**Indian Journal of Science and Technology,** *Vol 10(33), 10.17485/ijst/2017/v10i33/117228, September 2017*
*ISSN (Print) : 0974-6846*
*ISSN (Online) : 0974-5645*

# Improved Classification of Phonocardiography Signal Using Optimised Feature Selection

**G. V. Hari Prasad, R. D. Bhavani, T. M. Ragni and V. S. Rani**

Department of Electronics and Communication Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada – 520008, Andhra Pradesh, India; gvh.prasad2k@gmail.com, r.d.bhavani7@gmail.com, thotamaryragni@gmail.com, vemulasirisha3@gmail.com

## Abstract

**Objectives:** To propose hybrid feature selection technique based on Particle Swarm Optimization and Genetic Algorithm for phonocardiography. **Methods/Statistical Analysis:** The system estimated using heart sounds corresponding to different heart conditions like 320 signals out of which 150-normal, 70-Mitral Valve Prolapse, 50-Ventricular Septal Defect and 50-Pulmonary Stenosis. Features are extracted using DWT. **Finding:** A phonocardiographic signal reflects the health status of the heart. Generally there exists two heart sounds, but further sounds indicate disease. Phonocardiography is non-invasive, low-cost and accurate method to detect heart disease. This work proposes a framework to extract information from phonocardiography signal to classify whether it is proper or improper. Discrete Wavelet Transform method is implemented to extract features followed by Singular value decomposition for feature selection. **Applications/ Improvements:** pplications/Improvements: An experimental result gives the improvement of the proposed method by increasing the efficacy of the corresponding feature selection technique.

**Keywords:** Binary Particle Swarm Optimization**,** Feature Selection**,** Phonocardiography, DWT, GA Introduction

## 1. Introduction

Phonocardiography (PCG) is the graphic representation of sounds produced in the heart vessels. PCG is an improvement for ECG. PCG is concerned with automatic acoustic recording and the processing of heart sounds by the use of electronic stethoscope[1,2]. PCG is useful for children, where ECG recording is difficult to implement. PCG signals are premature indicator for heart issues, as before worsening of the issue, an appropriate detection can be done. Later, other techniques were implemented to get an improved view of the problem[3]. Signal processing tools are used to increase the detection value of the PCG information[4].

A detailed study of PCG signal was done and features of waveform were observed[5]. Specific findings that have a potential to distinguish between classes could be named as feature. Basically two heart sounds, S1 and S2 named as Fundamental Heart Sounds (FHS). The remaining two sounds S3 and S4 are generally not audible. A heart cycle is the interval of S1 and S2. Here, Systole and diastole can be observed from the heart cycle. The signal between S1 and S2 is flat, low in strength and high in time under normal conditions but various pathologies like Ventricular Septal Defect, Mitral Regurgitation etc. cause sounds in these areas which are higher in strength and frequency.

**Author for correspondence*

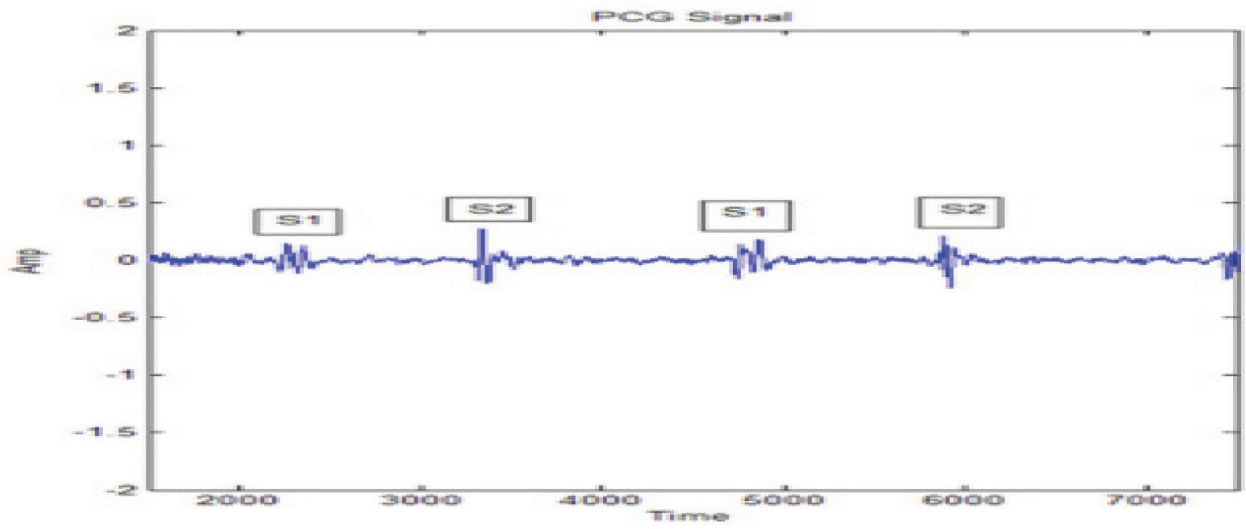


**Figure 1.** PCG Signal.

Features Extraction came into existence when the input data to any process or algorithm is too large and notoriously redundant. The input data is transformed into the reduced sets which contains the valuable properties and features of the data. So all the relevant information contained by the data's are also present in the extracted features, which can be process under any technique. Before extraction of features of the PCG signal it is necessary to extract the cardiac cycle of the one period, so called this process as pre-processing technique because a signal is processed before extracting the features. The most basic and principle element of an Artificial Neural Network (ANN) is called nodes. Nodes are act like artificial neurons which is generally same as natural neurons. . Features Extraction came into existence when the input data to any process or algorithm is too large and notoriously redundant. The input data is transformed into the reduced sets which contains the valuable properties and features of the data. So all the relevant information contained by the data's are also present in the extracted features, which can be process under any technique. Before extraction of features of the PCG signal it is necessary to extract the cardiac cycle of the one period, so called this process as pre-processing technique because a signal is processed before extracting the features. The most basic and principle element of an ANN is called nodes. Nodes are act like artificial neurons which is generally same as natural neurons.

This idea is initiated by real neurons present in the nervous system of the human body system[6]. ANN for classification of cardiac murmurs and got accuracy of 60. 68%. Classification is executed over normalized features sets. It was seen by trial and verifies that this method that neural network with 120 hidden nodes is giving the best result[7].

Feature selection technique is incorporated in discrete optimization problems. The main targets of feature selection technique are:

- To circumvent over fitting and increase prediction performance.
- To gain a better insight for underlying processes that generated the data.
- To offer faster and more cost-effective models.

ACO algorithm, GA and swarm intelligence are commonly used for finding the minimal feature subset. ACO is a system depends on agents, which mimic the natural behavior of ants, as well, as the GAs are stochastic global search techniques that mimic the natural biological evolution. The GA is minimization algorithm with a probabilistic component that search poorly understood, irregular spaces. GAs work with a multiple points. How to represent the problem as a string of elements is a critical factor in successfully applying a GA to a problem[8].

In[9] presented a cost effective and real-time system for heart auscultation monitoring and hearing. This paper studied the performance of the system. An analysis of

heart sound patterns for numerous diseases was conducted. The efficiency of the system was evaluated by comparing the execution time of algorithms in the proposed digital signal processor based phonocardiogram system with PC based phonocardiogram system.

In[10] discussed a method for selection of dynamic features, to find systolic murmurs from Phonocardiographic recordings and investigated the dynamic properties of spectral power during heart sound. The performance has been compared with different transformation methods. The implemented method can be applied as an investigation tool for primary health-care purposes as it has a high accuracy discriminating between normal and abnormal beats.

In[11] aimed to realize a feature subset for classification of heart murmurs. PCG signals were obtained from 36 patients with mitral insufficiency or physiological sounds, aortic valve Stenosis. The data were scrutinized with the techniques such as quantification analysis, Shannon energy and wavelets. They were used to extract more than 200 features. The derived feature was set better than the various feature sets and robust to noisy data.

In[12] developed a not harmful, portable and inexpensive care system using fetal phonocardiography. The fPCG technique has the potential to provide low-cost and long-term diagnostics to the under-served population. The fPCG signal contains valuable diagnostic data regarding health of the fetal during antenatal period.

In[13] presented software of fetal phonocardiographic signals comparing different fetal physiological states and diverse recording conditions. This software can be beneficial as a teaching tool for medical students and others.

In[14] presented auto diagnosis medical systems depend on the analysis of PCG signals. Noise can be induced by respiration and digestive sounds, movements or even signals from the surrounding environment and it is specified by wide frequency and intensity spectrum. The work presented a wavelet denoising algorithm for filtration of PCG signal disruptions from signals logged by mobile devices in a noisy environment.

In this work, a hybrid feature selection technique based on Binary Particle Swarm Optimization (BPSO) and Genetic Algorithm (GA) is proposed. To avoid early convergence and to merge the coordinates to achieve more convergence speed, the classical BPSO is modified using GA.

## 2. Methodology

A feature selection method is proposed, for classifying PCG signals. The proposed selection method is based on hybrid algorithm using Genetic algorithm and BPSO.

An implementation of a GA usually starts with a population of random chromosomes. These structures evaluate and give a better solution for the target problem, which is given more chances to reproduce than those chromosomes with defective solutions. The current population is the initial population in first generation. Once the selection process finishes and creates the next population from the intermediate population. Crossover is applied with a probability $P_m$ to randomly paired strings and forms two new strings that are inserted into the next population. After recombination, mutation operator is applied with low probability Pm for each bit in the population and next population can be evaluated. The procedure of evaluation, selection, recombination and mutation forms a generation in the accomplishment of a GA and named as Simple GA[15]. Table 1 shows Genetic Algorithm Pseudo code.

**Table 1.** GA Pseudo code

```
Begin
        Generate an initial population of solutions randomly.
        Calculate the initial population's Fitness.
Repeat
        Select a pair of parents based on fitness.
        Create two offspring using crossover.
Apply mutation to each child.
Evaluate the mutated offspring.
All the offspring will be the new population, the parents will die.
        Until a stop condition is satisfied
End.
```

Solutions are represented in binary as strings of 0s and 1s, and other encodings are also realizable. The evolution begins from a population of randomly created individuals and then generations are created. In every generation, the fitness of each individual in the population is calculated and more individuals are carefully chosen from current population based on their fitness, and it is changed to form a new population. New population is used for the

next iteration of algorithm. Commonly, the algorithm can be terminated when a maximum number of generations are created, or a fitness level has been got for the population[16].

The algorithm based on BPSO is easy for implementation and it successfully applied to solve a wide range of minimization problems in many application fields such as signal and image processing which includes image separation. Image segmentation is considered as a low-level image processing. The result of image segmentation shows a set of regions that collectively covers the entire image. Image segmentation methods can be classified into numerous categories based on region and thersholding of segmentations[17]. The BPSO algorithm can be explained basically as follows:

**Table 2.** Basic Algorithms for PSO

```
Randomly disperse particles into solution space
for i = 0 to numiterations do
for all particles in swarm do
        Compute fitness of current locations
          Upgrade X_Pbest if necessary
            Upgrade X_Gbest if necessary
          Upgrade velocity
                  Upgrade position
end for
end for
Mutual Information
```

The BPSO algorithm is an iterative, synchronous algorithm and it consists of two parts are as follows:

- Cognitive component: Particle considers about its previous best position.
- Social component: Particle considers about the best position among the particles [18].

In BPSO, each potential solution is identified as a particle. Two properties such as position x and velocity v are associated with each particle. Suppose if x and v of the $i^{th}$ particle are given below,

$$x = (x_{i1}, x_{i2}, \cdots, x_{iN}) \quad (1)$$

$$v = (v_{i1}, v_{i2}, \cdots, v_{iN}) \quad (2)$$

Where 'N' is the dimensions of the problem. In each iteration, a fitness function can be estimated for all the particles in the swarm. The velocity of each particle is upgraded by keeping track on best positions as follows:

- A particle has traversed so far and so named as "pBest".
- If any adjacent particle has traversed so far then it is a neighborhood of best so named as "nBest".
- When a particle acquires the whole population as its neighborhood, then best among neighborhood becomes the global best and named as "gBest".

Hence, a particle's velocity and position can be upgraded as follows

$$v = \omega \cdot v + c_1 r_1 (pBest - x) + c_2 r_2 (nBest - x) \quad (3)$$

$$x = x + v\Delta t \quad (4)$$

Where ω stands for "inertia weight" that controls the impact of the earlier velocity of the particle on its Present stage. The c1and c2 are parameters of positive constants, called as "acceleration constants" and the parameters r1 and r2 are random numbers uniformly distributed with the interval of (0, 1). These random numbers are upgraded every time when it appears. The parameter Δt is the given time-step. The population of particles is then moved accordingly and tends to cluster together from various directions. The BPSO algorithm runs along these processes iteratively until the termination criterion gets satisfied.

The particle's velocity and position are upgraded as follows:

$$v_i^d = wv_i^d + c_1 r_1 \left( p_i^d - x_i^d \right) + c_2 r_2 \left( p_g^d - x_i^d \right)$$

$$x_i^d = x_i^d + v_i^d$$

Where d represents dimensions; i is size of population; r1 and r2 are random values (0 - 1). The parameters in the BPSO affect the performance of the algorithm significantly[19]. The value of convergence is based $c_1$ and $c_2$ values.

Convergence is feasible when $1 > w > \frac{1}{2}(\phi_1 + \phi_2) - 1$

where $\phi_1 = c_1 r_1$, $\phi_2 = c_2 r_2$ also for stochastic $\phi_1$ and $\phi_2$, convergences results when $\phi_{ratio} = \frac{\phi_{crit}}{c_1 + c_2}$ is close to 1 and $\phi_{crit} = \sup \phi \left[ 0.5\phi - 1 < w,\ \phi \in (0, c_1 + c_2) \right]$

In the proposed hybrid technique, GA generates a population by encoding BPSO parameters. On GA algorithm termination, obtained BPSO parameters are upgraded into BPSO algorithm.

## 3. Results and Discussion

The system estimated using heart sounds corresponding to different heart conditions like normal, Mitral Valve Prolapse, Ventricular Septal Defect and Pulmonary Stenosis. 320 signals consisting of 150- normal heart sound, 70- MVP, 50- VSD and 50- PS. Features are extracted using DWT. The proposed GAPSO technique is compared with Mutual Information and PSO feature selection methods.Classification is done with Naïve Bayes, kNN, C4.5 and SVM for the selected features.

**Table 3. Experimental Results**

| | Accuracy | Precision | Sensitivity |
|---|---|---|---|
| Naive Bayes | 79.38 | 0.7708 | 0.7833 |
| C4.5 | 81.54 | 0.7882 | 0.8133 |
| KNN | 78.46 | 0.7604 | 0.7783 |
| SVM | 86.46 | 0.8404 | 0.8683 |
| Binary PSO | | | |
| Naive Bayes | 89.54 | 0.8747 | 0.8717 |
| C4.5 | 87.69 | 0.8501 | 0.855 |
| KNN | 88.99 | 0.8697 | 0.8652 |
| SVM | 92.31 | 0.9026 | 0.9 |
| Proposed hybrid GAPSO | | | |
| Naive Bayes | 91.38 | 0.8979 | 0.89 |
| C4.5 | 90.15 | 0.8807 | 0.8833 |
| KNN | 90.77 | 0.8901 | 0.8867 |
| SVM | 94.15 | 0.9327 | 0.9233 |

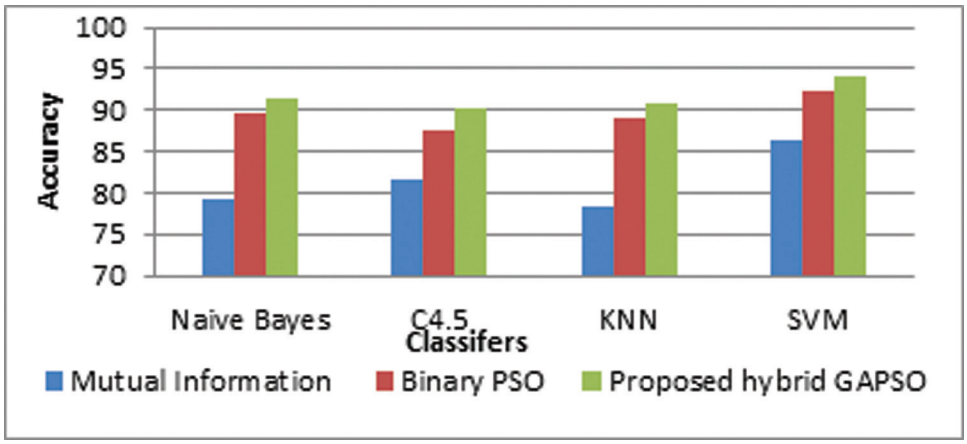


**Figure 2.** Accuracy.


**Figure 3.** Precision.

From the Figure 3 observation, the proposed technique improves the efficiency appreciably. The accuracy is enhanced in the range of 1.98% - 14.55%. SVM classifier achieves the maximum accuracy of 94.15%.

From the Figure 4 observation, the designed technique enhances the precision of the classifiers. The precision is enhanced in the range of 2.31% - 15.72%.

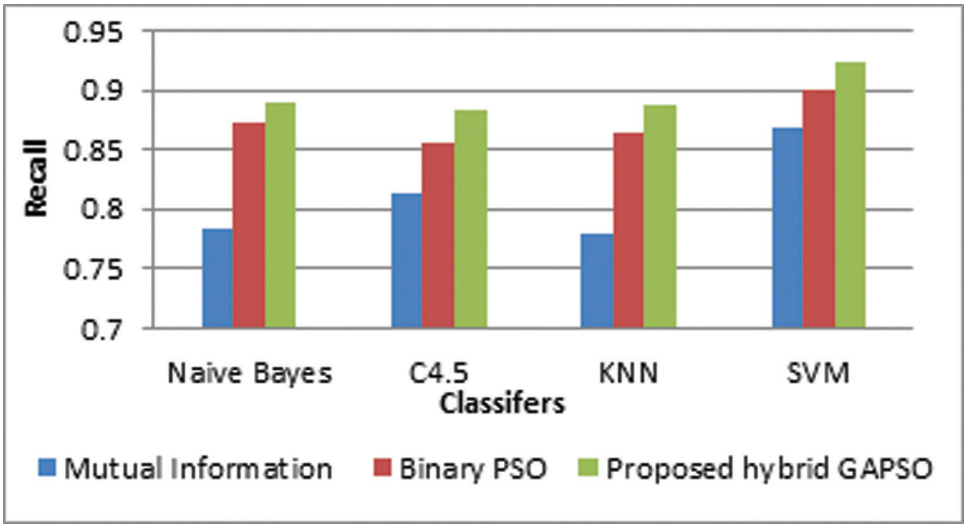


**Figure 4.** Sensitivity.

From the Figure 5 observation, the designed technique enhances the Sensitivity of the classifiers. The Sensitivity is enhanced in the range of 2.08% - 13.02%.

## 4. Conclusion

Hybrid selection technique is designed to add the most valuable properties of GA and BPSO and to overcome the problem of early convergence. Both techniques gave good results for some particular aspects but not for other ones separately. DWT is used for feature extraction and SVD for feature selection. This work proposes a hybrid GA-BPSO for feature selection. Results show that the proposed selection technique enhances classifier efficiency by enhancing accuracy, precision and Sensitivity.

## 5. References

1. Javed F, Venkatachalam PA. A signal processing module for the analysis of heart sounds and heart murmurs. Journal of Physics: Conference Series 2006 April;34(1):1098. Crossref.
2. Varady P. Wavelet-based adaptive denoising of phonocardiographic records. Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Turkey. 2001. 2 p.1846–9. Crossref.
3. Singh M, Cheema A. Heart Sounds Classification using Feature Extraction of Phonocardiography Signal. International Journal of Computer Applications. 2013;77(4):13–17.
4. Ahlström C. (2006). Processing of the Phonocardiographic Signal: Methods for the intelligent stethoscope [PhD thesis]. Sweden: Linkoping University; May 2006.
5. Cheema A, Singh M. Steps Involved in Heart Sound Analysis-A Review of Existing Trends. International Journal of Engineering Trends and Technology. 2013 July; 4(7):2921–5.
6. Reynolds DA, Rose RC. Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Transactions on Speech and Audio Processing. 1995;3(1):72–83. Crossref.
7. Badghare SS. Analysis of (Doctoral dissertation, National Institute of Technology, Rourkela).
8. Pahuja H, Chhabra J, Khokhar A. Ant colony optimization-based selected features for Text-independent speaker verification.
9. Balasubramaniam D, Nedumaran D. Design and Development of Digital Signal Processor based Phonocardiogram System. IEEE International Conference on Signal Acquisition and Processing, Bangalore.2010. p. 366–70. Crossref.
10. Quiceno-Manrique AF, Godino-Llorente JI, Blanco-Velasco M, Castellanos-Dominguez G. (2010). Selection of dynamic features based on time–frequency representations for heart murmur detection from phonocardiographic sig-

nals. Annals of biomedical engineering, 2010; 38(1):118–37. Crossref. PMid:19921435
11. Ahlstrom C, Hult P, Rask P, Karlsson JE, Nylander E, Dahlström U, Ask P. Feature extraction for systolic heart murmur classification. Annals of biomedical engineering, 2016; 34(11): 1666–77. Crossref. PMid:17019618
12. Chourasia VS, Tiwari AK, Gangopadhyay R. A novel approach for PCG signals processing to make possible fetal heart rate evaluations. Digital Signal Processing, 2014; 30: 165–83. Crossref.
13. Cesarelli M, Ruffo M., Romano M, Bifulco P. Simulation of foetal phonocardiographic recordings for testing of FHR extraction algorithms. Computer methods and programs in biomedicine. 2012; 107(3): 513–23. Crossref. PMid:22178069
14. Gradolewski D, Redlarski G. Wavelet-based denoising method for real PCG signal recorded by mobile devices in noisy environment. Computers in biology and medicine, 2014; 52: 119–29. Crossref. PMid:25038586
15. Premalatha K. Hybrid PSO and GA for Global Maximization. International Journal of Open Problems Computational Mathematics.2009 Dec; 2(4): 597–608.
16. Medha VW. Image Registration Techniques: An overview. International Journal of Signal Processing, Image Processing and Pattern Recognition. 2009 Sep; 2(3):11–28.
17. Fahd MA Mohsen. A new Optimization-Based Image Segmentation method By Particle Swarm Optimization, International Journal of Advanced Computer Science Application, Special Issue on Image Processing and Analysis.11–18.
18. Kai-Shiuan Shih. Observer-Based Adaptive Fuzzy Robust Controller with Self-Adjusted Membership Functions For A Class of Uncertain Mimo Nonlinear Systems: A PSO-SA Method, International Journal of Innovative Computing, Information and Control. 2012 Feb; 8(2):1419–37.
19. Van Den Bergh, F. An analysis of particle swarm optimizers [Doctoral dissertation]. Pretoria:University of Pretoria. 2006.