

# Ambient Cardiac Expert: A Cardiac Patient Monitoring System using Genetic and Clinical Knowledge Fusion

#Iqbal Gondal, Shoaib Sehgal, Mudasser Iqbal, Joarder Kamruzzaman  
GSIT, Faculty of IT, MONASH University, Churchill, 3842 Australia  
[Iqbal.Gondal, Shoaib.Sehgal, Mudasser.Iqbal,  
Joarder.Kamruzzaman]@infotech.monash.edu.au

## Abstract

*Cardiac patients can be regularly monitored using low cost sensor networks which can save many lives and valuable time of experts. This monitoring can be more effective if in addition to standard clinical parameters genetic information is used because of its ability to predict hereditary diseases like cardiac problems. Current clinical practices, however, only stress on physiological observation to predict heart failure rate which could miss the important information which could lead to fatal consequences. This paper presents Ambient Cardiac Expert (ACE) which combines physiological parameters observed using sensor networks with gene expression data to predict the heart failure rate. The system uses well established Support Vector Machines (SVM) for class prediction and uses Wrapper Evolutionary Algorithm based on Gaussian Estimation of Distribution Algorithm (EDA) to determine cardiac patient's criticality. Results suggest that ACE can be successfully applied for cardiac patient monitoring and has ability to integrate the information from both clinical and genetic sources.*

**Index Terms**—Ambient Intelligence, Fusion, Class Prediction, Support Vector Machines, Evolutionary Algorithm.

## 1. Introduction

Patients who show clinical symptoms (high blood pressure and high cholesterol etc) in older age of cardiac disease are required to have continuous monitoring of their condition by regular checks-up. This is not only expensive and time consuming exercise but also, at times introduces fatal consequences due to unavailability of expert at the required time. One of the

possible solutions is to use low cost sensors to monitor these patients continuously and when based on clinical information the risk of heart failure extends beyond certain threshold then expert should be contacted. This system will not only save precious lives but also will help to save valuable time of cardiac experts.

This paper presents such low cost monitoring system, *Ambient Cardiac Expert* (ACE), machine learning methods to intelligently make decisions regarding criticality level of cardiac patient and sensor networks to collect their physiological parameters. In addition to using simple clinical parameters, monitored using sensor networks ACE, incorporates genetic information. Because research has shown that mutations in genes can determine a patient's general risk of getting any particular disease. This genetic data is also helpful in determining rate of progression or the patient's response to the treatment. In order to address these issues, much interest has focused on microarray analysis, which is a method of determining the level of expression in a tissue sample of many genes simultaneously. Change in genes expression are as a result of some genes being up-regulated and producing increased amounts of messenger RNA (mRNA) whereas other genes being down-regulated to produce less mRNA [1]

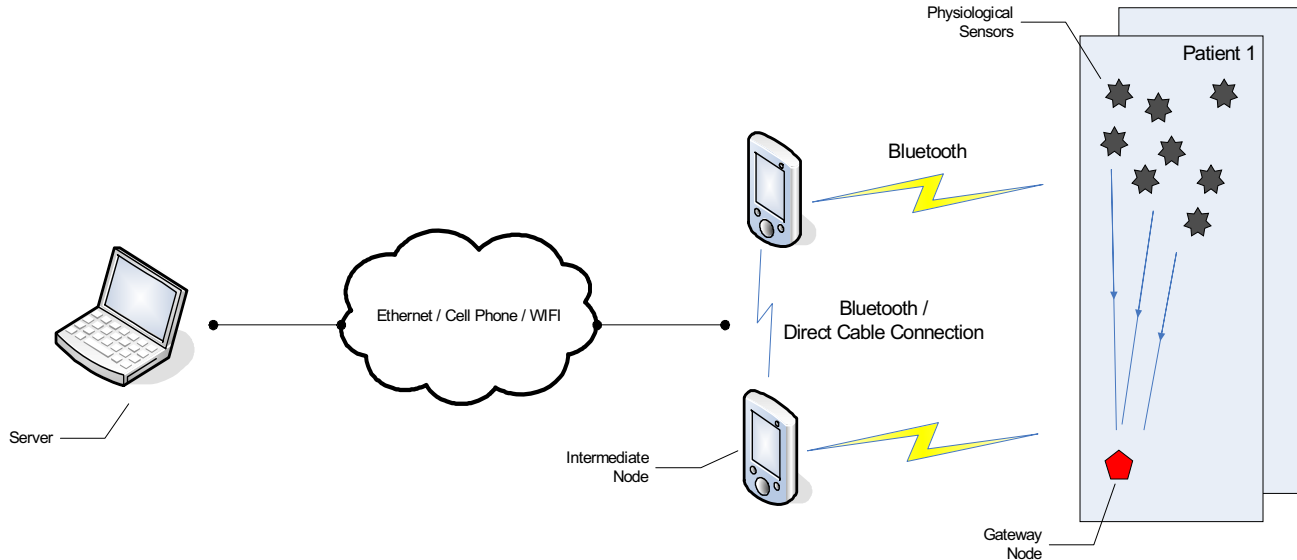
It is known that genetic mutations are associated with cardiac abnormalities [2]. To detect these mutations several machine learning algorithms have already proven their significance in the molecular classification of DNA microarray. Golub et al [1], Shipp et al [2], Pomeroy et al [3], Bhattacharjee et al [4] and Ramaswamy et al [5], all provide a broad range of examples where machine learning has been applied to leukemia, lymphoma, brain cancer, lung cancer and multiple primary tumour classification. *Support Vector*

*Machines* (SVM) has shown promising results in a variety of biological classification tasks, including gene expression microarrays [6, 7]. For example, Byvatov et al, [8] proposed the use of SVM over Back Propagation Neural Networks in identifying small organic molecules that potentially modulate the function of G-protein coupled receptors. Due to this proven performance ability in this paper ACE uses SVM to classify both normal and heart patient data based on both clinical parameters monitored using sensor networks and genetic expression data generated using microarrays.

Complete ACE architecture is demonstrated in Figure 1. Physiological sensors listen to the physiological parameters mentioned in Table 1 (Section 2). These parameters are then transferred via Bluetooth link to PDA and then to the server which has pre-recorded

genetic data of particular patient. ACE intelligent decision making module then integrates the physiological and genetic knowledge and then sends back the heart failure rate computed using ACE intelligent module. Based on current abnormal patient conditions and centralized genetic mutation information, the severity risk factor of heart disease would be established and then communicated to emergency services. This system could save valuable medical resources by weeding out false abnormal heart conditions and could improved drug tail process.

Rest of the paper is organized as follows: Section 2 will present ACE system in greater detail. Analysis of results will be presented in Section 3, while conclusions are drawn in Section 4.



**Figure 1: ACE Architecture for Cardiac Patient Monitoring.**

## 2. Ambient Cardiac Expert (ACE)

The complete ACE framework is demonstrated in Figure 2. Both genetic  $Y_g$  and clinical samples  $Y_c$  are classified using separate SVMs trained for both genetic and clinical data. The decision of SVM is then fused together using:

$$\gamma = \alpha.D_g + \beta.D_c \quad (1)$$

where  $\alpha$  and  $\beta$  are respective weights calculated using *Dynamic Wrapper Evolutionary Algorithm* (Section 2.A) while  $D_g$  and  $D_c$  are the respective distance values computed using SVM. To compute these values the SVM converts input vector space  $\mathbb{R}^n$  to higher dimensional space and attempt to locate a separating hyperplane [9-11]. The data

$Y_g$  is transformed into higher dimensional space. The separating hyperplane in higher dimension space satisfies:

$$W.Y_g + b = 0 \quad (2)$$

The distance between separating hyperplane and input vector  $Y_g$  is computed by:

$$D_g = \frac{|W^T Y_g + b|}{\|W\|^2} \quad (3)$$

Similarly, distance value  $D_c$  is computed and then these values are used in (1) to compute  $\gamma$  to finally predict the heart failure rate.

SVM maximize the margin between the classes using:

$$\text{Max} \frac{1}{\|W\|^2} \quad (4)$$

Subject to the condition

$$f_i (W.Y_{gi} + b) \geq 1 \quad (5)$$

Using the Kuhn-Tucker condition [11] and LaGrange Multiplier methods (4) is equivalent to solving the dual problem

$$\text{Max} \left[ \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i=1}^l a_i a_j a_i f_j K(Y_{gi}, Y_{gj}) \right] \quad (6)$$

where  $0 \leq a_i \leq C$ ,  $l = \text{number of inputs}$ ,  $i = 1, \dots, l$  and

$\sum_{i=1}^l a_i f_i = 0$ . In (6)  $K(Y_{gi}, Y_{gj})$  is the kernel function and in

our experiments, a linear, polynomial and RBF functions are used.  $C$  is the only adjustable parameter in the above calculations and can be computed using simple grid search method.

Following Sub-Section will present algorithm to compute weights and used in (1) to predict the odds of heart failure.

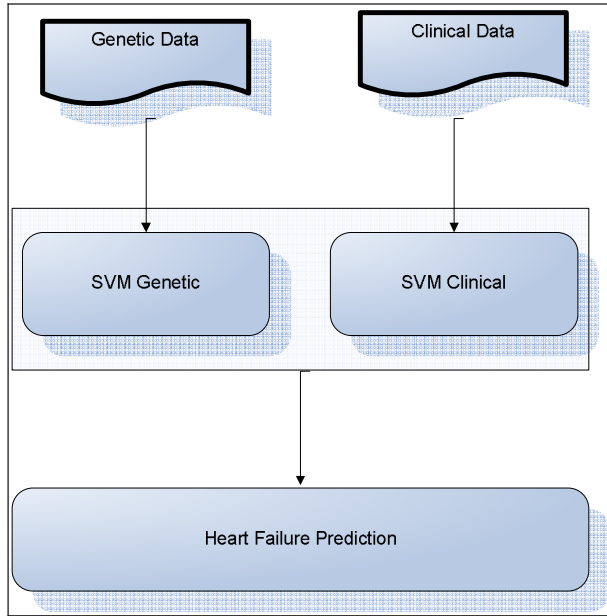


Figure 2: The ACE Framework

## 2.1 Dynamic Wrapper Evolutionary Algorithm for Adaptive Weight Calculation

To compute weights and in (1) both genetic  $Y_g$  and clinical  $Y_c$  data are first divided into two sets of  $k$  folds and then DWE processes the data for  $k$ -iterations (STEP 1, Figure 3).

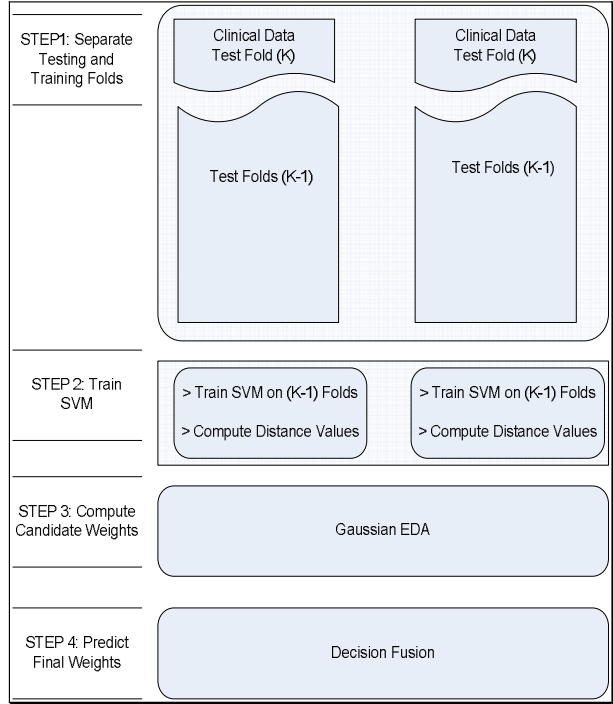


Figure 3: The DWE algorithm for dynamic computation of weights using Gaussian EDA.

For every iteration a  $k^{th}$  test fold is selected, rest of the  $k-1$  folds in  $Y_g$  and  $Y_c$  are used for the training of SVMs (SVM Genetic and SVM Clinical, Figure 2) respectively (STEP 2) such that, the probability  $P_i$  of the training subset being selected as a training data for a particular iteration is:

$$P_i = \frac{(k-1) \times \eta}{N \times L} \quad (7)$$

where  $N$  = total data items per class,  $k$  = number of folds,  $L$  = number of classes and  $\eta$  = number of samples in each subset. The remaining one fold is used for testing (*leave one out*) such that the selection probability  $P_v$  of each fold to become a part of validation data for a particular iteration is:

$$P_v = \frac{\eta}{N \times L} \quad (8)$$

After both Genetic SVM and Clinical SVM are trained on  $k-1$  folds the  $k^{th}$  fold is classified using (1) where the weights are computed using algorithm represented in Figure (4).

### Method: ComputeWeights

*Precondition:* Gene expression matrix  $Y_g$ , Clinical data matrix  $Y_c$ , Population size  $N_p$ , Number of Generations  $G_r$   
*Postcondition:* Weights and

### Algorithm:

1. Select 70% of the folds as training fold,  $T_r$  from  $(k-1)$  folds.
2. Select rest of the 30%  $T_r$  folds.
3. Train SVMs on  $T_r$ .
4. Compute distance  $D_{wg}$ ,  $D_{wc}$  values using (3).
5. For  $noOfGenerations = 1:Gr$ 
  - a. Initialize a weight population  $P_p$  of  $N_p$  individuals randomly.
  - b. Compute class using (1) for all and in  $P_p$ .
  - c. Compute prediction accuracy.
  - d. Rank  $P_p$  in descending order based on their prediction accuracy.
  - e. Select first  $S_p$  individuals  $B$  from  $P_n$  such as  $S_p \leq P_n$ .
  - f. Estimate  $n$ -dimensional *Probability Density Function* of  $B$  using  $P(Y) = P(Y|B)$ .
  - g. Sample new  $P_n$  individuals from  $P(Y)$  by partially replacing current population.
6. END
7. Use the last population of and and test on  $T_t$
8. Select the value and with the maximum accuracy.

STOP

**Figure 4: Gaussian EDA Algorithm to Compute Weights**

The choice of EDA algorithm is made due to its better performance for different optimization problem over traditional evolutionary algorithm e.g. Genetic Algorithms [12]. It has advantage over traditional evolutionary algorithms in a way that both crossover and mutation operators are substituted by estimation process.

### 3. Discussion of Results

#### 3.1 Test Data

To test ACE system clinical data, generated by following four different groups was used:

1. Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.
2. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.
3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.
4. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.

The data had 120 samples from healthy persons while there were 102 samples from different patients. Each sample comprised of 13 clinical parameters shown in Table 1. Genetic data on the other hand had much larger attributes (54,676 genes) while much lower number of samples for both normal and idiopathic heart failure samples. The data is downloaded from CardioGenomics project by Harvard University

“[http://cardiogenomics.med.harvard.edu/groups/proj1/pages/download\\_Hs-idiopathic.html](http://cardiogenomics.med.harvard.edu/groups/proj1/pages/download_Hs-idiopathic.html)” and was generated using Affymetrics HgU133 Plus 2.0 arrays and contains 12 samples each for normal and heart failure patients.

To provide real-time sensor network environment parameters to ACE, we developed a device to emulate the body area network – that includes sensors attached to a patient, the sensors in the environment and a gateway node. The device, named Virtual Patient (VP), comprises of an Atmel AtMega 128 8bit RISC based micro controller. The Device features a 2 line by 16 character standard alpha numeric LCD display. The VP also includes 1 Promi ESD blue tooth module. The Blue tooth module presents as an AT compatible modem and is controlled via a series of AT commands in ASCII. The blue tooth module has its own blue tooth stack and can perform discovery as well as be discovered. When connected to a peer device this module provides blue tooth serial services and appears as a COM port.

The next level of emulation is at the home network, which comprises the intermediate nodes in the network and the network server (SNS). In the experimental testbed, we used an HP iPAQ 6915 series PDA for this purpose, which has Microsoft Windows Mobile 5.0 for Pocket PC Phone Edition and GSM/GPRS/EDGE WLAN 802.11b, Bluetooth 1.2, integrated high sensitivity GPS receiver with HP iPAQ Quick GPS connection utility and IrDA. Rest of the framework components include the hospital network that has the diagnosis facilities and the physician’s terminal. We used a Toshiba Tecra series notebook with Intel Core Solo 1.66 GHz T1300 Processor with 1.5GB RAM, IEEE 802.11b, Bluetooth and USB interfaces to run ACE.

The VP was setup to use Bluetooth channel to communicate with the intermediate node (PDA), while WIFI was used between the PDA and the Genome & Network Server (notebook).

**Table 1: Clinical Data Attributes**

Feature ID	Clinical Parameters	Unit
1	Age	Years
2	Gender	Value 0: Female Value 1: Male
3	Chest Pain Type	Value 1: Typical Angina  Value 2: Atypical Angina  Value 3: Non-Angina Pain  Value 4: Asymptomatic
4	Resting Blood	mm hg

	Pressure	
5	Resting Electrocardiographic Results	Value 0: Normal  Value 1: Having ST-T Wave Abnormality (T Wave Inversions And/Or ST Elevation Or Depression Of > 0.05 mv)  Value 2: Showing Probable Or Definite Left Ventricular Hypertrophy By Estes' Criteria
6	Maximum Heart Rate Achieved	
7	Exercise Induced Angina	Value 1: Yes Value 0: No
8	ST Depression Induced By Exercise Relative To Rest	
9	The Slope of The Peak Exercise ST Segment	Value 1: Up Sloping Value 2: Flat Value 3: Down Sloping
10	Number of Major Vessels	(0-3)
11	Serum Cholesterol	mg/dl
12	Fasting Blood Sugar > 120	mg/dl (1 = True; 0 = False)
13	Thal	Value 3: Normal Value 6: Fixed Defect Value 7: Reversible Defect

With the above mentioned devices in place, the software components of the framework were written and deployed in respective devices. The DNA microarray and the corresponding clinical data of 31 normal and diseased patients each were obtained and the Genome server on the notebook was trained. The ACE was programmed in C and MATLAB to diagnose the disease based on the DNA and clinical data and produce a DNA-Risk Factor (in the range of 1-100) depicting the level of a cardiac risk.

### 3.2 Results

Table 2, shows that linear kernel performed better in terms of individual classification accuracies for both genetic and clinical data, tested using  $k$  fold leave one out cross validation. Also, interestingly linear kernel demonstrated better performance while classifying genetic data than clinical data which highlights the importance of the use of genetic information in predicting the heart failure rate. Also, it can serve as a metric of *ComputeWeights* algorithm which should give higher weight to genetic than clinical information due to better prediction ability of genetic data.

To test the overall accuracy of the system, normal samples from both genetic and clinical system were organized into  $k$  folds where each fold contains one sample from genetic and clinical data respectively. The outcome was considered to be correct if the ACE predicted result was normal. Similar organization was used for heart failure data for both genetic and clinical data. The overall accuracy of the system was 87.50%.

**Table 2:** Individual SVM Classification Accuracy

Kernel Type	Genetic	Clinical
<b>Linear</b>	91.67	87.50
<b>Polynomial</b>	91.67	87.89
<b>RBF</b>	50	71.76

## 4. Conclusions

Cardiac patient monitoring can save many lives. Current clinical practices only stress on physiological observation to predict heart failure rate which could miss the important information which eventually could lead to fatal consequences. This paper presents *Ambient Cardiac Expert* (ACE) which combines information from both well established clinical studies and also, microarray gene expression data to predict the heart failure rate. The system uses well established *Support Vector Machines* (SVM) for class prediction. It integrates information from both clinical and genetic data using *Wrapper Evolutionary Algorithm* based on *Gaussian Estimation of Distribution Algorithm* (EDA) to determine cardiac patient's criticality. Results suggest that ACE can be successfully applied for cardiac patient monitoring by integrating the information from both clinical and genetic sources.

### ACKNOWLEDGEMENTS

This project is supported by Monash University, research grant. The authors would also, like to acknowledge NeuNet team for compiling clinical data, originally produced by the authors mentioned in Section 3.

## References

- [1] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *Science*, pp. 286(5439):531-537, 1999.
- [2] M. A. Shipp, K. N. Ross, P. Tamayo, A. P. Weng, J. L. Kutok, R. C. Aguiar, M. Gaasenbeek, M. Angelo, M. Reich, G. S. Pinkus, T. S. Ray, M. A. Koval, K. W. Last, A. Norton, T. A. Lister, J. Mesirov, D. S. Neuberg, E. S. Lander, J. C. Aster, and T. R. Golub, "Diffuse large B-cell lymphoma outcome prediction by gene expression profiling and supervised machine learning," *Nat Med*, vol. 8(1), pp. 68-74, 2002.
- [3] S. L. Pomeroy, P. Tamayo, M. Gaasenbeek, L. M. Sturla, M. Angelo, M. E. McLaughlin, J. Y. Kim, L. C. Goumnerova, C. L. P. M. Black, J. C. Allen, D. Zagzag, J. Olson, T. Curran, C. Wetmore, J. A. Biegel, T. Poggio, S. Mukherjee, R. Rifkin, A. Califano, G. Stolovitzky, D. N. Louis, J. P. Mesirov, E. S.

- Lander, and T. R. Golub, "Prediction of central nervous system embryonal tumour outcome based on gene expression," *Nature*, vol. 415(24), pp. 436-442, 2002.
- [4] A. Bhattacharjee, W. G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E. F. Mark, E. S. Lander, W. Wong, B. E. Johnson, T. R. Golub, D. J. Sugarbaker, and M. Meyerson, "Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses," *Proc. Natl. Acad. Sci.*, pp. 13790-13795, 2001.
- [5] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C. H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander, and T. R. Golub, "Multiclass cancer diagnosis using tumour gene expression signatures," *Proc. Natl. Acad. Sci.*, pp. 98(26):15149-15154, 2001.
- [6] M. P. S. Brown, W. N. Grundy, N. C. D. Lin, C. Sugnet, T. S. Furey, M. Ares, and D. Haussler, "Knowledge-based analysis of microarray gene expression data using support vector machines," *Proc Natl. Acad. Sci.*, pp. 262-267, 1997.
- [7] S. Mukherjee, P. Tamayo, D. Slonim, A. Verri, T. Golub, J. P. Mesirov, and T. Poggio, "Support vector machine classification of microarray data," *Technical Report, Artificial Intelligence Laboratory, Massachusetts Institute of Technology*, 2000.
- [8] E. Byvatov and G. Schneider, "Support vector machine applications in bioinformatics," *Applied Bioinformatics*, vol. 2, pp. 67-77, 2003.
- [9] T. Evgeniou, M. Pontil, and T. Poggio, "Regularization networks and support vector machines," *Advances in Computational Mathematics*, vol. 13, pp. 1-50, 2000.
- [10] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121-167, 1998.
- [11] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*: Cambridge University Press, 2000.
- [12] Q. Lu and X. Yao, "Clustering and learning Gaussian distribution for continuous optimization," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 2, pp. 195-204, 2005.