Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, July 31 - August 4, 2005

Bayesian ANN Classifier for ECG Arrhythmia Diagnostic System: A Comparison Study

Dayong Gao, Michael Madden, Des Chambers, and Gerard Lyons
Department of Information Technology
National University of Ireland
Galway, Ireland
{Dayong.Gao, Michael.Madden, Des.Chambers, Gerard.Lyons}@nuigalway.ie

Abstract— **This paper outlines a system for detection of cardiac arrhythmias within ECG signals, based on a Bayesian Artificial Neural Network (ANN) classifier. The Bayesian (or Probabilistic) ANN Classifier is built by the use of a logistic regression model and the back propagation algorithm based on a Bayesian framework. Its performance for this task is evaluated by comparison with other classifiers including Naive Bayes, Decision Trees, Logistic Regression, and RBF Networks. A paired t-test is employed in comparing classifiers to select the optimum model. The system is evaluated using noisy ECG data, to simulate a real-world environment. It is hoped that the system can be further developed and fine-tuned for practical application.**

I. Introduction

Various Machine learning and data mining methods have been applied to improve the accuracy for the detection of ECG arrhythmias. Once a data mining task is identified, appropriate methods have to be selected for execution of this task. Method selection depends highly on the application context as given by initial task analysis, on the properties of the data on which the analysis is being performed, on previous experience with similar domains, and on user-specified requirements for the results [1].

An important factor influencing the success of classification is the choice of an appropriate machine learning technique for the selected task. Since an ever-increasing amount of techniques are available today, ever more specialized knowledge and experimental analysis is necessary to support the selection process.

The electrocardiogram (ECG) is the most important biosignal used by cardiologists for diagnostic purposes. The ECG is, of course, a nonlinear signal generated from a nonlinear system, the human body. Detection of abnormal ECG signals is a critical step in administering aid to patients. Usually, patients with heart conditions are continuously hooked up to cardiac monitors in hospital. Due to the large number of patients in intensive care units and the need for continuous observation of them, several methods for automated arrhythmia detection have been developed in the past few decades to attempt to assist with the monitoring task. These include Bayesian [2] [3] and heuristic approaches [4], expert systems [5], Markov models [6], self-organizing map [7], and Artificial Neural Networks [8] [9]. In general, different techniques might be applicable to similar tasks. Which one is selected depends to large extent on the exact requirements posed by the user. In the application being considered here, the most important requirements are that the system have high accuracy, that it can operate on a device with low computational power in real-time (though training may be performed off-line) and that diagnoses are not tied to the individual patient.

According to published results, existing approaches generally tend to suffer from problems that result from high sensitivity to noise included in the data, and unreliability in dealing with new or ambiguous patterns. In this domain, classifiers must be developed that are able to deal with nonlinear discrimination between classes, incomplete or ambiguous input patterns, and suppression of false alarms. It is necessary to develop detection schemes with a high accuracy, or equivalently, low false-positive and false-negative statistics, for them to be useful in a practical deployment. Guvenir *et. al.* [10] developed a supervised machine learning algorithm, called VF15 for Voting Features Intervals, which was applied to the task of arrhythmia analysis using on the UCI Arrythmia dataset [11], which has missing features and unlabeled classes. This classifier achieved an accuracy of 62%, which obviously is not sufficiently good for clinical use. It remains a problem to select an optimum model which can meet the requirements in practical application. So far, implementations of ML-based ECG classification schemes have, in general, been focused on problems within narrow clinical domains.

In this paper, we review several machine learning methods, *i.e.*, Naive Bayes, Decision Trees, Logistic Regression and RBF Network, and present a Probabilistic ANN for automatic detection of arrythmias from ECG signals. The purpose of the study is to improve our understanding of various techniques and evaluate their performance, in order to determine the optimum methods for our system. Practical problems that arise in implementing our system are also explored, including dealing with noisy ECG signals and suppressing false alarm signals.

II. Problem and Dataset

In order to assess the ability of techniques considered in this work to deal with incomplete or ambiguous biosignal data from multiple patients in a real-world setting, we use the UCI Arrythmia dataset [11] developed by Guvenir *et. al.*

[10] for our simulation experiment. This dataset consists of 452 ECG recordings from different patients and includes about 0.33% missing attribute values and 22 unclassified cases, so the prediction accuracy of any model built using it cannot be perfect. However, such characteristics make it quite similar to the the type of data that will be acquired using special-purpose hardware being developed in in our project in parallel with the work described here, as noisy data may well be acquired. Therefore, we believe that the results of this study should be transferable to actual ECG signals. Each record consists of a set of clinical parameters measured on ECG signals and some personal information about the subjects. The dataset falls into two classes, Normal and Abnormal, with 245 cases in the Normal group and 207 cases in the Abnormal group. The ECG signals used as system inputs include 5 parameters: QRS duration, PR interval, QT interval, T interval and P interval. The personal information available includes age, height, weight and sex.

Since there is a limited amount of data available, it is necessary to be sure that all the information included in the data can be used to evaluate the performance of a classifier. Ten-fold cross-validation is used in this study. In each fold, 90% of the dataset (407 ECG recordings) is used for training and the remaining 10% (45 ECG recordings) is used for testing. Then the average error across all trials is computed.

III. Classification by Machine Learning

Machine learning classification algorithms aim to automatically generate expressions that can assign each instance to a particular class. Numerous machine learning methods have already been applied for supporting decision making in medical domains, based on probabilistic or statistical models and various algorithmic approaches [14]. Of these methods, Naive Bayes, Decision Trees, Logistic Regression and Neural Networks are some of the most relevant to our research as they have been used in related applications.

Naive Bayes: The Naive Bayes classifier assumes that all other variables are conditionally independent of each other given the classification variable. Several studies have shown Naive Bayes to be competitive with more sophisticated classifiers [15].

Considering the task of arrythmia classification in this study, we assume that the ECG data was generated by a parametric model and use the training data to calculate Bayes-optimal estimates of the model parameters. Then, equipped with these estimates, the classifier classifies new signals using Bayes' rule to calculate the posterior probability that a class would have generated the signals. This classifier computes the conditional probabilities of the different classes given the values of attributes and then selects the class with the highest conditional probability. Supervised discretization is used to convert numeric attributes to nominal ones.

Logistic Regression: Logistic Regression may be used to estimate the probability of a certain event occurring. As well as being used to predict a dependent variable on the basis of values of independent variables, it may be used to determine the percentage of variance in the dependent variable explained by the independent variables, to rank the relative importance of independent variables, to assess interaction effects and to explore the impact of covariate control variables.

Considering the task in this study, a logistic regression model with a ridge estimator [16] is built. The Quasi-Newton Method is used to search for the optimized values of the variables. For estimating the log-likelihood probability, the ridge value is used and its value is set to 10^{-8} in all the experiments.

Decision Tree: Quinlan's C4.5 algorithm [17] for greedy top-down induction of decision trees is one of the most widely applied techniques for classification. Practical advantages of decision tree algorithms include their low computational complexity and ability to deal with numeric and categorical variables and missing variable values. However, the key feature of decision trees is that they are capable of providing an interpretable representation of a complex decision-making process [18], at least in principle.

In a medical diagnosis task, different cases may be described in terms of symptoms or results of diagnostic tests at different levels of precision, *e.g.*, a patient may be described as having cardiac arrhythmia without specifying the precise type of arrhythmia. The C4.5 algorithm, with its default settings, is used in this study as it has previously been shown to be useful for the detection of cardiac arrhythmia [19].

Neural Networks: An artificial neural network is a mathematical model for information processing based on a connectionist approach to computation. Back Propagation (BP) networks and Radial Basis Function (RBF) networks are both well-known variants. Both can learn arbitrary mappings or classifications.

RBF networks continue to attract interest for engineering applications. In an RBF network, the activation of a hidden unit is determined by the distance between the input vector and the center vector of the hidden unit. The weights connecting the hidden layer and output layer can be determined by a Linear Least Square (LLS) method [2]. Tjoa *et. al.* [20] concluded that RBF networks can give slightly higher classification accuracy for cardiac diagnosis compared to BP networks. This study implements a RBF network with the the K-means clustering algorithm to provide the basis functions.

For complex practical situations, where the goal is to estimate the probability of arrythmia, mixture models may be particularly appropriate because of the iterative nature of fitting such density functions. We present such a mixture model, the Bayesian ANN Classifier, in the next section.

IV. Bayesian ANN Classifier

Since detailed descriptions of BP neural networks can be found in the literature, *e.g.* [21], here we mainly focus on the description of the Bayesian ANN Classifier. Its configuration is shown schematically in Fig. 1, where $i = 1, 2, \cdots, I$ is the number of input nodes, $j = 1, 2, \cdots, J$ is the hidden nodes, $k = 1, 2, \cdots, K$ is the class labels of the output node.

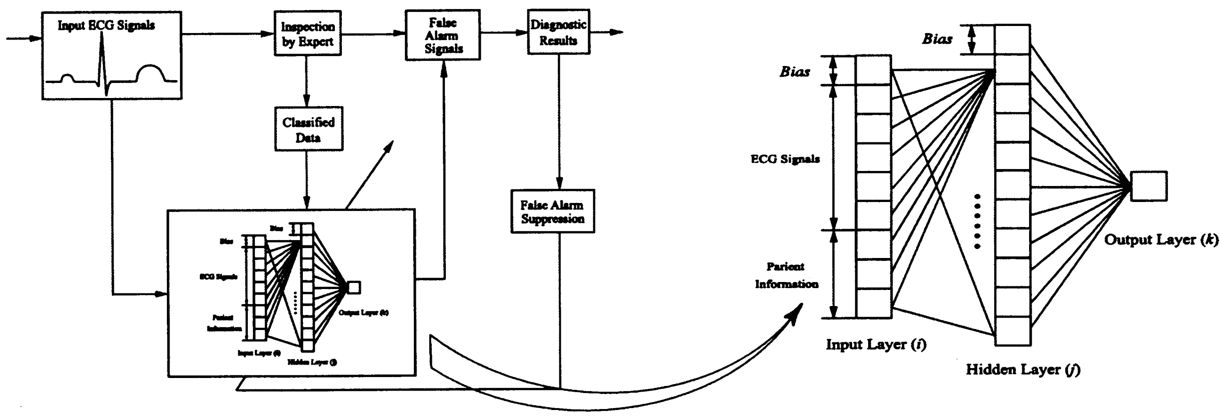

Fig. 1. Block Diagram of ANN-Based Arrhythmia Diagnostic System with ECG Signals and Configuration of Neural Network Classifier with input layer ($i = 1, 2, \cdots, I$), hidden layer ($j = 1, 2, \cdots, J$) and output layer ($k = 1, 2, \cdots, K$)

Assume we have a training set $\mathcal{D}$, consisting of N input-output pairs:

$$\mathcal{D} = [(X^n, y^n) \mid n = 1, 2, \cdots N] \tag{1}$$

where X is an input vector consisting of I elements and y is the corresponding class label consisting of K classes. The objective is to use an ANN to model the input-output relation ($y = k \mid X$). Here, the class label is binary-valued y = (1, 0), corresponding to Normal (no arrythmia) and Abnormal (arrythmia), respectively.

An alternative target for our task is to use a logistic regression model based on a Bayesian method, and estimate the class probability for a given input by:

$$\mathcal{P}(y = k \mid X), \qquad k = 1, 0 \tag{2}$$

The outputs for the summation operation and sigmoidal activation function in the hidden and output neurons, denoted by S_j and S_k, respectively, can be written as follows:

Hidden layer:

$$S_j = \tanh(\sum_i \omega_{ji} X_i + \omega_{j0}) \tag{3}$$

Output layer:

$$S_k = \sum_k \omega_{kj} S_j + \omega_{k0} \tag{4}$$

where tanh is the tangent hyperbolic function, a conventional sigmoid function. Here, ω_{ji} denotes the weight matrix in the input layer and ω_{kj} the weight matrix in the output layer.

To allow the outputs to be interpreted as probabilities, logistic regression is used to model the risk (or probability) of occurrence of arrhythmia. Let P ($y = k \mid X$) be the probability of the event $y = 1$, given the input vector X. This is modeled as a function of network output y by:

$$\mathcal{P}(y = k \mid X) = \frac{1}{1 + \exp(-S_k)} \tag{5}$$

The logistic regression model is simply a non-linear transformation of the linear regression. The logistic distribution is an S-shaped distribution function which is similar to the standard-normal distribution but easier to work with in most applications. It constrains the estimated probabilities to lie between 0 and 1.

The system is a multi-layer perceptron neural network, trained using the BP algorithm. The network is optimized using a log-likelihood cost function, given by

$$\mathcal{C}(\omega) = -\frac{1}{k} \sum_k \sum_i y_i(k) \ln[\mathcal{P}(y = k \mid X)] \tag{6}$$

where $\omega = [\ \omega_{ji},\ \omega_{kj}\ ]$ is the vector of network weights.

To minimize the cost function between the actual and desired outputs of the network, the BP algorithm passes information from the output neuron backwards to all hidden units to form error terms which are used to update the weights of the multi-layer network. In this study, a three-layer fully connected network with 10 hidden units and a single output unit is used.

V. Empirical Results

A. Classification Results

Clinical research often investigates the statistical relationship between symptoms (or test results) and the presence of disease. When significant associations are found, it is useful to express the data in ways which are clinically relevant. Therefore, Sensitivity and Specificity are used to measure the accuracy of the model and verify that the classifier acquires the underlying dynamics of the system from the data.

Let:

- FP = false positives;
- FN = false negatives;
- TP = true positives;
- TN = true negatives; and
- UN = classified as "uncertain", if applicable.

Then, Sensitivity is defined as the probability that a symptom is present (or a test predicts that the person has the disease) given that the person has the disease.

$$Sensitivity = \frac{TP}{FN + TP} \tag{7}$$

TABLE I

CLASSIFICATION RESULTS (MEAN±STDEV WITH 95% CONFIDENCE INTERVAL) FOR NAIVE BAYES, LOGISTIC REGRESSION, DECISION TREE, RBF NETWORK AND BAYESIAN ANN CLASSIFIER (BANN). $\mathcal{FDR}$: FALSE DISCOVERY RATE.

Sensitivity	Specificity	$\mathcal{FDR}$(%)
Naive Bayes		
0.59±0.03	0.19±0.03	29.54±1.11
Logistic Regression		
0.58±0.02	0.23±0.02	31.24±0.52
Decision Tree		
0.76±0.08	0.14±0.05	18.61±3.01
RBF Network		
0.70±0.03	0.40±0.02	35.13±0.53
Bayesian ANN		
0.76±0.04	0.15±0.02	19.31±1.67

Specificity is defined as the probability that a condition is not present (or a test predicts that the person does not have the disease) given that the person does not have the disease.

$$Specificity = \frac{TN}{FP + TN} \tag{8}$$

As well as Sensitivity and Specificity, another widely used measurement, the False Discovery Rate ($\mathcal{FDR}$) is used to examine how well the Bayesian ANN classifier performs in recognizing arrhythmias and to compare it with other machine learning methods. $\mathcal{FDR}$ quantifies the expected *proportion* of false predictions in the set of predictions, rather than quantifying the chance of any false positives.

$$\mathcal{FDR} = \frac{FP + FN}{FP + TP + FN + TN + UN} \tag{9}$$

Table I shows results for Naive Bayes, Logistic Regression, C4.5, RBF Network and Bayesian ANN. The results in Table I are calculated with 95% confidence interval and have been rounded into 2 decimal places. From Table I, it may be observed that in general, the scores of the Bayesian ANN Classifier for Sensitivity, Specificity and $\mathcal{FDR}$ are higher or equivalent to those of the other methods.

B. Method Evaluation and Model Selection

It is essential in Machine Learning to be able to compare results from different algorithms or variations statistically, to decide which is best for a given application. The paired t-test is used in this study to determine whether a significant difference exists between the means of two distributions. We perform pairwise comparisons between the Bayesian ANN classifier and each other one. While comparing any two result sets, our null hypothesis is that there is no statistically significant difference between them.

Because of the high degree of overlap between training sets due to the 10-fold cross-validation procedure, a corrected resampled t-test, as recommended by [22], is employed. The t-statistic is defined as:

$$t = \frac{\frac{1}{n}\sum_{j=1}^{n} x_j}{\sqrt{(\frac{1}{n} + \frac{n_2}{n_1})\hat{\sigma}^2}} \tag{10}$$

where x_j is the difference of the accuracy between the comparative algorithm, measured on run j ($1 \leq j \leq n$). In each run, n_1 instances are used for training, and n_2 instances are used for testing. $\hat{\sigma}^2$ is the estimates of the variance of the n differences.

TABLE II

EVALUATION EXPERIMENT FOR PAIRED T-TEST BASED ON THE PROPOSED BAYESIAN ANN (BANN) CLASSIFIER WITH NAIVE BAYES (NB), LOGISTIC REGRESSION (LR), DECISION TREE (DT) AND RBF NETWORK (RBF). $\mathcal{FDR}$: FALSE DISCOVERY RATE. P(T⇐T) IS THE PROBABILITY OF THE AVERAGE ACCURACY IS DIFFERENT. SIGDIFF: SIGNIFICANT DIFFERENCE.

	BANN vs NB	BANN vs LR	BANN vs DT	BANN vs RBF
Sensitivity				
P(T⇐t)(%)	0	0	1.6776	0.0271
SigDiff	True	True	False	True
Winner	BANN	BANN	Draw	BANN
Specificity				
P(T⇐t)(%)	0.0034	0	1.1783	0
SigDiff	True	True	False	True
Winner	BANN	BANN	Draw	BANN
$\mathcal{FDR}$				
P(T⇐t)(%)	0	0	0.7601	0
SigDiff	True	True	False	True
Winner	BANN	BANN	Draw	BANN

Table II presents the results of the paired comparisons. It is found that according corrected resampled t-test, the Bayesian ANN Classifier is superior on this dataset to the classifiers based on Naive Bayes, Logistic Regression and RBF Network, and equivalent to Decision Tree classifier. However, an advantage of the Bayesian ANN classifier is that, rather of a categorical classification, it can return a probability distribution. Because the noisy data are used to simulate the practical environment in the experiment, a relatively high error rate exists across all classifiers. Therefore, the property that the proposed Bayesian ANN Classifier can produce a probability output becomes very useful in a practical application, as the decision threshold based on the probability prediction can be adapted. This is discussed in the following section.

C. Classification Result Analysis

To simplify, a sample from the spited data (90% of the whole dataset for training, the rest for testing) is calculated by Bayesian ANN Classifier to be used for further analysis. The classification results for training and testing phases are shown in Fig. 2 and Fig. 3, respectively. For decision-making, the probability output from the classifier is rounded to 1 or 0, depending on the classification probability threshold. Examining results in detail, it was observed that misclassifications were mainly concentrated on the area where probabilities are estimated to be between 0.4 and 0.7. It is therefore concluded

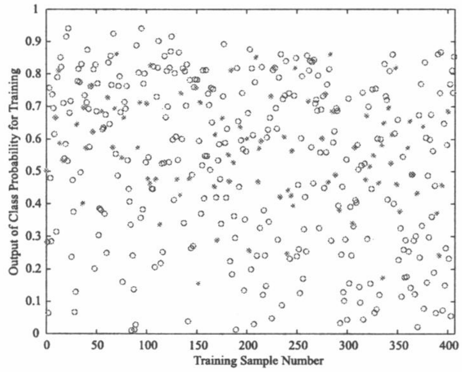

Fig. 2. Dispersal Plot of Classification Results for Training. The true and false classification results are marked by ∘ and ∗ respectively.

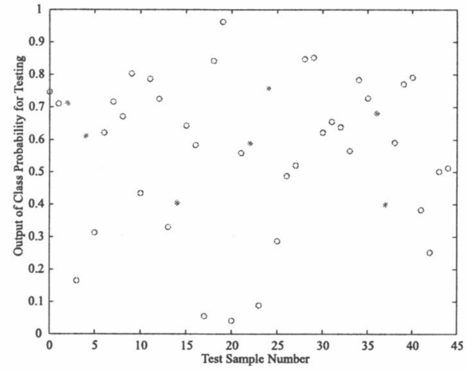

Fig. 3. Dispersal Plot of Classification Results for Testing. The true and false classification results are marked by ∘ and ∗ respectively.

that the classification results around probability = 0.5 have higher risk of misclassification. This also suggests that if the misclassifications between 0.4 and 0.7 could be ignored, the misclassification would be reduced substantially.

The randomness of the occurrence of arrhythmic beats suggests that a static analysis, based only on the features of the current beat, might be appropriate. The high intra- and inter-patient variability of the beat shape suggests an approach that takes the patient as a reference of himself or herself. However, in clinical domain, it is hard for a physician to build a new model for every patient, especially when we want to monitor a patient' condition in the real-time but have no his/her data to train our classifier and build our model. Therefore, the use of inter-patient data included high noisy component and incomplete information becomes very important in practical domain.

D. Threshold Adaption

A well-known difficulty with medical monitoring devices that raise alarms to alert the physician of problems is the "cry wolf" dilemma: such devices typically follow a conservative strategy of raising false alarms rather than risking a situation arising where a patient requires attention but no alarm is raised, but the physician may end up ignoring all alarms if many of them are false.

A dual threshold method is proposed here in an attempt to address this problem and suppress false alarm signals. The basic idea is to adopt a dual threshold, whereby an uncertainty criterion is introduced for high-risk classification outputs lying between Threshold 1 and Threshold 2.

IF:

Output Probability > Threshold 1

THEN:

y = Classification 1 (Normal)

ELSE IF:

Output Probability < Threshold 2

THEN:

y = Classification 2 (Abnormal)

ELSE:

y = Uncertain Class

To evaluate the performance of the dual threshold method, a sample of diagnostic results without dual threshold and with a dual threshold of 0.4 and 0.7 is calculated based on Bayesian ANN classifier and shown in Table III. Without dual threshold, the False Discovery Rate for training and testing phases is 22.36% and 15.56%, respectively. The False Discovery Rate for training and testing phases by the use of dual threshold is reduced to as little as 6.39% and 6.67%, respectively. Therefore, an obvious improvement in prediction accuracy is obtained by the use of dual threshold. However, we get a classification response of "uncertain" between probabilities of 0.4 and 0.7, meaning that such cases should be reviewed manually by the supervising physician for a final decision.

When the thresholds are varied, different levels of classification accuracy and uncertainty are obtained. This can be used to control the diagnostic strategy. When suitable thresholds are chosen in building the Bayesian ANN Classifier, an optimum balance between False Discovery Rate and uncertainty may be found. The dual threshold method presented here could be applied to suppress false alarm signals and generate rules about the suppression of alarms. These rules may also be reviewed by a physician and potentially incorporated into the training phase of the system.

VI. Conclusion

In this paper, a arrhythmia detection system with ECG signals based on a Bayesian ANN Classifier is presented and its performance is compared with that of other classifiers, specifically Naive Bayes, C4.5 Decision Trees, Logistic Regression and RBF Networks. A corrected resampled paired t-test is employed to evaluate the results of difference models. The results show that the Bayesian ANN Classifier is one of the optimum models. This classifier appears to acquire arrhythmia properties from the underlying dynamics of the system, even when the dataset includes incomplete information, such as missing feature values and unclassified classes. This approach is potentially useful for generating a pattern recognition model based on given {input, output} sets to classify future input sets for arrhythmia diagnosis. The capability of uncertainty management with the dual threshold method described here

TABLE III

ARRHYTHMIA DIAGNOSTIC RESULTS BY ECG SIGNALS FOR ANN-BASED DIAGNOSTIC SYSTEM WITH AND WITHOUT DUAL THRESHOLD METHOD. $\mathcal{FDR}$: FALSE DISCOVERY RATE.

	Case of Classes	Classification Results			Results with Dual Threshold Method			
		Normal	Abnormal	$\mathcal{FDR}$(%)	Normal	Abnormal	Uncertain	$\mathcal{FDR}$(%)
Training								
Normal	216	183	33	15.28	104	4	108	1.85
Abnormal	191	58	133	30.36	22	113	56	11.52
Total	407	241	166	22.36	126	117	164	6.39
Testing								
Normal	29	27	2	6.90	13	0	16	0.00
Abnormal	16	5	11	31.25	3	10	3	18.75
Total	45	32	13	15.56	16	10	19	6.67

could be used to control a diagnostic strategy and suppress false alarm signals.

Our future work will focus on evaluating the performance of the system based on a large amount of real-world data, to be gathered using custom hardware currently under development. As well as ECG monitoring, the system will include pulse-ox sensors and a blood pressure monitor. The pulse-ox sensor will monitor the oxygen saturation in the blood stream. It is based on a finger pulse oximeter and will do measurements in programmable intervals. The blood pressure monitor consists of a control unit and an inflatable cuff, which will be wrapped either around the upper arm of the wrist. Measurements will be taken in programmable intervals. The output of this sensor is the systolic, the diastolic and the mean arterial pressure.

The design of low-cost, high-performance, simple to use, and portable equipment for ECG signal monitoring, that offers a combination of diagnostic features, seems to be a goal that is highly worthwhile. We hope that this system can be further developed and fine-tuned for practical applications.

ACKNOWLEDGMENT

This research is supported by Enterprise Ireland, Strategic Research Project number *ARTP 2002 INF 210A*.

REFERENCES

[1] W. Klosgen and J. M. Zytkow, *Handbook of Data Mining and Knoweledge Discovery*, Oxford University Press, 2002.

[2] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, New York, USA, 2001.

[3] J. L. Willems and E. Lesaffre, "Comparison of multigroup logisitic and linear discriminant ECG and VCG classification," *J. Electrocardiol.*, vol. 20, pp. 83-92, 1987.

[4] J. L. Talmon, *Pattern Recognition of the ECG*, Berlin, Germany: Akademisch Proefscrift, 1983.

[5] 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, 1997.

[6] A. D. Coast, R. M. Stern, G. G. Cano, and S. A. Briller, "An approach to cardiac arrhythmia analysis using hidden Markov models," *IEEE Trans. Biomed. Eng.*, vol. 37, pp. 826-835, 1990.

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

[8] R. Silipo and C. Marchesi, "Artificail Neural Networks for Automatic ECG Analysis," *IEEE Trans. Signal Processing*, vol. 46 (5), pp. 1417-1425, 1998.

[9] Yang Wang, Yi-Sheng Zhu, Nitish V. Thakor, and Yu-Hong Xu, "A Short-Time Multifractal Approach for Arrhythmia Detection Based on Fuzzy Neural Network," *IEEE Trans Biomed. Eng.*, vol. 48 (9), pp. 989-995, 2001.

[10] H. A. Guvenir, B. Acar, G. Demiroz, A. Cekin, "A Supervised Machine Learning Algorithm for Arrhythima Analysis," *In: Proc. Computers in Cardiology Conference*, Lund, Sweden, 1997, vol. 24, pp. 433-436.

[11] C. L. Blake, C. J. Merz, UCI Repository of machine learning databases, [http://www.ics.uci.edu/~mlearn /MLRepository.html], Irvine, CA: University of California, Department of Information and Computer Science, 1998.

[12] M. Hintz-Madsen, L. K. Hansen, J. Larsen, and K. Drzewiecki, "A Probabilistic Neural Network Framework for Detection of Malignant Melanoma," *Artificial Neural Networks in Cancer Diagnosis, Prognosis and Patient Management*, CRC Press, pp. 141-183, 2001.

[13] Pierre Baldi and Soren Brunak, *Bioinformatics: The Machine Learning Approach*, 2nd Edition, ISBN: 0-262-02506-X, A Bradford Book, The MIT Press, Cambridge, Massachusetts, London, England: 2001.

[14] Stuart Russell, Peter Norvig, *Artificial Intelligence: A Modern Approach*, First Edition, pp. 531.

[15] George H. John and Pat Langley, "Estimating Continuous Distributions in Bayesian Classifiers," *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Mateopp, pp. 338-345, 1995.

[16] le Cessie, S. and van Houwelingen, J.C., "Ridge Estimators in Logistic Regression," *Applied Statistics*, vol. 41, No. 1, pp. 191-201, 1992.

[17] J. R. Quinlan, "Introduction on Decision Trees," *Machine Learning*, vol. 1, pp. 81-106, 1986.

[18] S. R. Safavin and D. Landgrebe, "A Survey of Decision Tree Classifier Methodology," *IEEE Trans. Syst, Man, Cybern.*, vol. 21, pp. 660-674, 1991.

[19] A. Ch. Stasis, E. N. Loukis, S. A. Pavlopoulos, and D. Koutsouris, "Using decision tree algorithms as a basis for a heart sound diagnosis decision support system," *Proceedings of the 4th Annual Conference on Information Technology in Biomedicine*, UK, 2003, pp. 354-357.

[20] M. P. Tjoa, D. Narayana Dutt, Y. T. Lim, B. W. Yau, R. C. Kugean, S. M. Krishnan and K. L. Chan, "Artificial neural networks for the classification of cardiac patient states using ECG and blood pressure data," *The 7th Austria and New Zealand Intelligence Information Systems Conference*, Perth, Western Australia, pp. 323-327, 2001.

[21] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning internal representation by error propagation, *In: Rumelhart, D. E., McClelland, J. L., and the PDP research group (Eds), parallel distributed processing*, (MIT press, Cambridge, MA, 1996) pp. 318-362.

[22] Remco R. Bouckaert and Eibe Frank, "Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms," *Advances in Knowledge Discovery and Data Mining, The 8th Pacific-Asia Conference, PAKDD 2004*, Springer, Sydney, Australia, pp. 3-12, 2004.