

# Feature Selection with Centre of Gravity Method using Ant Colony Optimization



Leena C Sekhar, R Vijayakumar, M K Sabu

**Abstract:** The high dimensional dataset with irrelevant, redundant and noisy features has much influence on the performance of machine learning problems. In this work, an existing Ant Colony Optimization (ACO) based feature selection algorithm is modified by attaching a dimensionality reduction method as a data pre-processing step. This is achieved by introducing the concept of Centre of Gravity (CoG) of a set of points. After reducing the dimension, the ACO algorithm is used to generate the optimal subset of features. The performance of the proposed algorithm is evaluated using Artificial Neural Network (ANN) classifier. The performance comparison using various dataset shows that the proposed method outperforms the existing ACO based feature selection methods.

**Keywords:** Ant Colony Optimization, Centre of Gravity, dimensionality reduction, feature selection

## I. INTRODUCTION

The process of selecting relevant feature subset from a dataset for building strong learning models is known as the feature selection (FS) technique. In data pre-processing, FS is used as a method for reducing the dimension of the dataset. The high dimension of the dataset greatly affects the performance of machine learning algorithms. The existing Ant Colony Optimization (ACO) [1] can be considered as a suitable algorithm for selecting the salient features from a dataset. The ACO algorithm is implemented by a completely connected graph called the construction graph. The nodes in the graph represents features and the edges represent the choice of selecting next feature. The best possible feature subset is then obtained by the traversal of the ant through the graph by visiting minimum number of nodes and satisfying the stopping criterion. In this work, initially, the original high dimensional dataset is modified by reducing the number of records considerably. This is achieved by applying a Mathematical concept of Centre of Gravity (CoG) of a set of points. CoG is considered as a representative vector for a group of objects. Here we consider a dataset with  $k$  classes and  $n$  objects. These  $n$  objects are grouped according to the class attribute value. After this grouping, the objects in each

group are further divided into several slices by selecting a suitable size for the slice. Each of these slices are then replaced with the corresponding CoG of that class. This is repeated for all  $k$  classes and the outcome of this process is a modified dataset with a considerable reduction in the number of records. This modified dataset is then used for feature selection. For selecting significant features from the modified dataset, ACO algorithm is used and artificial neural network (ANN) classifier is used to measure the significance of selected features. To compare the efficiency of the proposed method, from the original dataset features are selected using the same ACO algorithm. The selected features are then evaluated with the help of ANN classifier used in the proposed method. The experimental result show that the dataset modification has a significant effect in selecting the features efficiently. The experiment is performed on six different datasets from UCI machine learning repository. The remaining sections of the paper is organized as follows. In section 2, a review of the feature selection procedure is given. Section 3 briefly overview the dimensionality reduction based on CoG. Feature Selection using Ant Colony Optimization is outlined in section 4. Section 5 illustrates the proposed work. Experimental results and Analysis is given in section 6. Discussion is given in section 7 followed by a conclusion.

## II. REVIEW OF LITERATURE

Feature selection problem is closely related to ‘the curse of dimensionality’. Before performing any datamining task, redundant features are to be removed from the given dataset in order to get a better result. The main objective behind feature selection is to map the original high dimensional dataset to a dataset with manageable size by preserving the semantics of the data. The end result is the better performance of the machine learning algorithms when they are used in generating knowledge. FS is considered as a pre-processing technique for the dimension reduction of a dataset. To determine the quality of each feature that is to be included in the feature subset, FS algorithms are classified into filter based approaches [5], wrapper based approaches [6] and hybrid approaches [7]. A predefined training algorithm is used to evaluate the quality of features and the corresponding classification accuracy is determined in wrapper approach. In filter approach, the FS process is independent of machine learning algorithms. The properties of both filter and wrapper methods are used in hybrid approach. The review of various feature selection methods are given in this section. In order to perform non-correlated feature selection, Naraei and Sadeghian [3] used principal component analysis (PCA) and ANN classifier.

Manuscript published on 30 September 2019

\* Correspondence Author

Leena C Sekhar\*, Department of Computer Application, M E S College, Marampally, Aluva, Ernakulam, Kerala, India. email: [csleena11@gmail.com](mailto:csleena11@gmail.com)

R Vijayakumar, School of Computer Science, Mahatma Gandhi University, Kottayam, Kerala, India, email : [vijayakumar@mgu.ac.in](mailto:vijayakumar@mgu.ac.in)

M K Sabu, Department of Computer Application, Cochin University of Science and Technology, Kochi, Kerala, India, email : [sabumk@cusat.ac.in](mailto:sabumk@cusat.ac.in)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

The analysis has been performed on traumatic brain injury patients and the results shows good accuracy with a less mean absolute error. Nasution *et al.* [4] used a feature reduction frame work to simplify high dimensional data to a dataset with less dimension by using PCA. Information Gain [8] is one of the earliest measures used to rank attributes for FS and is measured based on the entropy of the system. Attributes that provide a better gain, receive a higher score and attributes that fall under a prespecified threshold are removed and the rest are kept.

Particle Swarm Optimization (PSO) [9] is an effective population-based method for feature selection problems. A binary PSO algorithm based on rough set theory was proposed by X Wang *et al.*[10] for feature reduction. Here the rough set dependency degree is used to evaluate each particle which represent prospective solution. Shunmugapriya and Kanmanib [11] proposed a hybrid algorithm called AC-ABC Hybrid for optimizing features. A feature selection method based on PSO along with a learning algorithm was proposed by Huang and Dun [12] and the optimization of features is achieved by combining the discrete and continuous PSO methods. A Unler *et al.*[13] proposed a feature subset selection method by combining the filter and wrapper approaches based on mutual information and PSO. Unsupervised probabilistic feature selection using ACO is proposed by Z D Behrouz *et al.*[14] in which the inter-feature information is utilized which reduces the redundancy in the final feature set. A criterion called maximum projection and minimum redundancy feature selection for unsupervised learning is proposed by W Shiping *et al.* [15]. Here FS is formalized with the use of projection matrices and then characterized equivalently as a matrix factorization problem. A novel feature selection algorithm based on ACO [16] consider the performance of the classifier and length of the feature vector as heuristic information to ACO. A novel feature selection technique based on rough sets and ACO is given by Yumin Chen *et al.* [17] which starts from the core and uses mutual information based feature significance to search through the feature space for optimal solutions. A graph clustering approach is introduced by Parham and Mehrdad [18] for feature selection. FS with fuzzy criteria are given by M Susana *et al.* [19]. Here the fuzzy objective functions are proven to be more accurate than classical objective functions.

### III. DIMENSIONALITY REDUCTION BASED ON COG

The efficiency of machine learning algorithms and analysis of the relationship between data or features are much influenced by the high dimension of the dataset. The process of reducing the dimension of the dataset is called dimensionality reduction. It is a pre-processing step in datamining to improve computational efficiency and accuracy. In this paper, an existing ACO based feature selection algorithm is modified by attaching a method to reduce the size of the dataset as a data pre-processing step. To reduce the size of the dataset, the concept of Centre of Gravity (CoG) of a set of points is introduced. The CoG is a representative vector for a group of objects. The CoG vector representing a set of points / slice  $\{o_1, o_2, \dots, o_n\}$  of a given class  $r$  is obtained by selecting all

feature values  $o_x(i), i = 1, 2, \dots, p$  from the selected set of objects which minimizes the following sum of differences.

$$\left| \sum_{j=1}^n (o_j(i) - o_x(i)) \right|, x = 1, 2, 3, \dots, n$$

ie

$$\text{CoG}(o_1, o_2, \dots, o_n) = (c_1, c_2, \dots, c_p)$$

where  $c_i = o_x(i)$  minimizing

$$\left( \left| \sum_{j=1}^n (o_j(i) - o_x(i)) \right| \right), x = 1, 2, 3, \dots, n \quad (1)$$

### IV. FEATURE SELECTION USING ANT COLONY OPTIMIZATION

In the area of machine learning, FS is an important problem. In real-world problems, because of the abundance of noisy, irrelevant or misleading features, it is necessary to perform feature selection. The motivation of FS in data mining and machine learning algorithms is to reduce the dimensionality of feature space [2]. ACO is considered as a modern algorithm for selecting significant features from a dataset. ACO was introduced by M Dorigo *et al.* [1] in 1992. It is an optimization technique and takes inspiration from the foraging behavior of ant colonies for their food. The communication between ants takes place through a volatile substance called pheromone. While moving towards the food source the ant leaves a pheromone trail on the ground. Since they can smell pheromone, when selecting their way, they select the path with high pheromone concentration. When an ant identifies a food source, it evaluates the quality and quantity of the food and during its return trip to the nest, the amount of pheromone that they deposit depends on the quality and quantity of the food. Thus the basic principle behind ACO is the pheromone laying and pheromone following behavior of real ants.

A construction graph shown in Fig.1 represents the problem domain for feature selection using ACO. In this graph, the features are represented by the nodes and choice of selecting the next feature by the distance between nodes. The computation of the shortest path with a minimum number of nodes and satisfying the optimality condition is equivalent to the search for the optimal feature set. In Fig.1 the ant is placed currently at node  $a_1$  and the ant has a choice of adding the next feature to it. The dotted lines represent the choice of selection. Using the probability rule given in (2), it adds features  $a_2$ , then  $a_3$  and finally  $a_4$ . When it reaches  $a_4$ , the current subset  $\{a_1, a_2, a_3, a_4\}$  is evaluated and reaches the stopping condition for ant traversal, ie, high classification accuracy is achieved with the feature subset. Once an ant completes the solution construction, the amount of pheromone to be deposited on the path is determined by that ant and is updated. This process is repeated until all iterations are completed and the best feature subset is generated.

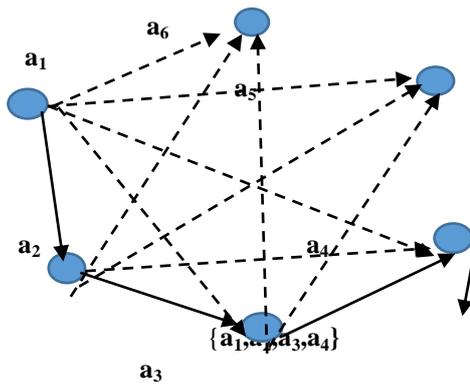


Fig 1. ACO problem representation

The probability rule [1] of an ant  $k$  at feature  $i$  is calculated using (2)

$$P_{ij}^k(t) = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha * [\eta_{ij}]^\beta}{\sum_{l \in N_i^k} [\tau_{il}(t)]^\alpha * [\eta_{il}]^\beta}, & j \in N_i^k \\ 0, & j \notin N_i^k \end{cases} \quad (2)$$

Where  $P_{ij}^k(t)$  is the probability of the  $k^{\text{th}}$  ant to move from node  $i$  to node  $j$ .  $N_i^k$  is the set of nodes in the neighborhood of the  $k^{\text{th}}$  ant in the  $i^{\text{th}}$  node.  $[\tau_{ij}(t)]^\alpha$  is the pheromone amount on the arc connecting node  $i$  and node  $j$ .  $[\eta_{ij}]^\beta$  is the heuristic value of the arc connecting node  $i$  and node  $j$ .  $\alpha$  and  $\beta$  are weight parameters that control the relative importance of pheromone and heuristic information.

When an ant made the solution after traversing the graph, the ant updates the pheromone on the edges of the selected path separately. The pheromone trail [20] of ant  $k$  at each edge  $(i, j)$  at time  $t + 1$  is updated as

$$\tau_{ij}(t+1) \leftarrow (1-\rho) \cdot \tau_{ij}(t) + \Delta\tau_{ij}(t), \quad 0 \leq \rho \leq 1 \quad (3)$$

where  $\Delta\tau_{ij}(t) = \sum_{k=1}^n (\gamma'(S^k) / |S^k|)$  (4)  
 $S^k$  is the feature subset constructed by  $k$  ants and  $\gamma'(S^k)$  is the classification performance.

## V. PROPOSED WORK

In this work, a novel ACO based FS process suitable for the multiclass high dimensional dataset is proposed. As the dimensionality of the dataset increases, data analysis becomes significantly harder. To reduce the dimension of the dataset horizontally a mathematical idea of CoG of a set of points is used in this proposed method. To apply this concept consider the high dimensional dataset having  $r$  classes  $(p_1, p_2, \dots, p_r)$ .  $(p_1, p_2, \dots, p_r)$  represents a partition of the Universe and each partition contains  $n$  objects  $\{x_1, x_2, \dots, x_n\}$ . Now, these  $n$  objects are further divided into  $q$  slices by selecting a suitable slice size  $c$ . If  $c$  is a divisor of  $n$ , then all the slices are of equal size otherwise the last slice is of less objects than  $c$ . By using the concept of CoG given in (1), each of these slices are replaced with the corresponding CoG of that slice. This process is repeated for all  $r$  classes and thus the horizontal dimension of the data set gets reduced considerably. After this slice reduction, the number of records in the modified dataset becomes  $\sum_{i=1}^r c_i$ , where  $c_i$  represents the number of CoG vectors in class  $i$ .

Then for selecting the significant features, the reduced dataset obtained using CoG is given to the ACO algorithm. By considering the features and its values contained in the preprocessed dataset, feature selection is performed using ACO algorithm. The algorithm initially assigns  $p$  ants where  $p$  is the total number of features in the dataset. By considering these features, a construction graph is considered to represent the problem domain. An ant is then placed randomly at a given node (feature) in the construction graph and it has the choice of adding next feature to its path based on the probability rule given in (2). Each ant starts constructing the path at different feature in the construction graph. The subset of features thus generated are evaluated using the ANN classifier. The best solution obtained in the first iteration is compared with the global best solution and interchange with the global best if the generated is best. The procedure is repeated until it satisfies the predefined condition. If the condition fails, then the pheromone is updated and a set of new ants are generated and the process is repeated.

The proposed CoGACO algorithm:

**Input** : High dimensional dataset with  $r$  classes

**Output** : Optimal Feature set with significant features.

Step 1 : Input the original high dimensional dataset and size of slice

Step 2 : Modify the dataset by applying steps 3 to 6 until all groups are processed

Step 3: Group all records based on class attribute value.

Step 4 : Slice each group based on slice size value

Step 5 : For each slice, generate CoG vector using (1)

Step 6 : Replace each slice with the corresponding CoG vector

Step 7 : Initialize all parameters of ACO

Step 8 : Repeat steps 9 to 14 until all iterations are over

Step 9 : Place all ants randomly to all features

Step 10: Construct all solutions using probability rule using (2).

Step 11 : Evaluate the generated solutions using ANN classifier.

Step 12: Select the Best solution and compare with a global solution

Step 13 : Update pheromone trail refer to (3)

Step 14: If the generated solution is best, then store it as the global best.

Step 15 : Output the best solution

## VI. EXPERIMENTAL RESULTS AND ANALYSIS

The effectiveness of the proposed method is tested using six real-world data sets taken from the UCI machine learning repository.

## Feature Selection With Centre of Gravity Method using Ant Colony Optimization

The various properties of these datasets like name, number of attributes, number of classes and size/instance are given in Table I. Table II shows the dataset with reduced size after applying the CoG approach, the dataset with the original size and the percentage of reduction in the dataset occurred by applying the proposed method. After reducing the size of the dataset using the proposed approach, it is applied in the existing ACO algorithm to generate the optimal feature set. The number of features in the optimal feature set obtained using the proposed method and the number of features in the optimal feature set from the original dataset are tabulated in Table III. Table IV shows the classification accuracy of the feature subset generated from the original dataset and the classification accuracy of the feature set generated using the proposed approach. This result shows that our proposed method is more accurate than the existing methods of feature selection.

Fig.2 shows the pictorial representation of the classification accuracies of the feature subset generated from the original dataset and using the proposed method. The percentage of reduction occurred in the resultant dataset and the number of instances in both original and reduced dataset are given in Fig.3.

**Table I: Details of the dataset used**

Name of Dataset	No. of Attributes	No. of classes	Size of dataset
Breast cancer	10	2	683
Heart disease	13	2	304
Pima Diabetes	8	2	768
Ionosphere	35	2	351
Vehicle	19	4	846
Vowel	13	11	990

**Table II: Size of original and reduced datasets with percentage of reduction**

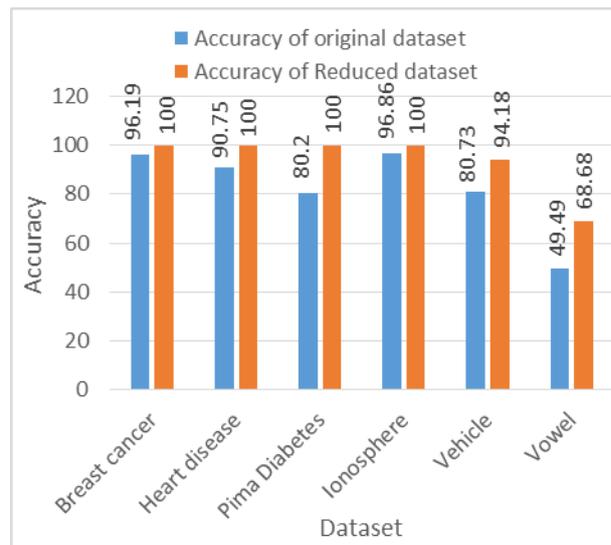
Name of Dataset	Size of Original dataset	Size of Reduced dataset	Percentage of reduction
Breast cancer	683	69	89.89
Heart disease	304	31	89.80
Pima Diabetes	768	77	89.97
Ionosphere	351	36	89.74
Vehicle	846	85	89.95
Vowel	990	99	90

**Table III: Number of features in the optimal feature subset of Original and Reduced dataset**

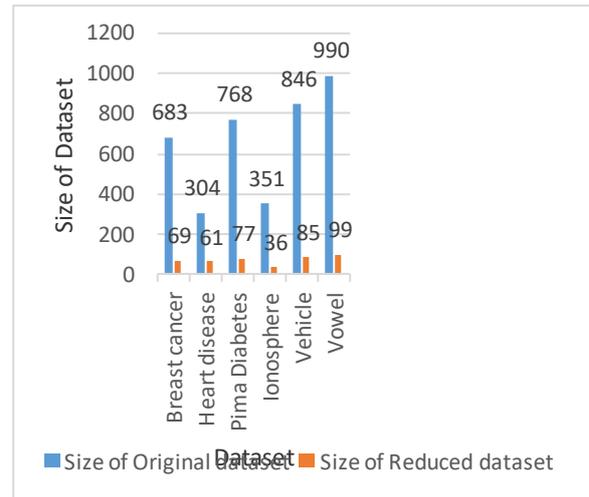
Name of Dataset	Number of features in the Original feature subset	Number of features in the Reduced feature subset
Breast cancer	6	5
Heart disease	9	4
Pima Diabetes	6	5
Ionosphere	19	13
Vehicle	14	12
Vowel	9	8

**Table IV: Classification accuracies of both reduced and original dataset**

Name of Dataset	Classification Accuracy of the original dataset	Classification Accuracy of the reduced dataset
Breast Cancer	96.19	100
Heart Disease	90.75	100
Pima Diabetes	80.20	100
Ionosphere	96.86	100
Vehicle	80.73	94.18
Vowel	49.49	68.68



**Fig. 2. Comparison of classification accuracies of original and reduced datasets**



**Fig. 3. Size of original and reduced dataset**

### VII. DISCUSSION

From the experimental results tabulated in Table IV and presented in Fig.2, it is observed that the classification accuracy of the proposed method is much greater than the accuracy obtained with the high dimensional dataset.



Using this proposed approach, from Table II and Fig.3, nearly 90% of data reduction is occurring for each of the dataset used in this experiment. This follows that the proposed method is appropriate in feature selection due to high classification accuracy.

### VIII. CONCLUSION

The efficiency of machine learning algorithms and analysis of the relationship between data or features are much influenced by the high dimension of the dataset. The process of reducing the dimension of the dataset is called dimensionality reduction. In this work, an existing ACO based feature selection algorithm is modified by attaching a dimensionality reduction method as a data pre-processing step. To reduce the size of the dataset, the concept of Centre of Gravity of a set of points is introduced.

This reduced dataset is then used for feature selection. For selecting significant features from this modified dataset, an ACO algorithm is used. To measure the significance of selected features an efficient ANN classifier is used. To compare the efficiency of the proposed method, features are selected from the original dataset using the same ACO algorithm and the selected features are evaluated using the ANN classifier. In this proposed method, we are getting high classification accuracy for the reduced dataset compared to the original dataset. The various dataset from the UCI repository is utilized in this work.

### REFERENCES

1. M Dorigo., T Stutzle, "Ant colony optimization", Encyclopedia of Machine Learning 2010 Edition.
2. H Liu., H Motoda , "Feature selection for knowledge discovery and data mining", Springer Science+ Business Media New York
3. P Naraei , and A Sadeghian, "A PCA based feature reduction in intracranial hypertension analysis", IEEE Int. Conf. on 30th Canadian Conference on Electrical and Computer Engineering (CCECE) pp 1-6.
4. M Z F Nasution, O S Sitompul and M Ramli, "PCA based feature reduction to improve the accuracy of decision tree c4.5 classification", 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE).
5. M Dash, and H Liu, "Feature selection for classification", Intelligent data analysis, vol. 1, pp. 131-156, March 1997.
6. I Guyon, A Elisseeff. , "An introduction to variable and feature selection", The Journal of Machine Learning Research vol. 3, pp. 1157-1182, March 2003.
7. H Liu, I Yu, "Toward integrating feature selection algorithms for classification and clustering", IEEE Transactions on Knowledge and Data Engineering vol. 17, pp. 491-502, April 2005.
8. J R Quinlan, "C4.5: Programs for Machine Learning", Morgan Kaufmann, San Francisco, 1993.
9. J Kennedy and R Eberhart, "Particle Swarm Optimization", Proceedings of the IEEE International Conference on Neural Networks, 4, 1942-1948.
10. .X Wang, J Yang,X Teng, W Xia and R Jensen, "Feature selection based on rough sets and Particle swarm optimization", Pattern Recognition Letters 28 (2007) 459-471.
11. P Shunmugapriya, S Kanmanib, "A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid)", Swarm and Evolutionary Computation 36 (2017) 27-36.
12. C L Huang and J F Dun, "A distributed PSO- SVM hybrid system with feature selection and parameter optimization", Applied Soft Computing 8 (2008) 1381-1391.

13. A Unler, A Murat, and R BChinnam, "mr<sup>2</sup>PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification", Information Sciences 181 (2011) 4625-4641.
14. Behrouz Zamani Dadaneh,Hossein Yeganeh Markid, Ali Zakerolhosseini, "Unsupervised probabilistic feature selection using ant colony optimization", Expert Systems With Applications 53 (2016) 27-42.
15. Shiping Wang, Witold Pedrycz, Qingxin Zhu, William Zhu. "Unsupervised feature selection via maximum projection and minimum redundancy", Knowledge-Based Systems, Volume 75, Issue C, February 2015, page 19-29.
16. .Hamidreza Rashidy Kanan, Karim Faez, "An improved feature selection method based on ant colony optimization (ACO) evaluated on face recognition system", Applied Mathematics and Computation 205 (2008) 716-725.
17. Yumin Chen, Duoqian Miao, Ruizhi Wang, "A rough set approach to feature selection based on ant colony optimization", Pattern Recognition Letters 31 (2010) 226-233
18. Parham Moradi, Mehrdad Rostami, "Integration of graph clustering with ant colony optimization for feature selection", Knowledge-Based Systems 84 (2015) 144-161.Susana M.Vieiraa, JoãoM.C.Sousaa, Uzay Kaymakb, "Fuzzy criteria for
19. feature selection", Fuzzy Sets and Systems 189 (2012) 1-18Shima Kashef, Hossein Nezamabadi, "An advanced ACO algorithm for feature subset selection", Neurocomputing 147(2015) 271-279.

### AUTHORS PROFILE



**Leena C Sekhar** completed her BSc degree from Newman College, Thodupuzha, Kerala, affiliated to Mahatma Gandhi University, Kottayam and MCA degree from Malnad College of Engineering, Hassan, Karnataka, affiliated to Mysore University in the year 1991 and 1995 respectively. She started her academic career as Lecturer in Department of computer science in the year 1996 and now working as associate professor, MES College, Marampally, Aluva, Kerala. Her research interest are Data Mining and Machine Learning



**Dr. R Vijayakumar** completed his BSc. Engineering degree and Ph.D from College of Engineering Trivandrum. He completed his MTech. From IIT Mumbai. He started his academic career in 1985 at NSS College of Engineering Palakkad, Kerala. Later he joined as Professor at School of Computer Sciences, Mahatma Gandhi University, Kottayam, Kerala. He served as the Director and Head of the school, Dean, Faculty of Engineering and Technology, Syndicate Member, Director, College Development Council etc of Mahatma Gandhi University. 10 scholars have been awarded Ph.D degree under his guidance and presented more than 100 research papers in referred journals and National and International conferences.



**Dr. Sabu M K** received his MCA degree from Government Engineering College Trichur, Kerala and Ph.D from Mahatma Gandhi University, Kottayam. He started his career in the year 1995 and presently working as Associate Professor and Head, Department of Computer Applications, Cochin University of Science and Technology, Kochi, Kerala. His research interest include Data Mining, Rough Set Theory, Optimization Technique and Machine Learning. He is having several research publications in referred journals and presented papers in National and International conferences.