IJCTA, 9(21), 2016, pp. 259-265

Classification on Atrial Fibrillation in Cardiac Signal Using Swarm Feature Selection

Bhagyashree Bhoyar* Pramod Patil ** Parag Kulkarni* Priyanka Abhang******

Abstract **:** There are advances in the field of healthcare system administration. The large amount of the data are collected and check through the system. There is a need of knowledge extraction and optimization for real time decision making in healthcare system. The proposed methods are used to optimize minimal set of features. The experimental result shows the classification accuracy of proposed methods are outperform than existing state of art method. The classification accuracy is improved by 8 to 10 %more than existing system. The proposed method is efficient in processing large data. It is also implemented through wireless network.

Keywords : Feature Selection, Swarm Intelligence, Big data, Metaheurstics, PSO, APSO

1. INTRODUCTION

One of the most advanced techniques in swarm is Particle swarm optimization (PSO) is a calculation method that advance an issue by over and over attempting to upgrade a hopeful arrangement as for a given measure of value. It conquer a troublesome issue by having a populace of competitor arrangements, here called particles, and moving these particles around in the pursuit space as per basic numerical formulae over the particle's place and speed. In component in grouping and insights, highlight choice, additionally called as variable choice, quality determination or variable subset choice is the way toward selecting a subset of relevant elements (variables, indicators) for use in model development. Highlight determination strategies are utilized for three reasons:

1. Researcher makes them easier to interpret by popularisation of models.
2. Training time will be shorter.
3. Enhanced simplification by sinking over fitting.

Big data [6-7] is data sets that are so complex that customary information handling applications are insufficient.. Challenges incorporate investigation, catch, term, seek, sharing, stockpiling, exchange, perception, questioning, and upgrading and data security. The term often refers simply to the use of prophetic analytics or definite other advanced data analytics approached that take out value from data, and infrequently to a particular size of data set. Accuracy in big data may show the way to more confident decision making, and better decision can result in greater operational efficiency, cost reduction and condensed risk [2].

Three main problematic issues are rise in big data are as follows :

1. Velocity is used to handle at a rising high speed problem that manifested to a huge amount of data.
2. Variety problem occur due to collection of data from a variety of source and that data are arranged or formatted in different way due to this data integration and processing are not easy for processing.

* Department of Computer engineering DYPIET, Pimpri, Pune, Maharashtra, INDIA Email- bhagyashree05bhoyar@gmail.com
** Department of Computer engineering DYPIET, Pimpri, Pune, Maharashtra, INDIA Email- pdpatiljune@gmail.com
*** Department of Computer engineering DYPIET, Pimpri, Pune, Maharashtra, INDIA Email- paragndia@gmail.com
**** Department of Computer engineering DYPIET, Pimpri, Pune, Maharashtra, INDIA Email- pabhang4@gmail.com

3. Volume problem occur due to storing process, processing of data and analysis for both computational and archiving [2].

Random forest is the collection or combination of multiple decision trees. It is learning method for classification and regression to take large numbers of decision tree at training time and provide output in the class's structure.It work on huge dataset. It find an efficient accuracy.

The purpose of the study is to explore the effectiveness of PSO and APSO in data mining structure and enhancing classification accuracy. To achieve this objective, the following tasks must be conducted:

1. To develop CFS/PSO and APSO Structure for feature extraction.
2. To analyze and minimize the Number of Features
3. To evaluate Classification Performance
4. To evaluate Features Accuracy
5. To develop Classification using Random Forest and Random Forest

2. RELATED WORK

This section discusses existing work done by the researchers for feature selections.

CART,D-tree [3] algorithm is used for feature selection method which gives the better result for classification. Best features are selected on the basis of particle dimension and velocity. But the previous method has take more time for selection it and it selection many features. This problem is solved by proposed method it takes less time to select features.

CCV [1] algorithm has been proposed for feature selection algorithm. This algorithm is based on from generalization and overfiting decision tree algorithm. In this algorithm, to choose good prediction accuracy on the basis of that selects appropriate features. Clustering Coefficients 2 of Variety (CCV) is a novel and proficient element determination technique. Results demonstrate that CCV beat them in all parts of arrived at the midpoint of exhibitions and velocity High accuracy of classification is achieved in less processing time. The proposed algorithm select a minimum number of features on basis of top ranked and it is classified in classes. Performance parameter are accuracy and time, feature are given to proposed algorithm.

PSO [5] algorithm has been proposed for multiobjective approach .Author described the PSO based multiobjective FS algorithm. The author described algorithm introduced the view of non dominated sorting into PSO to reference FS. The author also described algorithm is used for crowding and mutation. The two multi-target calculations are contrasted and two routine element determination strategies, a solitary target highlight choice technique, a two-phase highlight choice calculation, and three surely understood transformative multi-target calculations on 12benchmark information sets. The test output demonstrate that the two PSO-based multi-target calculations can consequently advance an arrangement of non ruled arrangements

A method of swarm search feature selection approach called swarm intelligence [2] which provides the better result for feature selection technique. This technique is also applicable to select optimal number of features and to solve NP-hard problems. Swarm selection and Feature selection is established to be a possible evaluating tool in get elevated accuracy in classification.

New Feature selection technique is used for feature selection and well known NP –hard problem [4].Brute force algorithms hardly difficult all achievable grouping of features takes according to the grapevine forever, stochastic optimization may be a solution.Metaheurstic algorithm used for swarm feature selection. Proposed algorithm evaluates the performance of each feature in class format and select best possible feature.

SVM, Accelerated PSO [6] computation has been proposed for grouping and relapse. It picks a most perfect component for gathering to get precision. In particular, molecule swarm upgrade is right now thoroughly used as a piece of dealing with unpalatable headway issues. It use a social event of a generally made accelerate PSO to plot a structure for getting it done headway issues. It first apply the future APSO-SVM to manufacture streamlining, and after that usage it for cash estimate and wander booking

3. PROPOSED SYSTEM

3.1. System Overview

The following figure 1 shows the architectural view of the proposed structure. The description of the architecture is as follows:

Upload Dataset : First, Data set of any size and of any format has to be uploaded by user.

Preprocessing : This raw data set is pre-processed and transformed into compatible format for further processing. String values are replaced by numerical values for next processing.

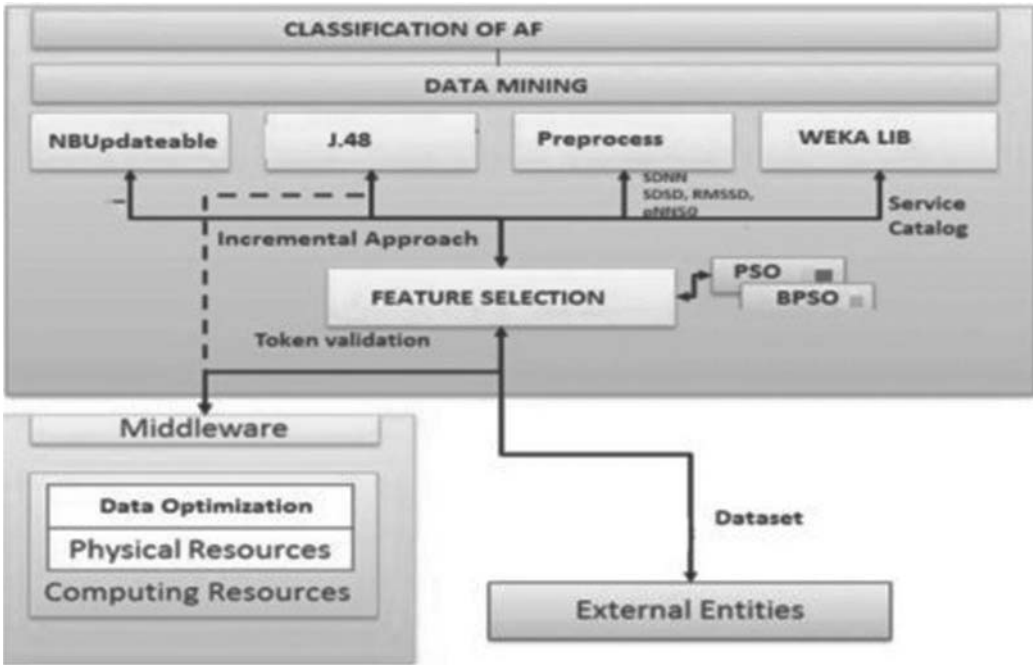

Fig. 1. System Architecture

Feature extraction : In proposed system, instead of considering all available features only small number of useful features are extracted for classification maintaining accuracy, reducing time and memory for classification.

Training data set : This step consists of training data set and classifier model which will be used to classify further data elements.

Testing data set : Classified trained data against which test data set is evaluated. Then this is used to further classify unknown label data.

Testing data set : Classified trained data against which test data set is evaluated. Then this is used to further classify unknown label data.

Result : Results of processing time, classified data, feature selected, accuracy, false positive rate, selected high ranked features are displayed.

3.2. Proposed Algorithm

This section discusses an algorithm of feature extraction PSO offer numerous comparability with transformative expansion procedures and Random Forest is used to classification

PSO : Feature Selection Algorithm

Steps :

1. Start the process.
2. Declare the position as p and velocities as v of all particle.
3. Estimate the wellness value of all particle.

4. for all particle set local best wellness= current wellness and local best position = current position.
5. Set global best wellness = min (local best fitness)
6. Revise velocities and position of each particle.
7. If current wellness < local best wellness then set local best wellness = current fitness
8. If current wellness < Global best wellness then set global best wellness current fitness
9. If the stopping criteria met then stop the process.
10. end if update *v* and *p* of all particle.
11. end if update *v* and *p* of all particle.
12. end if Estimate the fitness value of all particle.

Random Forest :

Steps :

1. Get input as X number of training set.
2. All the values of each case taken at random with replacement
3. This sample will be the training set for rising the tree.
4. Suppose Y is input variables, a number X <Y for all node, X variables are selected at random out of the Y.
5. X is used to split the node. The value of X is held constant while it rising the forest.
6. Each tree is rise to the most important level possible and there is no pruning.
7. Forecast new data by aggregating the predictions of the ntree trees.

3.3. Mathematical Algorithms

Pre-processing

1. High Pass Filter :

$$Y1[n] = \frac{1}{M}\sum_{m=0}^{M-2} x[n-M] \quad ...(1)$$

$$Y2[n] = x\left[n-\frac{M+1}{2}\right] \quad ...(2)$$

$$Y[n] = Y1[n] + Y2[n] \quad ...(3)$$

Where M is window size *i.e.*, 5 or 7 is optimal.

$x[n]$ is input data

m is the Filter length.

2. Low Pass Filter :

$$Y0 = Y0^2 + Y1^2 + Y2^2 + ... Y29^2 \quad ...(4)$$

$$Y[n] = Yn^2 + Yn-1^2 + ... Yn-29^2 \quad ...(5)$$

Here the window size is 30.

Feature Selection

1. New Velocities

$$Zi(h+1) = Zi(h) + Y1i(Pi - Xi(h)) + Y2i(G - Di(h)) \quad ...(6)$$

Where i is particle index.

h is discrete time index.

Zi is velocity of ith particle.

Di is position of ith particle.

Pi is best position founded by ith particle (personal).

g is best position founded by swarm.

G(1,2)i is interval [0,1] applied to ith particle.

2. New Position $Di(h+1) = Di(h) + Zi(h+1)$ ...(7)

Where *i* is particle index.
h is discrete time index.
Zi is velocity of ith particle.
Di is position of ith particle.
Pi is best position founded by ith particle (personal).
g is best position founded by swarm.
G(1,2)i is interval [0,1] applied to ith particle.

Classification

$$\overleftrightarrow{C}^{B}_{rf}(x) = \text{majority vote } \overline{C}(x)^{B}_{1} \quad ...(8)$$

Where Cb(*x*) is class prediction of both random forest tree.

4. RESULT

4.1. Data Set

For evaluating performance of system, Arrhythmias Database At first this database contained 452 examples and 279 qualities. This datasets to add up to 377 cases and 166 qualities, conveyed into 6 Classes with Class an alluding to "typical" ECG, Classes Q to T alluding to various classes of arrhythmia and Class U alluding to whatever is left of unclassified ones

Table 1. Arrhythmia classes and corresponding number of instances in the dataset

Data	*Class P*	*Class Q*	*Class R*	*Class S*	*Class T*	*Class U*	*Total*
Initial	237	37	24	16	48	18	380
Train	119	19	12	7	24	9	190
Test	118	19	12	7	24	9	189

All the PSO based techniques taken here were prepared by half of the aggregate datasets picked reasonably from the principle datasets guaranteeing representation of all classes present in the required rate.

4.2. Feature Selection

Only top features are selected. This is done on the basis of ranking mechanism in which rank is given to each attribute.

PVC	NSR	AF
NHR	Dioxin	APC
PFB	Diapers	Xyloprim
NHR	Pronestyl	Nitropaste

4.3. Performance Evaluation

To evaluate performance of system, four parameters are chosen which are given below:

1. **True positive (TP):** It is normal instances correctly classified as normal,.
2. **False positive (FP):** It is normal instances incorrectly classified as anomaly
3. **True negative (TN):** It is abnormal instances correctly classified as anomaly.
4. **False negative (FN):** It is abnormal instances incorrectly classified as normal.

Using all above four parameters, classification accuracy and false positive rate (FPR) are calculated using below formulae for existing and proposed system. Classification

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{False Positive Rate} = \frac{FP}{TN + FP}$$

4.4. Comparison between Existing System and Proposed System

4.4.1. Existing System

Classifier	*Approach*	*# features*	*Accuracy*	*Preprocessing Time*
NaïveBayes	PSO + NBay.	7	85.56	3000
	PSO + CFS	7	85.19	3001
	GA + CFS	7	84.78	3000
	All Inputs	13	83.70	2998
DecisionTree	PSO + DT	4	83.33	2967
	PSO + CFS	7	80.74	2701
	GA + CFS	7	76.08	2688
	All Inputs	13	76.09	2500

4.4.2. Proposed system

Incremental Classifier	*Accuracy*	*Precision*	*Recall*	*Preprocessing Time(s)*	*%Select Feature*
CFS-PSO	95.91	0.9588	0.9588	2269	24
APSO/BPSO	97.84	0.9781	0.781	1944	18

The Following figure shows the accuracy, preprocessing time and feature selection in graphical format.

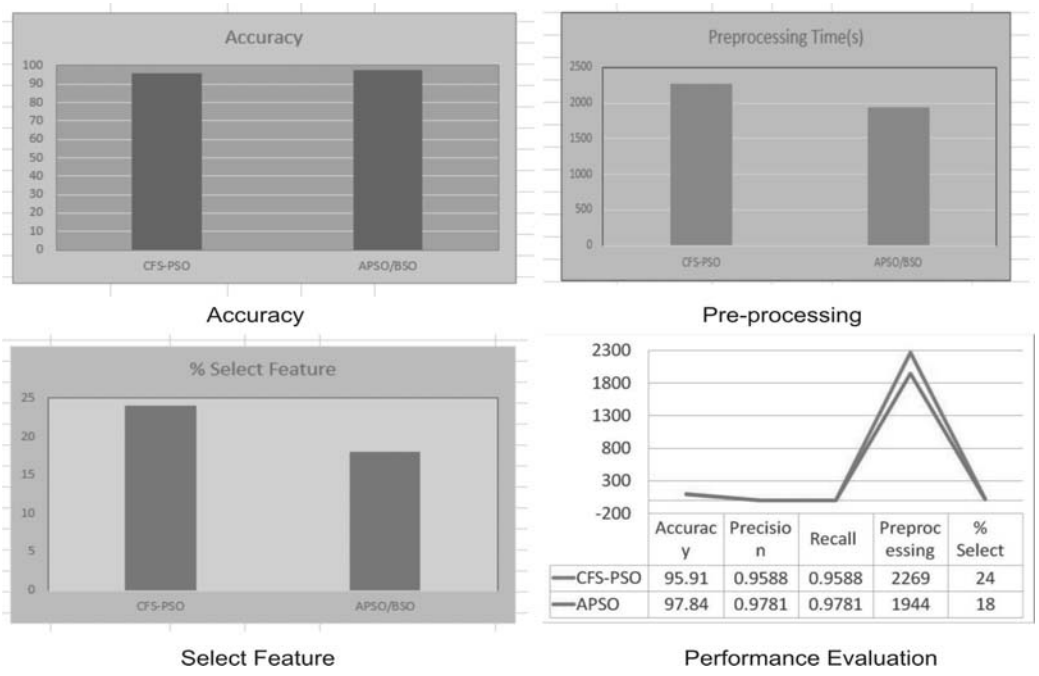

Fig. 1.

5. CONCLUSION

PSO and Random Forest perform well for classification and feature selection. Previous research this area gives rise to reduce problem occurring during feature selection and classification. In proposed system PSO is used to select best particle based on particle velocity and dimension. The evaluation strategy used for incremental method secure a higher gain in accuracy per second incurred in the pre-processing. The proposed method is used to optimize minimal set of features. The experimental result shows the classification accuracy of proposed method is outperform than existing state of art method. The classification accuracy is improved by 8 to 10 %more than existing system. This approach also fits better with real-world applications where their data arrive in streams. This approach will used in medical industries for heart patients. It is also implemented through wireless network.

6. REFERENCES

1. Fong, S., Liang, J., Wong, R., Ghanavati, M., "A novel feature selection by clustering coefficients of variations", 2014 Ninth International Conference on Digital Information Management (ICDIM), Sept. 29, 2014, pp.205-213.
2. Fong, Simon, et al. "Feature Selection in Life Science Classification: Metaheuristic Swarm Search." IT Professional (2014): 24-29.
3. Simon Fong ,Raymond wong, A. V.Vasilakos" Accelerated PSO Swarm Search feature Selection for Data stream Mining Big Data"IEEE,2015.
4. Simon Fong,Xin-she Yang,Suash Deb "Swarm Search for Feature Selection in Classification" IEEE,2013 .
5. Bing Xue,mengjie zhang,will N Browne "Particle Swarm optimization for feature selection in classification : A Multi Objective Approach"IEEE,2012 .
6. Xin-She Yang, Suash Deb, Simon Fong, "Accelerated Particle Swarm Optimization and Support Vector Machine for Business Optimization and Applications, "The Third International Conference on Networked Digital Technologies (NDT 2011), Springer CCIS 136, 11-13 July 2011, Macau, China, pp.53-66.
7. Fong, S., Liang, J., Wong, R., Ghanavati, M., "A novel element choice by grouping coefficients of varieties", 2014 Ninth International Conference on Digital Information Management (ICDIM), Sept. 29, 2014, pp.205-213 .
8. Wei Fan, Albert Bifet, "Mining Big Data: Current Status, and Forecast to the Future", SIGKDD Explorations, Volume 14, Issue 2, pp.1-5 .
9. Bifet, Albert, and Ricard Gavalda. "Gaining from Time-Changing Data with Adaptive Windowing." SDM. Vol. 7. 2007.
10. Eberhart, R., and Kennedy, J. (1995). Another enhancer utilizing molecule swarm hypothesis. Paper displayed at the Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on.
11. Kennedy, J., and Eberhart, R. 1995. Molecule swarm improvement. Paper exhibited at the Neural Networks, 1995. Proceedings.,IEEE International Conference on.