Wavelet Feature Selection Approach for Heart Murmur Classification

G. Venkata Hari Prasad, P. Rajesh Kumar

***Abstract*—**Phonocardiography is important in appraisal of congenital heart disease and pulmonary hypertension as it reflects the duration of right ventricular systoles. The systolic murmur in patients with intra-cardiac shunt decreases as pulmonary hypertension develops and may eventually disappear completely as the pulmonary pressure reaches systemic level. Phonocardiography and auscultation are non-invasive, low-cost, and accurate methods to assess heart disease. In this work an objective signal processing tool to extract information from phonocardiography signal using Wavelet is proposed to classify the murmur as normal or abnormal. Since the feature vector is large, a Binary Particle Swarm Optimization (PSO) with mutation for feature selection is proposed. The extracted features improve the classification accuracy and were tested across various classifiers including Naïve Bayes, kNN, C4.5, and SVM.

***Keywords*—**Phonocardiography, Coiflet, Feature selection, Particle Swarm Optimization.

I. INTRODUCTION

PHONOCARDIOGRAPHY is a technique used for tracing heart sounds and recording of cardiac acoustic vibration by a microphone-transducer. To understand the phonocardiography signal, it is essential to develop a competent tool for analysis and processing which can increase and optimize cardiac clinical diagnostic approach [1]. Waveforms of the heart sounds are recorded by Phonocardiography. As a reference to phonocardiography [2], [3], ECG is also measured.

The auditory signals produced from a phonocardiogram can be classified into four heart sounds [4] occurring due to
- Flow of blood from atria to ventricles
- When semilunar valves close
- Cessation of Ventricular filling
- Contraction of atria

First Heart Sound (Lub): Blood flows from atria into ventricles on closure of mitral and tricuspid valves. This process happens at the end of atrial contraction and start of the ventricular contraction. It happens approximately 0.05 second after onset of QRS complex and before ventricular systole. It has deep pitch, loud and is booming in character. It has a longer duration, lower in frequency (30 - 45Hz) and bigger intensity than the second sound. Duration is 50 to 100msec. It is best heard at apex of mid pericardium (auscultatory area) [5].

Second Heart Sound (Dub): It occurs at end of ventricular systole when semilunar valves (aortic and pulmonary aortic valves) close, in the arteries leading out of ventricles. It happens at 0.03-0.05 sec after the end of T wave. Its pitch is higher than the first sound. Frequency is 50 - 70Hz. Duration is 25 to 50msec. It is best heard in aortic and pulmonary areas (auscultatory area) [6].

Third Heart Sound: It is heard due to cessation of ventricular filling. It is heard in children and patients who have left ventricular failure because of rapid inflow of blood from atria into the ventricles. The accrued blood from atria and veins causes distention and vibration of ventricles. The frequency is below 30Hz. Duration is 0.1 to 0.2sec. It happens 0.12–0.18 sec after the onset of second heart sound. It is best heard, after lifting the legs, at apex and left lateral position (auscultatory area).

Fourth Heart Sound or Atrial Heart Sound: Produced by the contraction of atria. Not audible as it has low amplitude and frequency of vibrations. It happens immediately before the first heart sound starting 0.12 – 0.18 second after the P wave. Duration is around 0.03 to 0.06 second. The frequency is 10-50Hz.

Phonocardiogram has been successfully used to detect various heart conditions including [7]:
- Detection of Rheumatic Valvular Lesions: Rheumatic valvular lesions occur because of Rheumatic fever, an autoimmune or allergic disease where the heart valves are damaged or destroyed.
- Murmur of Aortic Stenosis: When the blood ejects from left ventricle through the aortic valve, due to resistance to the ejection, pressure in the left ventricle rises to as high as 350mm of Hg. This turbulent blood impinging produces a loud murmur. The sound can be heard quite a few feet away from the patient.
- Murmur of Mitral Regurgitation: During systole, blood flows backward through mitral valve.
- Murmur of Aortic Regurgitation: Here, the blood flows backward from aorta into the left ventricle which causes "blowing murmur", during diastole.
- Murmur of Mitral Stenosis: Blood passes with difficulty from left atrium into the left ventricle because of the pressure difference.

Automatic murmur classification has been extensively studied to aid medical diagnoses. Challenges include the noisy signal generated in the time domain. The process of automated murmur classification starts with feature extraction followed by feature selection and classification. Using the right feature extraction and selection technique is important as sound is a highly complex signal carrying many features mixed together. Other information sources are undesirable noise whose effect

Venkata Hari Prasad G and P. Rajesh Kumar are with Andhra University, Andhra Pradesh, India (e-mail: venkatahariprasadG@gmail.com, dr.rajeshkumarp@rediffmail.com).

World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:9, No:3, 2015

is minimized. The second reason is mathematical and relates to the curse of dimensionality [8], phenomenon implying that number of needed training vectors escalates exponentially with dimensionality. Also, low dimensional representations result in computational and storage savings [8]. Feature extraction transforms signals to a set of feature vectors with the aim of getting a new representation that is more compact, less redundant, and suits statistical modeling and distance calculation or any other score.

With training samples being a finite set, monotonously increasing features adds to computational burden and leads to overfitting training data. Feature selection is indispensable in many applications to get good generalization capabilities. Feature selection methods use subset selection or list of features to construct models describing the original data. It aims to reduce dimensionality, remove irrelevant and redundant features, reduce data needed to learn, improve algorithms predictive accuracy, and increase constructed model's comprehensibility. Feature selection methods explore data intrinsic properties using statistics or information theory.

Feature selection methods are divided into wrapper based, filter based and embedded methods [9]. Filter based techniques select features subsets as a preprocessing step, independent of predictor, attempting to find a predictive features subset by ranking simple statistics computed from empirical distribution, like correlation between input features and class labels. Wrapper based techniques use an algorithm that builds final classifier and searches all feature subsets space. Embedded methods perform feature selection during training and are specific to given learning machines, where search is guided by learning. A wavelet packet feature selection derived using multilayered Neural Network (NN) for speaker signal identification was proposed by Lung [10]. The concept of a multilayered NN without using gradient method was used. Each hidden unit's outputs are algebraically determined by error backpropagation. Then, weight parameters are determined using an exponentially weighted least squares method. Results show that the proposed technique improved performance in terms of recognition percentages.

The aim of this work is developing objective signal processing tools to emphasize and extract information from phonocardiography signals to predict signal as either normal or abnormal. Coiflet is used for feature extraction followed by feature selection using Information Gain (IG). A novel Binary Particle Swarm Optimization (BPSO) with mutation for feature selection is proposed and compared with the statistic feature extraction technique using various classifiers including Naïve Bayes, kNN, C4.5, and SVM.

II. LITERATURE SURVEY

Brennan et al. [11] proposed a novel phonocardiogram technique using mobile phone. Investigations were also carried out using iPhone, and Support Vector Machine was used for classification. Gavrovska et al. [12] proposed various techniques of feature extraction and statistically analyzed the characteristics for pediatric PCG. Singh and Cheema [13]

proposed ECG gating that can be practically implemented in electronic stethoscopes. Sinha and Rakesh [14] used K-Nearest Neighbor approach to classify Phonocardiogram signals. Ramos et al. [15] used segmentation techniques for feature extraction.

III. METHODOLOGY

In this section, a brief introduction to all the methods used is described followed by the proposed feature selection technique.

A. Feature Extraction Using Coiflet

Wavelet function has 2N moments equal to 0 and scaling function has 2N-1 moments equal to 0. Both functions have support of length 6N-1. General characteristics: Compactly supported wavelets with highest vanishing moments for both phi and psi for a support width [16].
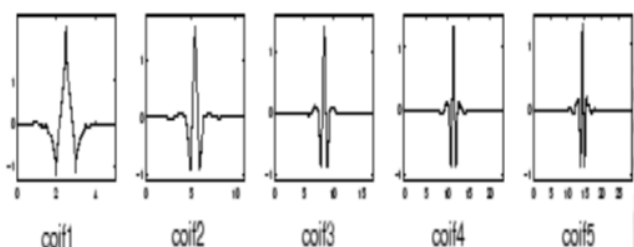

Fig. 1 Coiflet Wavelet Families

Choice of wavelet function is crucial for image compression performance. There are many reasons to decide the choice of wavelet for image compression. As wavelet produces all wavelet functions in transformation through scaling and translation, it determines the resulting wavelet transform's [17] characteristics.

Some families of orthogonal wavelets that are compactly supported are Daubechies and Coiflet wavelets. They correspond to Finite-Impulse Response (FIR) filters leading to efficient implementation. When symmetry and compact support in wavelets are needed, then orthogonality condition should be relaxed to allow nonorthogonal wavelet functions. Biorthogonal wavelets reveal the property of linear phase, needed for signal and image reconstruction [18], using two wavelets, one for reconstruction and the other for decomposition. Connected to sampling problems this property is used, when calculating difference between expansion over a signal and its sampled version instead of same single one when interesting properties are derived. A disadvantage of wavelets is asymmetry, which cause artifacts at wavelet sub-bands borders.

Coiflets are discrete wavelets designed with scaling functions and vanishing moments. The wavelet is near symmetric with N/3 vanishing moments and N/3-1 scaling functions. The function Ψ has 2N moments equal to 0 and function φ has 2N-1 moments equal to 0. Support length of both is 6N-1. The coifN Ψ and φ are considerably symmetric than dbNs. The coifN is compared to db3N or sym3N when considering support length. When considering number of

World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:9, No:3, 2015

vanishing moments of Ψ, coifN is compared to sym2N or db2N. If s is an adequately regular continuous time signal, for large j coefficient $<s,\Phi j,k>$ is approximated by $2^{-j/2} s(2^{-j} k)$. If s is a polynomial of degree d, $d \leq N - 1$, then approximation becomes equal.

Coiflets are Ingrid Daubechies designed discrete wavelets with scaling functions and vanishing moments. The wavelet is near symmetric, wavelet functions have N/3 vanishing moments and scaling functions N/3−1, and was used in applications using Calderon-Zygmund Operators. Both scaling function (low-pass filter) and wavelet function (High-Pass Filter) are normalised by a factor 21. Below are coefficients for scaling functions for C6-30. Wavelet coefficients are derived by reversing the scaling function coefficients and reversing the sign of every second one.

Mathematically,

$$B_k = (-1)^k C_{N-1-k} \qquad (1)$$

where k - coefficient index, B - wavelet coefficient, C - scaling function coefficient, N - wavelet index, i.e. 6 for C6. In this work, feature coefficients are extracted using Coiflet.

B. Information Gain

Information Gain (IG) is a popular statistical technique used for feature selection. IG is information obtained for category prediction by knowing the presence/absence of a document term [19]. Let $m\{C_i\}_{i=1}^m$ denote set of categories in target space. IG of term t is defined by (2)

$$G(t) = -\sum_{i=1}^m p_r(c_i) \log p_r(c_i) +$$
$$p_r(t)\sum_{i=1}^m p_r(c_i \mid t) \log p_r(c_i \mid t) +$$
$$p_r(\overline{t})\sum_{i=1}^m p_r(c_i \mid \overline{t}) \log p_r(c_i \mid \overline{t}) \qquad (2)$$

In a training sample, IG is computed for each feature selected attribute. When IG is less than a predetermined threshold, they are removed from feature space [20], [21].

C. Proposed Feature Selection Technique

PSO is a population-based stochastic optimization technique developed by Kennedy and Eberhart in 1995 [22]. A particle in search space is considered an "individual bird in a flock." To find best solution PSO uses local and global information. Fittest solution is found by using particles fitness function and velocities. Particle position in PSO gets updated using local and global positions of particles around its neighbor. Particles move across problem space searching for optimal particles (features). The process is repeated for a specific number of periods or till a minimum error is attained [23].

A population in PSO is called a swarm, with candidate solutions encoded as particles in search space. PSO starts with a particles population random initialization. The whole swarm moves in search space to locate best solution by updating each particle's position based on experience of own and neighboring particles. During movement, current particle

position i is represented by a vector $x_i = (x_{i1}, x_{i2}, ... x_{iD})$, where D is search space dimensionality. Particle velocity i is represented as $v_i = (v_{i1}, v_{i2}, ..., v_{iD})$, limited by a predefined maximum velocity, v_{max} and $v_{id}^t \in [-v_{max}, v_{max}]$. A particle's best previous position is recorded as personal best pbest, and best position obtained by population so far is called gbest. From the values of pbest and gbest, PSO searches for an optimal solution by updating velocity and each particle's position according to:

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1} \qquad (3)$$

$$v_{id}^{t+1} = w*v_{id}^t + c_1*r_{1i}*\left(p_{id} - x_{id}^t\right) + c_2*r_{2i}*\left(p_{gd} - x_{id}^t\right) \qquad (4)$$

where t denotes t^{th} iteration and $d \in D$ denotes d^{th} dimension in search space, w is inertia weight, r_{1i} and r_{2i} are random values uniformly distributed in [0, 1], c_1 and c_2 are acceleration constants, p_{id} and p_{gd} represent pbest and gbest elements in d^{th} dimension.

Kennedy and Eberhart introduced a PSO binary version. But, PSO binary version had problems jumping out of good local optima. This was handled by Binary PSO with Bit change mutation [24] whereas in binary PSO, particle's personal best and global best are updated as in continuous PSO. The difference between binary PSO and continuous PSO is that particles velocity uses probability that a bit takes on 0 or 1. Using this definition, a velocity is restricted within a range [0, 1]. In BPSO, equation updating velocity remains unchanged, but that below for updating a position is re-defined by (5)

$$If \left(rand() < S\left(v_n(t+1)\right)\right) then \ x_n(t+1) = 1$$
$$Else \ x_n(t+1) = 0 \qquad (5)$$

where S (.) is a sigmoid function used to transform velocity to probability constrained to interval [0, 1] and rand () is a random number selected from interval [0, 1]. The sigmoid function is given by

$$F(sigmoid) = \frac{1}{1+e^{-x}} \qquad (6)$$

When BPSO starts iteration to find an optimum solution, velocity tends to go into v_{max} or $-v_{max}$ by velocity update equation according to the corresponding target position which is one or zero, respectively. If velocity converges near v_{max} or $-v_{max}$, it is hard to change corresponding position with a small velocity variation making it tough to escape from a good local BPSO optimum. To handle this undesired position, large velocity movement is needed, which is unrelated to p_{best} and g_{best}. To accomplish this, the following is inserted between velocity update and position update in BPSO process [25].

World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:9, No:3, 2015

$$If \quad rand() < r_{mute}) \text{ then } v_n(t+1) = -v_n(t+1) \qquad (7)$$

where r_{mute} is a probability. If this is executed when velocity is near v_{max} and $-v_{max}$, then position changes from one to zero or zero to one, respectively. Features subsets are encoded as initial population randomly. The algorithm is iterated, and population fitness evaluated. Once the g_{best} is found, random mutation is initiated to find better solutions to the current solution. If better solutions are found then $g_{bes\_mutatet}$ becomes the new g_{best}. The algorithm can be described by

> *Initialize the particles $(features)$ randomly.*
>
> *Compute pbest and gbest*
>
> *Start mutation*
>
> *Change gbest if better solution*
>
> *Update the velocity and particles position*
>
> *Continue or terminate if termination criteria met*

The extracted features are classified using a probability based, distance based and decision tree based classifier described in the next section.

D. Naive Bayes Classifier

Naïve Bayes classifiers are Bayes theorem based statistical classifiers using a probabilistic approach to predict a data's class, by matching it to class with highest posterior probability [26]. Following are the algorithms used in Naïve Bayes:

$$P(C_i|V) = \frac{P(V|C_i)P(C_i)}{P(V)} \qquad (8)$$

where $V = (v_1, \ldots, v_n)$ is document represented in n-dimensional attribute vector and $C_1, \ldots, C_m$ represents m class. Naïve assumption of class conditional independence is made as it is computationally expensive to compute $P(V|C_i)$ to reduce computation. Thus,

$$P(V|C_i) = \prod_{k=1}^{n} P(x_k|C_i) \qquad (9)$$

Naïve Bayes classifier is a simple instance of a probabilistic classifier. Output $Pr(C|d)$ of a probabilistic classifier is probability that document d belongs to class C. Every document has terms which are probabilities based on occurrence within particular documents. With supervised training, Naïve Bayes learns pattern of examining a test documents set that is well-categorized and hence compares contents in all categories by building a words list as also their occurrence.

E. k-Nearest Neighbour (kNN) Classifier

k-NN classifier is based on a premise that vector space model is same for similar instances [27]. Training instances are indexed and associated with corresponding label. A test instance, when submitted, is treated as a query retrieving instances from training set similar to test instances. Test instances, class label assignment are based on k-NN distribution. Class label is refined by adding weights. So, higher accuracy is got through tuning. k-NN method is simple and easy to implement.

$$p(x) \cong \frac{k}{NV} \qquad (10)$$

Similarly, probability density function $p(x|Hi)$ of observation x conditioned to hypothesis H_i is approximated. If it is assumed that N_i is number of patterns associated to hypothesis $H_i, i = 1 \ldots C$, so that $N1 + \cdots + NC = N$.

k-NN is calculated by calculating Euclidian distance though other measures are also available, but Euclidian distance results in a mixture of ease, efficiency, and productivity. Example is classified by determining majority of labels samples for k-NN. This method is easy to implement, for example "x" has k nearest examples where feature space and where most have same label "y", and then "x" belongs to "y". The k-NN method depends on furthermost theorem when considering theory. When decision course is considered, a small number of nearest neighbor are considered. So, when this method is used, example disproportion problem is solved. Limited nearest neighbor number are considered by k-NN, not decision boundary. k-NN can classify an example set of boundary intercross that overlap. Euclidian distance is calculated as follows. When two vectors x_i and x_j are given where $x_i = (x_i^1, x_i^2, x_i^3, x_i^4, x_i^5, \ldots, x_i^n)$ and $x_j = (x_j^1, x_j^2, x_j^3, x_j^4, x_j^5, \ldots, x_j^n)$

The difference between x_i and x_j is

$$D(x_i, x_j) = \sqrt{\sum_{k}^{n=1} ((xik - xjk))^2} \qquad (11)$$

F. C4.5 Classifier

C4.5 algorithm generates decision trees and is an extension of ID3 algorithm overcoming its disadvantages [28]. C4.5 made changes to improve ID3 algorithm some of which are:

- Handling training data with omitted values of attributes
- Handling differing cost attributes
- After its creation Pruning the decision tree
- Handling attributes with continuous and discrete values

Let training data be a set $S = s_1, s_2 \ldots$ of classified samples. Each sample $S_i = x_1, x_2 \ldots$ is a vector where $x_1, x_2 \ldots$ represent sample's attributes/features. Training data is a vector $C = c_1, c_2 \ldots$, where $c_1, c_2 \ldots$ represent class to which every sample belongs. At each tree node, C4.5 chooses a data attribute that effectively splits data set of samples S into subsets that are one class or other. Normalized IG (difference in entropy) results from choosing an attribute to split data. Attribute factor with highest normalized IG is considered to make a decision. C4.5 algorithm continues on to smaller sub-lists having next highest normalized IG.

World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:9, No:3, 2015

G. Support Vector Machine (SVM) Classifier

SVMs are related supervised learning methods for classification and regression belonging to a family of generalized linear classification. A special SVM property is that it simultaneously minimizes empirical classification error and maximizes geometric margin. So SVM is called Maximum Margin Classifier [29]. SVM is based on Structural Risk Minimization (SRM). SVM maps input vector to higher dimensional space where maximal separating hyperplane is constructed. Two parallel hyperplanes are constructed on either side of a hyperplane that separates data. The separating hyperplane is that which maximizes distance between two parallel hyperplanes. An assumption is that the larger the margin or distance between parallel hyperplanes; the better generalization error. Data points of form are considered

$$\left\{ \begin{array}{l} (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)................., \\ (x_n, y_n) \end{array} \right\} \qquad (12)$$

where $y_n = 1/-1$ a constant denoting class to which a point x_n belongs. n = number of samples. Each x_n is p-dimensional real vector. Scaling is important to watch against variable (attributes) with larger variance.

IV. EXPERIMENTAL RESULTS

The system was evaluated using heart sounds corresponding to four different heart conditions: normal, Mitral Valve Prolapse (MVP), Ventricular Septal Defect (VSD), and Pulmonary Stenosis (PS). 325 signals consisting of 150 normal heart sound, 75 MVP, 50 VSD and 50 PS. Features are extracted using Coiflet. The proposed BPSO Feature Selection is compared with the IG, and the result of compared solution is shown in Table I.

TABLE I
EXPERIMENTAL RESULTS

Techniques	Classification accuracy %	Precision	Recall	F-Measure
IG_Naive Bayes	83.38	0.8064	0.8133	0.8098
IG_C4.5	84.62	0.8179	0.8333	0.8255
IG_KNN	81.85	0.7963	0.7967	0.7965
IG-_SVM	86.46	0.8404	0.8683	0.8541
BPSO_Naïve Bayes	89.54	0.8747	0.8717	0.8732
BPSO_C4.5	88.99	0.8697	0.8652	0.8674
BPSO_KNN	87.69	0.8501	0.855	0.8525
BPSO_SVM	92.31	0.9026	0.9	0.9013

Fig. 2 shows the classification accuracy obtained for various classifiers using IG and BPSO as feature extraction technique.
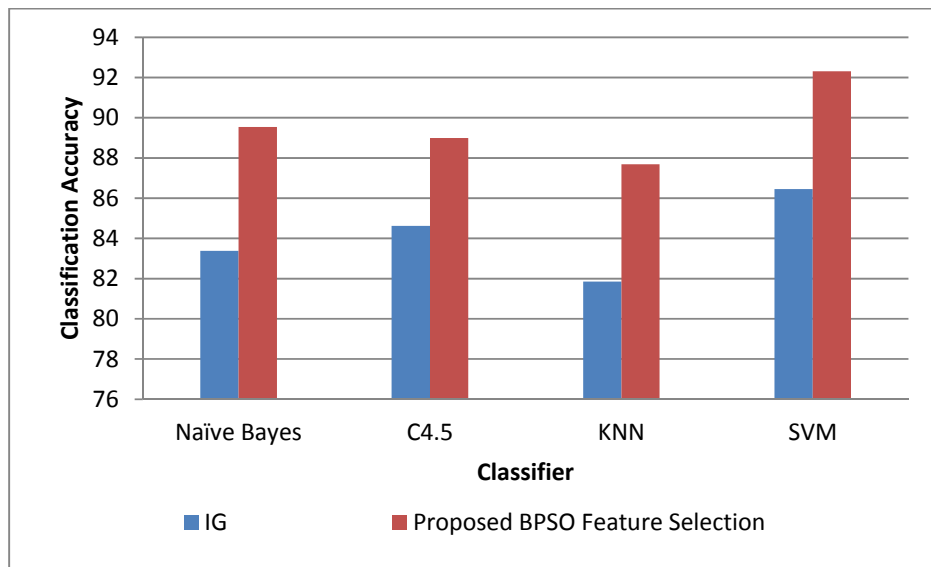

Fig. 2 Classification Accuracy for various classifiers

It is observed from the graph that the proposed feature selection improves the classifier efficiency significantly. The classification accuracy is improved in the range of 5.03% to 7.12%. SVM classifier achieves the best accuracy of 92.31%.

Fig. 3 shows the precision obtained. It is observed from the graph that the proposed feature selection improves the precision of the classifiers. The precision is improved in the range of 6.14% to 8.13% for BPSO based feature selection. Fig. 4 shows the recall obtained for various classifiers using IG and BPSO as feature extraction technique.

It is observed from the graph that the proposed feature selection improves the recall of the classifiers. The recall is improved in the range of 3.59% to 7.06%. Fig. 5 shows the F-measure which shows the joint relationship between precision and recall.

It is observed from the graph that the proposed feature selection improves the overall efficiency of classifiers. The F-measure is improved in the range of 4.95% to 7.53%. SVM classifier achieves the best F-measure of 0.9013.
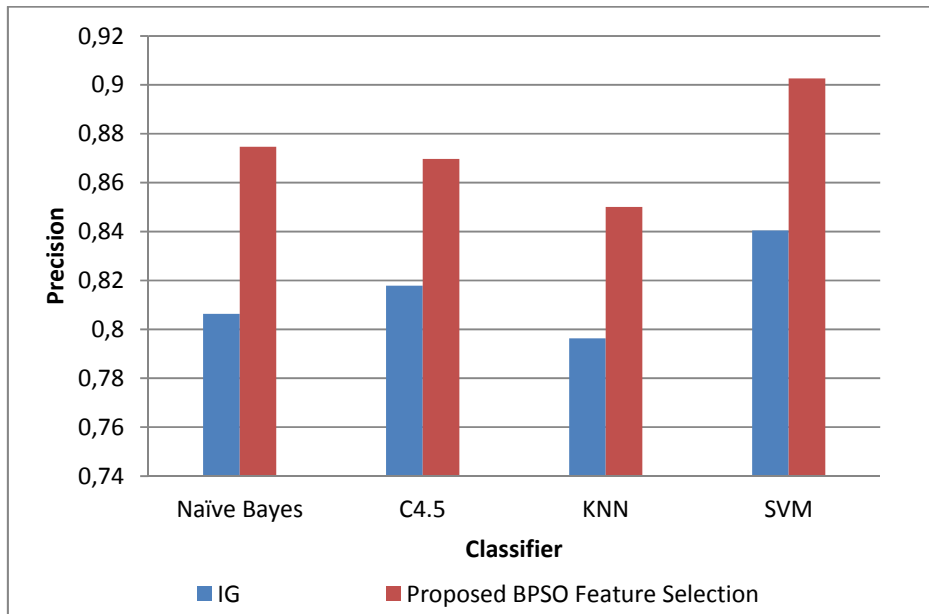
World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:9, No:3, 2015

Fig. 3 Precision obtained for various classifiers

Fig. 4 Recall obtained for various classifiers

World Academy of Science, Engineering and Technology
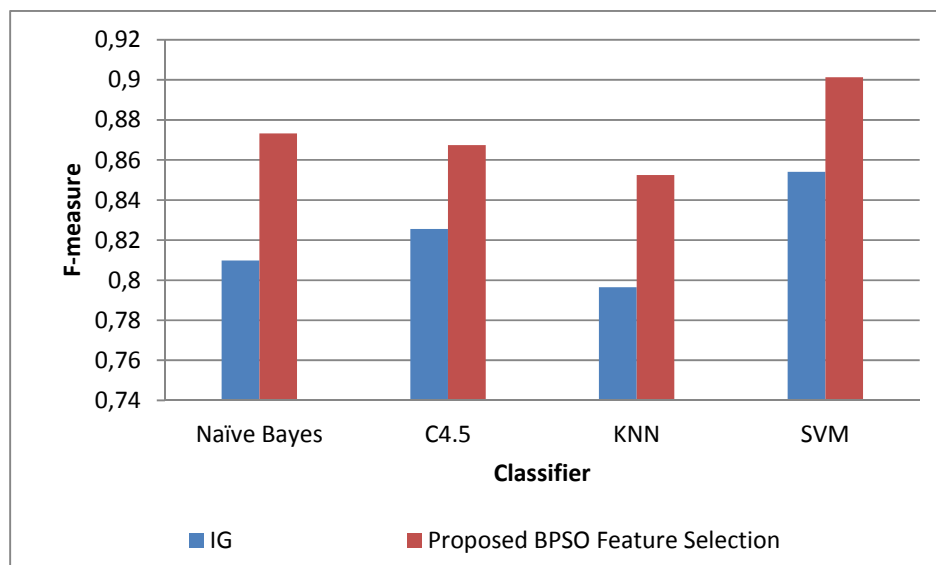International Journal of Biomedical and Biological Engineering
Vol:9, No:3, 2015
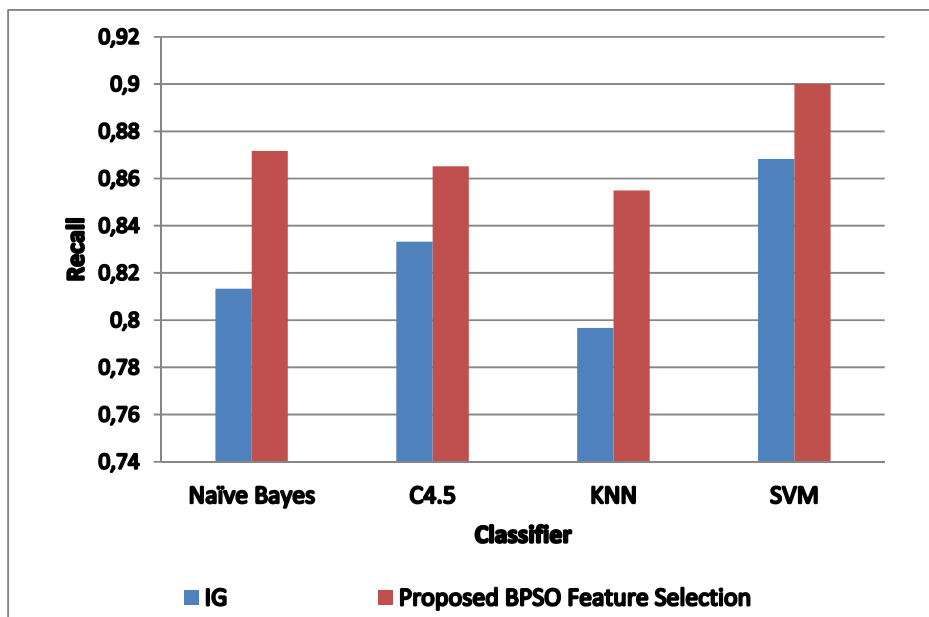
Fig. 5 F-Measure obtained for various classifiers

V. CONCLUSION

This work proposed developing an objective signal processing tool to extract information from phonocardiography signal to predict whether signal is normal or abnormal. Coiflet is used for feature extraction and Information Gain for feature selection. This work proposed a Binary PSO with mutation for feature selection. Experiment results prove that the new feature selection improves classifier efficiency by increasing accuracy, precision, and recall.

REFERENCES

[1] Roy, D. L. (2003). The paediatrician and cardiac auscultation. Paediatrics& child health, 8(9), 561.
[2] Donnerstein, R. L., & Thomsen, V. S. (1994). Hemodynamic and anatomic factors affecting the frequency content of Still's innocent murmur. The American journal of cardiology, 74(5), 508-510.
[3] Barschdorff, D., Femmer, U., &Trowitzsch, E. (1995, September). Automatic phonocardiogram signal analysis in infants based on wavelet transforms and artificial neural networks. In Computers in Cardiology 1995 (pp. 753-756). IEEE.
[4] Shino, H., Yoshida, H., Yana, K., Harada, K., Sudoh, J., &Harasewa, E. (1996). Detection and classification of systolic murmur for phonocardiogram screening. In Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE (Vol. 1, pp. 123-124). IEEE.
[5] Durand, L. G., &Pibarot, P. (1995). Digital signal processing of the phonocardiogram: review of the most recent advancements. Critical Reviews™ in Biomedical Engineering, 23(3-4).
[6] Bogdanović, V., Božić, I., Gavrovska, A., Stojić, V., &Jakovljević, V. (2013). Phonocardiography-based mitral valve prolapse detection using an artificial neural network. Serbian Journal of Experimental and Clinical Research, 14(3), 113-120.
[7] Rangayyan, R. M., &Lehner, R. J. (1986). Phonocardiogram signal analysis: a review. Critical reviews in biomedical engineering, 15(3), 211-236.
[8] Reynolds, D. A., & Rose, R. C. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. Speech and Audio Processing, IEEE Transactions on, 3(1), 72-83.
[9] Lung, S. Y. (2007). Efficient text independent speaker recognition with wavelet feature selection based multilayered neural network using supervised learning algorithm. Pattern Recognition, 40(12), 3616-3620.
[10] Guyon, I., &Elisseeff, A. (2003). An introduction to variable and feature selection. The Journal of Machine Learning Research, 3, 1157-1182.
[11] D. B., Brennan, T., Zuhlke, L. J., Abdelrahman, H. Y., Ntusi, N., Clifford, G. D. & Tarassenko, L. (2014, May). Signal quality classification of mobile phone-recorded phonocardiogram signals. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014 (pp. 1335-1339).
[12] Gavrovska, A., Zajić, G., Reljin, I., & Reljin, B. (2013). Classification of prolapsed mitral valve versus healthy heart from phonocardiograms by multifractal analysis. Computational and mathematical methods in medicine, 2013.
[13] Singh, M., & Cheema, A. (2013). Heart Sounds Classification using Feature Extraction of Phonocardiography Signal. International Journal of Computer Applications, 77(4).
[14] Sinha, Rakesh Kumar, et al. "K-Nearest Neighborhood Approach to Identify Level of Left Ventricular Ejection Fraction From Phonocardiogram." Journal of Clinical Engineering 38.2 (2013): 75-78.
[15] Ramos, J. P., Carvalho, P., & Coimbra, M. (2013, July). Towards a time-feature independent phonocardiogram segmentation. Annual International Conference of the IEEE in Engineering in Medicine and Biology Society (EMBC), 2013 35th (pp. 2116-2119).
[16] Dixit, A., &Majumdar, S. (2013). Comparative Analysis Of Coiflet And Daubechies Wavelets Using Global Threshold For Image De-Noising. International Journal of Advances in Engineering & Technology,6(5).
[17] Kumar, V. S., & Reddy, M. I. S. (2012). Image Compression Techniques by using Wavelet Transform. Journal of information engineering and applications, 2(5), 35-39.
[18] Kaur, E. R., & Singh, J. Comparative Study Of DWT Based Image Compression Using Haar, Daub & Coif Wavelets.
[19] Zhao, F., Liu, J., Liu, J., Guibas, L., & Reich, J. (2003). Collaborative signal and information processing: an information-directed approach. Proceedings of the IEEE, 91(8), 1199-1209.
[20] Weiss, L. E., Sanderson, A. C., & Neuman, C. P. (1987). Dynamic sensor-based control of robots with visual feedback. Robotics and Automation, IEEE Journal of, 3(5), 404-417.
[21] Sugumaran, V., V. Muralidharan, and K. I. Ramachandran. "Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing." Mechanical systems and signal processing 21.2 (2007): 930-942.
[22] Kennedy, J. (2010). Particle swarm optimization. In Encyclopedia of Machine Learning (pp. 760-766). Springer US.
[23] Ding, S., & Chen, L. (2010). Intelligent Optimization Methods for High-Dimensional Data Classification for Support Vector Machines. Intelligent Information Management, 2, 354.
[24] Sharma, L., Pathak, B. K., & Sharma, N. (2012). Breaking of simplified data encryption standard using binary particle swarm optimization. IJCSI International Journal of Computer Science Issues, 9(3), 1694-0814.
[25] Das, S., Abraham, A., & Konar, A. (2008). Particle swarm optimization and differential evolution algorithms: technical analysis, applications

World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:9, No:3, 2015

and hybridization perspectives. In Advances of Computational Intelligence in Industrial Systems (pp. 1-38). Springer Berlin Heidelberg.

[26] Ting, S. L., Ip, W. H., & Tsang, A. H. (2011). Is Naive Bayes a good classifier for document classification?. International Journal of Software Engineering and Its Applications, 5(3), 37.

[27] Kataria, A., & Singh, M. D. A Review of Data Classification Using K-Nearest Neighbour Algorithm.

[28] Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R., & Honrao, V. (2013). Predicting Students' Performance Using ID3 and C4. 5 Classification Algorithms. arXiv preprint arXiv:1310.2071.

[29] Srivastava, D. K., & Bhambhu, L. (2010). Data classification using support vector machine.

Venkata Hari Prasad G is with Andhra University. He is currently pursuing his doctorate in India.

P. Rajesh Kumar is currently working in Andhra University, India.