# Supervised Classification Models to Detect the Presence of Old Myocardial Infarction in Body Surface Potential Maps

H Zheng, H Wang, CD Nugent, DD Finlay

School of Computing and Mathematics, University of Ulster, Northern Ireland

## Abstract

*In this study we have investigated the classification of old myocardial infarction through the analysis of 192 lead Body Surface Potential Maps (BSPM). Following an analysis of the most prominent features based on a signal to noise ratio ranking criterion the top 6 features were selected. These features were subsequently used as inputs to a series of supervised classification models in the form of Naïve Bayes (NB), Support Vector Machine (SVM) and Random Forest (RF)-based classifiers. Following 10-fold cross validation it was found that the best performance for each classifier was 81.9% for NB, 82.8% for SVM and 84.5% for RF. The results have indicated the ability of the approach to successfully classify the recordings based on a non standard subset of recording sites from the BSPM.*

## 1. Introduction

The electrocardiogram (ECG) has been established as one of the most widely utilized tools for the non-invasive assessment of cardiac status. Although a number of standard techniques exist in clinical practice it is appreciated that there is benefit in considering alternative electrode placements in an attempt to provide a more representative picture of cardiac activity and hence improve the overall process of patient diagnosis.

One of the most widely accepted standard techniques is the 12-lead ECG. Although a commonly used tool in clinical practice it is appreciated that there are deficiencies when diagnosing a number of cardiac abnormalities [1]-[3]. This is related to the fact that the recording sites are limited to a small area on the precordium.

To offer a potentially improved approach would require techniques with denser spatial sampling for example BSPMs. BSPMs have the ability to record cardiac information from as many as 200 recording sites on both the anterior and posterior surfaces of the torso and hence capture much more information than traditional techniques such as the 12-lead ECG. This provides the potential for improved cardiac diagnosis, however, the large amounts of information which must be acquired and subsequently processed poses practical limitations. In addition, the different format for presenting the recorded data (contour maps as opposed to scalar traces) can act as a barrier in clinical acceptance as practitioners are extremely comfortable with the traditional techniques.

In the current study we build upon the knowledge gained within the realms of BSPM analysis and aim to identify which isointegral measurements can be used to effectively identify diseased subjects. Analysis of the spatial distributions of the selected isointegral measurements also infers some information about which recording sites are best suited for this type of diagnosis. Specifically, within this work, we adopt a data driven approach both in terms of selecting the most appropriate features and developing classification models. Based on this approach we examine the ability of a number of classification models developed through processes of supervised learning.

## 2. Methods

### 2.1. Study population

The current study utilized a set of 192 lead BSPMs. The data set contained recordings from 116 subjects, 57 of which exhibited evidence of old myocardial infarction (MI) and the remaining 59 were deemed to be normal. The data were recorded at the University of Utah, Salt Lake City, under the supervision of Professor Robert Lux. The recording procedure has previously been described in [4]. During the acquisition process 16 columns of 12 electrodes were placed on the subject's torso. The columns were spaced equally around the thoracic circumference. Figure 1 depicts a schematic of the electrode array.

Simultaneous recordings were taken from the 192 recording sites over a number of seconds all of which were sampled with respect to the Wilson's central terminal. Following the recording of the data for each patient, the data was averaged to render one cardiac cycle. The recordings for each patient were further processed to provide QRS, STT and QRST isointegral values. This provided a set of 576 (3*192) features per patient.

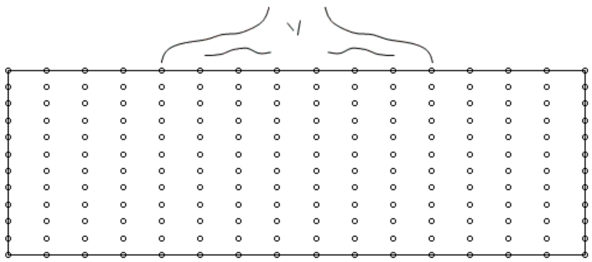

Figure 1. Schematic of 192 electrode array used during the acquisition process. Electrodes are arranged in 16 columns of 12 electrodes equally spaced around the thoracic circumference. For the sake of future discussion these electrodes are numbered from 1 to 16 across the top row (from left to right) with subsequent rows numbered in ascending order (e.g. row 2 = electrodes 17-32, row 3 = electrodes 33-48 and so on).

## 2.2. Feature processing

In the first instance the feature set was ranked based on its ability to discriminate two classes using signal-to-noise ratio (SNR) defined as:

$$SNR_i = (\mu_{i1} - \mu_{i2})/(\sigma_{i1} + \sigma_{i2}) \quad (1)$$

Where $\mu_{i1}$ and $\mu_{i2}$ are the mean values of feature $i$ for the samples from classes 1 and 2, and $\sigma_{i1}$ and $\sigma_{i2}$ are the respective standard deviations.

The value of $SNR_i$ is correlated with the class distinction of interest. A high value indicates there is a strong correlation between the feature value and the class discrimination. Based on this approach a subset of features with high values of $SNR_i$ were selected.

## 2.3. Methods for supervised classification

Three different supervised classification models were investigated: Naïve Bayes (NB), Support Vector Machine (SVM) and Random Forest (RF). These models were implemented within the framework provided by the *Weka* package [5].

The NB classifier is based on Bayesian theorem. In practice NB classifiers have been surprisingly successful and have often outperformed more sophisticated algorithms in varying practical applications [6].

With SVM classification is performed by constructing an *N*-dimensional series of hyperplanes that optimally separates samples into their respective categories. Through the use of a kernel function, an SVM can handle more complex classification problems. In this study, the implementation of the SVM is based on the sequential minimal optimization algorithm developed by Platt [7]. Two kernel functions: polynomial and radial basis functions were used in the SVM-based classification models.

The RF-based classifier is one of the most successful ensemble methods in classification. It consists of a series of classification trees [8]. The final classification of unseen entities is based on the majority votes across all the trees in the forest. In the given study 10 classification trees were deployed.

The quality of each classifier was assessed in terms of four statistical indicators: *overall classification accuracy* (*AC*), *precision* (*Pr*), *sensitivity* (*Se*) and *specificity* (*Sp*). In order to estimate how well classifiers perform on unseen data, a 10-fold cross validation was carried out, i.e. the entire dataset is randomly divided into 10 subsets, 9 folds are put together for training and the other fold is used as the test set.

# 3. Results

## 3.1. Feature ranking

The SNR was calculated for each feature in terms of its discriminative power to separate the two classes in the dataset: subjects with MI and those considered as normal. The 10 top-ranked features are listed in Table 1. Interestingly, there were no QRS features within the top 10 ranked features. Upon closer examination of the results of the entire ranking it was revealed that the first QRS feature was ranked as 199 (QRS153, SNR = 0.459).

Table 1 The description of top-ranked features in terms of the value of SNR

| Ranking | SNR | Feature |
|---|---|---|
| 1 | 0.856 | QRST123 |
| 2 | 0.847 | STT107 |
| 3 | 0.847 | STT123 |
| 4 | 0.824 | QRST139 |
| 5 | 0.814 | STT108 |
| 6 | 0.812 | QRST1 |
| 7 | 0.792 | STT36 |
| 8 | 0.792 | STT124 |
| 9 | 0.789 | QRST155 |
| 10 | 0.789 | STT139 |

The positioning of the electrodes associated with the top 6 features is illustrated in Figure 2.

## 3.2. Supervised classification

To explore the feasibility of using supervised classification models in the diagnosis of MI based on

BSPMs the three previously described supervised models of NB, RF and SVM were evaluated. Table 2 shows the prediction results with the top 6 ranked features. Following 10 fold cross validation it was found that the classification accuracy for each classifier with the top 6 features was 81.9% for NB, 82.8% for SVM and 77.6% for RF. This was in comparison to the same classifiers based on the full feature vector of 576 features attaining accuracies of 77.6%, 75.0% and 78.4% respectively. This suggests that an acceptable level of prediction results could be obtained by using only a small selected subset of features.

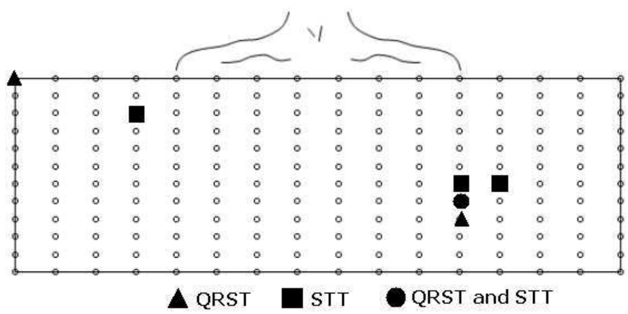


Figure 2 Electrode array indicating the positioning of the 6 selected electrodes following application of the variable ranking approach. The key indicates which isointegrals were placed in the various locations.

Table 2 Classification results for three classifiers using the 6 top-ranked features (QRST123, STT107, STT123, QRST139, STT108, and QRST1) based on 10-fold cross validation

| Model | *AC* (%) | MI Class | | | Normal Class | | |
|---|---|---|---|---|---|---|---|
| | | *Pr* (%) | *Se* (%) | *Sp* (%) | *Pr* (%) | *Se* (%) | *Sp* (%) |
| NB | 81.9 | 81.0 | 82.5 | 81.4 | 82.8 | 81.4 | 82.5 |
| SVM | 82.8 | 86.3 | 77.2 | 88.1 | 80.0 | 88.1 | 77.2 |
| RF | 77.6 | 77.2 | 77.2 | 78.0 | 78.0 | 78.0 | 77.2 |

## 3.3. Feature subset selection

To investigate the effect of the number of features used as inputs to a classification model, classifiers where exposed to ranked subsets of features ranging from the top 3 ranked features to the entire set of 576 features. For each selected subset of features, three classifiers were implemented (NB, SVM and RF). Figure 3 shows the prediction results of the three classifiers based on varying sizes of ranked feature subsets.

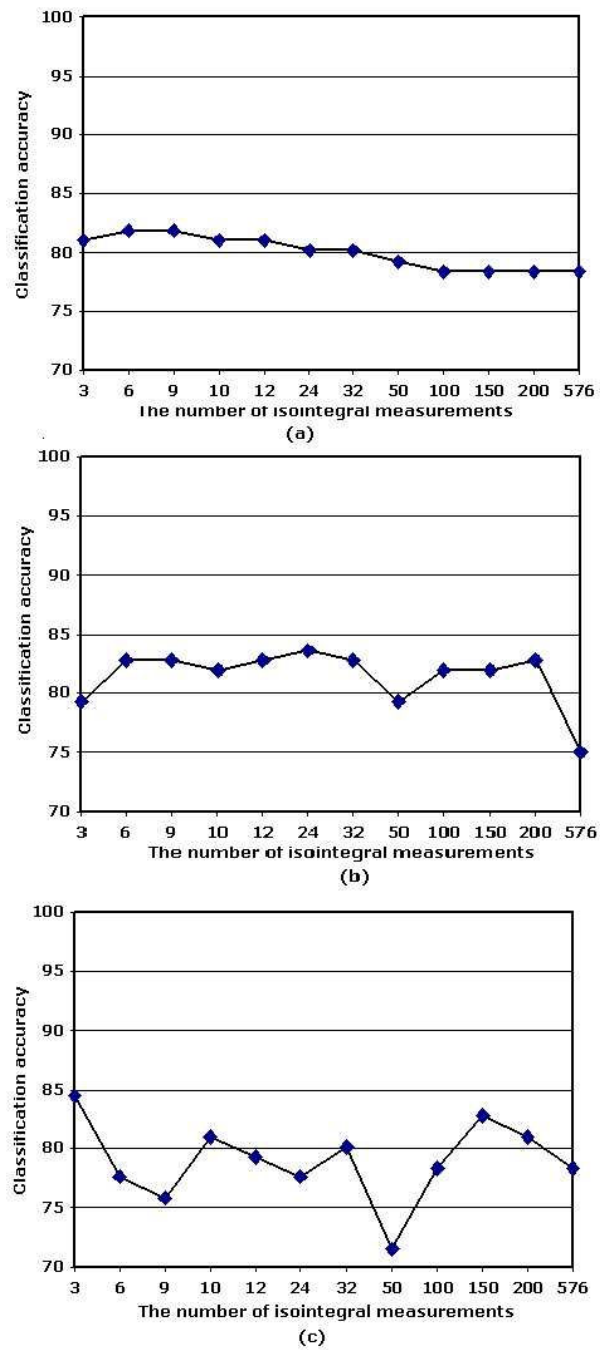


Figure 3 The impact of the number of isointegral measurements on the prediction results of three classifiers: (a) NB; (b) SVM; (c) RF. The classification accuracy is depicted on the y axis. The x-axis represents the number of isointegral measurements used as inputs to each classifier.

Each classifier exhibits different behaviors when exposed to varying numbers of input features. RF-based classification is more sensitive to the number of features than the other approaches. The prediction accuracy of RF is between 71.6% and 84.5%, while NB displays less variation in classification accuracy ranging from 77.6% to 81.9%.

An important observation in this study is that a high performance can be achieved for each prediction model with a small, selected subset of features. For example, both NB and SVM obtained the best prediction results when the number of features was set to 6 (81.9% for NB and 82.8% for SVM). RF with the 3 top-ranked features achieved the highest classification accuracy (84.5%).

## 4. Discussion and conclusions

In the current study we have investigated the application of supervised classification techniques and feature selection ranking processes to ascertain if a data driven approach can be effective in the selection of an alternative subset of recording sites without a compromise in classification performance. The results attained demonstrated that it was possible to attain acceptable levels of classification with an alternative lead set configuration. Such a result provides an indication that there is indeed benefit to consider the use of a smaller, non standard subset of recording sites from the BSPM for classification without a compromise in classification accuracy.

The three classification models investigated exhibited different levels of performance. The RF approach achieved the highest classification accuracy with the top 3 features (84.5%). The solutions based on NB had less variation in prediction results when the number of input features changed (ranging from 77.6% to 81.9%). These have provided positive preliminary results and warrant the investigation of other supervised classification techniques for example Neural Networks.

Another crucial issue that needs to be addressed is the impact of the number of features on the classification performance. As shown in Figure 3, there is no consistent relationship between the performance of each classifier and the number of selected features. This could be partly caused by using the same learning parameters within the classification models for different numbers of input features. The combination with other machine learning techniques to dynamically select the optimal learning parameters needs to be part of future research.

Based on the value of SNR, a selected subset of top-ranked features was selected as an input to each classifier. A deficiency of this feature selection technique is that the selected features could be highly correlated among themselves. For example, the average Pearson correlation value between the top three features (QRST123, STT107, and STT123) is 0.90. Such a high correlation indicates a substantial redundancy among the input features and may have a negative influence on prediction analysis. Incorporating more elaborate feature selection techniques such as sequential selection techniques [9] or maximum-relevance-minimum-redundancy-based approaches [10] would be an important task of our future work.

## Acknowledgements

The authors would wish to acknowledge the kind support of Professor Robert Lux in providing the clinical data used in this study.

.

## References

[1] Horwitz LI. Current clinical utility of body surface mapping. J Invasive Cardiol 1995;7:265.
[2] Mirvis D. Current status of body surface electrocardiographic mapping. Circulation 1987;75:684.
[3] Carley SD. Beyond the 12 Lead: Review of the use of additional leads for early electrocardiographic diagnosis of acute myocardial infarction. Emergency Medicine 2003;15:143.
[4] Lux RL, Smith CR, Wyatt RF, Abildskov JA, Limited lead selection for the estimation of body surface potential maps in electrocardiography. IEEE Transactions on Biomedical Engineering, vol. 25, no. 3, pp. 270-276, 1978.
[5] Witten IH, Frank E, Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
[6] Rish I, An Empirical Study of the Naive Bayes Classifier. In the Proc. of IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, 41-46, 2001.
[7] Platt J. Fast Training of Support Vector Machines using Sequential Minimal Optimization. Advances in Kernel Methods - Support Vector Learning, B. Schoelkopf, C. Burges, and A. Smola, eds., MIT Press, 1998.
[8] Robnik-Sikonja M, Improving Random Forests. In the Proc. of 15th European Conference on Machine learning 2004, 359-370, 2004.
[9] Finlay DD, Nugent CD, Donnelly MP, Lux RL, McCullagh PJ, and Black ND, Selection of optimal recording sites for limited lead body surface potential mapping: A sequential selection based approach. BMC Journal of Medical Informatics and Decision Making, vol. 6, no. 9, 1-9, 2006.
[10] Ding C, Peng H, Minimum Redundancy Feature Selection from Microarray Gene Expression Data. In the Proc. of the Computational Systems Bioinformatics (CSB'03), 523-528, 2003

Address for correspondence

Chris Nugent
School of Computing and Mathematics
University of Ulster at Jordanstown
Shore Road, Newtownabbey
County Antrim, Northern Ireland, BT37 0QB
cd.nugent@ulster.ac.uk