machine (RVM) for classification of supply and non-supply ischaemic ST episodes. In addition, we introduce the use of Genetic Algorithm (GA) for feature selection. The remainder of this paper is structured as follows: The next section contains a brief description of the technique used to generate and select the features used to train the RVM to differentiate between supply and non-supply ischaemic ST events. The third section presents the experimental results of applying this technique. The last section focuses the discussion and conclusion.

## 2. The methodology

The LTST DB consists of 86, two or three leads, 21 to 24 hour, Holter ECG recordings sampled at 250Hz. The gold standard for annotating ST episodes was based not only on ECG waveforms but also on detailed clinical information from the subjects and the decision of expert annotators of the database. In the LTST DB non-supply ischaemic ST events such as heart-rate related events, body position and conduction changes were also annotated [2].

An episode begins when the magnitude of the ST deviation first exceeds a lower annotation detection threshold,  $V_{lower}$ . The deviation then must reach or exceed an upper annotation detection threshold  $V_{upper}$  throughout a contin-

uous interval of at least  $T_{min}$  seconds. The episode ends when the deviation becomes smaller than  $V_{lower}$ , provided that it remains below  $V_{lower}$  in the following  $T_{sep} = 30$  seconds. Episode annotations of the LTST DB are available in three variant annotation protocols:

1. Protocol A  $V_{upper} = 75 \mu\text{V}$ ,  $T_{min} = 30$  seconds.
2. Protocol B  $V_{upper} = 100 \mu\text{V}$ ,  $T_{min} = 30$  seconds.
3. Protocol C  $V_{upper} = 100 \mu\text{V}$ ,  $T_{min} = 60$  seconds.

In this work, we used annotation protocol B to differentiate between supply and non-supply ischaemic ST episodes as the limits of this protocol are very similar to what clinician used in clinical practise. Annotation protocol B consists of 1130 ischaemic ST events, 234 heart-rate related ST events and 2388 axis-shift ST episodes. In order to balance the two classes of ischaemic and non-supply ischaemic episodes, only 886 of the axis shift episodes were chosen (at random) in addition to the 234 heart rate-related episodes. The available data was divided into two subsets; the training and test dataset. The training dataset consisted of 60% of the data and the remaining 40% was allocated to the test dataset. Before training a classifier, it is beneficial to transform the features to have the same sample mean and variance and this was done using zero-mean unit-variance normalisation.

To estimate how the values of diagnostic and morphological parameters change during ST episodes, mean values for each of the parameters, over 20 seconds interval, located 20 ms before the beginning ( $I_1$ ), 20ms after the beginning ( $I_2$ ), of each ST episode were used. The changes in the mean feature value in interval  $I_1$  with respect to  $I_2$  were computed and included as indices to be used for discriminant analysis. This means that for each feature, three measures were computed and subscript by the 1, 2,  $\Delta 12$ ) when describing the features. The performance metrics used to test the performance of the technique were: sensitivity (Se), specificity (Sp), positive predictivity value (+PV), negative predictivity value (-PV) and accuracy (Acc).

## 2.1. Feature extraction

Before generating the features, the ECG was preprocessed as follows. The high-frequency powerline noise was removed using a notch filter, R-peaks were detected using a QRS detector, baseline wander was removed using cubic spline interpolation to enable easy location of the isoelectric line. Abnormal beats were removed using a template matching algorithm that kept the beats only if the correlation coefficient between the beats and the template was greater than 0.95. Delineation of the ECG was achieved using a thresholding technique. After extracting the features, a GA was used for feature selection to select the most relevant subset of features and this is described in the next section.

*Repolarization features* - The features used to build a discriminant analyser between supply and non-supply ST episodes were inspired by [3]. The ST segment morphology features extracted from the ST episodes were the ST segment deviation and slope. The ST level,  $ST$ , was measured at the point  $J + 60(80)$  (ms) while ST segment slope,  $SL$ , was measured from  $J$  to  $J + 20ms$  using  $SL_i = J_i + 20ms$ , where  $i$  is the beat number. In addition the root mean square (RMS) of the ST segment shape change was used to estimate the morphology change of the entire ST segment from  $J$  to  $J + 120$  ms. Lastly, the Mahalabonis distance of the first five Karhunen Loève Transform (KLT) coefficients ( $ST_{MD}$ ) from the ST segment were used [8].

*Depolarization features* - As changes also occur in the depolarization phase (QRS complex) of the ECG during acute ischaemia, features from the depolarization phase were also extracted. The upslope ( $QRS_U$ ) and downslope ( $QRS_D$ ) of the R-wave were extracted from the ECG using a modified technique used by [9]. The features that were selected for estimating the QRS morphology change along the ST episodes were the RMS value of the signal from the onset to the peak of the QRS complex ( $QR_{RMS}$ ), and the RMS value of the signal from the peak to the offset of the QRS complex, ( $RS_{RMS}$ ). The duration of the QRS complex,  $QRS_{DUR}$  complex, was taken to be from the onset to the offset of the QRS complex. In addition the Mahalabonis distances of the first five KLT coefficients from the QRS segment ( $QRS_{MD}$ ) were used.

*Heart rate-related features* - Heart rate was extracted to add to the feature vector to use to build a classification system. In addition, heart rate variability (HRV) measures were computed, and the two measures to be included in the feature vector were Standard Deviation of Sequential Five-Minute R-R Interval Means (SDANN) and LF/HF-ratio. These measures were computed using a five minute window before the start of an episode. The LF/HF-ratio was obtained from the power spectral density estimated using the Lomb-Scarlge periodogram which requires no resampling of unevenly-sampled signals [10].

## 2.2. Feature selection

GA is a family of heuristic algorithms used as an optimization technique that mimics the mechanisms of evolution observed in nature selecting individuals with respect to a given fitness function and 'breeds' them over many generations to find the best set of parameters with respect to the given fitness function [11].

The GA starts from a population of randomly generated individuals. Each genome is evaluated by first fitting a multivariate linear regression model to fit the training set and validated on the validation set (which in this case is taken as a 30% subset of our training data). As the re-

sults may vary according to how the training set is drawn from the data, the fit function used for the selection of best genomes is defined by the median of the root mean square error (RMSE). The selection process used in this work is elitist such that the 10 best genomes with the lowest RMSE are passed down to the next generation unaltered.

The next step is to generate a second-generation population of solutions from those selected through genetic operators: crossover, and mutation. Each of the 35 genomes is crossed with one of the first 35 genomes that maximize their Hamming distance between them and generates 70 children of the next generation. A mutation rate of 1.5% was applied randomly to 20% of the 35 children generated by cross-over. The new population is then used in the next iteration of the algorithm. The algorithm terminates after 100 iterations or an  $RMSE < 0.05$  in terms of the classification error on the validation set is reached.

### 2.3. Classification of ST events

The Support Vector Machine (SVM), although is a powerful classifier has a disadvantage that its output is a binary classification decision and not the class membership posterior probability. Tipping introduced the RVM, which is a special case of a sparse Bayesian learning algorithm [12]. RVM does not only classify a new input variable, but can also provide a degree of uncertainty for the classification.

Consider a two-class problem with training data with  $n$  number of samples represented by  $X = (x_1, \dots, x_n)$  having class labels  $C = (c_1, \dots, c_n)$  with  $c_i \in (0, 1)$ . Define a classifier function  $y(x) = w^T \phi(x)$ , where  $\phi$  is a continuous feature-space transformation and  $w$  is a weight vector. The RVM has the special form (similar to the SVM algorithm) given by  $\sum_{n=1}^N w_n \cdot k(x, x_n) + b$  where  $k(\cdot)$  is the kernel function and  $b$  is a bias parameter. Based on the Bernoulli distribution, the likelihood is computed as:

$$p(c/w) = \prod_{i=1}^n \sigma(y(x_i))^{c_i} [1 - \sigma(y(x_i))]^{1-c_i} \quad (1)$$

Where  $\sigma(y)$  is the logistic sigmoid function given by  $\sigma(y(x)) = \frac{1}{1 + \exp(-y(x))}$ . To obtain  $p(c/w)$ , an iterative method is used. Let  $\alpha_i^*$  denotes the maximum a posteriori estimate of the hyperparameter  $\alpha_i$ . The maximum a posteriori estimate of the weights ( $W_m$ ) can be obtained by maximizing the following objective function:

$$f(w_n) = \sum_{i=1}^n \ln p(c_i/w_i) + \sum_{i=1}^n \ln p(w_i/\alpha_i^*) \quad (2)$$

Where the first summation term corresponds to the likelihood of the class labels, and the second term corresponds to the prior on the parameter  $w_i$ . In the final solution, the gradient of objective function  $f$  with respect to  $w$

is calculated and only those training data points having nonzero coefficients (relevance vectors) contribute to the decision function. The posterior probability is approximated around  $W_m$  by a Gaussian approximation with covariance  $\Sigma = -(H|_{W_m})^{-1}$  and mean  $\mu = \Sigma \Phi^T B$ , where  $H$  is the Hessian of  $f$ , matrix  $\Phi$  has elements  $\phi_{ij} = k(x_i, x_j)$  and  $B$  is a diagonal matrix with elements defined by  $\sigma(y(x_i))[1 - \sigma(y(x_i))]$ . An iterative procedure is used to find the set of weights to maximize the function (2) in which the hyperparameters  $\alpha_i$ , associated with each weight are updated. A radial basis function kernel was used to train the RVM.

### 3. Experimental results

To ensure that the optimal subset of features was selected, the GA was run 50 times and the best model was selected as the final model. The 9 features selected by the GA are shown in table 1 along with their mean values, standard deviation and p-values obtained using one-way Analysis of Variance (ANOVA). The performance metrics of the best model on the test set using a linear regression model were: Sp = 83.4% , Se= 80.7%, +PV= 80.1%, -PV= 80.5% and Acc =82.1% .

Table 1. List of the selected features by GA along with their mean ( $\mu$ ) and standard deviation ( $\sigma$ ) along with ANOVA p-values according to the annotation protocol B.

Variables	Supply ( $\mu \pm \sigma$ )	Non-supply ( $\mu \pm \sigma$ )	p-values
$STRMS_1$	85.3±69.6	133.6±101.7	$3.4 \times 10^{-18}$
$HR_{\Delta 12}$	12.3±12.5	10.6±10.4	$1.4 \times 10^{-12}$
$QRSMD_{\Delta 12}$	352.3±231.2	259.6±114.8	$7.0 \times 10^{-11}$
$SL_{\Delta 12}$	99.2±67.3	129.2±81.2	$3.9 \times 10^{-9}$
$ST_2$	223.2±166.6	151.5±138.1	$1.7 \times 10^{-9}$
$QRSU_{\Delta 12}$	88.9±47.2	131.1±66.2	$4.2 \times 10^{-8}$
$STMD_{\Delta 12}$	129.2±97.7	81.0±47.7	$2.2 \times 10^{-5}$
$RSRMS_{\Delta 12}$	271.3±129.5	306.6±129.8	$1.8 \times 10^{-4}$
LF/HF-ratio	0.5±0.4	0.7±0.5	$2.5 \times 10^{-1}$

The RVM was first trained using the full set of 35 features to discriminate between supply and non-supply ischaemic ST episodes using a linear kernel. In addition to using a full set of features, the 9 most relevant features as selected by GA were also used to train and test the RVM. The RVM models were tested using the test set and the results achieved are shown in Table 2. The results show that using 9 selected features achieved better performance than using the full set of 35 features.

Table 2. Performance of RVM models trained using 9 subset of features selected by GA and the full set of features.

	Se (%)	Sp (%)	+PV (%)	-PV (%)	Acc (%)
RVM-35	78.5	81.8	72.6	81.7	80.1
RVM-9	88.7	86.8	86.0	86.6	87.4

The posterior probability of the RVM was then used to

select the decision boundary with minimum overlap between correctly and incorrectly classified ST episodes for both ischaemic and non-supply ischaemic classes as shown in Figure 1. The decision boundary for ischaemic and non-supply ischaemic classes were chosen to be at the point of overlap,  $p = 0.35$  and  $0.61$ , respectively.

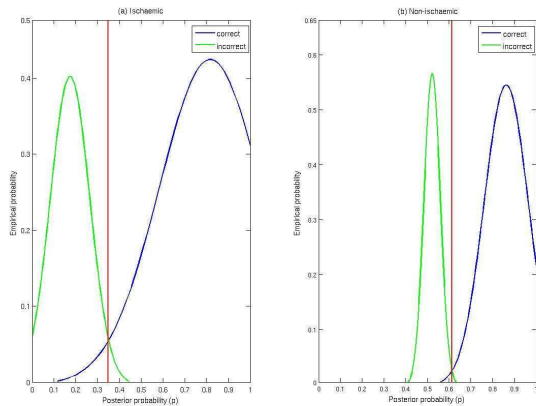


Figure 1. Gaussian approximations to the empirical probability distributions for the posterior probability of correctly and incorrectly classify ST episodes as ischaemic(a) or non-supply ischaemic(b) classes along with lower decision boundaries (red vertical lines at  $p = 0.35$  and  $0.61$  respectively) to indicate if the classification should be trusted.

#### 4. Discussion and conclusion

In this paper, the problem of distinguishing between supply and non-supply ischaemic ST events using features extracted from the ECG was addressed. A subset of the most relevant features was chosen using a GA and the selected features were then used to train an RVM to classify supply and non-supply ischaemic ST episodes. The results achieved using RVM with a reduced subset of features are comparable with results achieved using other techniques in the literature. The top scoring entry of the PhysioNet challenge [4] achieved a performance of Se and +PV of 99.0% and 88.8% respectively on the training data set. Mincholé *et al.* [3] used all the protocol B ST events and employed several features such as heart-rate changes and morphology-change features from the ST segment along with linear discriminant analysis. Their technique obtained a Se of 74.5% and a Sp of 93.2% when discriminating between ischaemic and non-supply ischaemic ST events. On the same dataset, Dranca *et al.* [5] used a mixture of time-domain analysis and various machine learning techniques such as decision trees to distinguish between ischaemic and heart-rate related ST events, and achieved a Se of 85.7% and +PV of 61.2% using protocol B.

The use of a feature selection tool allowed us to remove features which only appeared to add noise to the analy-

sis. Moreover, the GA revealed that a particular HRV metric, the LF/HF-ratio (but not the SDANN) was also of use. Since this metric is known to be susceptible to changes in the autonomic nervous system, it may be hypothesised that the inclusion of this metric was related to pain or stress rather than being highly correlated with ischaemic episodes. The main advantage of the RVM algorithm is that it provides the posterior probability of the class membership. Using a Bayesian interpretation of probability, the probability of an event can be interpreted as the degree of uncertainty associated with such an event. This uncertainty can be used as an additional guide to the clinician to decide whether or not to re-test a subject.

#### References

- [1] Gettes L, Cascio W. The Heart and Cardiovascular System. 2 edition. New York: Raven Press Ltd, 1992; 2021–2054.
- [2] Jager F, Moody GB, Taddei A, Antolic G, Emdin M, Smerdel A, Glavic B, Marchesi C, Mark RG. A long-term ST database for development and evaluation of ischemia detectors. In Proc. Computers in Cardiology. 1998; 301–304.
- [3] Mincholé A, Jager F, Laguna P. Discrimination between ischemic and artifactual st segments events in holter recordings. Biomedical Signal Processing and Control 2010;5:21–31.
- [4] Langley P, Bowers EJ, Wild J, Drinnan MJ, Allen J, Sims AJ, Brown N, Murray A. An algorithm to distinguish ischaemic and non-ischaemic ST changes in the holter ECG. In Proc. Computers in Cardiology. 2003; 239–242.
- [5] Dranca L, Goi A, Illarramendi A. Real-time detection of transient cardiac ischemic episodes from ECG signals. Physiol Meas Sep 2009;30(9):983–998.
- [6] Faganeli J, Jager F. Automatic classification between transient ischaemic and transient non-ischaemic heart-rate related ST segment episodes in ambulatory ECG records. Physiological Measurement 2010;5:323–337.
- [7] Moody GB, Jager F. Distinguishing ischemic from non-ischemic ST changes: the PhysioNet/Computers in Cardiology Challenge 2003. In Proc. Computers in Cardiology. 2003; 235–237.
- [8] Jager F, Mark R, Moody G, Divjak S. Analysis of transient ST segment changes during ambulatory monitoring using the Karhunen-Loève transform. In Computers in Cardiology. 1992; 691–694.
- [9] Pueyo E, Sornmo L, Laguna P. QRS slopes for detection and characterization of myocardial ischemia. IEEE Transaction in Biomedical Eng 2008;55(2):468–477.
- [10] Clifford GD, Tarassenko L. Quantifying errors in spectral estimates of HRV due to beat replacement and resampling 2005;52(4):630–638.
- [11] Goldberg D. Genetic algorithms in search, optimization and machine learning. Addison-Wesley, 1989.
- [12] Tipping M. Sparse Bayesian learning and the Relevance Vector Machine. Journal of Machine Learning Research 2001;1:211–244.