

A Patient-Adaptable ECG Beat Classifier Using a Mixture of Experts Approach

Yu Hen Hu,* *Senior Member, IEEE*, Surekha Palreddy, and Willis J. Tompkins, *Fellow, IEEE*

Abstract—We present a “mixture-of-experts” (MOE) approach to develop customized electrocardiogram (ECG) beat classifier in an effort to further improve the performance of ECG processing and to offer individualized health care. A small customized classifier is developed based on brief, patient-specific ECG data. It is then combined with a global classifier, which is tuned to a large ECG database of many patients, to form a MOE classifier structure. Tested with MIT/BIH arrhythmia database, we observe significant performance enhancement using this approach.

Index Terms— ECG beat classification, MIT/BIH database, mixture of experts, neural network, patient adaptation.

I. INTRODUCTION

COMPUTERIZED electrocardiography is now a well-established practice, after several years of significant progress. Many algorithms have been proposed over years for electrocardiogram (ECG) beat detection and classification. In a clinical setting, such as an intensive care unit, it is essential for automated systems to accurately detect and classify electrocardiographic signals on a real-time basis. Since several arrhythmia are potentially dangerous and life threatening, if not detected within a few seconds to a few minutes of its onset, automated electrocardiographic monitoring assumes a challenging role. Several algorithms have been proposed in the literature for detection and classification of ECG beats and reported results, that leave room for improvement. They include signal processing techniques; such as frequency analysis, template matching, and other parameter extraction methods. Artificial neural networks were also employed to exploit their natural ability in pattern-recognition tasks for successful classification of ECG beat [2], [3], [6]–[8], [23]–[25], [28]–[31].

One major problem faced by today’s automatic ECG analysis machine is the wild variations in the morphologies of ECG waveforms of different patients and patient groups. An ECG beat classifier which performs well for a given training database often fails miserably when presented with a different patient’s ECG waveform. Such an inconsistency in performance is a major hurdle preventing highly reliable, fully automated ECG processing systems to be widely used clinically.

Manuscript received September 13, 1995; revised May 5, 1997. Asterisk indicates corresponding author.

*Y. H. Hu is with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706 USA (e-mail: hu@engr.wisc.edu).

S. Palreddy and W. J. Tompkins are with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706 USA.

Publisher Item Identifier S 0018-9294(97)06116-8.

One obvious approach to alleviate this problem is to use as much training data as possible to develop the ECG classifier. This is the approach taken by all the vendors of ECG processing devices: A large in-house ECG database is developed and maintained to test each ECG processing algorithm to be incorporated into the product. However, such an approach suffers several pitfalls.

- 1) No matter how large this database may be, it is not possible to cover every ECG waveform of all potential patients. Hence, its performance is inherently limited.
- 2) The complexity of the classifier grows as the size of the training database grows. When a classifier is designed to correctly classify ECG from millions of patients (if it ever becomes possible), it has to take numerous exceptions into account. The result is a complicated classifier which is costly to develop, maintain, and update.
- 3) It is practically impossible to make the classifier learn to correct errors during normal clinical use. Thus, it may be rendered useless if it fails to recognize a specific type of ECG beats which occurs frequently in certain patient’s ECG records.

The answer, we believe, is to allow the classifier to be “patient-adaptable.” That is, to let the classification algorithm adaptable to the special characteristics of each patient’s ECG records. For example, we may include the training algorithm and the database used to develop the classifier to be delivered to the users, so that the classification algorithm can be fine-tuned to each patient. Unfortunately, this is impractical for several reasons.

- While it is possible to turn over training algorithms and databases to the users in an academic environment, it is unlikely that any commercial ECG machine vendor is willing to risk revealing their proprietary information to their competitors. Moreover, in-house database often contains millions of ECG records which could be costly to distribute.
- Users often do not want to be bothered by implementation details of an ECG algorithm. Thus, few users will be able to take advantage of this patient-adaptation feature even if it is available.
- Even if a user is willing to perform the patient customization, he or she still have to provide sufficient number of patient-specific training data in order to perform patient-adaptation. Manually editing ECG record is a time consuming, labor intensive task. Hence, the size of

patient-specific training data must be tightly controlled.

In this study, we propose a novel approach to patient-adaptation while avoiding these difficulties: 1) We do not require the factory-trained ECG classifier to provide training algorithms or training databases. Instead, all we need is that this classifier gives both its classification results, as well as an estimate of posterior probability of the feature vector as is drawn from each particular class. Hence, no company proprietary information is needed. 2) A patient-specific classifier will be developed using an automated procedure, without human supervision. 3) Only a brief manually edited patient ECG record (2–5 min) is needed to achieve significant performance improvement.

This proposed approach is based on three popular artificial neural network (ANN)-related algorithms, namely, the self-organizing maps (SOM), learning vector quantization (LVQ) algorithms, along with the mixture-of-experts (MOE) method. SOM and LVQ together are used to train the patient-specific classifier, and MOE is a paradigm which facilitates the combination of the two classifiers (original and patient-specific) to realize patient-adaptation. In MOE, the two classifiers are modeled as two experts on ECG beat classification. The original classifier, called the *Global expert* (GE) in this work, knows how to classify ECG beats for many other patients whose ECG records are part of the in-house, large ECG database. The patient-specific classifier, called the *local expert* (LE) in this work, is trained specifically with the ECG record of the patient. A gating function, based on the feature vector presented, dynamically weights the classification results of the GE's and the LE's to reach a combined decision. The process is analogous to two human experts arriving at a consensus based on their own expertise.

Section II reports the results of literature survey and Section III discusses data acquisition with preprocessing. Section IV discusses the proposed algorithms and the development of experts. Section V reports the results of the classifier on the database records and discusses the results. Section VI is a summary of the findings of this paper.

II. PRELIMINARIES

A. ECG Beat Classification Techniques

Automated ECG beat classification was traditionally performed using a decision-tree-like approach, based on various features extracted from an ECG beat [1], [4], [5], [13], [20], [22]. The features used include the width and height of QRS complex, RR interval, QRS complex area, etc. One of the difficulties is that these features are susceptible to variations of ECG beat morphology and temporal characteristics. As such, the classification rate reported in these earlier efforts are rather moderate.

Artificial neural networks (ANN's) have been widely accepted for pattern recognition tasks. Their abilities to learn from examples and extract the statistical properties of the examples presented during the training sessions, make them an ideal choice for an automated process that imitates human logic. Several efforts have been made to apply ANN's for

the purpose of ECG beat detection and classification. Previous reported efforts include [2], [3], [6]–[8], [23]–[25], and [28]–[31].

Hu *et al.* [7] reported the development of an adaptive multilayer perceptron (MLP) for classification of ECG beats. They have achieved an average recognition accuracy of 90% in classifying the beats into two groups; normal and abnormal. In an attempt to classify the beats into 13 groups according to the MIT Database annotations, they have reported an average recognition accuracy rate of 65%. An hierarchical system of the MLP networks which first classify the beat into normal or abnormal, and then classify it into the specific beat type, is developed, which improved the recognition accuracy to 84.5%.

B. Self-Organization Map (SOM) and Learning Vector Quantization (LVQ)

SOM and LVQ are both clustering based algorithms proposed by Kohonen [14], [15]. SOM is an unsupervised on-line clustering technique. In SOM, each cluster center (prototype or code word) is represented by the weights of a neuron which is assigned to a coordinate in the feature map. The SOM training algorithm forces adjacent neurons in the feature map to respond to similar feature (input) vectors. In a way, this feature map is analogous to the spatial organization of sensory processing areas in the brain. Let $m_i(t)$ be denoted as the weights (code word) or the i th neuron in SOM during the time instant t , the weights of SOM then are updated according to the following simple formula:

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)] \quad (1)$$

$h_{ci}(t)$ is the so-called *neighborhood kernel*, which determine the size of neighborhood of the i th neuron within which all neighboring neurons will be updated in response to the present feature vector $x(t)$. Initially, the neighborhood is large. The size reduces as clustering converges, until no neighboring neurons will get updated.

LVQ is a supervised, clustering-based classification technique which classifies a feature vector $x(t)$ according to the label of the cluster prototype (code word) into which $x(t)$ is clustered. Classification error occurs when the feature vectors within the same cluster (hence, assigned to the same class label) are actually drawn from different classes. To minimize classification error, the LVQ algorithm fine tunes the clustering boundary between clusters of different class labels by modifying the position of the clustering center (prototype or code word). This method is called “learning vector quantization” because this clustering based classification method is similar to the “vector quantization” method used for signal compression in the areas of communication and signal processing.

According to Kohonen, there are three different LVQ algorithms, called LVQ1, LVQ2, and LVQ3 developed at subsequent stages to handle classification problems with different natures. In this study, the optimized learning-rate LVQ1 and LVQ3 algorithms were used for the training and fine-tuning of the code book respectively. In LVQ1, for a given input vector $x(t)$, a code word y_c is found such that

$$c = \arg \min_i \{\|x - y_i\|\}. \quad (2)$$

The *code word* is then updated as follows:

$$y_c(t+1) = [(1 - s(t)\alpha_c(t))y_c(t) + s(t)\alpha_c(t)x(t)] \quad (3)$$

where $s(t) = +1$ if the classification is correct [i.e., $x(t)$ and y_c have the same class label] and $s(t) = -1$, otherwise. $\alpha_c(t)$ is a time-varying learning rate. Other code words in the code book remain unchanged. LVQ3 differs from LVQ1 in how the code words are updated: Assuming that $x(t)$ falls within a *window* between two adjacent clusters with corresponding code words y_i and y_j . Suppose that x and y_j belong to the same class, and x and y_i belong to different classes, then both these code words will be updated in LVQ3:

$$y_i(t+1) = y_i(t) - \alpha(t)[x(t) - y_i(t)] \quad (4a)$$

$$y_j(t+1) = y_j(t) + \alpha(t)[x(t) - y_j(t)]. \quad (4b)$$

On the other hand, if both y_i and y_j belong to the same class as $x(t)$, and $x(t)$ fall in a window centered at the cluster boundary of these two classes, then

$$y_k(t+1) = y_k(t) + \varepsilon\alpha(t)[x(t) - y_k(t)], \\ k \in \{i, j\}, 0.1 < \varepsilon < 0.5. \quad (5)$$

The optimal value of ε depends on the size of the window, being smaller for narrower windows. This algorithm is self-stabilizing, and optimal placement of the y_i does not change in continual training.

Software packages of both SOM and LVQ are available in the public domain,¹ and the application of these packages to the ECG beat classification problem is straight forward. The adaptation parameters in these packages (SOM_PAK and LVQ_PAK) were carefully fine tuned while developing the classifiers. As such, the development of the code book and eventually decision boundary can be made completely transparent to the user. Moreover, performance obtained using these package is very competitive compared to other approaches. In this research work, we first apply SOM to a set of training feature vectors. The resulting code book (prototypes) then will be submitted to the LVQ_PAK to facilitate fine tuning and classification.

III. MIXTURE OF EXPERTS (MOE)

This user adaptation problem bears certain resemblance to the incremental learning problem in that new data are to be incorporated to improve existing classifier's performance. However, the black-box model of the existing classifier prevents us from directly modifying the classifier structure as incremental learning algorithms do. Instead, we propose a different method called the MOE, to circumvent this problem.

The MOE approach was proposed by Jacobs *et al.* [9]–[12], [16], [26], [27]. The basic notion is that linear combinations of several statistical estimates can perform better than any individual estimate. This strategy is not new. It is a well known fact that a panel of experts often arrive at a better diagnosis than any single expert, because each expert is able to contribute from his/her own expertise.

The basic idea is to leave the existing black-box classifier intact. Instead, we use the given small, user-specific training data set to develop a LE classifier. Then we invoke a modified MOE approach to combine these two classifiers, hoping to achieve better performance.

To apply the MOE approach to solve the customization problem, we employ two experts: a GE and a LE. The GE represents the ECG beat classifier developed in factory. Thus, it is trained to classify all types of ECG beats present in the in-house ECG database. The LE represents a specialized ECG beat classifier, trained on a small segment of annotated ECG beats taken from the specific patient. As such, the GE and the LE are endowed with complementary knowledge bases, and can work together to reach a better decision than any one can reach individually.

The expert network is a combination of the GE and LE classifiers. Let $y_1(x)$ and $y_2(x)$ be the output (row) vectors of the two respective GE and LE classifiers. Each element of each vector indicates the degree of proximity of an unknown ECG beat to a predefined ECG beat class (category). In the MOE method, the combined i th output vector of both the experts is given by

$$h_{ic}(x) = g_{1c}(x)y_{1c}(x) + g_{2c}(x)y_{2c}(x) \quad (6)$$

where x is the input feature vector, g_{jc} , $j = 1, 2$ are the weighting vectors for each expert from a *gating network* and are defined by

$$g_{jc}(x) = \frac{\exp(v_{jc}^t x)}{\sum_{j=1}^2 \exp(v_{jc}^t x)} \quad (7)$$

where v_{ij} 's are the weight vectors of the gating network. Note that $\sum_{j=1}^2 g_{jc}(x) = 1$.

Theorem 1: Define $R[y_i(x)] = \{x | x \in X, \text{ and } y_i(x) = z(x)\}$, $i = 1, 2$ to be the subregion in the feature space where the classifier $y_i(x)$ makes correct classification of x and let $R[y(x)]$ be defined the same way. Assume $y_i(x) \in \{0, 1\}$ and $z(x) \in \{0, 1\}$, then

$$R[y(x)] \subseteq R[y_1(x)] \cup R[y_2(x)]. \quad (8)$$

Proof: We need only to prove that if both $y_1(x)$ and $y_2(x)$ misclassify a given feature vector x , then $y(x)$ cannot give correct classification on x . Since the correct classification output $z(x)$, the combined output $y(x)$, and individual classifier output $y_1(x)$ and $y_2(x)$ are all binary vectors of the same dimension, if both classifiers misclassify a given feature vector x which belongs to class c , we must have, for the c th elements of these binary vectors

$$z_c(x) \oplus y_{1c}(x) = z_c(x) \oplus y_{2c}(x) \\ = 0$$

where “ \oplus ” is the “exclusive-OR” operator in Boolean algebra. Since from (7), $g_{1c}(x) + g_{2c}(x) = 1$, we conclude $y_c(x) = 0$, if $y_{1c}(x) = y_{2c}(x) = 0$, and $y_c(x) = 1$, if $y_{1c}(x) = y_{2c}(x) = 1$. Hence, $z_c(x) \oplus y_c(x) = 0$. In other words, $y(x)$ must also misclassify the same feature vector x regardless the choice

¹University of Helsinki, Finland, URL: [ftp://cochlea.hut.fi/pub/](http://cochlea.hut.fi/pub/)

of $g_{1c}(x)$ and $g_{2c}(x)$. This is to say, if $x \in R[y_1(x)]$, and if $x \in R[y_2(x)]$, then $x \in R[y(x)]$. ■

The implication of Theorem 1 is that the maximum performance enhancement of a MOE approach occurs when $R[y_1(x)] \cap R[y_2(x)] = \phi$ (empty set). An example is to designate each classifier to be responsible for classifying a particular class. The assumption that $y_i(x) \in \{0, 1\}$ is essential in this theorem. If $y_i(x) \in [0, 1]$ (interval between zero and one), it is possible to find a counter example. Let $y_1(x) = [0.1 \ 0.5 \ 0.4]$, $y_2(x) = [0.5 \ 0.1 \ 0.4]$, and $z(x) = [0 \ 0 \ 1]$. Then $y(x) = [0.1g_1 + 0.5g_2 \ 0.5g_1 + 0.1g_2 \ 0.4]$. If $g_1 = g_2 = 0.5$, then $y(x) = [0.3 \ 0.3 \ 0.4]$ which yields correct classification.

On the other hand, whether $y_i(x)$ takes binary values or not, if both classifiers make correct classification, so will the combined classifier.

Theorem 2: With the same definitions as in Theorem 1, and $y_i(x) \in [0, 1]$

$$R[y_1(x)] \cap R[y_2(x)] \subseteq R[y(x)]. \quad (9)$$

Proof: Assume x [class c^* , and $y_{1c^*}(x) = \max_c y_{1c}(x)$, $y_{2c^*}(x) = \max_c y_{2c}(x)$]. Then

$$\begin{aligned} h_{c^*}(x) &= g_1(x)y_{1c^*}(x) + g_2(x)y_{2c^*}(x) \\ &= \max_c [g_1(x)y_{1c}(x) + g_2(x)y_{2c}(x)]. \end{aligned} \quad (10)$$

Thus, the output $y_{c^*}(x)$ is correctly classified. ■

From Theorem 2, it is clear that if both classifiers #1 and #2 correctly classify a pattern x , then the combined classifier will also correctly classify the same pattern. Hence, this pattern can be excluded from the user-adaptation training set as it will not affect the result.

Adaptation Algorithm: Based on the result indicated in Theorems 1 and 2, the design objective of the MOE network in (3) is to devise a training algorithm to estimate the parameter vectors $\{v_{ic}; i = 1, 2\}$. Given that $y_1(x)$ and $y_2(x)$ are fixed classifiers, this problem can be solved by a gradient procedure as follows: Let us assume $\{x_k; 1 \leq k \leq K', x_k[X]\}$ be a set of training data used for searching the optimal gating functions $g_{1c}(x)$ and $g_{2c}(x)$, such that the square error at the output $E_c = (1/2K') \sum_{k=1}^{K'} \|z_c(k) - y(v_{1c}, v_{2c}, x_k)\|^2$ is minimized. A gradient search algorithm can be devised as follows:

$$v_{ic}(t+1) = v_{ic}(t) - \mu \nabla_{v_i} E_c. \quad (11)$$

The initial values of v_{1c} and v_{2c} are set to be the centroids of the regions $R[y_1(x)]$ and $R[y_2(x)]$, respectively, for x in the user-specific training data set. The gradient of E_c with respect to v_{ic} can be calculated as

$$\begin{aligned} \nabla_{v_i} E_c &= \frac{1}{2K'} \sum_{k=1}^{K'} [y_c(v_{1c}, v_{2c}, x_k) - z_{kc}] \\ &\quad \bullet \nabla_{v_i} y_c(v_{1c}, v_{2c}, x_k) \\ &= \frac{1}{2K'} \sum_{k=1}^{K'} \left\{ [y_c(x_k) - z_{kc}] \bullet f'[h_c(x_k)] \right. \\ &\quad \left. \bullet \left[\sum_{j=1}^2 y_{jc}(x_k) \bullet \nabla_{v_i} g_{jc}(v_{1c}, v_{2c}, x_k) \right] \right\} \end{aligned} \quad (12)$$

where $h_c(x) = g_{1c}(x) \bullet y_{1c}(x) + g_{2c}(x) \bullet y_{2c}(x)$. In (13), we assumed the transfer function f is a differentiable threshold function, and is applied to the vector, element by element. Finally, with (13), we have as shown in (14) at the bottom of the page. Hence, for $i, j \in \{1, 2\}$, $i \neq j$, we have

$$\begin{aligned} \nabla_{v_i} E_c &= \frac{1}{2K'} \sum_{k=1}^{K'} \sum_{c=1}^C x_k \{ [y_c(x_k) - z_{kc}] \bullet f'[h_c(x_k)] \} \\ &\quad \bullet [y_{ic}(x_k) - y_{jc}(x_k)] \bullet g_{1c}(x_k) \bullet g_{2c}(x_k). \end{aligned} \quad (15)$$

Note that in above derivation, the error E is accumulated over the entire epoch (K' feature vectors). The summation over k may be removed if we use on-line update of v_i 's for each sample. This yields the following expression for $i, j \in \{1, 2\}$, $i \neq j$:

$$\begin{aligned} \nabla_{v_i} E_c &= x_k \bullet [y(x_k) - z_k]^t \bullet \text{diag} \{ f'[h(x_k)] \} \\ &\quad \bullet [y_{ic}(x_k) - y_{jc}(x_k)] \bullet g_{1c}(x_k) \bullet g_{2c}(x_k). \end{aligned} \quad (16)$$

Clearly, we have $\nabla_{v_1} E_c = -\nabla_{v_2} E_c$. This is not surprising with two parameter vectors arriving at a decision hyperplane $(v_{1c} - v_{2c})^t x = 0$.

Until now, we have assumed that the user-specific ECG beat classifier $y_{2c}(x)$ is readily available. However, in reality it needs to be trained with the user-specific training data set. Also, the combined classifier $y_c(x)$ needs to be trained by the same data set in order to determine the gating network parameters. Therefore, if $y_{2c}(x)$ is trained to 100% accuracy

$$\begin{aligned} \nabla_{v_i} g_{jc}(v_{1c}, v_{2c}, x_k) &= \frac{\nabla_{v_i} \exp(v_{jc}^t x_k)}{\sum_{m=1}^2 \exp(v_{mc}^t x_k)} \\ &= \frac{\nabla_{v_i} [\exp(v_{jc}^t x_k)] \bullet \sum_{m=1}^2 \exp(v_{mc}^t x_k) - \exp(v_{jc}^t x_k) \bullet \nabla_{v_i} \exp(v_{ic}^t x_k)}{\left[\sum_{m=1}^2 \exp(v_{mc}^t x_k) \right]^2} \\ &= (-1)^{i-j} x_k \bullet g_{1c}(x_k) \bullet g_{2c}(x_k) \end{aligned} \quad (14)$$

on the user-specific data set, then the gating network of choice may be $g_{2c}(x) \neq 1$ and $g_{1c}(x) \neq 0$. In light of the results of Theorems 1 and 2, we devised the following strategy to alleviate this problem: First, we construct the user-specific training data set to contain only those feature vectors which the original classifier misclassified. We further partition this training data set into two subsets: one for the training of the user-specific classifier $y_{2c}(x)$, and the other for estimating $g_{1c}(x)$ and $g_{2c}(x)$.

IV. EXPERIMENT

The purpose of this experiment is to demonstrate the usefulness of the proposed user-adaptation procedure. In particular, we will show that an ECG beat classifier trained on general patient records does not perform well when presented with patient records which contain rare beat types. Moreover, we show that the performance of the MOE classifier is able to gain significant performance enhancement with a small amount of annotated patient specific training data.

A. Data Preparation

In this study, we concentrate on the classification of ventricular ectopic beats (VEB's). The 48 records (tapes) from MIT/BIH ECG arrhythmia database [17], [19] are used for the development and evaluation of the classifier. The availability of annotated MIT/BIH database has enabled the evaluation of performance of the proposed beat classification algorithm. The American Association of Medical Instrumentation (AAMI)-recommended practice [18] has provided a protocol for a reproducible test with realistic clinical requirements, emphasizing tape-by-tape presentation of results that estimate an algorithm's ability to detect events of clinical significance.

Accompanying each tape in the MIT/BIH database is an annotation file in which each ECG beat has been identified by expert cardiologist annotators. These labels are referred to as "truth" annotations and are used in training (developing) the classifiers and also to evaluate the performance of the classifiers (experts) in testing phase. According to the AAMI-recommended practice, records containing the paced beats (four records) can be excluded from the reporting requirements. Since this study is to evaluate the performance of a classifier that can identify a premature ventricular contraction (PVC), certain records in the database with no PVC's (11 records) were excluded from the study, leaving 33 records of interest. These excluded records are listed in Table I. Data from channel 1, down-sampled to 180 samples/s were used in this study. The selected files consist of 13 records (numbered from 100–124, inclusive, with some numbers missing) and 20 records (numbered from 200–234, inclusive, with some numbers missing). The first group is intended to serve as a representative sample of a variety of waveforms and artifacts which an arrhythmia detector might encounter in routine clinical use. Records in the second group include complex ventricular, junctional, and supraventricular arrhythmias and conduction abnormalities. Several of these records are expected to present significant difficulty to arrhythmia detectors because of the features of the rhythm, QRS morphology

TABLE I
RECORDS OF MIT/BIH DATABASE THAT WERE EXCLUDED FROM THE STUDY

Excluded files from MIT/BIH Arrhythmia Database	
Paced beats	102, 104, 107, 217
No PVCs	101, 103, 112, 113, 115, 117, 122, 212, 220, 222, 232

TABLE II
FOUR CATEGORIES OF INTEREST INTO WHICH THE
ECG BEATS OF THIS STUDY ARE CLASSIFIED

Beat Labels (AAMI recommended practice)	
N	any beat that does not fall into the V, F or Q categories (normal beats, supraventricular ectopic beats, and others)
V	a ventricular ectopic beat (VEB): a ventricular premature beat, R-on-T ventricular premature beat, or ventricular escape beat.
F	fusion of ventricular and normal beat
Q	unclassifiable beats

variation, and signal quality. These records were reported to have gained considerable notoriety among database users [18].

In this experiment, we use the first group of files as the training data to develop a GE classifier which is able to classify typical ECG beats. The second group of 20 records is used to simulate the ECG records of 20 patients, which are to be classified by the GE classifier. Since these records consist of less-frequently seen beats, it is expected that the GE classifier will not perform well. If this GE classifier were a commercial device, it will be deemed not-applicable (due to low performance) to many of these 20 test records. However, with the MOE approach, we will adapt this GE classifier with a LE classifier to gain significant performance enhancement at low cost.

The beats in the MIT/BIH database are of several different types. In this study, we are interested in identifying four different categories, as indicated in Table II. Each of the four categories included beats of several types as shown in Table III. The AAMI convention was used to combine the beats into four classes of interest.

B. Training and Testing Procedure

In this study, a GE classifier was developed with SOM and LVQ algorithms using the data from the records of the first group (100–124). Before testing the records, a LE classifier was developed for each of the records in the second group using the first 2.5 min of data. The rest of the record is then tested using the mixture of global and LE's as explained before. Since each record in the MIT/BIH database is of length 30 min, the 2.5 min segment account for 1/12th of total available patient specific data and contains approximately 150 ECG beats. In practice, the attending cardiologist or any expert in ECG beat annotation will have to annotate a brief segment of patient-specific ECG in order to take advantage of the MOE approach. We believe that this is a reasonably small cost compared to the potential gain in performance enhancement. In future, we will explore a more effective method to further reduce the amount of required annotated patient-specific data.

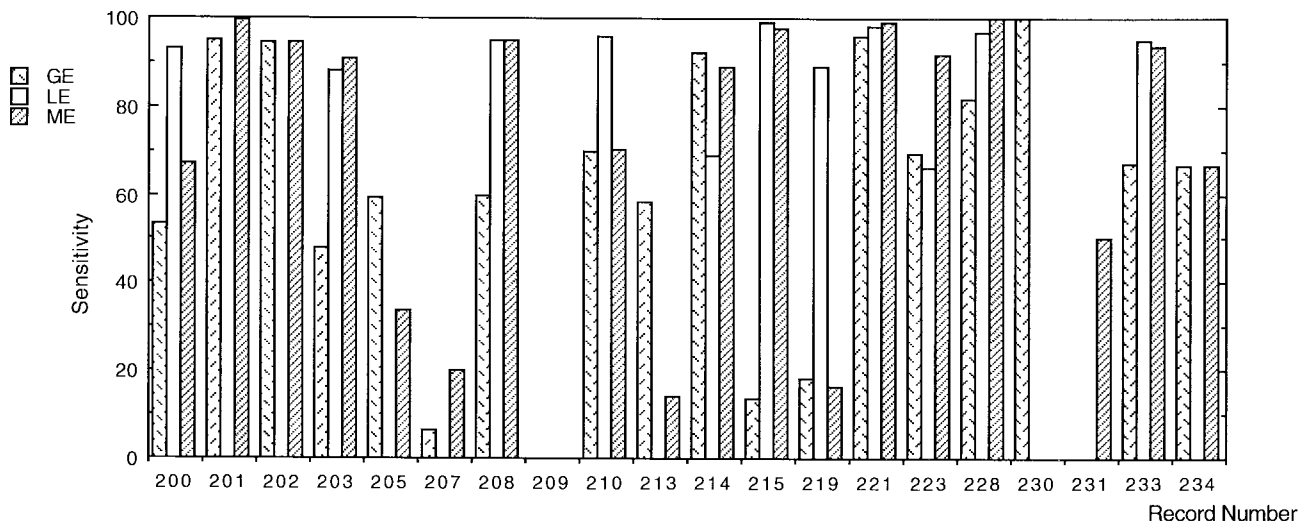


Fig. 1. Record by record comparison of sensitivity of three methods: GE, LE, and MOE.

TABLE III
BEATS OF MIT/BIH DATABASE CLUBBED INTO FOUR
CATEGORIES BASED ON AAMI-RECOMMENDED PRACTICE

Descriptor	MIT label	Test label
Normal beat	1	1
Left bundle branch block beat	2	1
Right bundle branch block beat	3	1
Bundle branch block beat (unspecified)	25	1
Nodal(junctional) premature beat	7	1
Atrial premature beat	8	1
Supraventricular premature beat(atrrial or nodal)	9	1
Aberrated atrial premature beat	4	1
Nodal(junctional) escape beat	11	1
Atrial escape beat	34	1
Ventricular premature beat	5	2
R-on-T ventricular premature beat	41	2
Ventricular escape beat	10	2
Fusion of ventricular and normal beat	6	3
Unclassifiable beat	13	4
Others	Other	4

The GE and LE classifiers were developed using the clustering algorithm implemented in SOM_PAK, and the fine-tuning algorithm implemented in LVQ_PAK. The MOE algorithm was implemented in MATLAB. The SOM's developed using all the data available in the training files had many of the nodes tuned to the normal beats providing a greater detail to the normal beats than to the abnormal ones. This lead to a successful recognition of most normal ECG beats and suboptimal recognition accuracies of abnormal beats. This bias was introduced due to the amount of data that falls into the category of normals was about ten times more than the data for other rhythms. Since the detail of the map is dependent upon the amount of data falling into that category, it is essential to provide equal amounts of data for each class. Therefore, normal beats were clustered (using SOM) and the prototype vectors developed were added to the dataset of beats from

other categories forming sensitized data. The sensitized data was then used for developing the GE.

1) *Preprocessing*: The objective of this paper is to classify the QRS beats into one of the four different categories. The QRS beats are obtained as 29 point templates. The position of annotation labels is used to identify the peak of the QRS waveform and 14 points on either side of the peak were picked up to form the template.

The 29-dimensional template is then reduced to a nine-dimensional vector using principal-component analysis, also known as the Karhunen–Loeve transformation. It is designed such that the data set may be represented by a reduced number of “effective” features and yet retain most of the intrinsic information content of the data. We may reduce the number of features needed for effective data representation by discarding those linear combinations that have small variances and retain only those terms that have large variances. The data vector \mathbf{x} is then approximated with the m largest eigenvalues of the correlation matrix \mathbf{R} , introducing an *approximating error*.

Temporal parameters such as the instantaneous RR interval, average RR interval, and the width of the QRS complex were also extracted. The instantaneous RR interval is calculated as the difference between the QRS peak of the present beat and the previous beat. The average RR interval is calculated as the average RR interval over the previous ten beats. The width of the QRS complex is calculated according to the Pan–Hamilton algorithm [21].

The information of each beat is stored as a 13-element vector, with the first nine elements representing the transformed morphological template, and the next three elements representing the temporal parameters. This leads to a 12-dimensional feature vector. The thirteenth element is the “label” of the beat from the annotation file, after suitable translation as described in Table III.

Several preprocessing steps were performed on the raw data to study their effects upon the performance of the classifiers. Specifically, subtracting the mean value from each template showed a remarkable improvement in the performance of the

TABLE IV
IDENTIFICATION OF TP, FP, TN, AND FN IN THIS STUDY.
N(n): NORMAL BEATS, V(v): PREMATURE VENTRICULAR
CONTRACTIONS, F(f): FUSION BEATS, Q(q): UNCLASSIFIABLE BEATS

TRUTH	Algorithm			
	n	v	f	q
N	TN	FP	TN	TN
V	FN	TP	FN	FN
F	TN	FP	TN	TN
Q	TN	FP	TN	TN

LVQ classifier. Even though the morphology of the beats belonging to the same category is similar, a baseline change can represent the data differently in the signal space. To avoid this problem, the mean value of the templates is subtracted. Templates were also scaled linearly between +1 and -1 before the expert classifiers are developed. Temporal information of the beats such as instantaneous RR interval, average RR interval over the past ten beats, and the width of QRS complex showed improvement in the classification of PVC beats.

2) *Training of the Global and Local Expert Classifiers:* For the GE classifier, the sensitized data from 13 MIT/BIH database tapes (#100–124) is used to develop a SOM of size 15×10 neurons. This is accomplished using SOM.PAK. The weights of each neuron form a code word in the code book of 150 code words. Each code word, or equivalently the associated neuron, then is labeled using annotated data. The label of the code word is assigned based on the label of annotated feature vectors assigned to that cluster.

Another classifier of 150 code words, based on LVQ algorithm, is developed using LVQ.PAK. The classification performance of the classifier developed using LVQ is superior for classes 1 and 3, whereas, the performance of the classifier developed using SOM is superior for classes 2 and 4. Therefore, the code books generated by LVQ and SOM were edited manually to select and combine those code words which yield superior performance. The resulting code book constitutes the GE classifier.

To enable the “soft combination” of the classifier output, it is desired that the outputs of each classifier be an estimate of the a posteriori probability of the feature vector belonging to that class. To facilitate this requirement, we assume that the posterior probability is a mixture of Gaussian distribution with each code word in the class being the mean of a Gaussian distribution with unity variance. This is a reasonable assumption since each code word is obtained using the SOM clustering algorithm based on the L_2 norm distance measure. Therefore, for large amount of samples, the posterior probability distribution of each class will converge to a Gaussian distribution asymptotically. For small samples such as those used for training a LE, a Gaussian distribution assumption seems to be an adequate approximation. Next the distance denoted by D_c ($c = 1, 2, \dots, 4$) between a feature vector $x(t)$ and the nearest code word of class c , is computed. The class c output of this GE classifier then is computed as $\exp(-D_c^2)$ which is proportional to the Gaussian density function $\exp[-\|x(t) - \mathbf{m}_c\|^2]$.

TABLE V
COMPARISON OF PERFORMANCE BETWEEN THE GE, LE, AND MOE CLASSIFIERS. ALL ENTRIES ARE IN PERCENT (%). FOR THOSE RECORDS WHERE FP = TP = 0, POSITIVE PREDICTIVITY IS ASSIGNED TO NaN (NOT A NUMBER) BECAUSE ITS DENOMINATOR IS ZERO

Record #	Classification Rate			Sensitivity			Specificity			Positive Predictivity		
	GE	LE	ME	GE	LE	ME	GE	LE	ME	GE	LE	ME
200	62.3	97.0	81.0	53.6	93.0	66.9	97.4	98.7	93.3	89.6	97.0	80.6
201	72.4	89.3	95.4	95.0	0.0	99.5	80.6	100	99.3	36.7	NaN	94.5
202	66.7	99.1	71.7	94.4	0.0	94.4	66.8	100	72.1	2.6	NaN	3.0
203	33.9	93.7	87.4	48.1	88.0	91.0	81.7	94.7	90.8	30.4	73.0	62.2
205	81.2	97.1	97.1	59.2	0.0	33.8	100	100	100	100	NaN	100
207	80.6	92.2	88.3	6.2	0.0	20.0	94.0	100	98.7	7.6	NaN	54.7
208	60.5	94.5	90.5	59.6	95.0	95.2	87.3	94.4	93.4	69.5	89.0	87.5
209	21.9	100	98.7	0	0.0	0	99.9	100	99.9	0	NaN	0
210	88.1	96.2	93.4	69.7	96.0	70.3	92.8	96.3	98.6	42.5	67.0	79.0
213	74.4	92.8	91.9	58.3	0.0	14.2	78.6	100	99.8	17.5	NaN	83.8
214	92.4	95.2	98.2	92.1	69.0	89.1	92.8	98.5	99.7	62.4	86.0	97.7
215	3.65	97.1	98.4	13.7	99.0	97.9	99.9	97.0	98.5	95.2	62.0	76.4
219	97.1	93.0	97.4	18.2	89.0	16.4	100	93.1	99.9	100	25.0	90.0
221	58.9	95.4	99.0	96.1	98.0	98.9	92.5	94.9	100	70.8	79.0	100
223	26.7	91.0	94.4	69.2	66.0	91.9	17.8	96.9	95.7	16.4	83.0	83.4
228	93.5	97.9	99.9	81.7	97.0	100	96.1	98.1	99.9	80.7	91.0	99.7
230	72.4	100	99.1	100	0.0	0	75.3	100	99.5	0.2	NaN	0
231	98.5	99.9	99.8	0	0.0	50.0	99.8	100	100	0	NaN	100
233	20.9	96.7	97.8	66.9	95.0	93.5	5.7	97.3	99.5	20.3	93.0	98.8
234	97.4	99.9	99.8	66.7	0.0	66.7	99.5	100	100	13.3	NaN	100
Avg	62.2	95.9	94.0	63.5	79.0	82.6	83.5	98.0	97.1	31.9	83.0	77.7

The LE classifier is developed in exactly the same manner as the global classifier, except that it uses only the first two and half minutes in the tape, and is constructed separately for each particular “patient tape” (tape #200–234) in the MIT/BIH database. We choose the first 2.5 min for training LE’s and the next 2.5 min of data to training the gating network of the MOE classifier. This practice is conformed to the AAMI-recommended procedure which allows to use of the first 5 min of data in each tape to fine tune the classifier. During testing with the combined MOE classifiers, only the last 25 min in each tape are used. Hence the testing data are never part of any training data through the entire process.

3) *Mixture of Experts (MOE) Classifier:* A gating network provides the scaling factors (g_{ic} ’s) for each class of both experts. The output of the gating network is a 2×4 matrix, with each row forming a scaling factor vector for each expert. The weights of the gating network are simply determined as the centroids of the regions as indicated by the code-book vectors of the corresponding expert.

The output of the classifier $h(x)$ is calculated as given by (6). Each input vector is classified into the class which has maximum output in the output vector $h(x)$. Through extensive experimentation, we further modified the computation of the gating network output so that $g_{2c}(x) = 1$ [i.e., $g_{1c}(x) = 0$], if $y_{2c}(x) > 0.9$ regardless of what was calculated from the gating network. This is intuitively convincing because it yields a decision for the LE when the LE classifier is certain about its diagnosis. We found that this modification improves the accuracy of the combined classifier and also improves the sensitivity.

TABLE VI
BEAT-BY-BEAT, RECORD-BY-RECORD TESTING RESULTS OF THE EXPERIMENT

Record #	Global Expert				local Expert				Mixture of Expert				Data in Record		
	TP	FP	FN	TN	TP	FP	FN	TN	TP	FP	FN	TN	# of Beats	Normal	PVC
200	408	48	353	1796	707	24	54	1820	509	124	252	1720	2605	1844	761
201	188	322	10	1338	0	0	198	1660	197	12	1	1648	1858	1660	198
202	17	671	1	1349	0	0	18	2020	17	564	1	1456	2038	2020	18
203	197	452	213	2020	360	132	50	2340	373	227	37	2245	2882	2472	410
205	42	0	29	2420	0	0	71	2420	24	0	47	2420	2491	2420	71
207	11	123	162	1919	0	0	173	2042	35	27	138	2015	2215	2042	173
208	551	241	374	1657	877	107	48	1791	881	125	44	1773	2823	1898	925
209	0	3	1	2856	0	0	1	2859	0	3	1	2856	2860	2859	1
210	129	167	56	2150	177	86	8	2231	130	32	55	2285	2502	2317	185
213	128	610	92	2241	0	0	220	2851	31	6	189	2845	3071	2851	220
214	226	136	19	1753	170	28	75	1861	218	6	27	1883	2134	1889	245
215	20	3	128	3021	147	92	1	2932	145	45	3	2979	3172	3024	148
219	10	0	45	2101	49	145	6	1956	9	2	46	2099	2156	2101	55
221	355	145	14	1788	360	98	9	1835	365	0	4	1933	2302	1933	369
223	324	1654	144	358	309	63	159	1949	430	87	38	1925	2480	2012	468
228	270	64	60	1584	319	31	11	1617	330	2	0	1646	1978	1648	330
230	1	569	0	1733	0	0	1	2302	0	12	1	2290	2303	2302	1
231	0	4	2	1871	0	0	2	1875	1	0	1	1875	1877	1875	2
233	522	2033	258	123	741	58	39	2098	729	11	51	2145	2936	2156	780
234	2	13	1	2561	0	0	3	2574	2	0	1	2574	2577	2574	3
Avg	3406	7243	1957	36654	4216	864	1147	43033	4430	1273	933	42624	49260	43897	5363

C. Results

The classifier was tested with the selected 20 records of the second group of the MIT database. The GE was left intact and is used as is for testing the 25 min of data from each 30-min testing record with first 5 min excluded as they are used to develop the LE and the “gating network.” The performance of the MOE classifier was compared to that of the GE and LE for each of the 20 records.

All detection statistics are founded on the mutually exclusive categories of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Since we are interested in estimating the performance of the classifiers based on the recognition of VEB’s (rhythm 2), the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) are defined appropriately as listed in Table IV.

Three statistics: sensitivity, specificity, and positive predictivity are used to compare the results. The respective definitions are as follows: *Sensitivity*: [$Se = TP/(TP + FN)$] is the fraction of real events that are correctly detected among all real events; *Specificity* [$Spec = TN/(TN + FP)$] is the fraction of nonevents that has been correctly rejected; and *Positive Predictivity*: [$PP = TP/(TP + FP)$] is the fraction of real events in all detected events. Another statistic *false positive rate* [$FPR = FP/(TN+FP)$] is the fraction of all nonevents that are not rejected. Since $FPR = 1 - Spec$, it is not listed here. Finally, the classification rate = $(TN + TP)/(TN + TP +$

$FN + FP)$. These three statistics, together with the percentage classification rates, are reported for each individual testing file as required by the AAMI-recommended practice [18]. The results are summarized in Table V (percentage) and Table VI (actual number of beats). A graph comparing the sensitivities of each record for the three methods are shown in Fig. 1.

D. Discussion

- 1) From Tables V and VI, we observe that the MOE approach is capable of significantly enhancing the performance of an ECG beat classifier over the global classifier. Moreover, we find that even with only 5 min of patient specific training data, the LE classifiers fare very well in all categories compared to both GE and ME classifiers. These observations confirmed our claim in this paper that patient-specific training data will significantly enhance the performance of a general purpose ECG classifier.
- 2) Comparing the LE and ME, we found that LE outperformed ME in terms of classification rate, mainly due to higher specificity (ability to correctly classify normal beats), but with lower sensitivity (ability to correctly classify PVC beats as PVC). Especially for those records where the first 5-min LE training data does not contain any PVC beats. Hence, although a LE classifier performs well, the availability of a global classifier does help to further enhance its performance.

- 3) In some cases, the improvement in classification rate is moderate; in others, significant improvements are observed. For example, in records 203, 209, 215, 223, and 233, the classification error rates of the ME classifier are all reduced by more than threefold below those of the GE. A closer examination of these ECG records indicates that patient-specific beat types are observed during the initial 5-min ECG records. For example, in record 215, the GE performs poorly because of the slight variation in morphology of the normal beats present in this record. However, the LE is able to pick up those patient-specific beats, and therefore, provide significantly enhanced performance (from 3.65% to 98.4%).
- 4) A potential drawback of this proposed method is the need to develop a LE classifier for each individual patient, even with only 5 min of patient's ECG record. Since this must be performed by a physician or a ECG specialist, potentially it would be very costly. We are currently looking into unsupervised learning method, hoping to reduce the number of beats a human expert need to examine in order to develop such a LE. It should be pointed out that in cases where patients' ECG records have been annotated previously by a human expert, the development of a LE would be quite easy and cost effective.

V. CONCLUSION

In this paper, we developed a novel approach to demonstrate the feasibility of having a patient-adaptable ECG beat classification algorithm. We outlined the basic requirements of such a system, namely accuracy, cost-effectiveness and protection of the device manufactures intellectual property rights. We presented a SOM/LVQ-based approach to illustrate that these requirements can be met. The potential benefit of patient adaptation is immense and is worth pursuing further. To the best of our knowledge, the application of the MOE approach to the patient-adaptation problem has never been done before. We believe it can be easily adapted to other automated patient-monitoring algorithms and eventually support decentralized remote patient-monitoring systems.

ACKNOWLEDGMENT

The authors would like to thank Dr. S. Luo at Burdick, Inc., Milton, WI, for many helpful discussions and suggestions. The SOM_PAK and LVQ_PAK developed by the University of Helsinki were used in this study.

REFERENCES

- [1] J. P. Abenstein, "Algorithms for real time ambulatory ECG monitoring," *Biomed. Sci. Instrum.*, vol. 14, pp. 73–79, 1978.
- [2] G. Bortolan, R. Degani, and J. L. Willems, "ECG classification with neural networks and cluster analysis," in *Proc. Computers in Cardiology*, 1991, pp. 177–180.
- [3] J. C. Chang, "Applying artificial neural network for ECG QRS detection," Master thesis, Univ. of Wisconsin—Madison, 1993.
- [4] E. L. Drazen and E. F. Garneau, "Use of computer-assisted ECG interpretation in the United States," in *Proc. Computers in Cardiology*, 1979.
- [5] C. Holzmann, U. Hasseldieck, E. Rosselot, P. Estevez, A. Andrade, and G. Acuna, "Interpretation module for screening normal ECG," *Med. Progress Through Technol.*, vol. 16, pp. 163–171, 1990.
- [6] Y. H. Hu, W. J. Tompkins, and Q. Xue, "Artificial neural network for ECG arrhythmia monitoring," in *Neural Networks for Signal Processing II*, S. Y. Kung F. Fallside, J. Aa. Sorenson, and C. A. Kamm, Eds. Piscataway, NJ: IEEE Press, 1992, pp. 350–359.
- [7] Y. H. Hu, W. J. Tompkins, J. L. Urrusti, and V. X. Alfonso, "Applications of artificial neural networks for ECG signal detection and classification," *J. Electrocardiol.*, 1994, accepted, in press.
- [8] Y. H. Hu, S. Palreddy, and W. J. Tompkins, Eds., "Patient adaptable ECG beat classification using mixture of experts," in *Neural Network for Signal Processing V*. Piscataway, NJ: IEEE Press, 1995, pp. 495–463.
- [9] R. Jacobs, "Method for combining experts' probability assessments," *Neural Computation*, vol. 7, no. 5, pp. 867–888, 1995.
- [10] R. A. Jacobs, M. I. Jordan, S. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, vol. 3, pp. 79–87, 1991.
- [11] M. I. Jordan and L. Xu, "Convergence properties of the EM approach to learning in mixture-of-experts architectures." Cambridge, MA: MIT Press, 1993, p. 9303.
- [12] M. I. Jordan and R. A. Jacobs, "Hierarchical mixtures of experts and the EM algorithm," *Neural Computation*, 1993.
- [13] P. Kinnias and H. A. Fozzard, "Rapid ECG analysis and arrhythmia detection," *Computer Techniques in Cardiology*. New York: Marcel Dekker, 1979, pp. 98–122.
- [14] T. Kohonen, *Self-Organization and Associative Memory*. Berlin: Springer-Verlag, 1984.
- [15] ———, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.
- [16] A. Krogh and J. Vedelsby, Eds., "Neural network ensembles, cross validation and active learning," in *Advances in Neural Information Processing Systems 7*. Cambridge MA: MIT Press, 1995.
- [17] R. Mark and G. Moody, "MIT-BIH arrhythmia database directory," Massachusetts Inst. Technol. (MIT), 1988.
- [18] R. Mark and R. Wallen, "AAMI-recommended practice: Testing and reporting performance results of ventricular arrhythmia detection algorithms," Association for the Advancement of Medical Instrumentation, Arrhythmia Monitoring Subcommittee, Tech. Rep. AAMI ECAR-1987.
- [19] G. B. Moody, "ECG database programmer's guide," Harvard-Massachusetts Inst. Technol (MIT) Division of Health Science and Technology, 1989.
- [20] F. M. Nolle, *et al.*, "Evaluation of a frequency-domain algorithm to detect ventricular fibrillation in the surface ECG," in *Proc. Computers in Cardiology*, 1988, pp. 337–360.
- [21] J. P. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, pp. 230–236, 1985.
- [22] H. V. Pipberger, "The ECG computer analysis system developed in the US Veterans Administration," in *Trends in Computer-Processed Electrocardiograms*, J. H. Van Bommel and J. L. Willems, Eds. Amsterdam, the Netherlands: North-Holland, 1977, pp. 42–48.
- [23] Y. Suzuki and K. Ono, "Personal computer system for ECG ST-segment recognition based on neural networks," *Med., Biol. Eng., Computing*, pp. 2–8, 1992.
- [24] Y. S. Tsai, B. N. Hung, and S. F. Tung, "An experiment on ECG classification using back-propagation neural network," in *Proc. Annu. Int. Conf. IEEE Engineering Medicine and Biology Society*, 1990, pp. 1463–1464.
- [25] R. Watrous and G. Towell, "A patient-adaptive neural network ECG patient monitoring algorithm," *Comput. Cardiol.*, 1995.
- [26] L. Xu and M. I. Jordan, "EM learning on a generalized finite mixture model for combining multiple classifiers," in *Proc. World Congress on Neural Networks*, vol. IV, Portland, OR, 1993, pp. 227–230.
- [27] L. Xu, M. I. Jordan, and G. E. Hinton, "An alternative model for mixtures of experts," in *Advances in Neural Information System 7*, J. D. Cowan, G. Tesauro, and J. Alspector, Eds. Cambridge, MA: MIT Press, 1995.
- [28] Q. Xue, Y. H. Hu, and W. J. Tompkins, "A neural network based adaptive matched filter for QRS detection of very noisy ECG signals," in *Proc. Dig. World Congress on Medical Physics and Biomedical Engineering*, 1991, p. 816.
- [29] Q. Z. Xue, Y. H. Hu, and W. J. Tompkins, "A neural network weight pattern study with ECG pattern recognition," in *Proc. IEEE 11th Int. Conf. Biomedical Engineering*, 1989, pp. 2023–2024.
- [30] ———, "Training of ECG signals in neural network pattern recognition," in *Proc. Annu. Int. Conf. IEEE Engineering Medicine and Biology Society*, 1990, pp. 1465–1466.
- [31] T. H. Yeap, F. Johnson, and Rachniowski, "ECG beat classification by a neural network," in *Proc. Annu. Int. Conf. IEEE Engineering Medicine and Biology Society*, 1990, pp. 1457–1458.



Yu Hen Hu (S'79-M'80-SM'87) received the B.S.E.E. degree from National Taiwan University, Taipei, Taiwan, R.O.C., in 1976. He received the M.S.E.E. and Ph.D. degrees in electrical engineering from University of Southern California, Los Angeles, in 1980 and 1982, respectively.

From 1983 to 1987, he was an Assistant Professor of the Electrical Engineering Department of Southern Methodist University, Dallas, TX. He joined the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, in 1987, as an Assistant Professor (1987-1989) and is currently an Associate Professor. His research interests include multimedia signal processing, artificial neural networks, fast algorithms, and design methodology for application specific micro-architectures, as well as computer-aided design tools for VLSI using artificial intelligence. He has published more than 150 journal and conference papers in these areas.

He is a former associate editor (1988-1990) for the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING in the areas of system identification and fast algorithms. He is currently Associate Editor of Journal of VLSI Signal Processing. He is a founding member of the Neural Network Signal Processing Technical Committee of IEEE Signal Processing Society and served as chair from 1993 to 1996. He is a former member of VLSI Signal Processing Technical Committee of the Signal Processing Society. Currently, he serves as the secretary of the IEEE Signal Processing Society.



Surekha Palreddy received the B.E. degree in biomedical engineering in 1990 from the College of Engineering, Osmania University, India. She received the M.S. degree in biomedical engineering in 1992 from the University of Akron, Akron, OH, and the Ph.D. degree in electrical engineering in 1996 from the University of Wisconsin-Madison.

She is now working as a design engineer on Implantable Cardioverter-Defibrillators at Guidant-CPI, St. Paul, MN.

Willis J. Tompkins (S'61-M'66-SM'77-F'92), for a photograph and biography, see p. 566 of the July 1997 issue of this TRANSACTIONS.