Robust Neural-Network-Based Classification of Premature Ventricular Contractions Using Wavelet Transform and Timing Interval Features

Omer T. Inan*, *Student Member, IEEE*, Laurent Giovangrandi, and Gregory T. A. Kovacs, *Member, IEEE*

***Abstract*—Automatic electrocardiogram (ECG) beat classification is essential to timely diagnosis of dangerous heart conditions. Specifically, accurate detection of premature ventricular contractions (PVCs) is imperative to prepare for the possible onset of life-threatening arrhythmias. Although many groups have developed highly accurate algorithms for detecting PVC beats, results have generally been limited to relatively small data sets. Additionally, many of the highest classification accuracies ($> 90\%$) have been achieved in experiments where training and testing sets overlapped significantly. Expanding the overall data set greatly reduces overall accuracy due to significant variation in ECG morphology among different patients. As a result, we believe that morphological information must be coupled with timing information, which is more constant among patients, in order to achieve high classification accuracy for larger data sets.**

With this approach, we combined wavelet-transformed ECG waves with timing information as our feature set for classification. We used select waveforms of 18 files of the MIT/BIH arrhythmia database, which provides an annotated collection of normal and arrhythmic beats, for training our neural-network classifier. We then tested the classifier on these 18 training files as well as 22 other files from the database. The accuracy was 95.16% over 93,281 beats from all 40 files, and 96.82% over the 22 files outside the training set in differentiating normal, PVC, and other beats.

***Index Terms*—Electrocardiogram (ECG), heartbeat classifier, neural networks, premature ventricular contraction, wavelets.**

I. Introduction

CARDIOVASCULAR disease (CVD) is prevalent among more than one third of the United States population [1]. The American Heart Association (AHA) reported in 2006 that over 70 million people were burdened by some form of CVD, of which seven million experienced myocardial infarction, or heart attack. Additionally, CVD was the underlying cause of death for 37.3% of the two and a half million people who died in 2003 [1]. Subsequently, thoroughly researching the causes, detection, prevention, and treatment of CVDs is a crucial task bearing a potentially vast impact on the health of many people. This paper focuses on improving the available methods for automatic detection and classification of cardiac arrhythmias which may lead to life-threatening cardiac conditions.

Electrocardiogram (ECG) analysis has proven to be the archetypal method in clinical settings for detection of dangerous cardiac conditions. ECG effectively presents valuable clinical information regarding the rate, morphology, and regularity of the heart while being a low-cost and non-invasive test [2]. Specifically, ECG analysis can be used to detect cardiac arrhythmias, such as premature ventricular contractions (PVCs).

PVCs result from irritated ectopic foci in the ventricular area of the heart. These foci cause premature contractions of the ventricles that are independent of the pace set by the sinoatrial node. Many studies have shown that PVCs, when associated with myocardial infarction, can be linked to mortality [3]. Consequently, their immediate detection and treatment is essential for patients with heart disease.

Automated systems provide clinicians the tools to be alerted in real time if life-threatening conditions surface in their patients. As a result, automatic detection and classification of cardiac electrophysiology using biomedical signal processing techniques has become a critical aspect of clinical monitoring.

A. Summary of Previous Work

Many methods for automatic detection and classification of various arrhythmias have recently been presented in literature including algorithms based on hidden Markov models [4], self-organizing maps [5], filter banks [6], and neural networks [7], [9], [14]. The most accurate published classifiers are those that have been tested over relatively small sets of data. Shyu, *et al.* used wavelet feature extraction in tandem with fuzzy neural network classification to achieve 97.04% accuracy for PVC beat classification. Their classification results covered seven files of the MIT/BIH arrhythmia database, two of which were in the training set of the neural network [7].

Senhadji, *et al.* also achieved a high accuracy (98%) using 25 beats for training and 28 beats for testing in their discrete wavelet transform, linear discriminants classifier [8]. Hosseini, *et al.* used a multilayer perceptron (MLP) neural network classifier and presented 88.3% accuracy over 10 files of the MIT/BIH arrhythmia database [9].

Other groups have tested their classifiers over larger data sets but have accomplished relatively lower classification accuracy. De Chazal's group [10] tested using 44 files from the MIT/BIH arrhythmia database achieving an accuracy of 89%. Their method of linear discriminant-based classification involved

Manuscript received November 23, 2005; revised May 21, 2006. This work was supported in part by the National Aeronautics and Space Administration (NASA) National Center for Space Biological Technologies. *Asterisk indicates corresponding author.*

*O. T. Inan is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: omeri@stanford.edu).

L. Giovangrandi is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: giovan@stanford.edu).

G. T. A. Kovacs is with the Departments of Electrical Engineering and Medicine, Stanford University, Stanford, CA 94305 USA (e-mail: kovacs@cis.stanford.edu).

Digital Object Identifier 10.1109/TBME.2006.880879

fusing heartbeat morphology with timing interval features for the training of their classifier. Hu, *et al.* arrived at an accuracy of 62% over 20 files of the MIT/BIH arrhythmia database using self-organizing maps and learning vector quantization. Their classifier accomplished much higher results (94%) when used as a local (patient-specific) rather than a global classifier [11].

In general, these results indicate that high accuracy over a large number of files is a very difficult problem to address. This claim is consistent with intuition as there is significant variation in heartbeat morphology among patients. Additionally, the issue of over-fitting versus generalization in a classifier must be addressed when targeting a larger set of data. The number of beats which are used to train must be very small in comparison to the total number of beats to be classified. If this is not the case, the classifier will perform very well on the given set of data, but very poorly on new data. In order to avoid over-fitting, we limited the percentage of total beats which were used for classifier training to less than 1% of the total beats to be classified (training and testing sets combined). To achieve higher accuracy over larger data sets, we used a method which isolated the relevant morphological and temporal information for each heartbeat.

B. Proposal of Novel Algorithm

Results from previous works suggest that the combination of waveform shape and timing interval features is critical for robust classification [10]. This is expected as there are arrhythmic beats whose proper classification depends more on timing properties than on waveform shape. For example, Wenckebach-type AV block waveforms look relatively similar to those of normal beats (especially when downsampled) while their timing information is vastly different and relevant in their proper classification as arrhythmic events.

The wavelet transform has been demonstrated as a tool for effectively isolating relevant properties of the waveform morphology from the noise, baseline drift, and amplitude variance of the original ECG signal [8], [12], [13]. As a result, groups using the downsampled wavelet transform of the ECG signal as their feature set rather than the original waveform have demonstrated high classification accuracy [7], [14]. Such an approach can also take advantage of the wavelet decomposition performed in an earlier detection stage, such as with the method described by Li, *et al.* [12] By re-using the wavelet decomposition for classification, a substantial gain in computation time can be achieved, as proposed by Shyu, *et al.* [7].

Consequently, we present a proficient classification algorithm for PVC beats which combines wavelet transform, timing interval features and neural network classification. We show that this method achieves both high accuracy and robust classification of PVC beats over a large data set.

II. Methods

This chapter summarizing our methods is subdivided into four sections: MIT/BIH arrhythmia database, signal processing for feature extraction, neural network classifier and performance metrics. The first outlines the ECG data used in this study and provides a brief description of the MIT/BIH arrhythmia database. The second contains a relevant introduction to wavelet transform theory and a description of the various feature sets used in training and classification. The third section details the design, training, and testing of the neural network classifier. Finally, the last section, performance metrics, elucidates the measures used to evaluate the performance of our classifier.

A. MIT/BIH Arrhythmia Database

In this paper, ECG data from the MIT/BIH arrhythmia database were used [15]. The database was created in 1980 as a reference standard for arrhythmia detectors [16]. Its inception allowed developers of arrhythmia analyzers an objective measure of accuracy, specificity, and sensitivity. The primary goal of providing this metric was to spur the automated arrhythmia detection and classification technology. Moody and Mark report that since 1980 it has been used worldwide in over 500 sites for this purpose [16].

The database is comprised of 48 files, each containing 30-min ECG segments selected from 24-hr. recordings of 47 different patients. Of the 48 files, 23 were randomly chosen and 25 were selected to include uncommon, threatening, arrhythmic heartbeat samples [16]. Each file contains two leads, with modified-lead II available in 45 files, V1 in 40 files, and II, V2, V4 and V5 distributed among 11 files. Data are bandpass-filtered at 0.1–100 Hz and sampled at 360 Hz. We used a total of 40 records (see Table III for list) from the database, focusing on modified-lead II signals except in two files (102 and 104), in which lead V5 was substituted since modified-lead II was not available. Though variations in classifier performance with lead selection are expected, such as shown by de Chazal's group, we did not investigate this parameter in this study [10].

The database is annotated both in timing information and beat classification. In this paper, we focused on classification only and used the annotation to locate beats in ECG signals. We did not integrate beat detection since accuracy of classification typically lags that of detection in reported algorithms. Many examples of highly accurate ($> 99\%$) detectors can be found in literature [12], [13], [17]. Of particular interest to this study are the detectors based on wavelet transforms, as both detection and classification could share a common wavelet transformation step.

B. Signal Processing For Feature Extraction

Feature sets were formed from the MIT/BIH arrhythmia database records by using various combinations of morphological data and timing information. The wavelet transform was used to extract morphological information from the ECG data. Six feature sets were formed using original ECG data and the first five scales of the dyadic wavelet transform concatenated with timing information. An additional set was formed using the fourth scale of the dyadic wavelet transform without timing information. This section presents an introduction to wavelet transform theory as well as the details of feature extraction for our study.

1) Wavelet Transform Introduction and Theory: The wavelet transform (WT) provides a description of a signal in a time-scale domain, analogous to a time-frequency domain, allowing the representation of temporal features at multiple resolutions. This is achieved by the decomposition of the signal over dilated (scale) and translated (time) versions of a prototype wavelet. In

its continuous form, the WT of a signal $x(t) \in L^2(R)$ is defined as

$$\mathrm{CWT}_a x(b) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{a}} \psi^* \left(\frac{t-b}{a} \right) dt \qquad (1)$$

where $\psi(t)$ is the prototype (or mother) wavelet ($\psi \in L^2(R)$), a the scale factor ($a \in R^+$), b the translation factor ($b \in R$), CWT the continuous wavelet transform operator and * the complex conjugate operator. For the transform to be complete and reversible, the prototype wavelet must satisfy a weak admissibility condition which imposes the wavelet to have a zero-mean, a unitary norm, and a rapid time-decay (hence their name).

Interestingly, such characteristics in the time domain correspond to the characteristics of a bandpass filter in the frequency domain. The wavelet transform in (1), which can be rewritten as a convolution product between the signal and the scaled wavelets, can thus also be interpreted as a filtering of the signal by bandpass filters whose center frequencies and bandwidths depend on the scaling factor. High scales translate into long, slow wavelets equivalent to narrow, low-frequency filters, while lower scales produce shorter, faster wavelets equivalent to wider, higher-frequency filters.

With these properties, WT achieves an ideal balance of time and frequency resolution: slow trends are represented with a high frequency resolution and a low time resolution, while fast components are well defined in time but less in frequency. Such inherent multiresolution capabilities make WTs very effective at detecting and representing singularities, and WTs have been applied in many occasions to ECG analysis. In particular, the specific transform used in this paper has been proposed earlier by Li, *et al.* and more recently by Martinez, *et al.* for the detection and delineation of QRS complexes [12], [13]. The remainder of this section focuses on the implementation of the WT used in this work; for more information on wavelet theory, the reader is referred, for instance, to Mallat [18].

The continuous WT as shown in (1) is a highly redundant representation. For fast implementation, scale and translation parameters of (1) are usually sampled to follow a dyadic grid pattern ($a = 2^k$, $b = 2^{kl}$, $k,l \in Z^+$), and the extra condition of orthogonality or bi-orthogonality is imposed on the wavelet design (see [18] for details). Such a discrete wavelet transform (DWT) suppresses the redundancy of the continuous WT, while conserving the property of reversibility. Moreover, the DWT can then be efficiently implemented using filter banks based on the cascading of a pair of low- and high-pass finite impulse response filters corresponding to the wavelet (denoted $H(\omega)$ and $G(\omega)$, respectively), followed by a decimation-by-2 step [18]. However, the time-invariance property of the continuous WT, important for pattern recognition, is lost in the DWT due to the dyadic sampling of the translation parameter. To keep the WT time-invariant, the dyadic sampling can be limited to the scale parameter only (at the expense of redundancy), and the WT is calculated for each point in time.

This transform is called a dyadic wavelet transform ($\mathrm{D}_y\mathrm{WT}$), and can still be efficiently implemented using a filter bank algorithm called "algorithme à trous" [18]. In this algorithm, the signal is also filtered using the same pair of low- and high-pass filters corresponding to the wavelet ($H(\omega)$ and $G(\omega)$ respectively), however, instead of decimating the signal between stages (as it is the case in the DWT), the frequency responses of the filters are successively compressed. Consequently, the filtering is performed on each sample, at any scale. The compression of the filters, $H(\omega)$ and $G(\omega)$ can easily be achieved by insertion of zeros in their respective impulse responses, $h(n)$ and $g(n)$ The arrangement of the successive low-pass and high-pass filters corresponding to this implementation is described in Fig. 1.

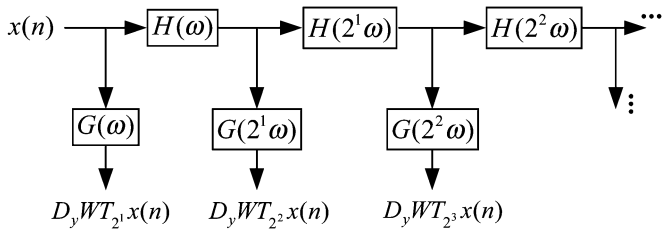

Fig. 1. Filter-bank implementation of a dyadic wavelet transform ($\mathrm{D}_y\mathrm{WT}$), also known as "algorithme à trous." The $\mathrm{D}_y\mathrm{WT}$ of a signal $x(n)$ is obtained by cascaded filtering with pairs of increasingly compressed low-pass finite impulse response filters. Low- and high-pass transfer functions $H(\omega)$ and $G(\omega)$ correspond to the wavelet prototype. At a scale 2^k, $k \in Z^+$, the filters take the form $H(2^{k-2}\omega)$ and $G(2^{k-1}\omega)$, except for $k < 2$, where $H(\omega)$ is not defined.

For each scale, $a = 2^k$, the resulting equivalent frequency response, $Q_k(\omega)$ corresponding to the cascaded filters can be written as

$$Q_k(\omega) = \begin{cases} G(\omega), & k = 1 \\ G(2\omega)H(\omega), & k = 2 \\ G(2^{k-1}\omega)H(2^{k-2}\omega)\cdots H(\omega), & k > 2 \end{cases} . \qquad (2)$$

It should be noted that these frequency responses, $Q_k(\omega)$ correspond for $k > 1$ to bandpass filters of increasingly narrower bandwidth and lower central frequency, as initially observed for the properties of daughter wavelets in the continuous wavelet transform.

Since one of the objectives of the approach is to facilitate the merging of the detection and classification steps, we used the same quadratic spline prototype wavelet, $\psi(t)$, that has already been successfully applied to QRS detection in Li, *et al.* [12] This bi-orthogonal wavelet has a compact support, one vanishing moment, and is defined by its Fourier transform (symbol ^) as

$$\hat{\psi}(\omega) = j\omega \left(\frac{\sin\left(\frac{\omega}{4}\right)}{\frac{\omega}{4}} \right)^4 . \qquad (3)$$

The corresponding finite impulse response (FIR) filters, $H(\omega)$ and $G(\omega)$ are given by

$$H(\omega) = e^{j\omega/2} \left(\cos \frac{\omega}{2} \right)^3$$
$$G(\omega) = 4je^{j\omega/2} \left(\sin \frac{\omega}{2} \right) \qquad (4)$$

and their impulse responses are

$$h(n) = \frac{1}{8} \cdot [\delta(n+2) + 3\delta(n+1) + 3\delta(n) + \delta(n-1)]$$
$$g(n) = 2 \cdot [\delta(n+1) + \delta(n)]. \qquad (5)$$

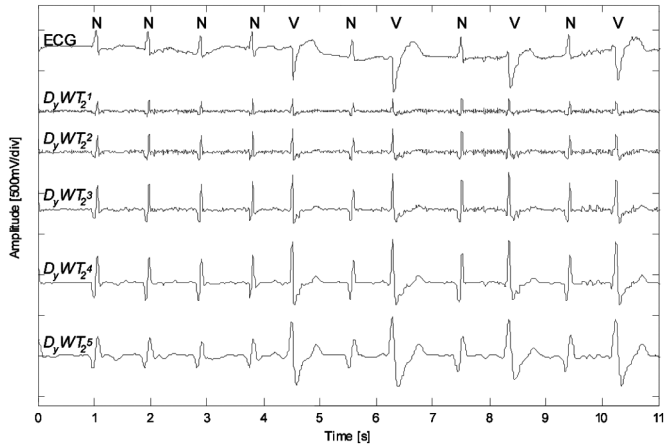

Fig. 2. Examples of normal QRS (annotated as "N") and PVC (annotated as "V") waveforms (lead II, file 200 from the MIT/BIH arrhythmia database) and their corresponding dyadic wavelet decompositions ($\mathrm{D}_y\mathrm{WT}$) for the first five scales 2^k, $k = 1, 2, 3, 4, 5$.

As mentioned, the impulse responses for $H(2^{k-1}\omega)$ and $G(2^{k-1}\omega)$ can easily be obtained by inserting $2^{(k-1)} - 1$ zeros in the impulse responses of $H(\omega)$ and $G(\omega)$.

We transformed ECG signals up to scale 2^5, after which the baseline wander became predominant. The corresponding -3 dB bandwidths of the equivalent $Q_k(\omega)$ filters for these first five scales are 60.5–180 Hz, 17.6–97 Hz, 8.4–45.7 Hz, 4.2–22.5 Hz, and 2.1–11.3 Hz, respectively. Note that the native sampling frequency of the MIT/BIH data is used (360 samples/s). Since a dyadic wavelet transform implies a scale progression in 2^k, we will use the denomination kth scale of the $\mathrm{D}_y\mathrm{WT}$ as a synonym for scale 2^k.

Examples of normal and PVC beats are shown in Fig. 2 with their dyadic wavelet transforms for the first five scales. The fourth scale appears to show very effective differentiation between the PVC and the normal beats with the least noise, as expected from the strong overlap between QRS frequency content and the bandwidth of the equivalent filter for this scale. The second, third, and fifth scales also appear to be competent discriminators between the two classes, and more quantitative evidence will be necessary in order to determine which of these scales will classify with highest accuracy.

2) Feature Extraction: We constructed six feature sets based on the original ECG data and its dyadic wavelet transform ($\mathrm{D}_y\mathrm{WT}$) at the first five scales, in conjunction with timing information. An additional set, based on the fourth scale of the $\mathrm{D}_y\mathrm{WT}$ but without timing information, was added to evaluate the effect of timing information on classification performance. We tested the classifier using each of these feature sets and compared the results. For all of the tests, the number of neurons in the hidden layer was set equal to the input vector length. Additionally, we held the following factors constant: inner layer decision functions, training method, feature set, and normalization process.

ECG waveform and wavelet transform features were extracted by selecting a window of -300 ms to $+400$ ms around the R wave as found in the database annotation. The 252-sample vectors were downsampled to 21, 25, 31, 42 or 63 samples (corresponding to $12x$, $10x$, $8x$, $6x$, and $4x$ decimation, respectively), and normalized to a mean of zero and standard deviation of unity. This reduced the DC offset and eliminated the amplitude variance from file to file.

In addition to morphology features, local timing information was also extracted. An RR-interval ratio (IR) reflecting the deviation from a constant beat rate was defined as

$$\mathrm{IR}_i = \frac{T_i - T_{i-1}}{T_{i+1} - T_i} \tag{6}$$

where T_i represents the time at which the R-wave for beat i occurs. The local RR-interval ratio information was chosen as opposed to other parameters such as average RR-interval or time series because it provides a convenient differentiator between normal beats ($\mathrm{IR}_i \sim 1$) and PVC beats ($\mathrm{IR}_i < 1$), and is normalized by definition ($\mathrm{IR}_i = 1$ at constant rate).

C. Neural Network Classifier

A neural network classifier was designed, trained and tested using the feature sets described above. In training, beats from different files of the MIT/BIH arrhythmia database were selected as representatives of various classes. The classifier was tested over a large set of ECG data for robustness and over a smaller set to demonstrate timing information relevance. This section describes the methods used in designing, training, and testing the neural network classifier.

1) Design: We implemented a feed-forward multi-layer perceptron (MLP) neural network with a single hidden layer for classification [19]. We explored various other implementations such as radial basis neural networks and multiple hidden layers, and empirically selected this particular network since it achieved the best performances. The number of neurons in the hidden layer was set equal to the number of features to be examined: For feature sets with timing, the neuron count was equal to the length of the waveform vector, plus one neuron for the RR-interval ratio (22, 26, 32, 43, 64 depending on the downsampling factor); for sets without timing, the neuron count was simply equal to the length of the waveform vector (21, 25, 31, 42, 63 depending on the downsampling factor). All computations were performed in Matlab® version 7.0 (The Mathworks, Natick, MA).

2) Training: The networks were trained with a total of 540 beats from 18 files of the MIT/BIH arrhythmia database. These files were selected as representatives of the following waveforms: normal, left bundle branch block (LBBB), right bundle branch block (RBBB), PVC, atrial flutter, and paced beats. Five files were selected to represent various forms of normal beats (100, 105, 106, 200, and 205), and five for PVC beats (119, 203, 208, 213, and 215). LBBB and RBBB waves (files 118, 212, and 214) were represented among unknown beats as they were closest to PVCs in terms of morphological features. Congruently, several files containing paced beats (102, 104, 107, and 217) were chosen to allow the networks to differentiate them from normal beats. Finally, file 201 contained atrial flutter beats which were necessary for the network to distinguish between RR-interval ratios in PVCs versus other arrhythmias.

From these 18 files, normal beats corresponded to beats annotated as "N," PVC beats to "V," and "other" beats to all other classifications of the database. The number of beats used for training and the files that produced the best results were determined iteratively. We constrained the number of beats used for training to less than 2% of the total beats that would be tested. This constraint was implemented to ensure the robustness of our classifier; using a higher percentage of training beats could have caused the network to be over-fitted to the training files rather than generalized on the more global features and aspects of PVC beats.

The training algorithm used was the Broyden, Fletcher, Goldfarb, and Shanno update, a quasi-Newton method which iteratively determines the weight matrix as well as the biases necessary to meet the desired error goal [20].

3) Testing: The classifier was tested in two venues to determine its performance. First, a total of 93 281 beats from 40 files of the MIT/BIH arrhythmia database were classified in order to verify high accuracy and robustness. Second, a smaller data set of seven records from the same database was selected for a demonstration of the value of timing interval to overall performance.

Of the 40 files chosen for the first experiment, 22 were completely foreign to the classifier. Additionally, only 2% (540) of the total beats in the 18 training files were used for training the neural network. As a result, 99.4% (92,741) of the beats presented to the neural network for classification did not overlap with the training beats.

In our second experiment, we classified PVC and non-PVC beats from the same files selected for study by Shyu, *et al.* [7] in order to compare results. These files were 111, 115, 116, 119, 221, 230, and 231. The purpose of this comparison was two-fold: first, since Shyu, *et al.* achieves very high accuracy using the wavelet transform and neural network method without timing information (97.04%), producing a higher result with our method would reinforce the value of timing information; second, by achieving a higher classification accuracy on a small data set (7 files) than on the larger data set (40 files were classified in this work) we could support our claim that, in general, classifier performance degrades as the size of the data set increases. This second point would justify our position that classification over a large number of files from the database is necessary in demonstrating robustness. Note that for this part, the network was trained using 225 PVC and normal beats from file 116 and 120 LBBB beats from file 111.

D. Performance Metrics

We quantified our classifier performance using the most common metrics found in literature: accuracy, sensitivity, and positive predictivity. The former expresses the overall system performance over all types of beats while the two latter quantities are specific to each class of beat. The methods used to quantify the performance of our classifier in each of these three areas are discussed below.

1) Accuracy, Sensitivity, and Positive Predictivity: The most crucial metric for determining overall system performance is usually accuracy. We defined the overall accuracy of the classifier for each file as follows:

$$A = 100 \cdot \left(1 - \frac{Ne}{Nb}\right). \tag{7}$$

In this equation, A is the accuracy, and the variables, Ne and Nb, represent the total number of classification errors and beats in the file, respectively. In addition to accuracy, two other measures of classifier performance used in previous works [21] are sensitivity (Se) and positive predictivity (Pp)

$$\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{8a}$$

$$\mathrm{Pp} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{8b}$$

In these equations, TP, FP, and FN denote true positives, false positives and false negatives, respectively. True positives are beats which have been correctly assigned to a certain class whereas false positives are beats which have been incorrectly assigned to that same class. A false negative occurs when a beat should have been assigned to that class but was missed and assigned to another class. Consequently, sensitivity measures how successfully a classifier recognizes beats of a certain class without missing them whereas positive predictivity measures how exclusively it classifies beats of a certain type.

2) Weighted Average Sensitivity and Positive Predictivity: We determined that average sensitivity and positive predictivity were not useful measures of overall performance because of the tremendous variation among files in number of normal, PVC, and unknown beats. The sensitivity for a file with only one PVC beat, for example, is much less relevant to overall system PVC sensitivity than the result from a file with 500 PVC beats. Consequently, we weighted the averages according to the number of beats of each class that were present in a file

$$\mathrm{Se}_{\mathrm{WA}}(\text{“}C\text{”}) = \frac{\sum_{i=1}^{Nf} Se_i(\text{“}C\text{”}) \cdot Nb_i(\text{“}C\text{”})}{\sum_{i=1}^{Nf} Nb_i\text{“}C\text{”})} \tag{9}$$

In this equation, $\mathrm{Se}_{\mathrm{WA}}$ is the weighted average sensitivity for the class "C" which is found by summing the product of the sensitivity, Se, and the number of beats, Nb, for file i over the total number of files, Nf. A similar equation was used to calculate the weighted average positive predictivity for each class.

III. Results

We determined the most effective feature set as well as the optimum vector length for high accuracy classification. With this feature set vector, we assessed the performance of our classifier in two tests: large dataset testing to verify system robustness, and smaller dataset testing to demonstrate timing information relevance. In both tests, the classifier performed at a very high level, achieving accuracies over 95%. The details and discussion of feature vector selection and both tests are in the sections that follow.

TABLE I
COMPARATIVE RESULTS FOR VARIOUS FEATURE SETS. "ECG" REFERS TO THE SET BASED ON THE ORIGINAL WAVEFORM, AND "$\mathrm{WT}\{k\}$" DENOTE THE SETS BASED ON THE kTH SCALE OF THE DYADIC WAVELET TRANSFORM ($\mathrm{D}_y\mathrm{WT}$). "IR" IS THE RATIO OF PREVIOUS TO NEXT RR-INTERVALS FOR A GIVEN BEAT. THE BEST PERFORMING SET IS SHOWN IN ITALICS

Feature Set		A	Normal		PVC		Other	
			Se	Pp	Se	Pp	Se	Pp
ECG	With RR-Int. Ratio (IR)	83.1	88.8	97.3	88.8	84.5	74.3	93.0
WT {1}		80.1	83.0	96.2	74.0	88.1	72.1	90.0
WT {2}		84.9	90.7	96.4	74.4	85.7	72.8	92.0
WT {3}		80.3	88.8	94.8	69.0	80.0	74.6	94.4
WT {4}		***95.2***	***98.1***	***97.0***	***85.2***	***92.4***	***87.4***	***94.5***
WT {5}		81.2	88.6	95.6	92.0	86.4	72.1	92.1
WT {4}	No IR	81.7	81.9	96.6	83.4	85.4	80.1	91.6

A. Feature Set Comparison

We tested our neural network classifier over the large data set (described above) using each of the seven feature sets: original waveform and first five scales of the $\mathrm{D}_y\mathrm{WT}$ with timing information, and fourth scale of the $\mathrm{D}_y\mathrm{WT}$ without timing information. The results from these variations are shown in Table I, where "ECG" refers to the set based on the original waveform, and "$\mathrm{WT}\{k\}$" denote the sets based on the kth scale of the $\mathrm{D}_y\mathrm{WT}$.

The feature set leveraging the fourth scale decomposition in tandem with RR-interval timing information surpasses the other six in performance. This result is consistent with our predictions as well as the findings of Shyu, *et al.*., who also used the fourth scale of a dyadic wavelet transform in their classification algorithm [7]. Interestingly, the sensitivity for PVC beats is higher in the fifth scale than in the fourth, likely due to the closer match of the fifth scale with the lower frequencies of PVCs. However, other than that particular result, the fourth scale of the $\mathrm{D}_y\mathrm{WT}$ with RR-interval information outperformed the other feature sets in every metric.

B. Determination of Vector Length

After selecting the most effective feature set, we determined the optimum decimation factor for downsampling of the waveform. This value was found empirically, and results are shown in Table II. We defined the optimum choice in this case as the number of samples that maximized classification accuracy.

As per these results, we determined that downsampling to 42 samples was optimal. We used the 42 sample downsampling for our classification tests described below.

C. Large Data Set Testing

We classified over the large data set using the fourth scale of the $\mathrm{D}_y\mathrm{WT}$ downsampled to a vector length of 42 concatenated with timing information. The average overall accuracy was 93.13% for the 18 files of the training set and 96.82% for the other 22 files. The average over all files was 95.16%. Using (9) above, we found that the weighted average sensitivity was 98.33% for normal beats, 82.57% for PVC beats and 86.76% for other beats. The values for positive predictivity were 97.58% for normal, 93.42% for PVC and 94.59% for other beats. The file-by-file comprehensive results are shown in Table III.

TABLE II
CLASSIFICATION ACCURACY AND TIME FOR VARIOUS VECTOR LENGTHS

No. of Samples	No. of Neurons	Classification Accuracy	Classification Time (s)
21	22	87.4	129.2
25	26	88.7	159.5
31	32	89.3	166.7
42	43	95.2	172.1
63	64	88.8	213.4

D. Timing Information Relevance

Using the same feature set as for the large data set testing described above, we achieved an accuracy of 99.78% over the seven records used by Shyu, *et al.* [7], 2.74% higher than their reported value of 97.04%. We found a classifier sensitivity of 98.32% for PVC and 99.58% for non-PVC beats. Positive predictivity was 98.57% for PVC and 99.88% for non-PVC beats. Using the fourth scale of the $\mathrm{D}_y\mathrm{WT}$ without timing information, the classification accuracy dropped to 96.40%. Sensitivity fell to 93.60% for PVC and 96.52% for non-PVC, and positive predictivity to 68.71% and 99.53%, respectively.

IV. DISCUSSION

In both tests, the classifier performed with very high accuracy ($> 95\%$). The results from the large data set test confirm the robustness of our system, and from the smaller data set test reinforce the value of timing information to overall performance. In the large data set test, we found that the classifier performance was worst with any files containing a large number ($> 10\%$ of the total beats in the file) of fused beats; both fused ventricular and normal as well as fused paced and normal provided low results. This finding is consistent with the results from de Chazal's group [10]. In file 104, which contains the largest percentage of fused beats (29.9%), the classification accuracy was only 78.61%. Files 208 and 213, which also have a high percentage of fused beats (12.6% and 11.1%, respectively) show very low sensitivity and positive predictivity in the "other" beats class. We believe that these results are due to the fact that fused beats are to be classified as "other," while their morphology and timing information can closely resemble normal beats.

Generally, the classification accuracy over files which were included in training is higher than the accuracy for the testing set. However, with our classifier, the accuracy was lower for our training set than our testing set. This discrepancy could be attributed to the training set containing 1717 more fused beats (the largest source of error in our classifier) than the testing set as well as only 2% of the beats in the training set being used for training the network. Also, it should be noted that although we limited the total number of training beats to less than 2% of all beats in the training set, our classifier is not optimal in terms of its generalization because of the internal structure of the network. Less hidden layer neurons might lead to better generalization of the classifier. As the focus of this study was primarily on

TABLE III
COMPREHENSIVE RESULTS FOR TRAINING AND TESTING FILES

			Normal			PVC			Other		
File	**Beats**	**Acc.**	**Beats**	**Se**	**Pp**	**Beats**	**Se**	**Pp**	**Beats**	**Se**	**Pp**
100	2271	98.55	2237	100.00	98.54	1	100.00	100.00	33	0.00	-
101	1864	99.62	1859	99.84	99.84	0	-	-	5	20.00	33.33
102	2185	94.23	99	27.27	77.14	4	75.00	6.12	2082	97.45	96.57
103	2083	99.90	2081	100.00	99.90	0	-	-	2	0.00	-
104	2227	78.61	163	76.69	22.60	2	50.00	10.00	2062	78.79	97.65
105	2572	93.19	2526	94.69	98.84	41	4.88	4.88	5	60.00	2.70
106	2027	96.84	1507	99.27	99.01	520	89.81	99.79	0	-	-
107	2136	99.86	0	-	-	59	94.92	100.00	2077	100.00	99.86
109	2530	86.00	0	-	-	38	65.79	7.37	2492	86.31	99.44
112	2537	99.57	2535	99.64	99.92	0	-	-	2	0.00	0.00
113	1794	98.10	1788	98.26	100.00	0	-	-	6	50.00	8.82
114	1879	93.13	1820	93.56	99.53	43	100.00	91.49	16	25.00	3.31
115	1952	100.00	1952	100.00	100.00	0	-	-	0	-	-
116	2411	98.30	2301	99.30	99.87	109	77.98	91.40	1	0.00	0.00
118	2277	98.55	0	-	-	16	43.75	36.84	2261	98.94	99.69
119	1987	100.00	1543	100.00	100.00	444	100.00	100.00	0	-	-
121	1862	99.30	1860	99.35	99.95	1	100.00	100.00	1	0.00	0.00
122	2474	100.00	2474	100.00	100.00	0	-	-	0	-	-
123	1517	99.87	1514	100.00	100.00	3	33.33	100.00	0	-	-
200	2600	96.30	1742	99.20	96.97	826	93.94	100.00	32	0.00	0.00
201	1963	93.88	1625	99.69	96.54	198	87.88	81.69	140	35.00	68.06
202	2135	97.70	2060	99.76	98.42	19	89.47	62.96	56	25.00	70.00
203	2979	87.07	2528	86.34	99.41	444	91.89	79.53	7	42.86	1.11
205	2655	99.40	2570	99.84	99.57	71	100.00	95.95	14	14.29	50.00
208	2953	87.39	1585	98.86	84.56	992	99.09	93.52	376	8.24	63.27
210	2648	95.62	2421	97.73	98.62	194	85.05	94.29	33	3.03	1.35
212	2747	99.60	922	99.02	99.89	0	-	-	1825	99.89	99.51
213	3249	84.54	2639	100.00	85.67	220	43.18	90.48	390	3.33	20.31
214	2260	85.52	0	-	-	256	47.66	92.42	2004	90.36	96.95
215	3361	94.91	3194	99.66	95.84	164	3.05	83.33	3	66.67	5.88
217	2208	87.72	244	96.31	55.56	162	91.98	68.98	1802	86.17	98.98
219	2154	98.61	2082	99.71	99.52	64	75.00	100.00	8	0.00	0.00
220	2046	96.14	1952	100.00	96.58	0	-	-	94	15.96	100.00
221	2427	99.92	2031	100.00	99.95	396	99.49	100.00	0	-	-
223	2604	85.82	2028	99.85	95.24	473	40.38	98.45	103	18.45	6.69
228	2053	96.00	1688	97.92	99.52	362	87.85	97.55	3	0.00	0.00
230	2255	97.69	2254	97.69	100.00	1	100.00	100.00	0	-	-
231	1570	99.17	314	100.00	96.02	2	100.00	100.00	1254	98.96	100.00
233	3077	91.48	2229	95.42	97.08	830	82.77	99.13	18	5.56	0.52
234	2752	98.18	2699	99.96	98.21	3	100.00	100.00	50	2.00	50.00
Total	93281	-	67066	-	-	6958	-	-	19257	-	-
Avg	2332.03	**95.16**	1676.65	-	-	173.95	-	-	481.43	-	-
Wt. Avg	-	-	-	**98.33**	**97.58**	-	**82.57**	**93.42**	-	**86.76**	**94.59**

demonstrating the relative effectiveness of our feature extraction algorithms, we did not optimize this aspect of the neural network classifier itself.

Our results from testing over the smaller data set demonstrate two points: first, timing information in conjunction with wavelet transform features improves overall classification accuracy; second, highly accurate classification is much more difficult to attain over a larger set of data. Using a similar approach to Shyu, *et al.*, with the exception of our approach adding RR-interval ratio to the feature set, we attain a higher classification

accuracy of 99.78% (+2.74% improvement). When the RR-interval ratio is not included in our feature set, classification accuracy drops to 96.40%, only slightly lower than the accuracy reported by Shyu, *et al.* These two findings demonstrate the value of timing information to classifier performance.

The greatest improvement (+18.97% accuracy) over results from Shyu, *et al.* was observed in file 111, which consists of 2123 LBBB beats and one PVC beat. We believe that this improvement in differentiating between LBBB and PVC beats is most likely a result of using RR-interval timing information. As Shyu, *et al.* note, the waveforms of LBBB and PVC beats are quite similar. As a result, simply including beats from file 111 in the training set is not sufficient for very high accuracy. Timing information from the original waveforms is necessary in order to effectively differentiate between the two similar waves. By including RR-interval timing information, we provide the neural network with a prominent distinction between the two types of beats.

Additionally, the result of 99.78% is higher (+4.62%) than the value we achieved for the larger data set of 40 files. This is consistent with our second claim, that the same method will, generally, produce lower results when applied to a greater number of files. The significant variation from one patient to another in the morphology of the beats creates one of the hardest obstacles in PVC classification. Consequently, a greater number of patients (files) cause larger variations, and lower accuracy of classification.

It should also be noted that the coupling of detection and classification would degrade the performance of the classifier. The detector would introduce false negatives, false positives and their associated erroneous timing information. Previous studies have proposed that these errors are negligible due to the large disparity between detection and classification accuracy.

However, in light of the high classification accuracy demonstrated in this work, the relative impact of these errors in detection is potentially larger. As a result, coupled detection and classification is the subject of on-going studies.

V. Conclusion

We accurately classified ECG beats for a large set of data by combining wavelet transforms and timing information for the neural network classifier. We found that the fourth scale of a dyadic wavelet transform with a quadratic spline wavelet together with the pre-/post RR-interval ratio is very effective in differentiating normal and PVC from other beats. An overall accuracy of detection of 95.16% was achieved over 40 files from the MIT/BIH arrhythmia database. When applying our method to the classification study presented in Shyu, *et al.* [7] we demonstrated improvement in the ability to distinguish between LBBB and PVC beats (+2.74% in overall accuracy). Without timing information, our accuracy dropped by over 3% over the same set of data. These two results reinforce the value of timing information to our classifier.

It should be noted that in addition to feature set selection (as discussed in this paper), classifier performance can also affect overall classification results. However, it is likely that the relative differences in accuracy among the feature sets will not change. In future studies, we will examine other classifier implementations which may provide higher overall results.

Future work should also focus on expanding the number of classes to include fused, LBBB, RBBB, and paced beats, as these were the types that were classified with lowest accuracy in our algorithm. In order to achieve this type of class globalization, more timing information must be accessible to the classifier. This can be implemented by including more timing information in the feature set, or by using memory-based classifiers. To include more timing information without sacrificing computational time and network complexity, the wavelet information could be reduced into a compact, dense vector using linear discriminants. Subsequently, more timing information such as running RR-interval average, ST interval, and QRS duration could be included in the feature set.

Another possible approach could involve using a classifier such as a delayed-input neural network, which looks at a series of beats. The timing information would then be inherent, as the waveform morphology for a series of beats will depend greatly on timing. These implementations could further accentuate the difference between bundle branch block and PVC beats which have similar waveform characteristics, but very different timing features.

References

[1] T. Thom *et al.*, "Heart disease and stroke statistics—2006 update. A report from the american heart association statistics committee and stroke statistics subcommittee," in *Circulation*, Jan. 11, 2006.

[2] D. Dubin, *Rapid Interpretation of EKG's*, 5 ed. Tampa, FL: Cover, 2000.

[3] I. Atsushi, M. Hwa, A. Hassankhani, T. Liu, and S. M. Narayan, "Abnormal heart rate turbulence predicts the initiation of ventricular arrhythmias," *Pacing Clin. Electrophysiol.*, vol. 11, pp. 1189–97.

[4] W. T. Cheng and K. L. Chan, "Classification of electrocardiogram using hidden markov models," in *Proc. 20th Annu. Int. Conf. IEEE EMBS*, 1998, vol. 20, pp. 143–146.

[5] M. R. Risk, J. F. Sobh, and J. P. Saul, "Beat detection and classification of ECG using self organizing maps," in *Proc. 19th Int. Conf. IEEE EMBS*, 1997, vol. 19, pp. 89–91.

[6] X. Alfonso and T. Q. Nguyen, "ECG beat detection using filter banks," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 2, pp. 192–202, Feb. 1999.

[7] L. Y. Shyu, Y. H. Wu, and W. C. Hu, "Using wavelet transform and fuzzy neural network for VPC detection from the holter ECG," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1269–1273, Jul. 2004.

[8] L. Senhadji, G. Carrault, J. J. Bellanger, and G. Passariello, "Comparing wavelet transforms for recognizing cardiac patterns," *IEEE Eng. Med. Biol. Mag.*, vol. 14, no. 2, pp. 167–173, Mar.-Apr. 1995.

[9] H. G. Hosseini, K. J. Reynolds, and D. Powers, "A multi-stage neural network classifier for ECG events," in *Proc. 23rd Int. Conf. IEEE EMBS*, 2001, vol. 2, pp. 1672–1675.

[10] P. D. Chazal and R. B. Reilly, "Automatic classification of ECG beats using waveform shape and heart beat interval features," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'03)*, 2003, vol. 2, pp. 269–272.

[11] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "Patient-adaptable ECG beat classifier using a mixture of experts approach," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 891–897, Sep. 1997.

[12] C. Li, C. X. Zheng, and C. F. Tai, "Detection of ECG characteristic points using wavelet transforms," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 1, pp. 21–28, Jan. 1995.

[13] J. P. Martinez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, "A wavelet-based ECG delineator: Evaluation on standard databases," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 570–581, Apr. 2004.

[14] R. V. Andreao, B. Dorizzi, P. C. Cortez, and J. C. M. Mota, "Efficient ECG multi-level wavelet classification through neural network dimensionality reduction," in *Proc. 2002 IEEE Signal Processing Workshop (Neural Networks for Signal Processing XII)*, 2002, pp. 395–404.

[15] R. Mark and G. Moody, MIT-BIH Arrhythmia Database 1997 [Online]. Available: http://ecg.mit.edu/dbinfo.html

[16] G. B. Moody and R. G. Mark, "The impact of the mit/bih arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May-Jun. 2001.
[17] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, pp. 230–236, 1985.
[18] S. Mallat, *A Wavelet Tour of Signal Processing*, 2 ed. San Diego, CA: Academic, 1999.
[19] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 1 ed. New York: Macmillan College, 1994.
[20] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
[21] N. Maglaveras, T. Stamkopoulos, K. Diamantaras, C. Pappas, and M. Strintzis, "ECG pattern recognition and classification using non-linear transformations and neural networks: A review," *Int. J. Med. Informatics*, vol. 52, pp. 191–208, 1998.

Omer T. Inan (S'06) received the B.S. and M.S. degrees, both in electrical engineering, from Stanford University, Stanford, CA, in 2004 and 2005, respectively. He is currently pursuing the Ph.D. degree in electrical engineering at Stanford University.

His research interests include design of electronic sensors and systems for biomedical applications. He has worked on projects involving Fourier-domain optical coherence tomography, low-cost monitoring of model organisms in space and terrestrial settings, and interactions of electromagnetic waves with biological tissues. Currently, he focuses on signal processing and system design for wearable electrocardiogram monitors.

Laurent Giovangrandi received the M.S. degree in microengineering and the Ph.D. degree in applied sciences from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1995 and 1999, respectively. His Ph.D. dissertation focused on microsystems for neuron-electronics interfaces and the control of cell-surface interactions.

From 1999 to 2000, he was a Research Associate on planar patch clamp devices at the Institute of Physical Chemistry (ICP), EPFL. In 2000, he joined Stanford University, Stanford, CA, to pursue postgraduate research on cell-based biosensors and biosignal analysis. Since 2004, he has been with the National Center for Space Biological Technologies (NCSBT), Stanford University. His current interests include the application of cell-based microsystems in bioassays, drug screening and space research, as well as the analysis and visualization of biomedical signals.

Gregory T. A. Kovacs (M'83) received the B.A.Sc. degree in electrical engineering from the University of British Columbia, Vancouver, B.C., Canada, the M.S. degree in bioengineering from the University of California, Berkeley, and the Ph.D. degree in electrical engineering and the M.D. degree from Stanford University, Stanford, CA.

He is a long-standing member of the Defense Sciences Research Council (DARPA), and has served as Associate Chair and Chairman. He also has extensive industry experience including co-founding several companies, including Cepheid in Sunnyvale, CA. He is a Professor of Electrical Engineering at Stanford University with a courtesy appointment in the Department of Medicine. His present research areas include biomedical instruments and sensors, miniaturized spaceflight hardware, and biotechnology. He is the Director of Medical Device Technologies for the Astrobionics Program at the NASA Ames Research Center, and Principal Investigator of the Stanford-NASA National Center for Space Biological Technologies. He helps direct a variety of projects spanning wearable physiologic monitors, biosensor instruments for detection of chemical and biological warfare agents and space biology applications, and free-flyer experiment payloads. In 2003, he served as the Investigation Scientist for the debris team of the Columbia Accident Investigation Board, having worked for the first four months after the accident at the Kennedy Space Center. In this role, he carried out physical, photographic, X-ray, chemical, and other analyses on selected items from the nearly 90 000 pounds of recovered debris and worked toward understanding the nature of the accident. He currently serves as Engineering/Medical Liason on the Spacecraft Crew Survival Integration Investigation Team (SCSIIT) of the Johnson Space Center.

Dr. Kovacs received an NSF Young Investigator Award, held the Noyce Family Chair, and was a Terman and then University Fellow at Stanford. He currently is the Thomas V. Jones Development Scholar in the School of Engineering. He is a Fellow of the American Institute for Medical and Biological Engineering and a Fellow National of the Explorers Club. He was a member of a NASA and National Geographic Society sponsored team that climbed Licancabur volcano (19 734 ft.) on the Chile/Bolivia border in November of 2003, serving as medical, physiologic research, and photography lead. In November 2004, he served the same role on a return expedition to Licancabur, and carried out medical research and underwater videography in the summit lake.