CAD DETECTION USING NEURAL NETWORK FUSION OF THE 12 LEAD STRESS ECG SYSTEM

Galal M. BinMakhashen*, Samer M. A. Arafat*, and Mary L. Dohrmann**
* King Fahd University of Petroleum and Minerals
Dahran, Saudi Arabia
Binmakhashen@kfupm.edu.sa; Arafat@kfupm.edu.sa
** University of Missouri Hospital Cardiovascular Medicine Clinic
Columbia, Missouri, USA
Dohrmannm@health.missouri.edu

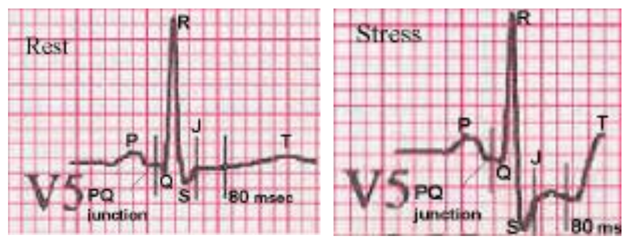
ABSTRACT
In this paper, we develop and test a system for integrating transformed information of the 12 lead stress ECG signals, at the classifier real-valued output level. A coronary artery disease data set was collected and utilized in this study. Four types of features were extracted using the discrete cosine transform, two levels of the discrete wavelet transform, and dimensionality-reduced data using principle component analysis. For each feature type, 12 neural networks were trained and tested using the backpropagation algorithm. Several experiments have been conducted to test this system. Results have demonstrated superior performance when using a fusion of 12 classifier output values, compared to single lead classifier systems. We observed that a 3-level discrete wavelet transform has computed 95-100% performance success rates, using sensitivity, specificity, or accuracy.

KEY WORDS
12 lead ECG system, coronary artery disease, classifier fusion, discrete wavelet transform, discrete cosine transform, principal component analysis, neural networks, pattern recognition.

1. Introduction

According to the American Heart Association, each year about 295,000 emergency medical services-treated out-of-hospital cardiac arrests occur in the US, alone. This fact suggests a need for an emergency automated system for the detection of heart disease.

The electrocardiogram (ECG) is an important biomedical signal that is used to understand the heart activity. The exercise stress ECG is used as an indirect method for assessing individuals for limitation of coronary blood flow due to obstructive coronary artery disease (CAD). The reliability of the changes in the exercise ECG in predicting patients who have significant coronary disease is dependent on the prevalence of the disease in the population being tested [1]. The repolarization phase of the ventricular depolarization, the ST segment, is used for assessment of ischemia. Traditionally, one millimeter of depression of the ST segment from the baseline lasting for 80 milliseconds following the J point (end of the ventricular depolarization or QRS complex) and occurring in at least three consecutive QRS complexes is indicative of ischemia, when comparing stress to rest ECG signals (see Figure 1).



(A) (B)

Figure 1. A normal ECG beat is shown at rest in (A), and an abnormal beat using stress ECG with horizontal ST depression that is 2 mm from the PQ segment is shown in (B)

Research publications have been conducted to detect abnormality in the ECG signals such as [2][3][4]. A 12 lead ECG system is a standard measurement of the heart activity using 12 leads. These leads are divided into two main groups: limb group (I, II, III, VF, VL, VR) and precordial group (v1, v2, v3, v4, v5, v6). The electrodes that correspond to the leads are attached to specific points on the patient body in order to record the electric activity in the heart. According to [5], there exists a relation between the individual 12 leads that are utilized to generate the entire set of 12 lead signals, such a subset of leads may be used to generate a diagnosis. There are research works that have addressed this subject, such as [6][7][8].

Information fusion has been successfully applied to pattern recognition problems such as biometric based identification, where the ultimate goal is to identify a person based on his or her physiological characteristics [9]. Another domain for information fusion is seizure detection. For example, research was been conducted with the objective of enhancing seizure recognition accuracy by integrating ECG and EEG signals [10][11].

Information fusion may be implemented using at least one of four different processing levels: raw data,

DOI: 10.2316/J.2014.216.818-0080

extracted feature vectors, matching or classification output, and/or the final labeling decision(s). At each level of fusion, the amount of information available for fusion decreases, as shown in Figure 2. Suppose we want to do a fusion of ECG lead signals at level 1 (raw data), then the amount of information available at this level would be the fully recorded raw signals [9]. If we consider the fusion at level two (feature vectors), then we would be combining together two (or more) extracted feature vectors. Fusion at level three combines together continuous values that correspond to output units, as detailed in Section 2.4, below. Finally, a level 4 fusion may be implemented using a logical operator on a single bit domain, i.e., 0 or 1. Figure 2 illustrates the relation between the levels of fusion, in terms of information size [12].

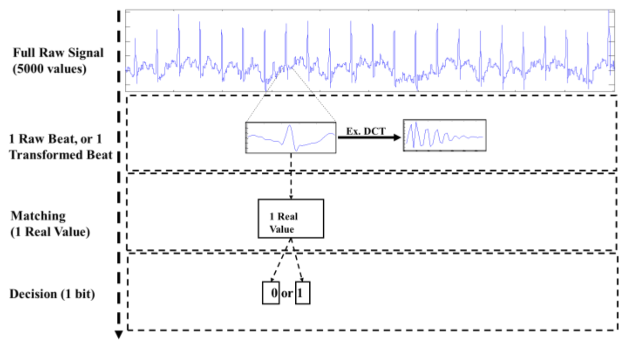


Figure 1. Information fusion levels

In this study, a CAD detection system using stress ECGs is introduced. The novelty of the system is comprised of the integration of the 12 lead ECG signals on the matching level. We have considered this level of fusion because the integration at this level is inexpensive, computation-wise, compared to earlier levels (raw data or feature vectors), and yet there is still a fairly enough amount information to be considered.

The rest of the paper is organized as follows: Section 2 describes the methodology. Experiments and results are introduced and discussed in Section 3, and finally, Section 4 concludes this paper.

2. Methods

The integration of the 12 lead stress ECG for CAD detection involves four main components: preprocessing and segmentation, feature extraction, model integration, and classification. Figure 3 illustrates an overview of the complete system with two phases. The first phase starts with acquiring the stress ECG signal bank using the standard 12 lead stress ECG system. The lead raw signals are fed into the preprocessing stage in order to improve signal quality by removing noise that may be presented to the signal. Next, the segmentation stage segments and selects one heart beat signal for every lead. In the third stage, the segmented ECG signal is passed through feature extraction in order to extract features that characterize beat information so that it becomes feasible for a trained classifier to distinguish between normal and pathological data. Next, a model selection algorithm determines an optimum neural network structure that would be used to train and test corresponding lead data. Next, the evaluation stage trains and tests the fusion of all neural network models used for each lead signal. Finally, neural networks are tested using test data. The remainder of this section describes system components in detail.

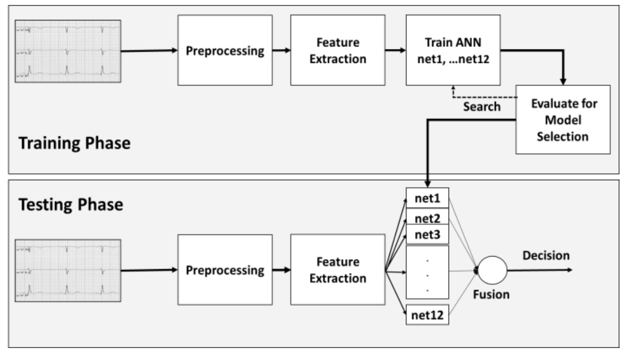


Figure 2. System overview

2.1 Database

A data set consisting of 65 patients was selected from a database of ECG stress test patient data that was collected during years 2002-2004 at the University of Missouri Hospital Cardiology Division. Patient ages ranged from 38 to 77. Most patients were men. ECG signals were selected such that 33 of the group were labeled normal and 32 abnormal.

ECG signals were labeled normal if their corresponding patients had no known history of coronary disease and if their rest and stress test ECG signals were both normal (no ST segment changes with exercise).

EGG signals were labeled abnormal if two conditions were met. First, their stress test EGG signals were abnormal, in comparison to rest, demonstrating at least 1 mm horizontal ST segment depression that lasted for 80 msec. Secondly, a patient must have had a prior history in at least one of the following three events: myocardial infarction, percutaneous coronary intervention, or coronary bypass surgery.

2.2 Pre-Processing

The targets of the pre-processing stage are to detect the QRS complex and to remove both baseline drifting and noise. For QRS detection, we have adopted the cubic spline technique (CST) to find the knots of the QRS complex [13]. In this technique, 3 points have to be determined: the starting point, endpoint, and the Fiducial point (the inflection point leading to the Q point valley) of each QRS complex. Next, CST builds a synthetic wandering curve by fitting the 3 points to a (third order) polynomial function.

To eliminate baseline wandering, CST computes the deference between the original ECG signal and the synthetic wandering curve. Figure 4 illustrates a few examples that show successful removal of the baseline wandering using CST.

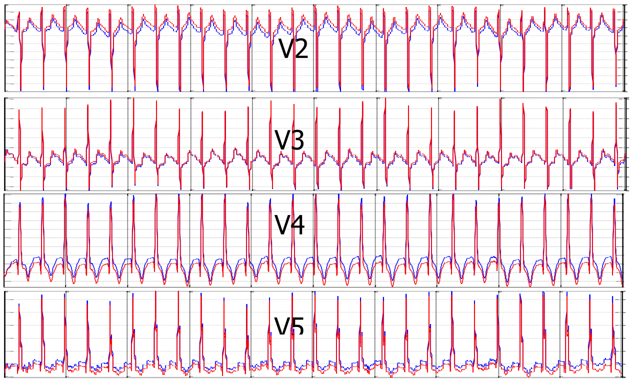


Figure 3. Examples that demonstrate the successful removal of the baseline drifting

In order to filter out noise, wavelet transformations are applied to ECG signals. A wavelet is a small signal or an excursion that has a limited time duration. We used a *Symlet* wavelet for a discrete wavelet transformation (DWT). DWT filters the ECG signal into low and high frequency parts that are also called approximation and detail components, respectively. More information and details about DWT are covered in Subsection 2.3.2, below.

In the pre-processing stage, DWT is used in order to remove high frequency noise, using the following three steps. First, DWT transforms an ECG signal using a low-pass filter (g) and the high-pass filter (*h*). Next, the low-pass filter response is kept for the reconstruction of the ECG signal, while the high-pass filter response (high frequency noise) is ignored. Finally, we took the inverse of the low-pass filter response, and this last response signal is treated as the new, reconstructed and filtered ECG signal.

Having removed the baseline wandering, lead signals are expected to have beats that are fitted to the baseline. We have adopted a strategy to pick one beat for the next stage of computation. In order to avoid the first and last few beats that may have some extra noise components, we have consistently picked the 5th beat in every ECG lead signal.

It is imported to point out that we had to deal with leads that have negative values. Basically, the pre-processing procedure has the same processing over all leads except for leads v1, v2, aVR and aVL, as shown in Figure 5. These leads have mostly negative values, and this indicates that the QRS peaks are inverted along the baseline. In order to deal with this computational exception, we inverted leads v1, v2, aVR and aVL signals before applying the same pre-processing procedure, described, above.

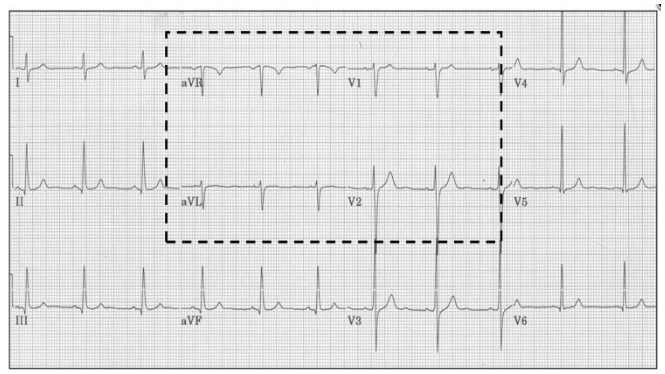


Figure 4. Negative-valued leads

2.3 Feature Extraction

This subsection elaborates on the two transformations used to extract features from ECG signals, namely, the discrete cosine and the discrete wavelet transforms. Moreover, we discuss principal component analysis, a known algorithm for dimensionality (features) reduction that we used in a separate set of experiments.

2.3.1 The Discrete Cosine Transform (DCT)

DCT is a well-known technique in signal and image processing. Basically, it has the power to compute signal energy, which is one of the most important characteristics of input signal information [14]. Moreover, research has shown that the first 20% of the DCT transformation has sufficient information that can be used to reconstruct the entire original signal [15]. Figure 6 shows as an example of one patient ECG signal's DCT.

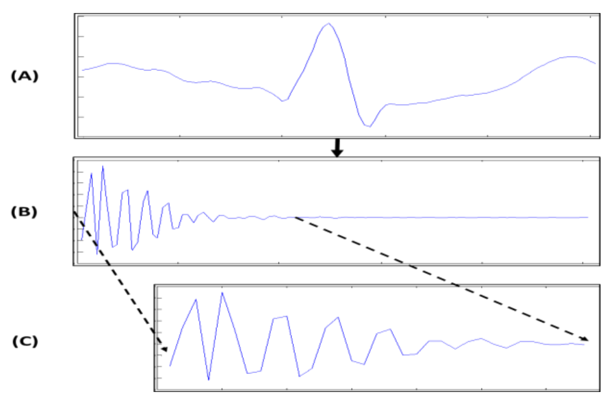


Figure 5. DCT Feature Extraction: A) the original ECG signal, B) the full DCT transformation, or y(k), C) The final extracted DCT feature values (signal)

DCT uses the following formula to compute its response (i.e. transformation), y(k):

$$y(k) = \omega(k) \sum_{n=1}^{N} x(n) \cos\left(\frac{\pi(2n-1)(k-1)}{2N}\right)$$

where $n = 1, 2, 3, \ldots N$, x(n) is the original ECG signal, and $\omega(k)$ are the DCT coefficients

$$\omega(k) = \begin{cases} \frac{1}{\sqrt{N}} & if \quad k = 1 \\ \sqrt{\frac{2}{N}} & if \quad 2 \le k \le N \end{cases}$$

2.3.2 The Discrete Wavelet Transform (DWT)

The discrete wavelet transform (DWT) is a known signal analysis technique that has been successfully applied in several research works for the extraction of discriminant signal features [16][17]. DWT is computed by convolving (and effectively, down sampling) the ECG signal using a *mother wavelet*, at a dyadic scale. A mother wavelet is the wavelet that is being considered at a certain transform computation. A dyadic scale refers to the down sampling scale and is computed using the base 2, such as $(2^1, 2^2, \ldots 2^n)$. Therefore, candidate scale values would be: 2, 4, 8, 16, 32, …etc.

The convolution of the signal with the DWT is defined as follows [16]:

$$Sig_{low}\ (k) = \sum_{i=1}^{n} x(i) * h(-i + 2 \times k)$$

$$Sig_{high}\ (k) = \sum_{i=1}^{n} x(i) * g(-i + 2 \times k)$$

where * is the convolution operator, x(.) is the signal and g(.) and h(.) are the low-pass and high-pass filters, respectively. Finally, the ψ wavelet is computed by dilating and shifting the mother wavelet:

$$\psi(a,b)\ (t) = 1/\sqrt{a} \ \times \psi((t-b)/a)$$

where a and b are the scaling and shifting parameters, respectively.

The wavelet is convolved with the original ECG signal in order to compute transformation coefficients that emphasize portions of the signal that have similar wavelet structure and characteristics. Transformation coefficients also deemphasize the other signal portions that are dissimilar to the wavelet. Therefore, the wavelet transformation coefficient values measure the similarity between a wavelet and signal parts [18]. Due to the similarity nature in the transformation, the output values of the wavelet transformation heavily depend on the mother wavelet selection [18]. For all of our DWT feature extraction, we have chosen the *symlet* 12-tap wavelet (shown in Figure 7), mainly because of its structure that is similar to a normal QRS complex structure.

In our experiments, we have considered two levels of DWT transformations via two passes. During the first pass, the ECG signal is filtered into the approximation and detail coefficient signals. We selected the first level approximation components to be our feature vector, named DWT1, that will be used in the classification stage, next.

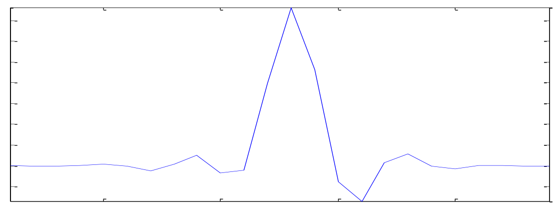


Figure 7. Symlet 12-tap wavelet

Bushra et al. [19] recommended that level 8 approximation components are a better choice for an ECG dataset. Therefore, we repeated the filtering process using the approximation components from level 1 into level 2 (scale value 4) in order to feed into the third level (scale value 8). Next, the approximation components of level 3 are named the DWT2 feature vector. Figure 8 illustrates the multi-level DWT feature extraction.

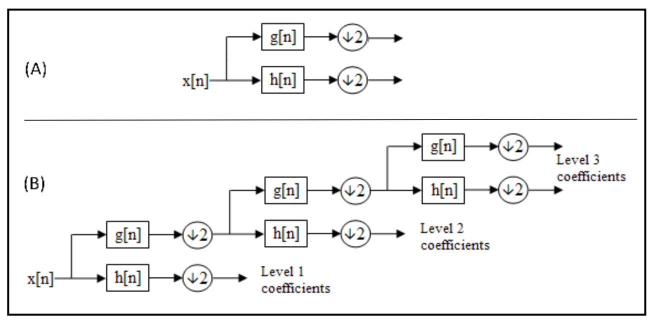


Figure 8. Multi-level DWT feature extraction: A) Level one, only ; B) Levels one, two, and three

2.3.3 Principal Component Analysis

Principal component analysis (PCA) is a mathematical method that stems from linear algebra and has been successfully applied to dimensionality reduction and/or feature selection of data that have correlated dimensions. PCA computes the uncorrelated variables of all the data using the correlation matrix as follow:

$$R = E(X.X^T)$$

where E is the expectation function and X is the full data matrix. So the reduced space using PCA is computed as:

$$\bar{X} = X.\lambda\_P$$

where λ_P is a subset of eigenvectors computed using the following formula:

$$R.\lambda\_q = \lambda\_q.v$$

λ_q is the eigenvector and q∈p , and v is the corresponding eigenvalue.

In a separate set of experiments, we compute PCA for the pre-processed ECG beats in order to compute PCA-reduced feature vectors, as described in Section 3, below.

2.4 Information Fusion and Classification

In order to classify patient data into either normal or abnormal ECG signals, we have integrated 12 multi-layer perceptron (MLP) classifiers that were trained using the backpropagation algorithm. MLP is a common artificial neural network technique that has been frequently used for classification problems. Although MLP has a mathematically sound algorithm – backpropagation - that computes the gradient at each step in order to approach a global minimum error, it is considered sometimes to be like a black-box classifier. This is due to the selection of several MLP learning parameters, such as the selection of training subset samples, network topology (number of hidden layers), the number of neurons for each layer, activation functions, and the learning algorithm. Such parameters and/or functions may not necessarily compute the same target discriminant boundary for the same problem, every time. However, once the overall classification system efficiently determines the MLP learning parameters, the system will reuse the parameters on the new testing data in order to distinguish between normal and abnormal signals, efficiently.

The MLP network topology was set to be the same for all ECG leads in all experiments. The names of the 12 MLPs correspond to the numbering of the 12 leads. For example, the system uses MLP1 for a neural network that classifies lead I data, MLP2 for lead 2 data, and so on. Experimentally, the MLP structure has been determined to consist of three layers: 1 input, 1 hidden, and 1 output layer. On the other hand, experiments have been carried out using different parameters for different MLPs, in order to target an enhanced classification of different lead signals. Figure 2, above, illustrates the combination of MLPs at the testing phase.

The algorithm directs the system to select its parameters from a predefined set of values as tabulated in Table 1.

The output of each MLP may be interpreted as a score value that predicts whether a given input belongs to the positive class (i.e. the disease). The higher the value of the score, the higher the probability that the input (transformed) signal exhibits an abnormal pattern.

There are several information integration metrics that have been studied, previously. The average sum rule is among the strong statistically sound information fusion technique that was proven to combine information efficiently [12]. Therefore, we have adopted an average sum rule fusion in order to combine the output of the 12 MLPs into a single value that is named the Fusion value.

Table 1
MLPs set of parameters

H.layer Neurons	***Act. Fun(H)***	***Act. Fun (Out)***	***Training Alg.***
10	Purelin	Tansig	Trainlm
	Tansig	Purelin	Traincgf
	Purelin	Logsig	Trainlm
	Logsig	Tansig	Traincgf
20	Purelin	Tansig	Trainlm
	Tansig	Purelin	Traincgf
	Purelin	Logsig	Traincgf
	Logsig	Purelin	Trainlm
	Logsig	Purelin	Traincgf
	Logsig	Tansig	Traincgf

The system computes the Fusion value as follows:

$$Fusion = \frac{1}{n}\sum_{i=1}^{n} y_i$$

where y_i is the prediction score of MLP_i , and i =1, 2 … 12.

The fusion score is fed to a one-bit threshold decision unit, i.e., has either one of 2 values: {0, 1}, using the following standard threshold rule:

$$D = \begin{cases} C_1, & Fusion \geq 0.5 \\ C_2, & Otherwise \end{cases}$$

where *D* is the system decision, *C1* and *C2* are the normal and abnormal classes, respectively. A decision with a 0 value implies absence of the disease, whereas a decision with value 1 implies presence of the disease.

3. Experiments and Results

This section shows and discusses the experiments and results. Every lead has 33 ECG samples representing the positive (CAD-labeled) class and 32 samples representing negative class. To train and test the system, we have divided the data into 70% training and 30% testing segments. The reported performance is computed using only the testing data.

We have done several experiments using four types of feature vector sets: DCT, DWT1, DWT2, and PCA as discussed, above. The dimensionality of each type is different. For the DCT feature type, the algorithm selected only 33 out of a hundred coefficients because around the third of the DCT response can be used to reconstruct the original signal, as illustrated in Figure 6, above. Both DWT1 and DWT2 computed 62 coefficients, each. These numbers were picked due to down sampling a given beat DWT response signal into an approximation and detail. Only the approximation is kept for the training and testing stage, as explained, above. PCA selects only a set of pre-processed data features ranging from 6 to 9 points for each beat. In this sense, at least 90% of the information variance was retained.

Figure 9 illustrates the DCT-related results. Lead-v5 shows the best results in comparison to the other single leads taken one at a time. The MLP performance using the fusion of 12 multiple leads has generally outperformed the lead v5 computed performance, or was at least matching it in one case for DCT data case, as shown, below. This demonstrates the significance of using the fusion technique

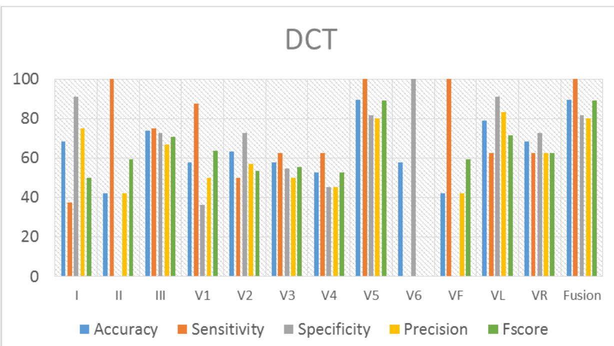


Figure 6. DCT performance

Figures 10 and 11 illustrate DWT1 and DWT2 related results. It can be observed that the performance using DWT1 and DWT2 on lead-v5 is better than those corresponding to other leads. Fusion in both cases has a higher performance compared to using any single lead, alone.

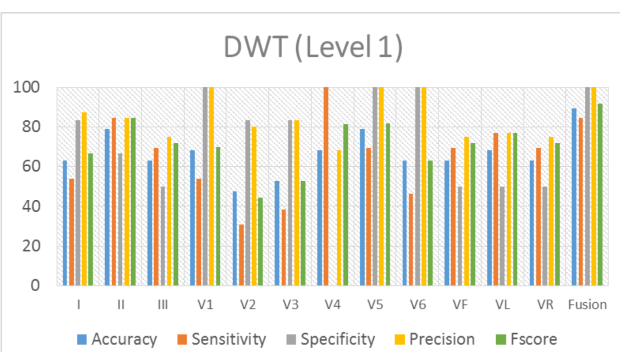


Figure 7. DWT1 performance

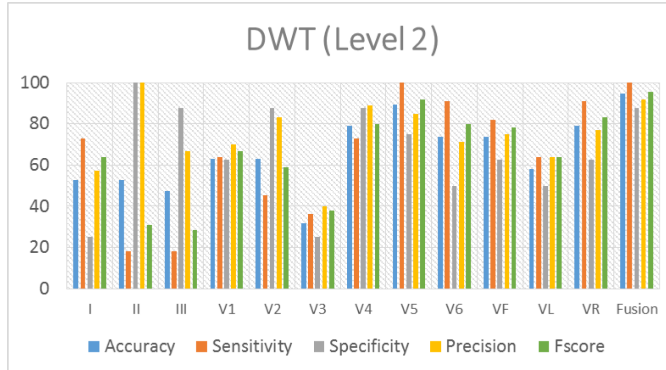


Figure 8. DWT2 performance

Figure 12 shows that using PCA on pre-processed data as described, above, lead-v6 has better performance over other leads. However, lead-v6 has poor specificity that reduces its general performance. Our interpretation of the fluctuation in the results of v5 and v6 leads for this PCA data set is that there is a measurable effect that is due to the reduction of the extracted feature vectors dimensionality. This reduction appears to have caused the loss a certain amount of information that was significantly needed during the training process.

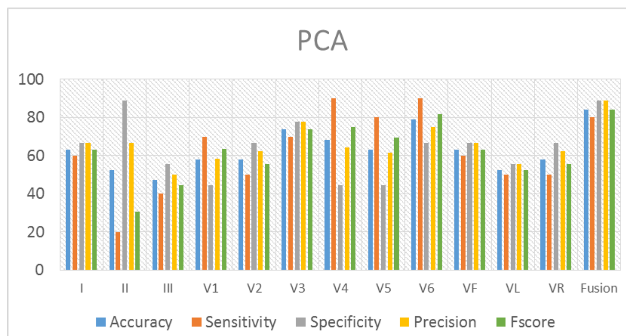


Figure 9. PCA performance

Overall, Fusion has a stable performance, compared to using data from any other single lead. Its sensitivity and specificity have almost identical high values. Therefore, we can say that the fusion of the 12 lead data at the matching level better models the complex dynamics of the CAD data set classification.

4. Conclusion

Experimentally, this paper has shown that the fusion of real-valued classifier outputs establishes a performance upper bound, compared to that of classifiers that correspond to individual leads in a 12 lead system. Results have shown that a combination of classifiers corresponding to 12 lead ECG recordings computes either an improved system performance or a performance that is at least comparable to classifiers corresponding to individual leads, using a coronary artery disease data set. Moreover, a fusion model is relatively more stable and more consistent when performance was measured using specificity, accuracy and sensitivity.

Wavelets have shown strength in extracting beat information that is similar to a mother wavelet. Of special interest was the 3-level discrete wavelet transform that has computed 95-100% success rates using sensitivity, specificity, or accuracy.

On the other hand, dimensionality reduction appears to show some risk and therefore, caution is advised when reducing data dimensionality.

Results also confirmed that v5 is a very important lead that may be used for CAD diagnosis, especially if a system is to base its diagnosis using only one lead.

Future work includes doing more experimentation using other data, such as arrhythmia, building a mathematical support for using a fusion of classifier real valued output, and exploring other feature extraction strategies.

Acknowledgements

The authors would like to acknowledge the support provided by the Deanship of Scientific Research at King Fahd University of Petroleum & Minerals (KFUPM).

Moreover, the authors would like to acknowledge the faculty and staff at the University Hospital Cardiology Clinic of the University of Missouri Hospital, for the help they offered in acquiring, organizing, and establishing the annotation of the MU ECG CAD data set.

References

[1] G. Diamond and J. Forester, "Analysis of Probability as an aid in the clinical diagnosis of Coronary-Artery Disease," New England Journal of Medicine, vol. 300, pp. 1350-1358, 1979.

[2] B. Vuksanovic and M. Alhamdi, “ECG based system for arrhythmia detection and patient identification,” in Proceedings of the ITI 2013 35th International Conference on Information Technology Interfaces (ITI), 2013, pp. 315–320.

[3] C. Ye, B. V. K. Vijaya Kumar, and M. T. Coimbra, “Combining general multi-class and specific two-class classifiers for improved customized ECG heartbeat classification,” in 2012 21st International Conference on Pattern Recognition (ICPR), 2012, pp. 2428–2431.

[4] A. Li, S. Wang, H. Zheng, L. Ji, and J. Wu, “A novel abnormal ECG beats detection method,” in 2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE), 2010, vol. 1, pp. 47–51.

[5] J. Malmivuo and R. Plonsey, Bioelectromagnetism Principles and Applications of Bioelectric and Biomagnetic Fields. Oxford University Press, 1995.

[6] J. A. Scherer and J. L. Willems, “Evaluation of 12-lead ECG synthesis using analysis measurements in 240 patients,” in Proceedings of Computers in Cardiology 1992, 1992, pp. 91–94.

[7] I. Tomasic and R. Trobec, “Electrocardiographic systems with reduced numbers of leads - synthesis of the 12-lead ECG,” Biomed. Eng. IEEE Rev. In, vol. Early Access Online, 2013.

[8] H. Atoui, J. Fayn, and P. Rubel, “A neural network approach for patient-specific 12-lead ECG synthesis in patient monitoring environments,” in Computers in Cardiology, 2004, 2004, pp. 161–164.

[9] F. Agrafioti and D. Hatzinakos, “Fusion of ECG sources for human identification,” in 3rd International Symposium on Communications, Control and Signal Processing, 2008. ISCCSP 2008, 2008, pp. 1542–1547.

[10] B. R. Greene, G. B. Boylan, R. B. Reilly, P. de Chazal, and S. Connolly, “Combination of EEG and ECG for improved automatic neonatal seizure detection,” Clin. Neurophysiol., vol. 118, no. 6, pp. 1348–1359, Jun. 2007.

[11] W. Deburchgraeve, P. J. Cherian, M. De Vos, R. M. Swarte, J. H. Blok, G. H. Visser, P. Govaert, and S. Van Huffel, “Automated neonatal seizure detection mimicking a human observer reading EEG,” Clin. Neurophysiol., vol. 119, no. 11, pp. 2447–2454, Nov. 2008.

[12] A. Ross, N. Karthik, and A. K. Jain, Handbook of Multibiometrics, vol. Vol. 6. Springer Berlin Heidelberg, 2006.

[13] M. Arif, I. A. Malagore, and F. A. Afsar, “Automatic Detection and Localization of Myocardial Infarction Using Back Propagation Neural Networks,” in 2010 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE), 2010, pp. 1–4.

[14] E.-S. M. El-Alfy and G. M. BinMakhashen, “Improved Personal Identification Using Face and Hand Geometry Fusion and Support Vector Machines,” in Networked Digital Technologies. Springer Berlin Heidelberg, 2012, pp. 253–261.

[15] V. A. Allen and J. Belina, “ECG data compression using the discrete cosine transform (DCT),” in Proceedings of Computers in Cardiology 1992, 1992, pp. 687–690.

[16] R. J. Martis, U. R. Acharya, and L. C. Min, “ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform,” Biomed. Signal Process. Control, vol. 8, no. 5, pp. 437–448, Sep. 2013.

[17] P. Gokhale, “ECG Signal Denoising using Discrete Wavelet Transform for removal of 50Hz PLI noise, "Int. J. Emerg. Technol. Adv. Eng., vol. 2, no. 5, pp. 81-85, 2012.

[18] S. Arafat and M. Skubic, “Modeling Fuzziness Measures for Best Wavelet Selection,” *IEEE Transactions on Fuzzy Systems*, October 2008.

[19] J. Bushra, L. Olivier, F. Eric, and B. Ouadi, “Detection of QRS complex in ECG signal based on classification approach,” in 2010 17th IEEE International Conference on Image Processing (ICIP), 2010, pp. 345–348.