

OPTIMAL FEATURE SELECTION ALGORITHM FOR HIGH DIMENSIONAL DATA SETS USING PARTICLE SWARM OPTIMIZATION

D.Sheela Jeyarani¹, Dr.Mrs.A.Pethalakshmi² and DR.Mrs.K. Jayapriya³

Abstract: In high dimensional data sets, there are large numbers of features in classification problems. But all of them are not necessarily beneficial for classification. Feature selection involves the choosing of a small number of relevant features in order to achieve similar or even better classification performance. When all the features are used, irrelevant and redundant features reduce the performance. Maximizing the classification performance and minimizing the number of features are the important objectives of feature selection. Fuzzy entropy facilitates the partition of the input space into decision regions. It enables the selection of appropriate features with good separability for the classification task. In the recent times, the observations in the swarm behavior of birds and fishes, has led to the study of the nature inspired heuristic algorithms called particle swarm optimization. It has been applied to find solution for many optimization applications. This paper proposes a hybrid method for feature subset selection based on Fuzzy entropy with particle swarm optimization. The proposed algorithm is applied to four different data sets. Further the performance metrics and the dimension metrics are calculated and compared with the existing FCBF (Fast correlation based filter solution) algorithm. The results show that the proposed algorithm can give significant results for the feature selection process.

Keywords: High Dimensional Data Set, Fuzzy Entropy, Particle Swam Optimization, Feature Section

I. INTRODUCTION

Huge data base have been accumulated in the recent times and it has created the need for good machine learning techniques. Consequently, newer approaches to feature selection has become inevitable. Hence feature selection has become a prominent area of interest. The volume of the data set becoming huge, unless the unwanted features are removed, it hinders the learning, in many applications. Simultaneously, when the number of irrelevant and redundant features are reduced, the running time of algorithm is minimized and the level of performance is maximized.

Fuzzy entropy enables the partition of the input space into decision regions in order to select appropriate features with good separability for the classification task. This concept has been defined in various ways and generalized in various applied fields such as communication theory, mathematics, statistical thermodynamics and economics. In the arena of contemporary global optimization algorithms, computational intelligence and soft computing, meta heuristic algorithms play a vital role.

The increase in search space explodes in size with reference to the number of available features in the data set^[1]. In such situations, a thorough search becomes practically unfeasible. In response to this need, a broad spectrum of methods have been introduced to feature selection such as greedy based sequential forward selection (SFS)^[2] and sequential backward selection (SBS)^[3]. In spite of that, there are continuing problems in the varied feature selection approaches like stagnation in local optima and high computational cost. So, an efficient global search technique is inevitable to meet this demand. For such an efficient global search, Evaluating computations (EC) are very popular. Particle swarm optimization (PSO)^[4,5] is a relatively recent EC technique based on swarm intelligence.

Kennedy and Eberhart developed the particle swarm optimization (PSO) in 1995. PSO is a population based stochastic optimization technique. It stimulates the social behavior of organisms such as bird flocking and fish schooling to describe an automatically evolving system. PSO has proved to be a very efficient technique in several fields including feature selection^[6,7,8]. It is cost effective and converge more quickly when compared with other EC algorithms such as genetic algorithm (GAs) and genetic programming (GP).

¹ Research Scholar, Department Of Computer Science Mother Teresa Women's University Kodaikanal, TamilNadu, India

² Head & Associate Professor, Department Of Computer Science M.V.M Government Arts College(W) Dindigul, TamilNadu, India

³ Visiting professor Department Of Computer Science NadarSaraswathi College of Arts & Science Theni, Tamil Nadu, India

This paper proposes an approach to find the feature selection based on fuzzy entropy and particle swarm optimization which gives significant result compared to the existing FCBF algorithm. . In section 2, related works are discussed. Section 3 deals with methodology, proposed architecture and algorithm. In Section 4, experimental analysis to support the proposed algorithm is discussed. Finally, section 5 summarizes the present study and draw some conclusion.

II. RELATED WORKS

Machine learning (ML) is a fast developing field. ML algorithm concentrates on the data aspects which are helpful for analysis and future prediction. Feature selection is fundamental to the Machine learning (ML) and it is also known as dimensionality reduction which identifies the goodness of features. Wrapper approaches and filter approaches are the two major categories in the classification of the existing feature selection algorithms. Wrapper approaches include a learning algorithm in the evaluation procedure while filter approaches do not. Although Filter approaches are considered as computationally cheaper and more general, wrapper approaches are known for accuracy^[9]. Among the two approaches used to select features, namely, the filter approach and wrapper approach, the former is efficient in analyzing the high dimensional data.

The two familiar wrapper feature selection methods are SFS and SBS . SFS starts with no feature and then candidate features are added sequentially to the initial feature subset until further addition does not increase the classification performance. SBS begins with all features and then candidate features are removed sequentially from the initial feature subset until further removal does not decrease the classification performance. However, these two methods have a drawback that once a feature is selected, it cannot be removed (selected) later and this is called as the resting effect^[10]. This limitation can be overcome by combining both SFS and SBS into one algorithm. Therefore, the “plus-1-take – away- r” performs l times forward selection, followed by r times backward elimination. The challenge is to determine the optimal values (l,r).

Pudil^[11] has proposed two floating feature selection algorithms in order to address this limitation namely sequentially forward floating selection (SFFS) and sequential backward floating selection (SBFS). SFFS and SBFS determine the values for (l,r) automatically. These two floating methods are regarded to be at least as good as the best sequential method, while they also suffer from the problem of stagnation of local optima.

Within filter approach, there are various feature selection algorithms and they are classified into two kinds, feature weighting algorithm and subset search algorithm. Among the feature weighting algorithms, the Relief algorithm^[12] is well known. It weighs feature according to the ability to discriminate instances under different targets based on distance related criteria function. Nevertheless, the disadvantage with Relief is that it is ineffective in removing redundant features^[13]. Relief F^[14], which is an extension of Relief, although is a competent method to deal with noisy and incomplete data set as well as multiclass problems, still it cannot identify redundant features.

Decision trees (Dts) adopt the relevant features which are essential for the classification of the training set while it removes all other features which are not needed. Cardie^[15] has introduced a filter feature selection algorithm using Dts to select a subset of feature for a nearest neighbor algorithm. The focus algorithm^[16] which belongs to the filter approach, thoroughly searches all possible feature subsets and chooses the smallest one. Because of the focus algorithm’s exhaustive search, it becomes ineffective.

Fuzzy entropy enables the partition of the input space into decision regions in order to select appropriate features with good separability for the classification task. This concept has been defined in various ways and generalised in various applied fields such as communication theory, mathematics, statistical thermodynamics and economics. An algorithm based on discernibility matrix and information gain to reduce attributes, feature subset selection based on Fuzzy entropy measures for handling classification problems, and also efficient feature selection technique using supervised Fuzzy information theory was proposed by B. Azhagusundari et al^[17,18,19].

Jen – De- Shie et al^[20] has offered a method which deals with feature selection based on Fuzzy entropy measure for handling classification problems. It considers boundary samples while selecting

features. Thus Fuzzy entropy method measures the impurity of a feature and does not measure the purity of data sets. This feature selection method selects relevant features to get higher average classification accuracy. A method for innovative feature subset selection based on Fuzzy entropy measure was proposed by BehrouzMinaciet al^[21]. S.Sethuramalingam et al^[22] has presented a method on hybrid feature selection for network intrusion. In this paper, it deals with a new algorithm based on hybrid method which combines information gain and genetic algorithm in order to select features. Clustering is carried out on selected features for classification.

PSO has received more attention recently, as an EC technique, for solving feature selection problems. Since rough set tackles imprecision, uncertainty and vagueness, Wang et al^[23] proposes a filter feature selection algorithm based on an improved binary PSO and rough set theory. Fuzzy set can also show the dependence between features and class labels. Chakrobarthy^[24] compares the performance of PSO with that of GA in a filter feature selection algorithm with a fuzzy set based fitness function. The results show that PSO is more effective than GA in terms of the classification performance. Azevedo et al^[25], using PSO and a support vector machine (SVM), proposes a wrapper feature selection algorithm for personal identification in key stroke dynamic systems.

In standard PSO, gbest is updated only when a better solution is found. Chuang et al^[26] developed a strategy for gbest in PSO for feature selection in which gbest will be reset to zero if it maintains the same value after several iterations. Liu et al^[27] introduced a multi swarm PSO algorithm to search for the optimal feature subset and optimize the parameters of SVM simultaneously. Based on PSO, Unler et al^[28] propose a feature selection algorithm with an adoptive selection strategy where a feature is chosen, not only ascending to the likelihood calculated by PSO but also to its contribution to the features already selected. Unler et al^[29] also present a hybrid filter-wrapper feature subset selection algorithm based on particle swarm optimization (PSO) for support vector machine classification. Modified binary particle swarm optimization (MBPSO) method is proposed by SM Vieira et al^[30] for feature selection with the simultaneous optimization of SVM kernel parameter setting, applied to mortality prediction in septic patients. A novel spam deduction method was proposed by zang et al^[31] that focused on reducing the false positive error of mislabeling non spam as spam. In the above method they used wrapper based feature selection, decision tree classifier model c4.5, cost matrix, k-fold cross validation, the binary PSO with mutation operator respectively. HH Inbarani et al^[32] presents new supervised feature selection methods based on hybridization of particle swarm optimization. A novel feature selection approach for the classification of high dimensional cancer micro array data is proposed by B Sahu et al^[33] using filtering technique and particle swarm optimization.

III.METHODOLOGY

This section is discussed as follows. In section 3.1, Fuzzy entropy based feature selection method is explained. In section 3.2, Particle Swarm Optimization is discussed. Our proposed architecture and algorithm is explained in section 3.3.

3.1 FUZZY ENTROPY BASED FEATURE SELECTION METHOD

In the outcome of random experiment entropy is a measure of uncertainty. The concept of fuzzy entropy is an extension to Shannon entropy in which the assessments of entropies are facilitated by the fuzzy sets. Based on the Fuzzy C-means clustering algorithm (FCM)^[34], fuzzy entropy is estimated by computing the membership. The fuzzy C-means algorithm proposed by Trivedi and Bezdeck^[35] is given by

$$L_m(U, v) = \sum_{k=1}^n \sum_{i=1}^c (U_{ik})^m |x_k - v_i|_A^2 P(x_j) \quad \text{Equ. 1}$$

$$\text{Where } u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_i\|_A}{\|x_k - v_j\|_A} \right)^{2/m-1}} \quad \forall i, k \quad \text{Equ. 2}$$

$$v_i = \frac{\sum_{k=1}^n (u_{ik})^m x_k}{\sum_{k=1}^n (u_{ik})^m} \quad \text{Equ. 3}$$

The membership of each feature in all clusters is computed by the FCM algorithm instead of each pattern. The clusters are represented by centroid. Rather than calculating probability of a specific feature which will be in each specific class and then normalizing them, we have the actual membership of this feature in the classes that also sum up to one. So the new term called match degree D_c is given by

$$D_c = \frac{\sum_{x_d \in c} U_c(x_d)}{\sum_{x_d \in C} U_c(x_d)} \quad \text{Equ. 4}$$

The fuzzy entropy of the elements of class c is given by

$$FE_c = -D_c \log D_c \quad \text{Equ. 5}$$

To compute $H(X)$, we have to compute fuzzy entropy along the universal set

$$FE = \sum_{c=1}^c FE_c \quad \text{Equ. 6}$$

A threshold r is set to remove those features that have no major contributions. Thus features are selected using Fuzzy entropy based on threshold value.

3.2. PARTICLE SWARM OPTIMIZATION (PSO)

Particle swarm optimization (PSO) is a population based stochastic optimization technique, introduced by Kennedy and Eberhart in 1995. PSO simulates the social behavior of bird flocking and fish schooling^[36]. In PSO each single candidate solution is ‘‘an individual bird of the flock’’, that is a particle in the search space. Each particle makes use of its individual memory and knowledge gained by the swarm as whole to find the best solution. Fitness values of all particles are evaluated by a fitness function to be optimized and the movement of the particles are directed by the velocities of all particles. During the iterative search process in the d -dimensional solution space, each particle (i.e., candidate solution) will adjust its flying velocity and position according to its own flying experience as well as those of the other companion particles in the swarm.

The initial swarm is generally created in such a way that the population of the particles is distributed randomly over the search space. Each particle is updated by following two ‘best’ values called $pbest_{pd}$ and $gbest_d$ at every iteration. Each particle keeps track of its coordinates in the problem space which are associated with the best solution (fitness) the particle has achieved so far. This fitness value called $pbest_{pd}$ is stored. The ‘best’ global position found over all trajectories travelled by the particles of the swarm is called $gbest_d$. During the search process, the particles move according to the following equations:

$$V_{pd}^{new} = W \times V_{pd}^{old} + c_1 \times rand_1 \times (pbest_{pd} - x_{pd}^{old}) + c_2 \times rand_2 \times (gbest_d - x_{pd}^{old}) \quad \text{Equ. 7}$$

$$X_{pd} = X_{pd} + v_{pd}$$

Equ. 8

Where $rand_1$ and $rand_2$ are random numbers between(0,1) so as to provide a stochastic weighting of the different components participating in the particle velocity definition. c_1 and c_2 are acceleration factors, usually $c_1 = c_2 = 2$. The inertia weight w is used as a tradeoff between the global and local exploration capabilities of the swarm. Large values of this parameter permit better global exploration, while small values lead to a fine search in the solution space. Particle velocities of each dimension are tried to a maximum velocity V_{max} . If the sum of velocities causes the total velocity of that dimension to exceed V_{max} , then the velocity of that dimension is limited to V_{max} . V_{max} is a user specified parameter. Equation 7 allows the computation of the velocity at iteration $T + 1$ for each particle in the swarm by combining linearly its current velocity (at iteration T) and the distances that separate the current particle position from its best previous position and the best global position, respectively. The particle position is updated with equ. 8. Both equ. 7 and equ. 8 are iterated until convergence of the search process is reached. Typical convergence criteria are based on the iterative behavior of the best value of the adopted fitness function(s) and simply on a user-defined maximum number of iterations.

3.3. PROPOSED WORK

This paper proposes an optimal feature selection algorithm using Fuzzy entropy and particle swarm optimization concept in high dimensional data. Features can be extracted using the concept of fuzzy entropy using fuzzy C-means algorithm and Symmetrical Uncertainty. Particle swarm optimization algorithm is one of the nature inspired optimization algorithm which can produce competitive solutions for a wide variety of problems. This work combines both fuzzy entropy and Symmetrical Uncertainty in the first Phase and particle swarm optimization in the second phase. It is also proved that this approach can give significant results in selecting features, particularly in the case of high dimensional data.

Figure 1 shows the proposed architecture for feature selection based on Fuzzy Entropy and PSO.

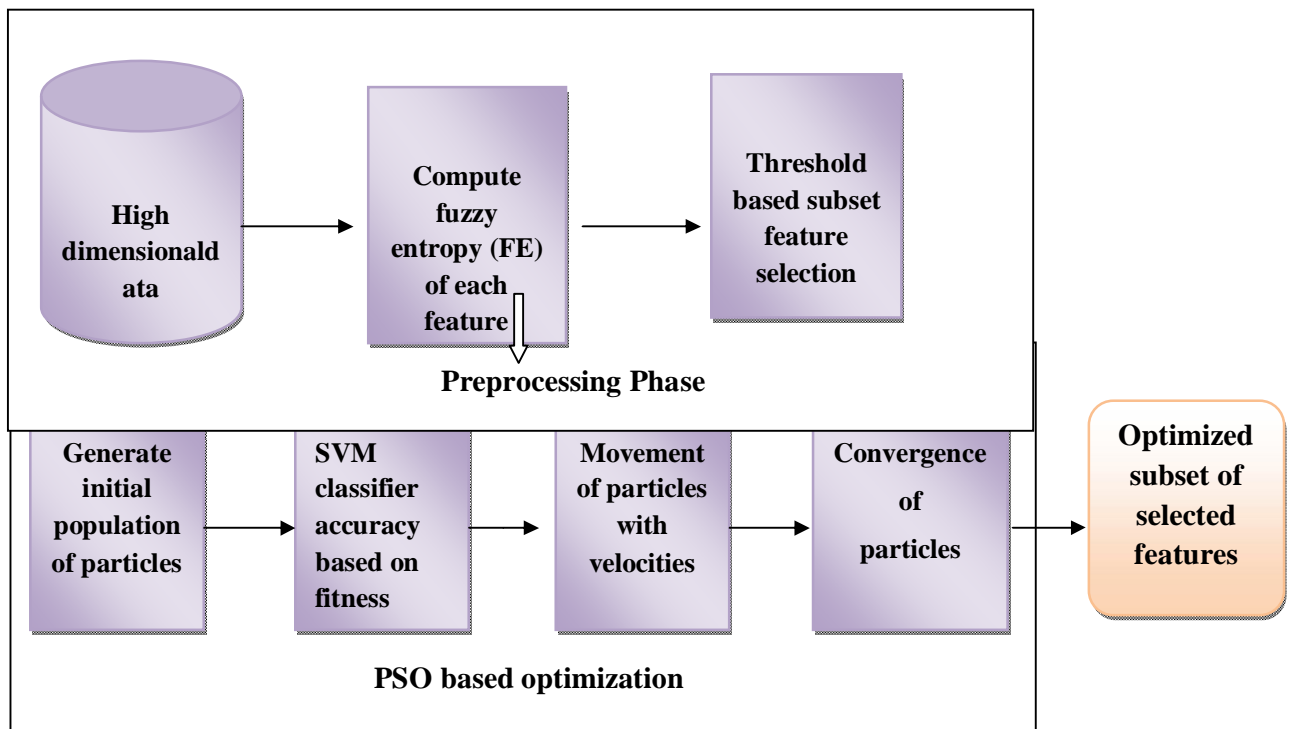


Figure1.proposed architecture for Feature Selection based on Fuzzy Entropy and PSO concept

It includes Preprocessing Phase and PSO based Optimization phase. Here, output of preprocessing phase is further processed by PSO based optimization phase. Finally, optimized subset of selected features is generated. The following is the Pseudo code of our proposed optimized Feature selection algorithm using Fuzzy Entropy and PSO.

Pseudo code of FMI-PSO Algorithm

Procedure

Input – N – set of all features

δ – threshold value

Begin

For $i = 1$ to N do begin

 calculate $FE_{i,c}$ for f_i ;

 if ($FE_{i,c} \geq \delta$)

 append f_i to S_1^1 list.

End;

For $i = 1$ to N do begin

 calculate $SU_{i,c}$ for each f_i ;

 if ($SU_{i,c} \geq \delta$)

 append f_i to S_2^1 list.

S^1 list $\leftarrow S_1^1 \text{ list} \cap S_2^1 \text{ list}$.

 If $S^1 \text{ list} = \emptyset$

$S^1 \text{ list} = S_1^1 \text{ list}$

End;

 Initialize population

$x_p (i = 1 \dots n)$ where $x_p \in$ subset of randomly selected features from S^1_{list} .

While (number of generations or the stopping criterion is not met)

 For $p = 1$ to number of particle

 If the fitness of X_p is greater than the fitness of $pbest_p$

 then update $pbest_p = X_p$

 For $K \in$ Neighbourhood of X_p

 If the fitness of X_k is greater than that of $gbest$ then

 Update $gbest = X_k$

 Next K

 For each dimension d

$$V_{pd}^{new} = W \times V_{pd}^{old} + c_1 \times rand_1 \times (pbest_{pd} - x_{pd}^{old}) + c_2 \times rand_2 \times (gbest_d - x_{pd}^{old})$$

 If $V_{pd} \notin (V_{min}, V_{max})$ then

$$V_{pd} = \max(\min(V_{max}, V_{pd}), V_{min})$$

$$X_{pd} = X_{pd} + V_{pd}$$

 Next d

 Next p

Next generation until stopping criterion

End

Given a data set with N features and a class c, the algorithm finds a set of predominant features S_{best} for the class concept. It consists of two parts. In the first part, it calculates fuzzy entropy value of each feature, selects relevant features into S_1^1 list based on the predefined threshold δ . Further it calculates the Symmetrical Uncertainty value of each feature, selects relevant features in to S_2^1 list. The common features from S_1^1 list and S_2^1 list constitutes S^1 list. If $S^1 \text{ list} = \emptyset$ (i.e if there are no common elements in S_1^1 list and S_2^1 list), then S^1 list contains features selected from S_1^1 list. In the second part, each particle represent a single bird. Select randomly the number of particles from S^1 list and allow it to move in the search space. Based on the rules of particle swarm optimization, particles

update their velocities and positions. Each particle renewal is based on its adaptive value. The best adaptive of each particle renewal is $pbest$ and the best adaptive value with in a group of $pbest$ is $gbest$. Once $pbest$ and $gbest$ are obtained, we can keep track of the features of particles with regard to their position and speed. After the particles reach the global best position, it still explore the search space for further search. The particles will stop after it found the optimal solution with highest fitness value. The particle which gives the highest fitness value with N number of features will be S_{best} .

IV. RESULT ANALYSIS OF OFEPSO BASED FEATURE SELECTION

4.1. EXPERIMENTAL DATA SETS

In this section the performance of our FMI-PSO algorithm is compared with the existing FCBF algorithm using four different data sets (LC, CTG, ORL and COLON). LC data set contains multivariate data set about lung cancer with number of instances 32 and number of integer attributes 56. CTG (Cardiotomography) is a monitoring of fetal heart rate and uterine contractions database consists in total of 552 intrapartum recordings with 21 attributes, which were acquired between April 2010 and August 2012 at the obstetrics ward of the University Hospital in Bruno, Czech Republic. ORL data set includes the database of faces contains set of face images taken between April 1992 and April 1994 at the lab. The database was used in the context of a face recognition project carried out in collaboration with the speech vision and robotics group of the Cambridge university Engineering department. The COLON tumor data set contains the expression of the 2000 genes with highest minimal intensity across the 62 tissues. The identity of the 62 tissues is given in file tissues. The numbers correspond to patients, a positive sign to a normal tissue, and a negative sign to a tumor tissue.

The results of all the four data sets with respect to Sensitivity, Specificity, Accuracy, and Features selected are analyzed in this section.

4.2 Accuracy Measure Analysis

Accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined.

$$Accuracy = \frac{\text{number of true positives and no of true negatives}}{\text{number of true positives} + \text{false positives} + \text{false negatives} + \text{true negatives}}$$

The following Figure 2 shows the comparative analysis of accuracy of our FMI-PSO algorithm against FCBF algorithm.

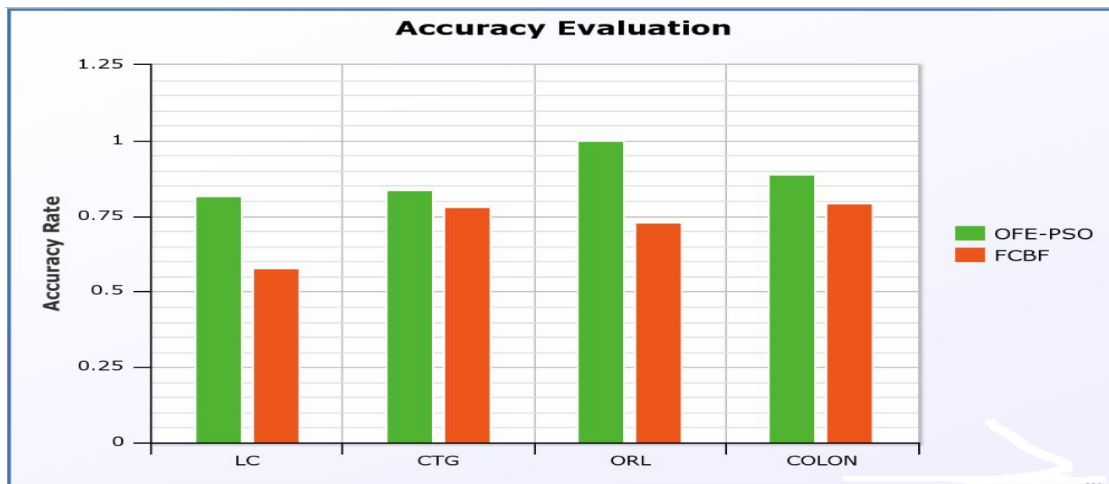
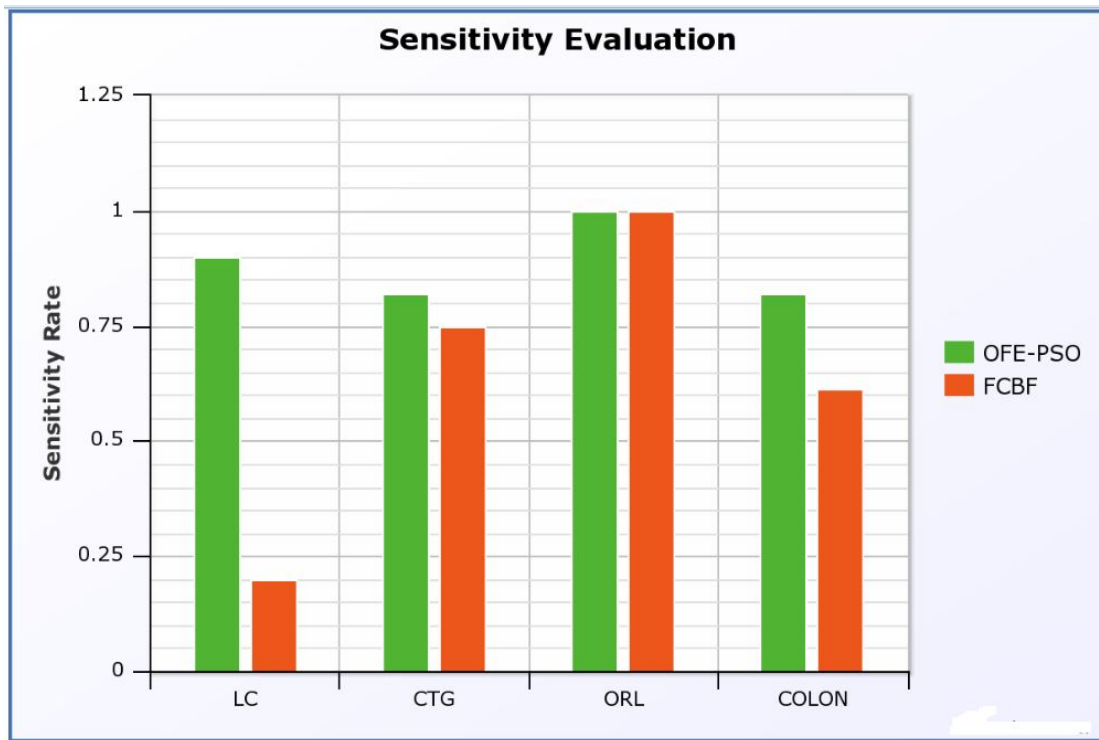


Figure 2: Accuracy Ratio**4.3 Sensitivity Measure Analysis**

Sensitivity also called the true positive rate, measures the proportion of positives which are correctly identified as such

$$\text{sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}$$

The following Figure 3 shows the comparative analysis of sensitivity ratio of our FMI-PS algorithm against FCBF algorithm.

**Figure 3: Sensitivity ratio (FMI-PSO and FCBF)****4.4. Specificity Measure Analysis**

Specificity (also called the true negative rate) measures the proportion of negatives which are correctly identified as such

$$\text{specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}}$$

The following Figure 4 shows the comparative analysis of specificity ratio of our FMI-PSO algorithm against FCBF algorithm.

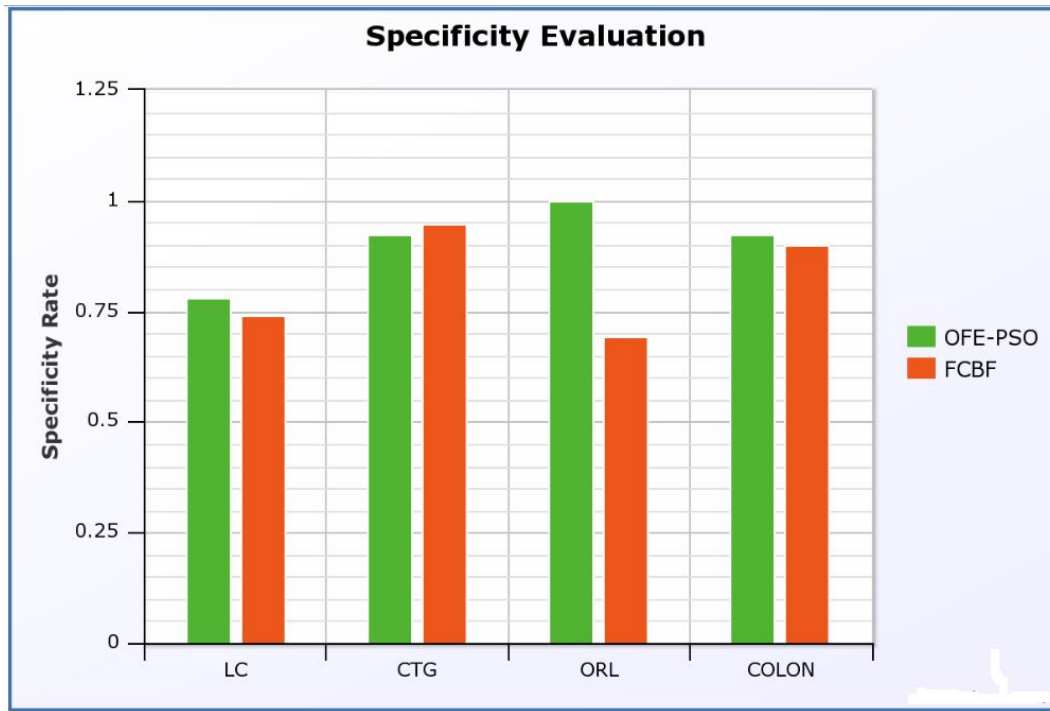
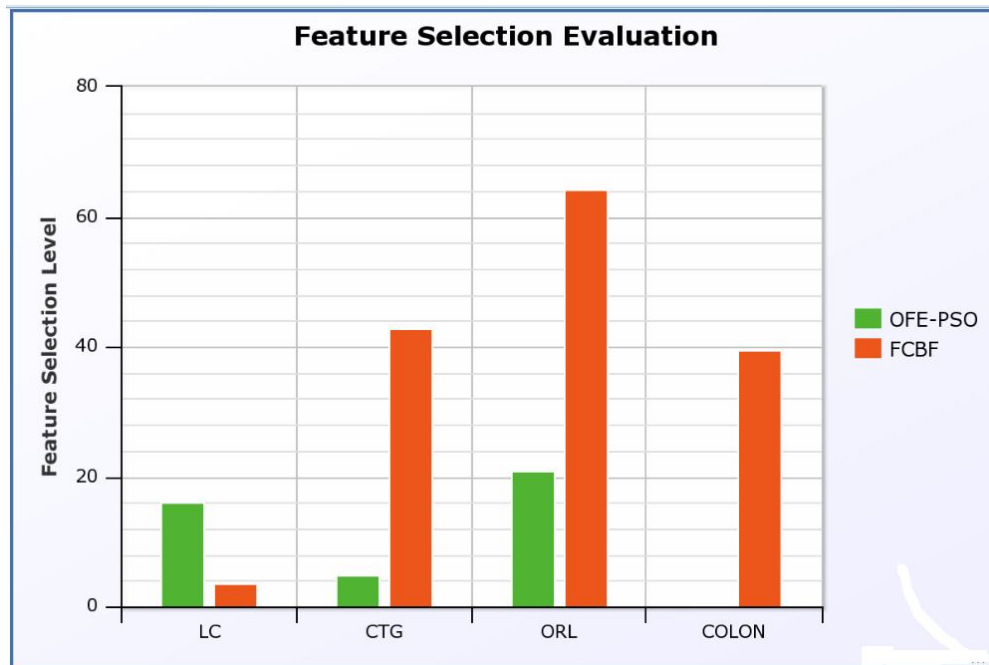


Figure 4: Specificity Ratio

4.5. Feature Selected Analysis

The following Figure 4 shows the comparative analysis of feature selection process.



4.6 Critical Findings & Observations

The results of all the four data sets with respect to Specificity, Sensitivity, Accuracy and Features selected are given in table 1.

		OFE PSO)	FCBF
LC	Sensitivity	0.9	0.2
	Specificity	0.7803	0.74242
	Accuracy	0.818	0.57721
	Features Selected in percentage	16.07	3.57
CTG	Sensitivity	0.81961	0.74873
	Specificity	0.92314	0.94582
	Accuracy	0.83531	0.77859
	Features Selected in percentage	4.76	42.86
ORL	Sensitivity	1	1
	Specificity	1	0.6929
	Accuracy	1	0.72846
	Features Selected in percentage	20.99	64.20
COLON	Sensitivity	0.822	0.61364
	Specificity	0.925	0.9
	Accuracy	0.8876	0.79385
	Features Selected in percentage	0.25	39.6

Table 1. Comparative Analysis of OFE PSO and FCBF

V .SUMMARY

It is very significant to build efficient classification model for classification problems with high dimensional data sets. Selection of the features and selection of the classification methods are the important tasks. This paper presents an optimized algorithm for feature subset selection based on fuzzy entropy and particle swarm optimization. Experimental results show that the above method proposed, simplified both the feature selection and the total number of parameters needed, effectively, thereby obtained higher classification accuracy compared to the other feature selection method. Thus FMIPSO algorithm not only selects relevant features but also improves the performance by discarding redundant, noisy, corrupted and unimportant features. The proposed algorithm can be compared with other meta heuristic algorithms and the results can be analyzed.

REFERENCES

- [1]. Guyon. I and Elisseeff. A, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, March 2003, vol. 3, pp. 1157–1182.
- [2]. Whitney.A, "A direct method of nonparametric measurement selection", *IEEE Trans. Comput.*, September 1971, vol. C-20, no. 9, pp. 1100–1103.
- [3]. Marill.Tand Green. D, "On the effectiveness of receptors in recognitionsystems," *IEEE Trans. Inf. Theory*, January 1963, vol. IT-9, no. 1, pp. 11–17.
- [4]. Kennedy.J andEberhart. R, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Netw.*, 1995, vol. 4, pp. 1942–1948.

- [5]. Shi .Y and Eberhart.R, “A modified particle swarm optimizer,” in *Proc.IEEE Int. CEC*, 1998, pp. 69–73.
- [6]. Unler. A and Murat.A, “A discrete particle swarm optimization methodfor feature selection in binary classification problems”, *ur. J. Oper. Res.*,November.2010, vol. 206, no. 3, pp. 528–539.
- [7]. Liu.Y, Wang.G, Chen.H, Dong.H, “An improved particle swarmoptimization for feature selection,” *J. Bionic Eng.*, June 2011, vol. 8, no. 2, pp. 191–200.
- [8]. Mohemmed.A, Zhang. M, and Johnston. M, “Particle swarm optimizationbased AdaBoost for face detection,” in *Proc. IEEE CEC*, 2009,pp. 2494–2501.
- [9]. Kohavi . R, and John. G.H , “Wrappers for feature subset selection,” *Artif.Intell.*, December 1997, vol. 97, no. 1/2, pp. 273–324.
- [10]. Yusta. S.C, “Different metaheuristic strategies to solve the feature selectionproblem,” *Pattern Recognit. Lett.*, April 2009, vol. 30, no. 5, pp. 525–534.
- [11]. Pudil.P, Novovicova.J, and Kittler.J.V, “Floating search methods infeature selection,” *Pattern Recognit. Lett.*, November 1994, vol. 15, no. 11, pp. 1119–1125.
- [12]. Kira. K and Rendell L.A, “A practical approach to feature selection,” in*Proc. 9th Int. Workshop Mach. Learn.*, 1992, pp. 249–256.
- [13]. Koller.D, and SahamimmM , “Toward Optimal Feature Selection,” *Proc. Int’l Conf. Machine Learning*, 1996, pp. 284-292.
- [14]. Kononenko. I, “Estimating Attributes: Analysis and Extensions of RELIEF,” *Proc. European Conf. Machine Learning*, 1994, pp. 171-182.
- [15]. Cardie.C , “Using decision trees to improve case-based learning,” in *Proc.10th ICML*, 1993, pp. 25–32.
- [16]. Almuallim.H ,Dietterich. T.G, “Learning Boolean concepts in thepresence of many irrelevant features,” *Artif. Intell.*, September 1994,vol. 69, no. 1/2,pp. 279–305.
- [17]. Azhagusundari.B, AntonySelvadosThanamani, “Feature selection based on Information Gain”, *Innovative Technology and Exploring Engineering* ,Vol 2,Issue 2 ,January 2013.
- [18]. Azhagusundari.B, Antony SelvadosThanamani, “Feature selection based on Fuzzy Entropy”, *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, ISSN 2278-6856, March - April 2013, Volume 2, Issue 2, pp.30-34.
- [19]. AzhaguSundari .B and Antony SelvadosThanamani, “An Efficient Feature Selection Technique using Supervised Fuzzy Information Theory”, *International Journal of Computer Applications* (0975 – 8887), January 2014, Volume 85, No 19.
- [20]. Jen-Da Shie. Shyi-Ming Chen (Feb 2007), Feature subset selection based on fuzzy entropy measures for handling classification problems, *ApplIntell* ,2008, 28,pp 69–82.
- [21]. Hamid Parvin, BehrouzMinaeiBidgoli,hosseinghaffarin , An Innovative Feature Selection Using Fuzzy Entropy, *Advances in Neural Networks – ISNN 2011*.
- [22]. Sethuramalingam. S, Naganathan E.R, “Hybrid Feature selection for Network Intrusion Detection”, *International Journal Of Computer Science and Engineering*, 2011, Volume 3, issue 5, pp.1773-1780.
- [23]. Wang.X, Yang.Y, Teng.X, Xia.W, and Jensen.R, “Feature selectionbased on rough sets and particle swarm optimization,” *Pattern Recognit.Lett.*, March 2007,vol. 28, no. 4, pp. 459–471.
- [24]. Chakraborty.B , “Feature subset selection by particle swarm optimizationwith fuzzy fitness function,” in *Proc. 3rd Int. Conf. ISKE*, 2008, vol. 1,pp. 1038–1042.
- [25]. Azevedo. G, Cavalcanti.G, and Filho.E, “An approach to feature selectionfor keystroke dynamics systems based on PSO and feature weighting,” in*Proc. IEEE CEC*, 2007, pp. 3577–3584.
- [26]. Chuang L.Y, Chang.H.W, Tu. C.J and Yang. C.H, “Improved binaryPSO for feature selection using gene expression data”, *Comput. Biol.Chem.*, Feb. 2008, vol. 32, no. 1, pp. 29–37.
- [27]. Liu.Y, Wang.G, Chen.H, and Dong.H, “An improved particle swarmoptimization for feature selection”, *J. Bionic Eng.*, June. 2011,vol. 8, no. 2, pp. 191–200.
- [28]. Unler .A and Murat.A, “A discrete particle swarm optimization methodfor feature selection in binary classification problems”, *Eur. J. Oper. Res*November. 2010,vol.206,no.3, pp. 528–539.
- [29]. Unler.A, Murat.A, Chinnam.R.B, “ mr^2 PSO: a maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification”, October 2011, Elsevier ,Volume 181, Issue 20, pp.4625–4641.

- [30]. Vieira.SM., Mendonça. L.F, Farinha.GJ , “Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients,” Applied Soft Computing, August 2013, Elsevier, Volume 13, Issue 8, pp. 3494–3504 .
- [31]. Y Zhang., S Wang, P Phillips, G Ji , “Binary PSO with mutation operator for feature selection using decision tree applied to spam detection” - Knowledge-Based Systems, 2014 , Elsevier , Volume 64, July pp. 22–31.
- [32]. HH Inbarani, AT Azar, G Jothi., “Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis”, January 2014, Elsevier, Volume 113, Issue 1, pp. 175–185.
- [33]. Sahu. B, Mishra. D, “A Novel Feature Selection Algorithm using Particle Swarm Optimization for Cancer Microarray Data”, 2012, Elsevier, Volume 38, pp. 27-31.
- [34]. Rami N.Khushaba, Adel Al-Jumaily and Ahmed Al-Ani, ”Novel Feature Extraction Method On Fuzzy Entropy and Wavelet Packet Transform for Myoelectric Control”, 2007 International Symposium on communications and information Technologies, pp 352-357.
- [35]. Trivedi .M.M and Bezdeck.J.C, "Low-level segmentation of aerial images with fuzzy clustering", IEEE Transactionson Systems, Man and Cybernetics, vol. SMC-16, 1986, pp. 589-598.
- [36]. Kennedy. J and Eberhart. R.C, *Swarm Intelligence*. San Mateo, CA:Morgan Kaufmann, 2001.