# HIGHLY ACCURATE HIGHER ORDER STATISTICS BASED NEURAL NETWORK CLASSIFIER OF SPECIFIC ABNORMALITY IN ELECTROCARDIOGRAM SIGNALS

*M. Sabry-Rizk, W. Zgallai, S. El-Khafif, E. Carson*, K. Grattan, P. Thompson+*
Electrical, Electronic and Information Engineering Department,
* Centre for Measurement and Information in Medicine,
+ Royal Free Hospital, London, UK
City University, Northampton Square, London EC1V OHB, UK.

## ABSTRACT

The paper describes a simple yet highly accurate multi-layer feed-forward neural network classifier (based on the back-propagation algorithm) specifically designed to successfully distinguish between normal and abnormal higher-order statistics features of electrocardiogram (ECG) signals. The concerned abnormality in ECG is associated with ventricular late potentials (LP's) indicative of life threatening heart diseases. LP's are defined as signals from areas of delayed conduction which outlast the normal QRS period (80-100 msec). The QRS along with the P and T waves constitute the heart beat cycle. This classifier incorporates both pre-processing and adaptive weight adjustments across the input layer during the training phase of the network to enhance extraction of features pertinent to LP's found in 1-d cumulants. The latter is deemed necessary to offset the low S/N ratio in the cumulant domains concomitant to performing short data segmentation in order to capture the LP's transient appearance. In this paper we summarize the procedures of feature selection for neural network training, modification to the back propagation algorithm to speed its rate of conversion, and the pilot trial results of the neural ECG classifier.

## 1. INTRODUCTION

The relationship between myocardial infarction (MI) and short-duration high-frequency components occurring around the terminal end of the QRS complex in the cardiac cycle of the ECG (Figure 1) has been investigated by a number of dedicated research workers ([1] and references therein). The high frequency components are associated with late potentials (LP's) emanating from areas of delayed conduction and outlast the normal QRS period (80-100 msec). LP's are linked with malignant ventricular tachycardia (VT) after a myocardial infarction (dead zone or scar tissue in the ventricular muscle). The later is highly correlated with sudden cardiac death.

Common methodology for detecting LP's in the time domain involves temporal scanning of the S-T region (Figure 1) of the cardiac cycle and relies upon accurate identification of the QRS boundaries [2]. The detection problem is exacerbated by the fact that LP's are relatively very weak (micro-volts) and very often below the noise floor. In the frequency domain, second- order statistics can offer only a limited success. The shapes of power spectra of normal and abnormal (malignant VT) ECGs are invariably broadly similar and without significant features above the noise floor, at approximately

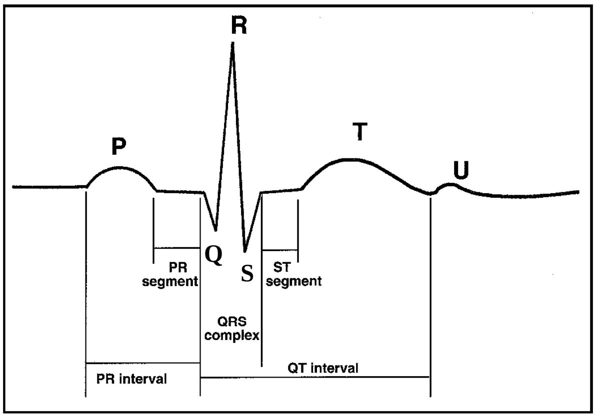



Figure 1: A diagrammatic illustration of the ECG waveforms and definitions.

-70 dB [3]. This is not surprising since LP's are essentially non-linear transient events [1] and consequently interact with the inherent non-linearity of the cardiac waves as well as certain class of recursive non-linearity attributed to external factors such as motion artifact [3,4,11].

We have shown in previous publications [3-6] that results obtained using higher-order statistics (HOS) indeed offer some empirical evidence that: (i) ECG signals contain intrinsic as well as quadratic and higher-order non-linearities, (ii) the QRS wave is predominantly linear non-Gaussian [7], the P and T waves are characterized by having quadratic and cubic non-linearities, (iii) the QRS wave can be totally resolved from the motion artifact in the bispectrum domains and (iv) disproportionately high-frequency non-linearity in the bicoherence squared is indicative of abnormality in an otherwise innocent looking ECG [3]. However, it is very important to emphasize that non-linear filtering and a high resolution technique such as the spectral MUSIC incorporates an optimized window must be applied to a short duration data sample (without compromising the variance), prior to the application of HOS [3,4]. Third-order Volterra filtering applied to raw data can be beneficial in isolating quadratic and cubic non-linearity in the higher-domains [3,10].

In this contribution, and in the same spirit as our recent works, the higher-order statistical features are selected and enhanced using sampled weights of a non-linear function based on a priori information about distinguished abnormality signatures in the higher domains. The function is modified adaptively during the training of the neural network which employs ten 1-d cumulants every 1000 or less cycles per patient. After this the updated version of the function parameters are fixed over the next 1000 cardiac cycles. Subsequently, a simple neural network classifier based on a modified version [8,9] of the back-propagation algorithms performs accurate LP’s and even ischemic ECG classification.

The paper is organized as follows: In section 2, a summary of the procedures of feature selection and enhancement for neural network training is given. Design of the neural network classifier which incorporates (a) block adaptive weight adjustment across the input layer, and (b) modification to the back-propagation algorithm to speed its rate of convergence is described in section 3. In section 4, experimental results obtained from conducting a pilot study on normal and abnormal subjects and employing the neural ECG classifier are documented. Finally to demonstrate the effectiveness of the neural classifier a brief comparative evaluation of the back-propagation- and recurrent back-propagation-based networks is concluded.

## 2. HIGHER-ORDER STATISTICS FEATURE SELECTION AND ENHANCEMENT

The experimental setup consists of an ECG monitor, interface card and a workstation. Raw ECG data are measured using three orthogonal surface electrodes, sampled at 500Hz and fed to the computer which performs the following operations.

1. Accurate on-line QRS detection [7]. This involves Volterra whitening filters in the time domain or/and the high-resolution spectral MUSIC in the frequency domain [7]. Positions of ECG peaks (R points in Figure 1) are pinpointed in the time domain (further details can be found in [7]).

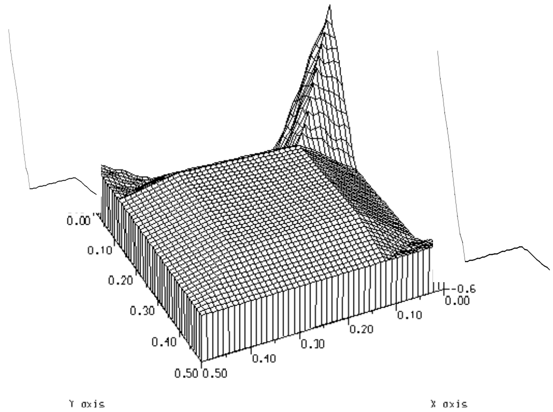
2. The MUSIC algorithm incorporates 2 sliding sets of three overlapping Kaiser windows and adaptive thresholding operations which not only pinpoint the high-level low-frequency QRS spectral peaks (LF**QRS**SPs) per cycle, but also performs the preliminary spotting of the low-level high-frequency late potential components over a range of frequencies from 100-250 Hz. Detection of late potential high-frequency spectral peaks is carried out off-line every 5 LF**QRS**SPs to allow appropriate segmentation between the R-R marking in the time domain processing which runs almost synchronizingly with the MUSIC routine. A detailed procedure for segmentation is beyond the scope of this paper [11] since it involves calculating the bicoherence squared and mapping a particular region for each individual segment to confirm existence of quadratic non-linearity before moving on to interrogate another segment or skip a few segments up to the next R peak. This controlled skipping helps to avoid the highly non-linear T wave of the present cycle and the P wave of the adjacent one (Figure 1).

3. The Volterra filtering can be used to partially suppresses motion artifact only in those cases of missing LF**QRS**SPs [7] and the MUSIC routine is repeated over the same cardiac cycle for confirmation of the presence or absence of QRSs. Again, this has been found to be only necessary in extreme cases and in the absence of QRS waves (ventricular fibrillation).

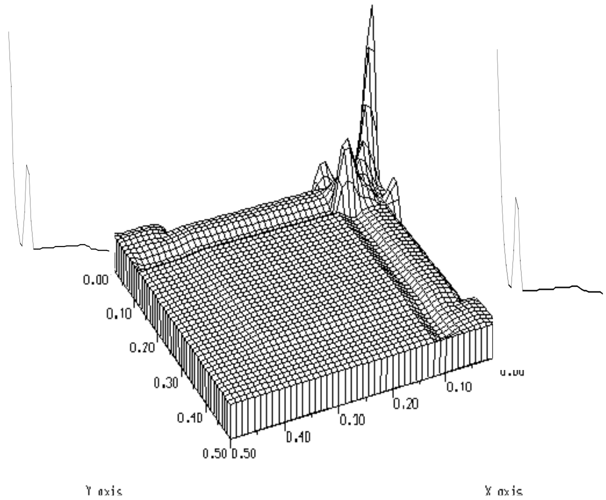
4. Off-line calculations of the cumulant diagonal and wall 1-d slices are performed on those segments suspected of having LP’s as depicted in Figure 2. It is clearly seen that abnormality is manifested in the eminent ‘petal pattern’[1] [11].

5. An arbitrarily chosen non-linear function modifies the envelope of the so-called ‘petal pattern’ to enhance its peculiarity against background artifact.

6. The non-linear function is then sampled across the input layer of the neural network to be described in the following section.



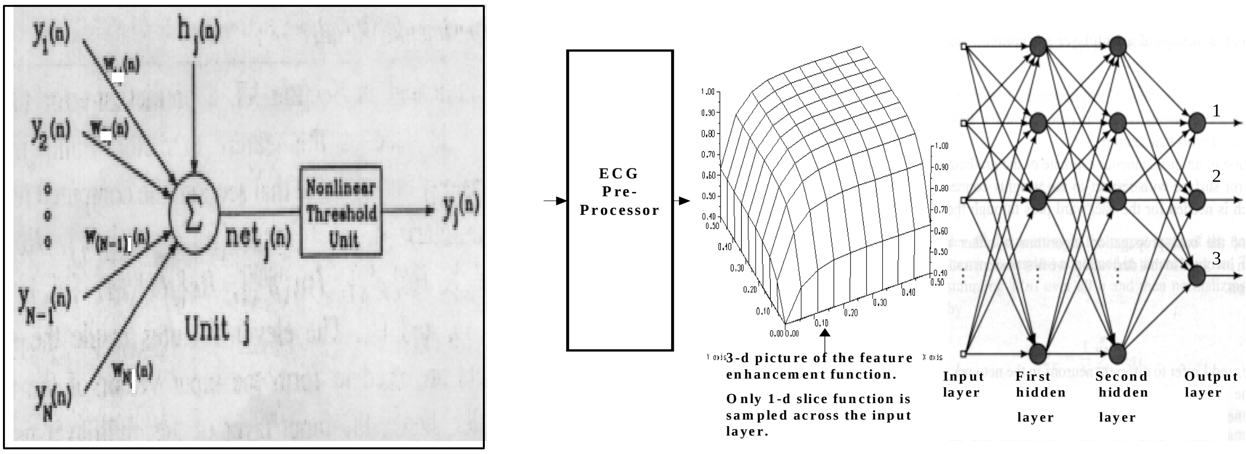

(a)




(b)

Figure 2: Typical third-order cumulants and their 1-d diagonal and wall slices shown in insets (left, right) of (a) a normal subject and (b) a subject confirmed of having infarction in the ventricular muscle.

[1] We are the first to discover the petal pattern (a horizontal slice has a petal shape) in the cumulant domains. Five thousand cardiac cycles of normal and abnormal ECGs were put to the test.




(a) (b)

Figure 3: Architecture of the 4-layerd neural network. (a) Neuron or processing unit in the network. (b) The four-layer neural network. Only 1-d slice of the weight function modifies the corresponding cumulant slice.

## 3. DESIGN OF THE NEURAL CLASSIFIER

In this section, we describe the 4-layer network (input layer, two hidden layers and output layer) used in the study reported here. Figure 3 shows the network preceded by a preprocessing unit which performs the difficult task of determining a set of meaningful and representative features in the HOS domains. Ordinarily, a sigmoid logistic function is used to describe the input-output relation of the non-linear device. The neural network is designed to classify two classes; normal and abnormal third-order cumulants. The combined use of skewness (from the third-order cumulants) and kurtosis (from the fourth-order cumulants) can provide more accuracy in difficult cases but are not considered here. In particular, we have found the utility of the diagonal slice of the fourth-order cumulant can be of more help when used in the desired response for the third output. The use of higher orders than the third statistics, however, adds more complexity to the network and is currently being investigated for other types of abnormality. Therefore, the use of the third output terminal does not apply in this communication.

### 3-a Block Adaptive Weight Adjustment

Initially we attempted to obtain the classification by feeding cumulant slices of short ST segments of the order of 10-30 samples at 500 Hz sampling rate. This attempt was 80% successful as the network missed low profile ‘petal patterns’ with low levels of signal-to-noise ratios in their vicinity as a result of short data segmentation. We then introduced the following function to strengthen the relative magnitude of the discriminant cumulant slice features.

$$f(x) = (1 - e^{-\alpha x}) \qquad 0 \le x \le \beta. \tag{1}$$

The above function is sampled across the input layer and its parameters ($\alpha$,$\beta$) can be adaptively changed for each cumulant slice fed during the training phase which usually takes up to 10 modified cumulant slices every 1000 cardiac cycles. Obviously the shape of function can be changed to cater for other types of abnormalities (will be reported in another publication due to lack of space).

### 3-b Modification to the back-propagation algorithm

The back propagation method [9] used in the supervised learning of a multi-layer neural network is basically a gradient descent method. Although this method has become the most popular learning algorithm for multi-layer networks, its rate of convergence is often found to be too slow for practical applications. Therefore, we have adopted two well established methods [8,9].

In the modified back-propagation method, every weight $w_{ij}$ in the network is given its own learning rate $\eta_{ij}$, and the training data set is divided into a number of epochs each containing K training patterns (training patterns are 1-d cumulants from overlapping segments of the ST region). The weight $w_{ij}$ and learning rate $\eta_{ij}$ are updated every time after an entire training epoch (10 cumulants) has been presented to the network. The weight and learning rate updating rules of the modified back-propagation algorithm can be summarized as follows [8,9]

$$w_{ij}(n+1) = w_{ij}(n) + \eta_{ij}(n+1)\sum_{k=1}^{K}\delta_{kj}(n)\,y_{ki}(n) + \tau\Delta w_{ij}(n-1) \tag{2}$$

$$\eta_{ij}(n+1) = \eta_{ij}(n) + \Delta\eta_{ij}(n) \tag{3}$$

$$\Delta\eta_{ij}(n) = \begin{Bmatrix} \Omega, & if & S(n-1)D(n) > 0 \\ -\phi\eta_{ij}(n), & if & S(n-1)D(n) < 0 \\ 0, & otherwise & \end{Bmatrix} \tag{4}$$

$$D(n)=\sum_{k=1}^{K}\frac{\partial \varepsilon_k(n)}{\partial w_{ij}(n)} \quad (5)$$

$$S(n)=(1-\Theta)D(n)+\Theta S(n-1). \quad (6)$$

The index n refers to the nth epoch in the training data; the index k refers to the kth pattern in an epoch containing K patterns; $\delta_{kj}$ is the modulated error signal of neuron j with the kth pattern in an epoch; $y_{ki}$ refers to the actual computed output of neuron I with the kth pattern in an epoch; $\varepsilon_k$ is the index performance to be minimized by the weight update rule with the kth input pattern; finally, $\Omega$, $\Phi$, and $\Theta$ (all of which have values between 0 and 1) are the control parameters.

## 4. EXPERIMENTAL PILOT RESULTS

Three orthogonal leads ECG were recorded from several subjects confirmed of having VT with a prior myocardial infarction (MI). Two subjects suspected of having MI but time- and frequency-domains analysis had not shown any abnormality, and several normal subjects. A total of 3,000 cardiac cycles for this pilot studies. Their feature extraction and enhancement were performed as described in section 3. The parameter $\alpha$ and $\beta$ of the exponential weight function applied across the input layer were chosen to fall in the region of 1-2 for $\alpha$, and 0.25 – 0.5 respectively.

Some aspects of the training phase are briefly described in section 3-a. The initial learning rate $\eta_{ij}(0)$ were all chosen to be 0.06. The momentum factor, $\alpha$, was fixed at 0.09. The control parameters $\beta$, $\Phi$ and $\Theta$ were chosen to be 0.03, 0.1 and 0.5 respectively.

The classifier described here achieved very high (96%) classification rate. The remaining 4% failure mainly arose because the MI suspected cases were not invasively examined and confirmed at the time of writing this paper.

## 5. CONCLUSIONS

In this paper, we have presented the results of a pilot study aimed at the non-invasive classification of a particular type of ECG abnormality, namely, late potentials. This has been achieved by the prudent use of their third-order cumulant 1-d slices. A four-layer neural network classifier based on modified back-propagation algorithm and incorporating adaptive feature enhancement weights applied to its input layer during its learning phase has been successfully tested. Classification rate obtained from 3000 cardiac cycles of normal, confirmed, and suspected abnormal subjects is 96%. In a separate study conducted on the same data a sophisticated recurrent back-propagation network achieved less that 80% success rate. However, the instability issues of the latter network has not been fully investigated.

## 6. REFERENCES


[1] P. Gomis, D. L. Jones, P. Caminal, E. J. Berbari and P. Lander, “Analysis of abnormal signals within the QRS complex of the high-resolution electrocardiogram,” IEEE Transactions on Biomedical Engineering, Vol. 44, No. 8, pp. 681-693, August 1997.

[2] C. Li, C. Zheng, and C. F. Tai, “Detection of ECG characteristic points using wavelet transforms,” IEEE Transactions on Biomedical Engineering, Vol. 42, January 1995.

[3] M. Sabry-Rizk, W. Zgallai, S. El-Khafif, E. Carson, and K. Grattan, “Higher-Order Ambulatory Electrocardiogram Identification and Motion Artifact Suppression With Adaptive Second- and Third-Order Volterra Filters,” SPIE98 Advanced Signal Processing Algorithms, Architectures, and Implementations VIII, paper #3461-43, San Diego, USA, 19-24 July 1998.

[4] W. Zgallai, M. Sabry-Rizk, P. Hardiman, and J. O'Riordan, "MUSIC-Based Bispectrum Detector: A Novel Non-Invasive Detection Method For Overlapping Fetal And Mother ECG Signals," IEEE EMBS-97, pp. 72-75, Chicago, USA, October 1997.

[5] M. Sabry-Rizk, W. Zgallai, P. Hardiman, and J. O'Riordan, “Third-order cumulant signature matching technique for non-invasive fetal heart beat identification,” ICASSP97, Vol. 5, pp. 3781-3784, Germany, April 1997.

[6] M. Sabry-Rizk, "Transabdominal fetal mother ECG signals deinterleaver using a neural network," NNESMED'96, 2nd International Conference on Neural Networks and Expert Systems in Medicine and Healthcare, UK, 1996.

[7] M. Sabry-Rizk, W. Zgallai, “Robust MUSIC-based Spectral Detection of Volterra Filtered Electrocardiogram QRS and Late Potentials,” submitted to IEEE Signal Processing Letters, 1998.

[8] R. A. Jacobs, “Increased rates of convergence through learning rate adaptation,” Neural Networks, Vol. 1, pp. 295-307, 1988.

[9] D. R. Hush and J. M. Salas, “Improving the learning rate of back-propagation with the gradient reuse algorithm,” in Proceedings of the second international conference on neural networks, Vol. 1, pp. 441-447, 1988.

[10] W. Zgallai, *The application of non-linear filtering, super-resolution spectral methods and higher-order statistics to non-invasive fetal Electrocardiogram (ECG) detection*, PhD Thesis, City University, to be submitted in 1998.

[11] M. Sabry-Rizk, S. El-Khafif, E. Carson and C. Morgan, “Non-linear filtering and super-resolution algorithms in higher-order statistical domains for the detection of myocardial ischemia,” to be submitted.

[12] A. Spiegl, P. Steinbigler, I. Schmueking, A. Knez and R. Haberl, “Analysis of beat-to-beat variability of frequency contents in the electrocardiogram using two-dimensional Fourier transform”, IEEE Transactions on Biomedical Engineering, Vol. 45, No. 2, pp. 235-241, February 1998.