Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features

Philip de Chazal*, *Member, IEEE*, Maria O'Dwyer, and Richard B. Reilly, *Senior Member, IEEE*

Abstract—A method for the automatic processing of the electrocardiogram (ECG) for the classification of heartbeats is presented. The method allocates manually detected heartbeats to one of the five beat classes recommended by ANSI/AAMI EC57:1998 standard, i.e., normal beat, ventricular ectopic beat (VEB), supraventricular ectopic beat (SVEB), fusion of a normal and a VEB, or unknown beat type. Data was obtained from the 44 nonpacemaker recordings of the MIT-BIH arrhythmia database. The data was split into two datasets with each dataset containing approximately 50 000 beats from 22 recordings. The first dataset was used to select a classifier configuration from candidate configurations. Twelve configurations processing feature sets derived from two ECG leads were compared. Feature sets were based on ECG morphology, heartbeat intervals, and RR-intervals. All configurations adopted a statistical classifier model utilizing supervised learning. The second dataset was used to provide an independent performance assessment of the selected configuration. This assessment resulted in a sensitivity of 75.9%, a positive predictivity of 38.5%, and a false positive rate of 4.7% for the SVEB class. For the VEB class, the sensitivity was 77.7%, the positive predictivity was 81.9%, and the false positive rate was 1.2%. These results are an improvement on previously reported results for automated heartbeat classification systems.

Index Terms—Electrocardiogram (ECG), heartbeat classifier, linear discriminant analysis, statistical classifier model.

I. INTRODUCTION

HEART arrhythmias result from any disturbance in the rate, regularity, and site of origin or conduction of the cardiac electric impulse [1]. Broadly speaking, arrhythmias can be divided into two groups. The first group includes ventricular fibrillation and tachycardia which are life-threatening and require immediate therapy with a defibrillator. Detection of these arrhythmias is well researched and successful detectors have been designed with high sensitivity and specificity [2]–[5].

This study investigated the second group which includes arrhythmias that are not imminently life-threatening but may require therapy to prevent further problems. The electrocardiogram (ECG) is a low-cost, noninvasive, and effective test for ar-

rhythmia analysis and has become the standard diagnostic tool. Some arrhythmias appear infrequently and up to a week of ECG activity may need to be recorded using a Holter ECG monitor to successfully capture them. Many arrhythmias manifest as sequences of heartbeats with unusual timing or ECG morphology. An important step toward identifying an arrhythmia is the classification of heartbeats. The rhythm of the ECG signal can then be determined by knowing the classification of consecutive heartbeats in the signal [6]. Classification of heartbeats can be very time-consuming and hence any automated processing of the ECG that assists this process would be of assistance and is the focus of this study.

Automated classification of heartbeats has been previously reported by other investigators (e.g., [7]–[12]) using a variety of features to represent the ECG and a number of classification methods. Features include ECG morphology [8], [9], heartbeat interval features [8]–[11], frequency-based features [7], higher order cumulant features [10], Karhunen–Loeve expansion of ECG morphology [11], and hermite polynomials [12]. Classifiers methods employed include linear discriminants [7], back propagation neural networks [8]–[10], self-organizing maps with learning vector quantization [11], and self-organizing networks [12].

Despite standards recommended for reporting performance results of cardiac rhythm algorithms by the Association for the Advancement of Medical Instrumentation (AAMI) [13], [14], only the work in [11] and [12] have utilized these standards. This makes it very difficult to assess the relative merits of the different algorithms. The AAMI standards are adopted in this study and our results have been compared to those of [11] and [12].

In [11], Hu *et al.* customized a heartbeat classifier to a specific patient (known as a local classifier) and then combined it with a global classifier designed from a large database of ECG signals. The two classifiers were then combined using a mixture of experts (MOE) approach. The local classifier requires a cardiologist to annotate a segment of a patient-specific ECG in order to implement the MOE approach. The global classifier achieved an accuracy of 62.2% and the MOE classifier achieving 94.0% accuracy for distinguishing ventricular ectopic beats (VEB) from non-VEBs.

Lagerholm *et al.* [12] described a method for clustering ECG heartbeats from a recording into 25 clusters and determined that on average 98.5% of the heartbeats in any one cluster were from the same heartbeat class. This translates to a classification performance of 98.5% if an expert can correctly identify the dominant beat of a cluster.

TABLE I
MAPPING THE MIT-BIH ARRHYTHMIA DATABASE HEARTBEAT TYPES TO THE AAMI HEARTBEAT CLASSES

AAMI heartbeat class	N	S	V	F	Q
Description	Any heartbeat not in the S, V, F or Q classes	Supraventricular ectopic beat	Ventricular ectopic beat	Fusion beat	Unknown beat
MIT-BIH heartbeat types	normal beat (NOR)	atrial premature beat (AP)	premature ventricular contraction (PVC)	fusion of ventricular and normal beat (fVN)	paced beat (P)
	left bundle branch block beat (LBBB)	aberrated atrial premature beat (aAP)	ventricular escape beat (VE)		fusion of paced and normal beat (fPN)
	right bundle branch block beat (RBBB)	nodal (junctional) premature beat (NP)			unclassified beat (U)
	atrial escape beats (AE)	supraventricular premature beat (SP)			
	nodal (junctional) escape beat (NE)				

TABLE II
HEARTBEAT TYPES ASSOCIATED WITH THE EXTRACTED BEATS FOR THE FULL DATABASE, DATASET 1 (DS1) AND DATASET 2 (DS2) FROM THE MIT-BIH ARRHYTHMIA DATABASE. HEARTBEAT TYPE AND CLASS ABBREVIATIONS ARE DEFINED IN TABLE I

	Heartbeat type	NOR	LBBB	RBBB	AP	aAP	NP	SP	PVC	fVN	AE	NE	VE	P	fPN	U	
	Heartbeat class	N	N	N	S	S	S	S	V	F	N	N	V	Q	Q	Q	total
Full database	number	75054	8074	7259	2544	150	83	2	7129	803	16	229	106	7028	982	33	109492
	% of total	68.5	7.4	6.6	2.3	0.1	0.1	0.0	6.5	0.7	0.0	0.2	0.1	6.4	0.9	0.0	100.0
	recs	40	4	6	27	7	5	1	37	17	1	5	2	4	3	6	
DS1	number	38104	3949	3783	808	100	32	2	3682	415	16	16	105	0	0	8	51020
	% of total	74.7	7.7	7.4	1.6	0.2	0.1	0.0	7.2	0.8	0.0	0.0	0.2	0.0	0.0	0.0	100.0
	recs	18	2	3	14	3	3	1	17	10	1	3	1	0	0	3	
DS2	number	36444	4125	3476	1736	50	51	0	3220	388	0	213	1	0	0	7	49711
	% of total	73.3	8.3	7.0	3.5	0.1	0.1	0.0	6.5	0.8	0.0	0.4	0.0	0.0	0.0	0.0	100.0
	recs	19	2	3	13	4	2	0	16	7	0	2	1	0	0	2	

Dataset 1 comprises data from recordings 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223 and 230.
Dataset 2 comprises data from recordings 100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233 and 234.
The four paced recordings (102, 104, 107 and 217) were not included in dataset 1 or 2.

II. AIM

The aims of this study were first to examine the heartbeat classification problem and, second, to design and test an automatic classification system using a comprehensive ECG database following AAMI recommended practice [13]. Methodology improvements on previous approaches included:

- trialing eight representations of the ECG morphology;
- use of the five heartbeat classes recommended in [13];
- comparison of 12 classifier configurations processing features obtained from single and multiple ECG leads;
- weighting the training examples to prevent the large classes from dominating the training process.

In Section III, the ECG database is discussed. Section IV presents the methodology adopted for this classification problem. Tests and experiments that were performed are also described in detail in this section. The results of the classification performance of different classifier configurations are presented in Section V. Section VI discusses the significant results and a comparison is made with previously published work. Some conclusions are drawn in Section VII.

III. ECG DATA

Data from the MIT-BIH arrhythmia database [15] were used in this study, which includes recordings of many common and life-threatening arrhythmias along with examples of normal sinus rhythm. The database contains 48 recordings, each containing two 30-min ECG lead signals (denoted lead A and B). In 45 recordings, lead A is modified-lead II and for the other three is lead V5. Lead B is lead V1 for 40 recordings and is either lead II, V2, V4, or V5 for the other recordings. Twenty-three of the recordings are intended to serve as a representative sample of routine clinical recordings and 25 recordings contain complex ventricular, junctional, and supraventricular arrhythmias [15].

The data are bandpass filtered at 0.1–100 Hz and sampled at 360 Hz. There are over 109 000 labeled ventricular beats from 15 different heartbeat types (Table I lists the heartbeat types). There is a large difference in the number of examples of the heartbeat types as shown in Table II. The largest class is "Normal beat" (NOR) with over 75 000 examples and the smallest class is "Supraventricular premature beat" (SP) with just two examples.

In agreement with the AAMI recommended practice, the four recordings containing paced beats were removed from
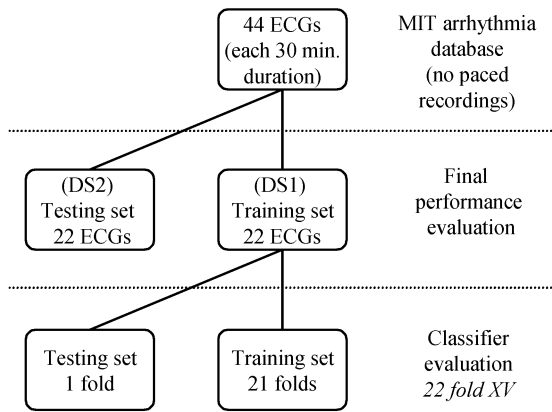
Fig. 1. Division of the MIT-BIH arrhythmia database into training and testing sets for classifier evaluation using the cross-validation scheme. Final performance evaluation is performed on an independent set.

the analysis. The remaining recordings were divided into two datasets with each dataset containing ECG data from 22 recordings with the same approximate proportion of beat types. Both datasets contain approximately 50 000 heartbeats and have a mixture of the routine and complex arrhythmia recordings. Table II shows the breakdown of each dataset by heartbeat type. The first dataset (DS1) was used to evaluate the performance of different candidate classifiers. The second dataset (DS2) was used for a final performance evaluation of the heartbeat classification system, as shown in Fig. 1.

The AAMI recommended practice was used to combine the MIT-BIH heartbeat types into five heartbeat classes which were used in all subsequent processing. Each class includes heartbeats of one or more types as shown in Table I. Class N contains beats originating in the sinus node (normal and bundle branch block beat types), class S contains supraventricular ectopic beats (SVEBs), class V contain ventricular ectopic beats (VEBs), class F contains beats that result from fusing normal and VEBs, and class Q contains unknown beats including paced beats.

IV. METHODS FOR AUTOMATED HEARTBEAT CLASSIFICATION

Fig. 2 depicts the stages of an automated system suitable for heartbeat classification which is based on the outcomes of this study. It consists of three stages: a preprocessing stage, a processing stage, and a classification stage. The digitized ECG is applied at the input to the preprocessing stage. The preprocessing stage utilizes a filtering unit to remove artifact signals from the ECG signal. These signals include baseline wander, power line interference, and high-frequency noise. The processing stage consists of heartbeat detection and feature extraction modules. The heartbeat detection module attempts to locate all heartbeats. The feature extraction module is concerned with forming a vector of measurements (feature vector) from each heartbeat that are processed by the classifier stage. The feature extraction modules are required because, although it is possible for the classification stage to process the ECG samples directly, greater classification performance is often achieved if a smaller number of discriminating features (than the number of ECG samples) are first extracted from the

ECG. The classification stage contains one or more classifier units which select one of the required classes in response to the input feature vector. The classifier units normally contain parameters which are set during the system development to optimize the classification performance. A combiner then unites the decisions of the classifier units to form the final decision of the system. The modules forming these stages are discussed in more detail below.

A. ECG Filtering

All ECG signals were filtered with two median filters to remove the baseline wander. Each signal was processed with a median filter of 200-ms width to remove QRS complexes and P-waves. The resulting signal was then processed with a median filter of 600 ms width to remove T-waves. The signal resulting from the second filter operation contained the baseline of the ECG signal, which was then subtracted from the original signal to produce the baseline corrected ECG signal. Unwanted power-line and high-frequency noise was removed from the baseline corrected ECG with a 12-tap low-pass filter. The filter was a finite impulse response filter with equal ripple in the pass and stop bands. The 3-dB point of the filter was 35 Hz. The filtered ECG signals were used in all subsequent processing.

B. Heartbeat Detection

This study did not investigate the problem of heartbeat detection from the ECG; instead we have utilized the heartbeat fiducial point times provided with the MIT-BIH arrhythmia database. The provided fiducial points occur at the instant of the major local extremum of a QRS-complex (i.e., either the time of the R-wave maximum or S-wave minimum). These fiducial points were first obtained automatically and then manually corrected on a beat-by-beat basis.

C. Heartbeat Segmentation

The ECG heartbeat segmentation program of Laguna *et al.*[1] was used to provide estimates of heartbeat segmentation points, i.e., the QRS onset and offset and T-wave offset times, a Boolean value indicating the presence/absence of a P-wave, and, if present, the P-wave onset and offset time for each heartbeat fiducial point. The program was applied separately to both ECG leads to form two sets of estimates of the segmentation points for each heartbeat.. The program has been validated on the Common Standards in Electrocardiography Multilead database [16] and the MIT-BIH QT database [17] and the accuracy of the system in determining heartbeat segmentation points was comparable with the interexpert variation.

D. Feature Extraction

Features relating to fiducial point intervals were calculated for each heartbeat. Features relating to heartbeat intervals and ECG morphology were calculated separately for the two ECG signals for each heartbeat (see the "Processing" stage of Fig. 2). Table III lists the features used in this study.

[1]"ecgpuwave": see http://www.physionet.org/physiotools/software-index.shtml.
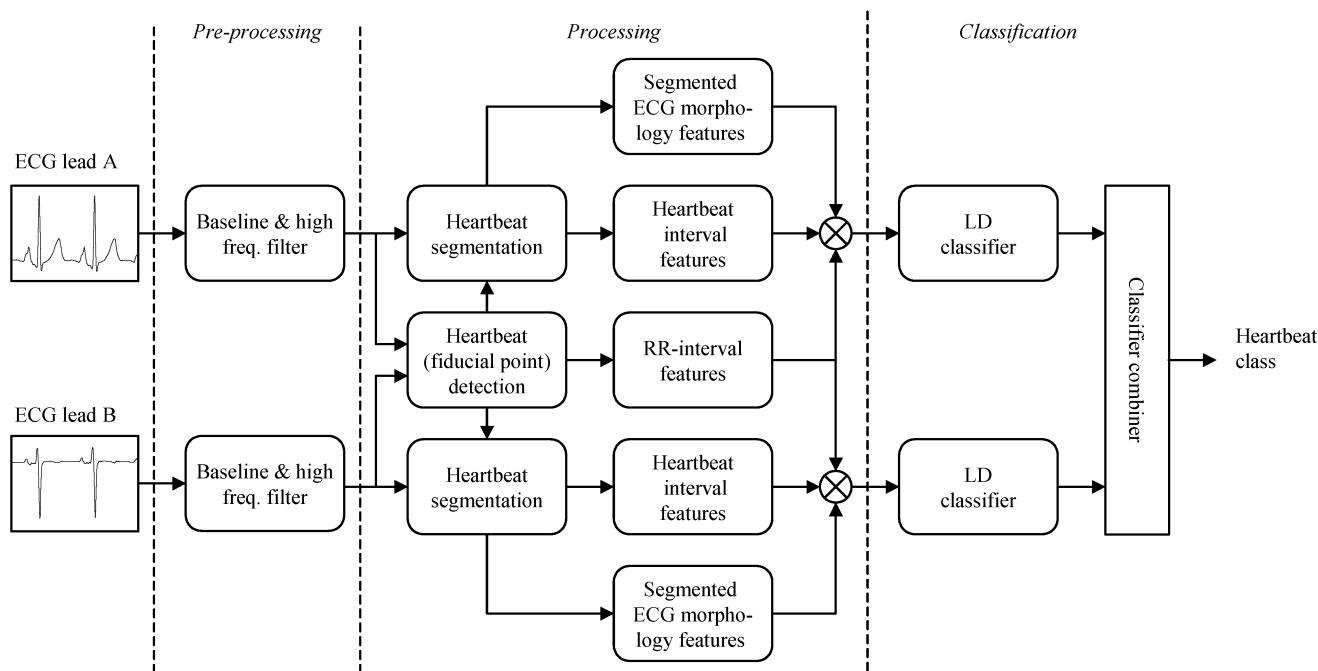
Fig. 2. Schematic representation of classifier configuration IX for the automated processing of the ECG for classification of heartbeats.

TABLE III
FEATURE GROUPS CONSIDERED IN THIS STUDY. WHERE FEATURES ARE
DERIVED SEPARATELY FROM THE TWO ECG LEADS, THE GROUP LABELS FOR
LEAD B FEATURES ARE SHOWN IN SQUARE BRACKETS

Group Label	Features
RR intervals	• Pre-RR interval • Post-RR interval • Average RR-interval • Local avg. RR-interval
Heart-beat intervals A [B]	• QRS duration (QRS offset -QRS onset) of lead A [B] • T-wave duration (T-wave offset - QRS offset) of lead A [B] • P wave flag for lead A [B]
Morphology 1A [1B]	• ECG morphology (10 samples) between QRS onset and QRS offset of lead A [B] • ECG morphology (9 samples) between QRS offset and T-wave offset of lead A [B]
Morphology 2A [2B]	• Normalised ECG morphology (10 samples) between QRS onset and QRS offset of lead A [B] • Normalised ECG morphology (9 samples) between QRS offset and T-wave offset of lead A [B]
Morphology 3A [3B]	• ECG morphology (10 samples) between FP-50ms to FP+100ms of lead A [B] • ECG morphology (8 samples) between FP+150ms to FP+500ms of lead A [B]
Morphology 4A [4B]	• Normalised ECG morphology (10 samples) between FP-50ms to FP+100ms of lead A [B] • Normalised ECG morphology (8 samples) between FP+150ms to FP+500ms of lead A [B]

1) RR-Interval Features: Heartbeat fiducial point intervals (henceforth called RR-intervals) were defined as the interval between successive heartbeat fiducial points. Four features (see Table III: RR-intervals) were extracted from the RR sequence.

The pre-RR-interval was the RR-interval between a given heartbeat and the previous heartbeat. The post-RR-interval was the RR-interval between a given heartbeat and the following heartbeat. The average RR-interval was the mean of the RR-intervals for a recording and had the same value for all heartbeats in a recording. Finally, the local average RR-interval was determined by averaging the RR-intervals of the ten RR-intervals surrounding a heartbeat.

2) Heartbeat Interval Features: Three features per ECG lead (see Table III: heartbeat intervals A and B) relating to heartbeat intervals were calculated after heartbeat segmentation. The QRS duration was the time interval between the QRS onset and the QRS offset. The T-wave duration was defined as the time interval between the QRS offset and the T-wave offset. The third feature was a Boolean variable indicating the presence or absence of a P-wave.

3) ECG Morphology Features: Features used in this study to represent the morphology of the ECG signals were divided into eight groups. All groups contained amplitude values of the ECG signal. The major distinction between the groups was the method used for sampling the signals. Four groups used a sampling rate determined by the heartbeat segmentation information and four groups used a fixed-interval sampling rate.

a) Segmented ECG morphology features: Four groups used a sample rate and sampling windows determined by the QRS onset and offset and the T-wave offset points and contained 19 features each. Fig. 3(a) depicts the sampling process. Two sampling windows were utilized. The boundaries of the first window were determined by the QRS onset and offset. Ten features were derived by uniformly sampling the ECG amplitude in this window. The second window was bounded by the QRS offset and the T-wave offset. Nine features of the ECG amplitude were formed by uniformly sampling in this window. As the ECG signals were already sampled, linear
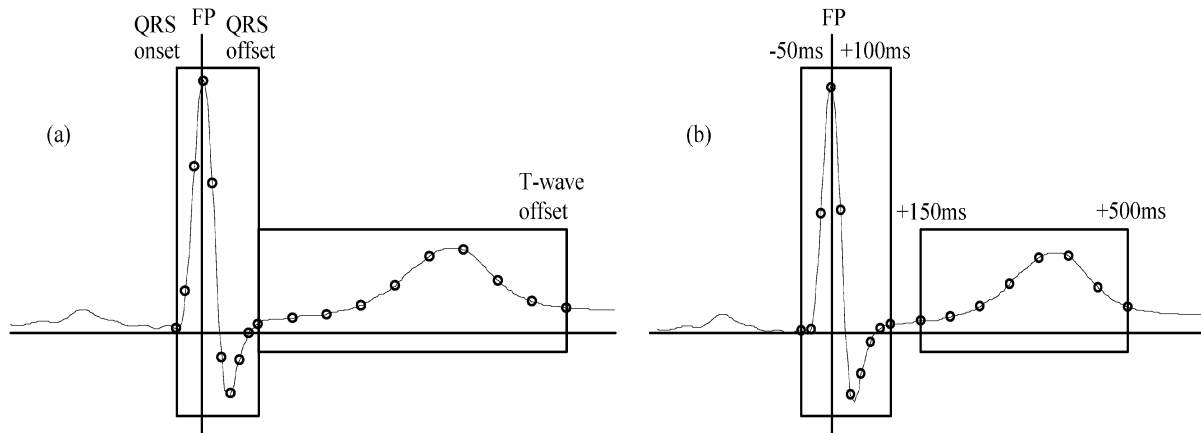
Fig. 3. Two time-sampling methods for extracting ECG morphology features. (a) After determining the fudicial point (FP), the QRS onset and offset and T-wave offset points are found. Ten evenly spaced samples of the ECG between the QRS onset and offset and nine evenly spaced samples of the ECG between the QRS offset and T-wave offset are extracted. (b) After determining the FP, nine samples of the ECG between FP-50 ms and $FP + 100$ ms and nine samples between $FP+150$ ms and $FP+500$ ms are extracted.

interpolation was used to estimate the intermediate values of the ECG amplitude.

The above sampling method was applied to lead A to form the Morphology 1A group (see Table III: Morphology 1A) and applied to lead B to form the Morphology 1B group (see Table III: Morphology 1B).

The Morphology 2A group (see Table III: Morphology 2A) and the Morphology 2B group (see Table III: Morphology 2B) were calculated in an identical way to the Morphology 1A and 1B groups, respectively, except that the entire ECG signal was scaled so that the standard deviation of all sample values was unity before the sampling windows were applied. This scaling operation resulted in a signal that was independent of the original ECG signal amplitude.

b) Fixed-interval ECG morphology features: The sample rate was fixed and the sampling windows were located by the heartbeat fiducial point (FP) for four groups. Fig. 3(b) depicts the sampling process. Two sampling windows were formed based on FP. The first window approximately covered the QRS-complex and contained the portion of the ECG between FP-50 ms and $FP+100$ ms. A 60-Hz sampling rate was applied to the ECG amplitude resulting in ten features. The second window approximately contained the T-wave and started at $FP+150$ ms and finished at $FP+500$ ms. The ECG signal amplitude was sampled at 20 Hz in this window, resulting in eight features. Lower sampling rates were used for T-wave sampling windows as the frequency content of this wave is lower than the frequency content of the QRS-complex. This reduced the number of feature values to be processed by the classifier.

The above sampling method was applied to lead A to form the Morphology 3A group (see Table III: Morphology 3A) and applied to lead B to form the Morphology 3B group (see Table III: Morphology 3B).

The Morphology 4A group (see Table III: Morphology 4A) and the Morphology 4B group (see Table III: Morphology 4B) were calculated in an identical way to the Morphology 3A and 3B groups, respectively, after scaling the ECG signal as described previously.

4) Feature Sets: The above features were combined into eight feature sets which are summarized in Table IV. Feature

TABLE IV
EIGHT FEATURE SETS INVESTIGATED IN THIS STUDY

Feature Set	Size	Feature Groups
FS1	26	RR-intervals, Heartbeat intervals lead A, Morphology 1A
FS2	26	RR-intervals, Heartbeat intervals lead A, Morphology 2A
FS3	22	RR-intervals, Morphology 3A
FS4	22	RR-intervals, Morphology 4A
FS5	26	RR-intervals, Heartbeat intervals lead B, Morphology 1B
FS6	26	RR-intervals, Heartbeat intervals lead B, Morphology 2B
FS7	22	RR-intervals Morphology 3B
FS8	22	RR-intervals, Morphology 4B

set 1 (FS1) contained 26 features including the RR features and the features from lead A requiring the determination of the heartbeat segmentation points. These features included the unscaled segmented ECG morphology group (Morphology 1A) and the heartbeat interval A group. Feature set 2 contained the same features as FS1 except that the scaled segmented ECG morphology group (Morphology 2A) replaced the Morphology 1A group. Feature set 3 (FS3) contained 22 features including the RR features and the unscaled fixed-interval morphology features from lead A (Morphology 3A). Thus, the determination of the heartbeat segmentation points was not required for this feature set. Feature set 4 contained the same features as FS3 except that the scaled fixed-interval ECG morphology group (Morphology 4A) features replaced the Morphology 3A group.

Feature sets 5–8 (FS5 to FS8) were similar to FS1 to FS4 except that features included were calculated from lead B (see Table IV).

The eight feature sets were used to examine the effect on classification performance of lead placement, alternative ECG morphology representations, and inclusion of heartbeat segmentation information.

E. Classifier Model

Classifier models based on linear discriminants (LDs) were utilized throughout this study. The model parameters were determined using "plug-in" maximum-likelihood estimates calculated from the training data.

For linear discriminants, the unweighted likelihood (L) function is defined as [18]

$$L = \sum_{k=1}^{c} \sum_{n=1}^{N_k} \log\left(f_k\left(\mathbf{x}_{kn}, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}\right)\right) \qquad (1)$$

where the number of classes is c, the number of training examples in class k is N_k, and $f_k(\mathbf{x}_{kn}, \boldsymbol{\mu}_k, \boldsymbol{\Sigma})$ is the value of the Gaussian distribution with mean $\boldsymbol{\mu}_k$ and common covariance $\boldsymbol{\Sigma}$ evaluated at training example $\mathbf{x}_{kn}$. The training process determines the parameter values of $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}$ that maximizes the value of L.

The relative proportions of the classes of the available training examples influences the performance of an LD classifier. If a few classes dominate the training data examples, then the classifier training process is heavily influenced by these classes [18]. One solution to avoid this problem is to include a subsample of the training examples of the larger classes in the training process but this is wasteful of training data. A better solution implemented in this study is to include all training examples but reduce the relative contribution of the training examples of the large classes to the training process. This is achieved by weighting the contribution of each training example to the likelihood function by a class dependent factor w_k.

A weighted likelihood function is

$$L = \sum_{k=1}^{c} w_k \sum_{n=1}^{N_k} \log\left(f_k\left(\mathbf{x}_{kn}, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}\right)\right) \qquad (2)$$

and the contribution of a training example to L is directly related to w_k.

This is maximized when the mean vectors are defined as

$$\boldsymbol{\mu}_k = \sum_{n=1}^{N_k} \mathbf{x}_{kn}/N_k \qquad (3)$$

and the covariance matrix is defined as

$$\boldsymbol{\Sigma} = \left(\sum_{k=1}^{c} w_k \sum_{n=1}^{N_k} \left(\mathbf{x}_{kn} - \boldsymbol{\mu}_k\right)\left(\mathbf{x}_{kn} - \boldsymbol{\mu}_k\right)^T\right) \Big/ \sum_{k=1}^{c} w_k N_k \ . \qquad (4)$$

In this study, values for w_k were determined as follows. All classes were weighted so that, as far as possible, each class contributed the equivalent of 400 examples to the training process. This threshold was chosen as it was felt that classes with less examples were not adequately represented and hence their contribution to the likelihood function needed to be reduced. If the number of examples in a class (N_k) was greater than 400, then the w_k was set to $400/N_k$. For the classes with less than 400 examples, no weighting was used (i.e., $w_k = 1$). The values of w_k for the five heartbeat classes were: N: $400/45\,868$; S: $400/942$; V: $400/3787$; F: $400/415$; and Q: 1.

F. Classifying and Combining Classifiers

After determining the $\boldsymbol{\mu}_k$'s and $\boldsymbol{\Sigma}$ from the training data, a feature vector $\mathbf{x}$ is classified by assuming values for the prior

probabilities π_k and calculated the estimated posterior probabilities $P(k \,|\, \mathbf{x})$ for the kth class using

$$P(k \,|\, \mathbf{x}) = \frac{\exp(y_k)}{\sum_{l=1}^{c} \exp(y_l)} \qquad (5)$$

where $y_k = -(1/2)\boldsymbol{\mu}_k^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k + \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}^{-1} \mathbf{x} + \log(\pi_k)$.

The prior probability of the N, S, V, and F classes was set to 10/41. As the number of unknown beats was a tiny fraction of all the heartbeats, the prior probability of the Q class was reduced relative to the other classes and was set to 1/41.

The final classification of a single feature set system was obtained by choosing the class with the highest posterior probability estimate from (5).

To obtain a classification based on processing information from multiple feature sets simultaneously, the posterior probabilities obtained from each feature set were combined across the separate classifier outputs. Assuming that the outputs from M classifiers are to be combined, the final posterior probability output $\bar{P}(k \,|\, \mathbf{x})$ was calculated from the individual classifier outputs $P_m(k \,|\, \mathbf{x})$ using the unweighted Bayesian product integration scheme [19] as follows:

$$\bar{P}(k \,|\, \mathbf{x}) = \prod_{m=1}^{M} P_m(k \,|\, \mathbf{x}) \Big/ \sum_{l=1}^{M} \prod_{m=1}^{M} P_m(l \,|\, \mathbf{x}) \ . \qquad (6)$$

As before, the final classification is obtained by choosing the class with the highest posterior probability estimate. By using information from all available signals, more efficient use of the available ECG diagnostic information was made.

G. Candidate Classifier Configurations

Twelve classifiers configurations were evaluated. Configurations I-VIII were single-lead configurations processing FS1-FS8, respectively. These configurations utilized one LD classifier. Configurations IX-VII were multiple-lead configurations that processed FS1 and FS5; FS2 and FS6; FS3 and FS7; and FS4 and FS8 respectively. These four configuration implemented two LD classifiers (each processing one feature set) and were combined using the integration scheme of (6).

H. Classification Performance Measures

The performance of the classifier configurations was calculated using a number of measures. Table V shows how the full classification matrix is used to calculate the performance measures. In accordance with the AAMI recommendations, two sets of performance measurements were calculated. These performance measures focus on the ability of algorithms to distinguish VEBs from non-VEBs [Table V(a)] and SVEBs from non-SVEBS [Table V(b)]. It is worth noting that the AAMI-recommended calculations shown in 4(a) do not reward or penalize a classifier for the classification of ventricular fusion (F) or unknown beats (Q) as VEBs. Similarly, in Table V(b), the calculations do not reward or penalize a classifier for the classification of unknown beats as SVEBs.

The performance measures listed in Table V(a) and (b) do not measure the ability of the classifier to separate the heartbeats into the multiple classes simultaneously, i.e., the multiway classification performance. While these measures are very useful

TABLE V

PERFORMANCE MEASURES USED IN THIS STUDY FOR (a) DISTINGUISHING VEBS FROM NON-VEBS, (b) DISTINGUISHING SVEBS FROM NON-SVEBS, AND (c) DISTINGUISHING THE FIVE AAMI HEARTBEAT CLASSES. THE PERFORMANCE MEASURES ARE ALL DERIVED FROM THE MUTUALLY EXCLUSIVE GROUPS OF TRUE POSITIVES (TP), TRUE NEGATIVES (TN), FALSE POSITIVES (FP), AND FALSE NEGATIVES (FN). THE CELLS ASSOCIATED WITH THESE GROUPS IN (a) AND (b) ARE SHADED TO AID IDENTIFICATION

(a)

Algorithm label

Reference label		n	s	v	f	q
	N	Nn	Ns	Nv	Nf	Nq
	S	Sn	Ss	Sv	Sf	Sq
	V	Vn	Vs	Vv	Vf	Vq
	F	Fn	Fs	Fv	Ff	Fq
	Q	Qn	Qs	Qv	Qf	Qq

$$TN_V = Nn + Ns + Nf + Nq + Sn$$
$$+ Ss + Sf + Sq + Fn + Fs + Ff$$
$$+ Fq + Qn + Qs + Qf + Qq$$
$$FN_V = Vn + Vs + Vf + Vq$$
$$TP_V = Vv$$
$$FP_V = Nv + Sv$$
$$VEB\ Se = TP_V / (TP_V + FN_V)$$
$$VEB + P = TP_V / (TP_V + FP_V)$$
$$VEB\ FPR = FP_V / (TN_V + FP_V)$$
$$VEB\ Acc = \frac{TP_V + TN_V}{TP_V + TN_V + FP_V + FN_V}$$

(b)

Algorithm label

Reference label		n	s	v	f	q
	N	Nn	Ns	Nv	Nf	Nq
	S	Sn	Ss	Sv	Sf	Sq
	V	Vn	Vs	Vv	Vf	Vq
	F	Fn	Fs	Fv	Ff	Fq
	Q	Qn	Qs	Qv	Qf	Qq

$$TN_S = Nn + Nv + Nf + Nq + Vn$$
$$+ Vv + Vf + Vq + Fn + Fv + Ff$$
$$+ Fq + Qn + Qv + Qf + Qq$$
$$FN_S = Sn + Sv + Sf + Sq$$
$$TP_S = Ss$$
$$FP_S = Ns + Vs + Fs$$
$$SVEB\ Se = TP_S / (TP_S + FN_S)$$
$$SVEB + P = TP_S / (TP_S + FP_S)$$
$$SVEB\ FPR = FP_S / (TN_S + FP_S)$$
$$SVEB\ Acc = \frac{TP_S + TN_S}{TP_S + TN_S + FP_S + FN_S}$$

(c)

Algorithm label

Reference label		n	s	v	f	q	sum
	N	Nn	Ns	Nv	Nf	Nq	ΣN
	S	Sn	Ss	Sv	Sf	Sq	ΣS
	V	Vn	Vs	Vv	Vf	Vq	ΣV
	F	Fn	Fs	Fv	Ff	Fq	ΣF
	Q	Qn	Qs	Qv	Qf	Qq	ΣQ
							Σ

$$TN = Nn$$
$$TP_V = Vv$$
$$TP_S = Ss$$
$$TP_F = Ff$$
$$TP_Q = Qq$$
$$Sp = TN / \Sigma N$$
$$VEB\ Se : see\ Table\ 4(a)$$
$$SVEB\ Se : see\ Table\ 4(b)$$
$$F\ Se : TP_F / \Sigma F,$$
$$Q\ Se : TP_Q / \Sigma Q$$
$$Acc : (TN + TP_S + TP_V + TP_F + TP_Q) / \Sigma$$

Abbreviations: *Acc*: Accuracy, *F*: Fusion beat class, *FPR*: False positive rate, *N*: Normal beat class, *+P*: Positive predictivity, *Q*: Unknown beat class, *Se*: sensitivity, *Sp*: Specificity, *S & SVEB*: Supraventricular ectopic beat class, *V & VEB*: Ventricular ectopic beat class.

clinically, they were not convenient for comparing the performance of different classifier configurations. To rank the performance of the different classifier configurations, the multiway classification accuracy and class sensitivities were calculated for each classifier [see Table V(c)] and compared.

All aggregate performance measures for a dataset were calculated by giving each heartbeat equal weighting (the so-called "gross" statistics in [13]).

I. Clustering Performance Measure

The method described in this study was compared to the heartbeat clustering method of [12]. Heartbeat clustering systems do not classify heartbeats but attempt to group all heartbeats from the same heartbeat class into a unique cluster. For these systems, a heartbeat is defined as misclustered if it assigned to a cluster where the dominant beat type is a different class. To determine the clustering performance of our method, each heartbeat was processed and the five possible outputs of the classifiers were considered as cluster labels rather than heartbeat classes. Once the cluster labels were determined for all heartbeats in a recording, the expert annotations were used to determine the dominant heartbeat class of each cluster. To compare our method to [12], a similar error measure (R^{LD}) to their R^{SOM} measure was defined.

R^{LD} was calculated by first determining the number of correctly clustered VEBs (CC_{VEB}) and correctly clustered non-VEBs $(CC_{non-VEB})$. A VEB was correctly clustered if a beat belonging to the VEB class was assigned to a cluster

where the dominant beat was the VEB class. A non-VEB was correctly clustered if a beat belonging to N, S, F, or Q beat classes was assigned to a cluster where the dominant beat was either N, S, F or Q beat class. R^{LD} was then determined using

$$R^{LD} = \frac{CC_{VEB} + CC_{non-VEB}}{total\ number\ of\ beats}. \tag{7}$$

J. Assessing the Performance of the Classifiers

Classifier performance was estimated using two methods in this study. In the first method, the cross-validation scheme [20], [21] was applied to DS1. The available data was divided into folds in two ways. In the first division (DD1), the data was assigned to 22 folds with each fold containing data from one recording. In the second division (DD2), all heartbeats were considered independent and randomly assigned to the 22 folds. The cross-validation method was used to assess the performance of different candidate classifiers and identify the best performing classifier for a second performance estimation.

The second method of performance estimation was to train the system using DS1 and test the system by processing DS2. Fig. 1 shows how the data was divided in this study into separate sets for performance assessment.

V. RESULTS

The multiway classification performance figures are shown in Table VI for the candidate classifier configurations considered

TABLE VI

MULTIWAY PERFORMANCE OF THE CANDIDATE CONFIGURATIONS ON DS1. (A) CONFIGURATION I USING THE BEAT-BASED DATA DIVISION SCHEME AND (B) SINGLE-LEAD AND (C) MULTI-LEAD CONFIGURATIONS USING THE RECORD-BASED DATA DIVISION SCHEME. THE SELECTED CONFIGURATION IS IN BOLD

	Config.	Data div. scheme	Feature set(s)	ECG scaling	Heartbeat segmenting	ECG signal	Acc (%)	Spec (%)	Sensitivities (%)			
									SVEB	VEB	F	Q
(a)	I	DD2	FS1	N	Y	A	86.1	87.0	77.6	76.4	87.0	0.0
	I	DD1	FS1	N	Y	A	79.0	80.6	59.9	65.3	74.5	25.0
	II	DD1	FS2	Y	Y	A	76.9	78.3	44.5	67.3	85.3	25.0
	III	DD1	FS3	N	N	A	81.8	86.5	56.5	39.1	16.1	12.5
(b)	IV	DD1	FS4	Y	N	A	77.8	80.9	32.6	57.0	23.1	12.5
	V	DD1	FS5	N	Y	B	74.1	76.6	32.6	55.6	62.2	12.5
	VI	DD1	FS6	Y	Y	B	74.9	77.2	29.8	60.7	51.6	0.0
	VII	DD1	FS7	N	N	B	83.0	88.1	30.4	39.9	32.1	0.0
	VIII	DD1	FS8	Y	N	B	78.6	82.5	29.0	42.8	80.7	0.0
	IX	**DD1**	**FS1,FS5**	**N**	**Y**	**A+B**	**84.5**	**86.7**	**53.3**	**67.3**	**71.6**	**12.5**
(c)	X	DD1	FS2,FS6	Y	Y	A+B	82.9	84.8	46.1	69.1	83.6	0.0
	XI	DD1	FS3,FS7	N	N	A+B	86.2	91.2	54.7	40.0	26.0	0.0
	XII	DD1	FS4,FS8	Y	N	A+B	84.2	87.3	45.3	58.7	64.8	0.0

in this study. All performance figures were determined by aggregating the test-set figures from the 22 test-set folds. Table VI(a) presents results for configuration I using cross-validation applied to DD2. Results in Table VI(b) and VI(c) are obtained using cross validation applied to DD1. Table VI(b) presents results for the single-lead configurations processing FS1-FS8. Table VI(c) presents results for the four multilead configurations. The objective of this section of the study was to select a classifier configuration with the best performance in separating all classes simultaneously. For this reason Table VI shows the multiway classification performance measures only.

After consideration of the results in Table VI, configuration IX was chosen as the best performing classifier. It was retrained using DS1 and tested on DS2. A schematic diagram of a system implementing configuration IX is shown in Fig. 2. In accordance with AAMI recommendations, the classification performance for each recording and the gross performance figures were calculated for DS2 (test-set) and are shown in Table VII(a). For comparison, the classification performance of a published system [11] is shown in Table VII(b). The clustering performance is shown in Tables VII(c) and VIII shows a summary table of beat-by-beat performance.

VI. DISCUSSION

The classification performance for configuration I was determined using the two data division schemes, DD1 and DD2. The results show that the classification performance figures resulting from DD2 [Table VI(a)] are notably higher than the same figures resulting from DD1 [Table VI(b)]. The multiway accuracy estimated using DD2 was 86.1% which was 7.1% higher than the same result estimated using DD1. The same trends can be seen for the specificity and sensitivity results in this table.

These results demonstrate that dividing data on a beat basis results in optimistically biased classifications results. This scheme is biased as intrasubject beats can be highly dependent and hence dependent beats may appear in the training and testing sets. This data division scheme has been previously favored by other investigators (e.g., [7], [9], [10]) and should be avoided. The DD1 scheme is a superior scheme to DD2

as the test data do not contain any heartbeats from subjects represented in the training data. For the rest of this study, the DD1 scheme was employed.

A. Performance of the Candidate Configurations

In terms of accuracy, the best-performing single-lead classifier was configuration VII with an accuracy figure of 83.0%. Inspection of the sensitivity and specificity results reveal that, while it achieved a high specificity (88.1%), the resulting sensitivities were very low (SVEB 30.4%, VEB 39.9%, F 32.1%, and Q 0.0%), and so this configuration would be of limited clinical use. Of more benefit is a system with more balanced specificity and sensitivity performance figures. On this basis, classifier configuration I achieved the best performance (accuracy 79.0%, specificity 80.6%, and sensitivities: SVEB 59.9%, VEB 65.3%, fusion (F) 74.5%, and unknown (Q) 25.0%) of the single-lead configurations.

The multilead configurations always resulted in higher accuracy performance than the single-lead configurations processing the same feature sets. For the four multilead configurations assessed, the increase in accuracy occurred because of an increase in specificity and the VEB sensitivity with no change or a slight decrease in the other specificities compared to the single-lead configurations. Thus, combining the information from multiple ECG leads was a good performance enhancing strategy.

Configuration XI, in terms of accuracy, was the best-performing classifier overall (86.2%). It achieved a high specificity of 91.2% but modest sensitivity figures for the SVEB (54.7%), VEB (40.0%), fusion (F-26.0%), and unknown (Q-0.0%) classes. As a result, this was not judged to be the best-performing configuration classifier. Instead, configuration IX was chosen as it had more balanced sensitivity and specificity performance figures (specificity 86.7%; sensitivities: SVEB 53.3%, VEB 67.3%, F 71.6%, and Q 12.5%) while maintaining a high overall accuracy of 84.5%. This classifier processed feature sets that utilized unscaled ECG morphology and included heartbeat segmentation points. Thus, the inclusion of the absolute amplitude of an ECG beat and the addition of heartbeat segmentation information to the feature-extraction process

TABLE VII
(A) CLASSIFICATION PERFORMANCE OF CONFIGURATION IX ON EACH RECORDING OF DS2 USING THE AAMI RECOMMENDED PERFORMANCE MEASURES, (B)
AGGREGATE CLASSIFICATION PERFORMANCE FOR THE SYSTEM IN [11], AND (C) THE CLUSTERING PERFORMANCE OF CLASSIFIER IX AND SYSTEM FROM [12]

		number of beats					SVEB				VEB				Clustering	
Rec	N	S	V	F	Q	Acc	Se	+P	FPR	Acc	Se	+P	FPR	R^{LD}	R^{SOM}
100	2239	33	1	0	0	99.3	51.5	100.0	0.0	100.0	100.0	100.0	0.0	100.0	
103	2082	2	0	0	0	99.5	50.0	10.0	0.4	100.0	-	-	0.0	100.0	
105	2526	0	41	0	5	99.8	-	0.0	0.2	98.0	97.6	44.0	2.0	98.4	
111	2123	0	1	0	0	100.0	-	-	0.0	99.9	100.0	33.3	0.1	100.0	
113	1789	6	0	0	0	98.8	100.0	21.4	1.2	100.0	-	-	0.0	100.0	
117	1534	1	0	0	0	99.9	100.0	50.0	0.1	100.0	-	-	0.0	100.0	
121	1861	1	1	0	0	99.7	0.0	0.0	0.3	100.0	100.0	50.0	0.1	100.0	
123	1515	0	3	0	0	99.5	-	0.0	0.5	99.8	0.0	-	0.0	99.8	
200[a]	1743	30	826	2	0	96.3	23.3	8.6	2.9	93.7	82.1	97.7	0.9	95.6	98.1
202[a]	2061	55	19	1	0	70.0	54.6	4.7	29.6	98.4	89.5	34.0	1.6	99.1	99.4
210[a]	2423	22	195	10	0	95.0	45.5	7.7	4.6	93.8	62.1	57.1	3.7	93.8	98.3
212	2748	0	0	0	0	98.7	-	0.0	1.3	98.8	-	0.0	1.2	100.0	
213[a]	2641	28	220	362	0	99.0	10.7	30.0	0.2	93.9	21.8	65.8	0.8	94.0	97.6
214[a]	2002	0	256	1	2	97.6	-	0.0	2.4	95.8	78.9	83.1	2.1	98.1	99.9
219[a]	2082	7	64	1	0	86.3	0.0	0.0	13.5	97.5	29.7	67.9	0.4	97.5	99.6
221[a]	2031	0	396	0	0	94.6	-	0.0	5.4	99.2	99.0	96.3	0.7	99.2	100.0
222	2274	209	0	0	0	77.3	82.3	24.6	23.2	98.6	-	0.0	1.5	100.0	
228[a]	1688	3	362	0	0	95.5	0.0	0.0	4.4	94.6	69.9	99.6	0.1	99.1	100.0
231[a]	1568	1	2	0	0	88.3	0.0	0.0	11.7	99.9	0.0	-	0.0	99.9	99.9
232	398	1382	0	0	0	85.1	80.9	99.8	0.5	90.1	-	0.0	9.9	100.0	
233[a]	2230	7	831	11	0	98.2	0.0	0.0	1.6	95.9	87.6	96.8	1.1	98.3	99.8
234[a]	2700	50	3	0	0	99.3	60.0	100.0	0.0	100.0	100.0	100.0	0.0	100.0	100.0
gross	44258	1837	3221	388	7	94.6	75.9	38.5	4.7	97.4	77.7	81.9	1.2	98.6	
gross[b]										96.4	77.5	90.6	1.1	97.5	99.2

(a) label at left of table.

(b)
	Acc	Se	+P	FPR
GE[b] [11]	75.3	69.6	34.6	16.8
MOE[b] [11]	93.6	78.9	76.0	3.2

(c) label at top right.

[a]Recordings in common with the test-set used in [11,12].
[b]Gross statistics calculated using recordings in common with [11,12].
[c]R^{SOM} calculated using equation 24 and Table VII in [12].

were worthwhile steps. This latter observation contradicts the findings of Millet *et al.* [22].

B. Final Testing

Table VII(a) shows the performance assessment, as recommended by the AAMI standards, of configuration IX on DS2. This assessment is unbiased as DS2 was not used at any point in the development of the classifier.

For the SVEB class, the gross sensitivity was 75.9%, the positive predictivity was 38.5%, and the false positive rate (FPR) was 4.7%. The gross sensitivity was 77.7%, the positive predictivity was 81.9%, and the FPR was 1.2% for the VEB class. Thus, the performance of our system in detecting SVEBs is not as high as the performance of the system in detecting VEBs. A likely reason for this is that there were 942 SVEBs in the training data which was notably less than the 3787 VEBs available.

The summary table of beat-by-beat performance (Table VIII) provides insight into how groups are being misclassified. These results show that two main errors are being made by the system based on configuration IX. First, 1904 normal (N) beats were misclassified as SVEB (S) beats, and, second, 3509 N beats were misclassified as fusion (F) beats. Future work needs to look at identifying features that more clearly distinguish normal beats from these two aberrant beat classes.

TABLE VIII
SUMMARY TABLE OF BEAT-BY-BEAT PERFORMANCE
ON DS2 OF CLASSIFIER IX

		Algorithm				
		n	s	v	f	q
Reference	N	38444	1904	303	3509	98
	S	173	1395	252	16	1
	V	117	321	2504	176	103
	F	33	1	7	347	0
	Q	4	0	3	0	0

Distinguishing normal beats from fusion beats is an inherently difficult problem as fusion beats are a union of ventricular and normal beats.

C. Comparison With Other Automated Heartbeat Processing Systems

The VEB classification results of this study were compared to the reported results of Hu *et al.* [11], as this study looked at the problem of distinguishing VEB from non-VEB heartbeats on the MIT-BIH arrhythmia database. Hu used a test-set of 20 recordings which excluded all recordings with no premature ventricular contractions (PVCs). There are 11 test-set recordings (see footnote "a" of Table VII) that are common to both

studies and the aggregate performance results on these recordings were compared directly. The gross performance of configuration IX on these recordings was: accuracy 96.4%, sensitivity 77.5%, positive predictivity 90.6%, and FPR 1.1%.

Using [11, Table VI], the gross sensitivity, positive predictivity, and FPR rate were recalculated for the 11 recordings in common and results are shown in Table VII(b). Our system is notably better than the fully automatic configuration (GE classifier) approach (accuracy 75.3%, sensitivity 69.6%, positive predictivity 34.6%, and FPR 16.8%) and is comparable to the results obtained with their semi-automatic MOE approach (accuracy 93.6%, sensitivity 78.9%, positive predictivity 76.0%, and FPR 3.2%). The significant difference is that our system does not require an expert to annotate part of the recording under test in order to train the local classifier for the MOE approach.

Lagerholm *et al.* [12] also used the same 20 recordings as Hu to determine the performance of clustering VEB and non-VEBs. Using the 11 recordings in common with our study, a comparison was made of the clustering performance of our system and Lagerholm's [see Table VII(c)]. Our system achieved a clustering accuracy (R^{LD}) of 97.5% which was below the 99.2% (R^{SOM}) achieved by Lagerholm. It should be noted that our system used five clusters whereas Lagerholm's system used 25 clusters, and so our system is attempting a more constrained and therefore more difficult clustering problem. A significant advantage of our system over Lagerholm's is that no annotating of the dominant beat of a cluster by an expert is required before the classification of a cluster can be made.

D. Automatic Heartbeat Detection

To fully automate the heartbeat classification method presented here, an automatic heartbeat detection module is required. Automatic heartbeat detection results in some errors in heartbeat detection (missed heartbeats, erroneously detected heartbeats, and errors in heartbeat fiducial point identification), and this will cause a reduction in the performance of the presented heartbeat classifier method for two reasons. First, a missed heartbeat or an erroneously detected heartbeat cannot be classified correctly. Second, there will be errors introduced into the RR-interval features due to heartbeat detection and fiducial point identification errors. A number of schemes exist that claim less than 0.5% error rate in detecting heartbeats e.g., [12], [23], [24]. This error rate is much less than the error rate of our heartbeat classification method (e.g., VEB error rate $= 100 - $ VEB Acc $= 3.6\%$), and it is anticipated that automating the heartbeat detection process with such a scheme would not degrade the heartbeat classification performance of our system appreciably.

VII. CONCLUSION

Using the MIT-BIH arrhythmia database, we have discussed the development of a method for the processing of the ECG for the classification of heartbeats into five groups: normal beats, VEBs, SVEBs, fusion of normal and VEBs, and unknown beat types. In this study, heartbeat fiducial points were determined manually. The classification performance of 12 classifier configurations were compared and the best configuration chosen for

an independent performance assessment. The chosen configuration processed feature sets derived from two ECG leads each utilizing features derived from RR-intervals, unscaled ECG morphology, and heartbeat segmentation information. The configuration implemented two linear discriminant classifiers (each processing one feature set) which combined the classifier outputs to form the final decision. The independent performance assessment of this configuration resulted in a sensitivity of 75.9%, a positive predictivity of 38.5%, and an FPR of 4.7% for the SVEB class. The VEB sensitivity was 77.7%, the positive predictivity was 81.9%, and the FPR was 1.2%. These results are an improvement on previously reported results for automated heartbeat classification systems.

REFERENCES

[1] M. S. Thaler, *The Only EKG Book You'll Ever Need*, 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins, 1999.

[2] K. Minami, H. Nakajima, and T. Toyoshima, "Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network," *IEEE Trans. Biomed. Eng.*, vol. 46, pp. 179–185, Feb. 1999.

[3] S. Evans, H. Hastings, and M. Bodenheimer, "Differentiation of beats of ventricular and sinus origin using a self-training neural network," *PACE*, vol. 17, pp. 611–626, 1994.

[4] R. Clayton, A. Murray, and R. Campbell, "Recognition of ventricular fibrillation using neural networks," *Med. Biol. Eng. Comput.*, vol. 32, pp. 217–220, 1994.

[5] S. Barro, R. Ruiz, D. Cabello, and J. Mira, "Algorithmic sequential decision-making in the frequency domain for life threatening ventricular arrhythmias and imitative artefacts: A diagnostic system," *J. Biomed. Eng.*, vol. 11, pp. 320–328, 1989.

[6] J. A. Kastor, *Arrhythmias*, 2nd ed. London, U.K.: W.B. Saunders, 1994.

[7] L. Senhadji, G. Carrault, J. J. Bellanger, and G. Passariello, "Comparing wavelet transforms for recognizing cardiac patterns," *IEEE Eng. Med. Biol. Mag.*, vol. 14, pp. 167–173, Mar.–Apr. 1995.

[8] T. H. Yeap, F. Johnson, and M. Rachniowski, "ECG beat classification by a neural network," in *Proc. Annu. Int. Conf. IEEE Engineering Medicine and Biology Soc.*, 1990, pp. 1457–1458.

[9] Y. H. Hu, W. J. Tompkins, J. L. Urrusti, and V. X. Afonso, "Applications of artificial neural networks for ECG signal detection and classification," *J. Electrocardiol.*, vol. 26, pp. 66–73, 1993.

[10] S. Osowski and T. L. Linh, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 1265–1271, Nov. 2001.

[11] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ECG beat classifier using a mixture of experts approach," *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 891–900, Sept. 1997.

[12] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sornmo, "Clustering ECG complexes using hermite functions and self-organizing maps," *IEEE Trans. Biomed. Eng.*, vol. 47, pp. 838–848, July 2000.

[13] *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*, 1998. Association for the Advancement of Medical Instrumentation.

[14] *Recommended Practice for Testing and Reporting Performance Results of Ventricular Arrhythmia Detection Algorithms*, 1987. Association for the Advancement of Medical Instrumentation.

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

[16] P. Laguna, R. Jané, and P. Caminal, "Automatic detection of wave boundaries in multilead ECG signals: Validation with the CSE database," *Comput. Biomed. Res.*, vol. 27, no. 1, pp. 45–60, 1994.

[17] R. Jané, A. Blasi, J. García, and P. Laguna, "Evaluation of an automatic threshold based detector of waveform limits in Holter ECG with the QT database," *Comput. Cardiol.*, vol. 24, pp. 295–298, 1997.

[18] B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge, U.K.: Cambridge Univ. Press, 1996.

[19] M. Heckmann, F. Berthommier, and K. Kroschel, "Noise adaptive stream weighting in audio-visual speech recognition," *Eurosip J. Appl. Signal Processing*, no. 11, pp. 1260–1273, 2002.

[20] R. Kohavi, "A study of cross validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artificial Intelligence*, 1995, pp. 1137–1143.

[21] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York: Oxford Univ. Press, 1995.

[22] J. Millet, M. A. Perez, G. Joseph, A. Mocholi, and J. Chorro, "Previous identification of QRS onset and offset is not essential for classifying QRS complexes in a single lead," *Comput. Cardiol.*, vol. 24, pp. 299–302, 1997.

[23] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen, and S. Luo, "ECG beat detection using filter banks," *IEEE Trans. Biomed. Eng.*, vol. 46, pp. 556–564, Feb. 1999.

[24] S. Kadambe, R. Murray, and G. F. Boudreaux-Bartels, "Wavelet transformed-based QRS complex detector," *IEEE Trans. Biomed. Eng.*, vol. 46, pp. 838–848, July 1999.

Maria O'Dwyer received the B.E. and M.Eng.Sc. degrees in electronic engineering from the National University of Ireland, Dublin, in 1999 and 2001, respectively.

She is currently an IC Design Engineer with Silicon & Software Systems, Cork, Ireland. Her research interests include biomedical engineering, signal processing, and the design of integrated circuits and systems.

Philip de Chazal (M'94) received the B.E. degree in electronic engineering and the M.Biomed.E. and Ph.D. degrees in biomedical engineering from the University of New South Wales, Sydney, Australia, in 1989, 1995, and 1999, respectively.

He was a Research Scientist with CSIRO, Sydney, from 1990 to 1992, a Biomedical Engineer with Med-Care Systems, Sydney, from 1993 to 1999, and a Research Fellow with the University College Dublin, Dublin, Ireland, from 1999 to 2003. He is cofounder and currently the Chief Technical Officer of BiancaMed, Ireland, a company providing intelligent computer-based analysis of medical signals. His research interests include signal processing and pattern recognition for biomedical applications and image processing for multimodal applications.

r de Chazal is a member of the IEEE Engineering in Medicine and Biology Society. He is a reviewer for the IEEE TRANSACTIONS ON MULTIMEDIA and IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING.

Richard B. Reilly (M'92–SM'04) received the B.E., M.Eng.Sc., and Ph.D. degrees in 1987, 1989, and 1992, all in Electronic Engineering, from the National University of Ireland.

In 1988, he joined Space Technology Ireland and the Dept. de Recherche Spatiale (CNRS group) in Paris, developing DSP-based on-board experimentation for the NASA satellites. In 1990, he joined the National Rehabilitation Hospital and in 1992 became a Postdoctoral Research Fellow at University College, Dublin, focusing on signal processing for speech and gesture recognition. Since 1996, he has been on the academic staff in the Department of Electronic and Electrical Engineering at University College, Dublin. He is currently Senior Lecturer and researches into neurological signal processing and multimodal signal processing.

Dr Reilly was the 1999/2001 Silvanus P. Thompson International Lecturer for the Institution of Electrical Engineers (IEE). In 2004, he was awarded a U.S. Fulbright Award for research collaboration with the Nathan Kline Institute for Psychiatric Research, Orangeburg, NY. He is a member of the IEEE Engineering in Medicine and Biology Society and Signal Processing Society. He is a committee member of the IEEE United Kingdom and Republic of Ireland Section. He is an Associate Editor for IEEE TRANSACTIONS ON MULTIMEDIA and also a reviewer for IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING and IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS.