IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
Vol. 3, Issue 5, May 2016

Agriculture Classification System Using Differential Evolution Algorithm

D. Asir Antony Gnana Singh[1], E. Jebamalar Leavline[2], V. Priyanka[3], V. Swathi[4]

Department of CSE, Anna University, BIT Campus, Tiruchirappalli, India[1]

Department of ECE, Anna University, BIT Campus, Tiruchirappalli, India[2]

Department of IT, Anna University, BIT Campus, Tiruchirappalli, India[3, 4]

Abstract: As it is known that the backbone the country's economy is agriculture. Nowadays, the youngsters are not giving importance to the agriculture since the agricultural firms are not fully computerized for enhancing productivity. Hence, the agricultural firms are aimed to computerize their operations in order to increase the productivity. This objective motivates this paper to develop an agriculture classification system with higher accuracy. Moreover, this paper proposes a deferential evolution technique with attribute selection to improve the accuracy in agriculture classification. This proposed system is tested on the various real world benchmark datasets. For evaluating the proposed system the different classification algorithms namely function based logistic classifier, instance classifier namely K-star and tree based logistic model trees (LMT) are used. The experimental results of the proposed system are promising with higher classification accuracy.

Keywords: Agriculture classification system, differential evolution algorithm, agriculture productivity.

I. INTRODUCTION

Agricultural classification systems enable the formers to predict or identify agriculture properties such as pasture, seed, leaf, growth of the crops and plants, diseases of the crops and plants, etc. The agriculture sectors contribute to the growth of the gross domestic product (GDP) of country and a considerable amount of people derive their livelihood from the agricultural sector. The application of information and communications technology (ICT) in agriculture is increasing day-by-day for improving the productivity. E-agriculture is an emerging field focusing on the enhancement of agricultural and rural development through improved information and communication processes. More specifically, e-agriculture involves the conceptualization, design, development, evaluation and application of innovative ways to use information and communication technologies (IT) in the rural domain, with a primary focus on agriculture.

Nowadays, the advancement of ICT and e-agriculture generate massive data related to the properties such as different attributes of pasture, various attributes of seeds, various attributes of leaf, various symptoms for the growth of the crops and plants, different test results for diagnosing the diseases of the crops and plants, etc. Therefore, classifying these massive data is a challenging task to develop the agriculture classification systems, since the irrelevant and redundant data reduces the classification accuracy in agricultural classification system. In the recent past, e-agriculture is extremely involved in predicting or identifying the agriculture properties such as pasture, seed, leaf, growth of the crops and plants, diseases of the crops and plants, etc. Therefore, the accuracy of the agricultural classification system plays a vital role in agriculture sector. The misprediction or misidentification leads to wrong decision in the process of agriculture sector that may lead to adverse effect on yield or crop losses.

Therefore, this paper aims to improve the accuracy in the agricultural classification system by removing the irrelevant and redundant attributes using the proposed differential evolution based variable selection method. This proposed method is tested on various real world agricultural datasets. The performance of the proposed method is also tested using different classification algorithm such as function based classifier namely library for support vector machines (LIBSVM), instance based classifier namely K-star, and tree based namely logistic model trees (LMT) in terms of classification accuracy with 10-fold cross validation test mode. The performance of the proposed method is also compared with the existing methods.

II. LITERATURE SURVEY

Differential evolution (DE) is an optimization method. It is capable of handling non differentiable, multimodal, and nonlinear objective functions. It is a simple, direct search, parallel, and easy to use method having excellent convergence and high-speed implementation properties [1]. The DE is used to solve the economic dispatch problem (ED) with transmission loss by satisfying the linear equality and inequality constraints [2]. Particle swarm optimization (PSO) proposed by Kennedy and Eberhart in 1997 simulates the behavior of bird flocks searching for targets for moving the particle around in the search space according to simple mathematical formulae of velocity and position. Differential evolution is better than particle swarm (PSO) in terms of solution quality, running time, chance of reaching to best solutions in variety of problems.

The quality of solution is most important factor of evolution algorithm. The runtime of the differential evolution algorithm is faster than particle swarm optimization [3]. The genetic algorithm was popularized by Holland, in early 1970's and further in the late 1980's by Goldberg [4]. Another value the genetic algorithm is that there can be many different constraints to the problem based on the specifics of the solution for which the searching genetic algorithm works and allows finding the solution in a fast manner. Differential evolution displays the better results than genetic algorithm [5]. Hence, differential evolution is a better algorithm than particle swarm optimization and the genetic algorithm.

Feature selection (FS) is an indispensable dimensionality reduction technique commonly used to reduce the data from the high-dimensional data. The FS techniques are used to build the models that describe the data. The reason behind using feature selection techniques include how to select a subset of attributes that are present in the high-dimensional space, removing irrelevant and redundant attributes, reducing the amount of data needed for learning, improving algorithm's predictive accuracy, and reducing the consumption of time in the construction of the models [6]. The existing evaluation measures utilized in feature selection techniques are divided into two categories namely filters and wrappers. Filter based feature selection methods are faster than wrapper based feature selection methods. The filter method can be adopted for the very high-dimensional space. When the number of attribute are considerable amount with less generality constrain the wrapper method is used [7].

Different types of seeds are categorized into classes on the basis of their morphological features that are called seed classification [8]. The classification of chickpea seeds varieties was made according to the morphological properties of chickpea seeds, by considering its 400 samples which includes its four varieties; Kaka, Piroz, Ilc, and Jam [9]. Pazoki and Pazoki presented a system to classify 5 varieties of rain fed wheat grain cultivar with artificial neural network to gain good accuracy [10]. Chen et al suggested a model as a combination of vision-based approach and pattern recognition techniques along with the neural networks to classify the five corn seed varieties [11]. Machine learning algorithms can be used to identify different varieties of wheat seeds to classify them according to their quality [12] and also the machine learning techniques are used to determine several growth periods of the crops such as cotton and classification of seed cotton [13].

The detection of diseased crops in agricultural environments is a challenging task to produce precise results in the agriculture. To detect the Ceratocystis wilts disease in eucalyptus crops, machine learning (ML) techniques are used to classify the three classes that is ground, healthy, and diseased plants [14]. Eucalyptus is the main forest species in some countries and it is planted in several areas [15]. The harmful effect of diseased crop damages the eucalypt plantations and made the economical losses [16]. Ghiasi and Amirfattahi proposed a segmentation framework to provide a safe landing using an aerial image. They manually extracted color and texture features from images, labeled each are asemantically and employed two K- nearest neighbor (NN) classifiers [17].

Most of the researches related to crop area estimation is associated with classification of landsat thematic mapper (TM) images of medium spatial resolution [18]. Several studies have been conducted using different conceptual approaches with high temporal-resolution data with coarser spatial-resolution. However, some methodologies did not prove to be useful to routine monitoring. Hence this paper presents crop area estimation with the moderate resolution imaging spectroradiometer (MODIS) crop detection algorithm (MCDA) [19]. David B et al presented and approach to estimate the subpixel functions for compute the crop area estimation with time series of MODIS data [20].

Classifying the plant leaves is a prime and complex task, especially for leaves, since the background of the leaves have some interferents and overlapping phenomena [21]. According to the theory of plant shape taxonomy, plants are basically classified according to the shapes of their leaves and flowers. Usually, leaves are approximately two-dimensional in shape and flowers are three-dimensional. It is difficult to analyse shapes and structures of flowers since they have complex three dimension (3D) structures [22]. Oide et al chose leaf shape images as neural networks input and applied hopfield model and a simple perceptron for soybean leaf classification [23]. Adopted accelerated douglas–peucker approximation algorithm for leaf shape approximation and used the modified dynamic programming algorithm for leaf shape matching [24].

E-agriculture is a growing field for enhancing existing agriculture and food security through enhanced processes for knowledge access and switch using information and communication technologies [25]. Peter Namisiko et al have proposed as a study which is conducted at majority of farmers who are not able to sell their product at market price due to lack of enough information available. And suggested the information communication technologies (ICT) tools for such farmers to produce their products and selling at market based competitive prices [26]. Marcel et discusses about service reuters market light (RML) that sent the short message service (SMS) to the former about the agricultural information [27]. Danasing Asir et al discussed some feature selection method [28-32].

III. PROPOSED SYSTEM

This paper proposes a differential evolution algorithm based feature selection for agriculture classification system. The workflow of this proposed system is illustrated in Figure 1. Initially, the real world dataset is given to the proposed differential evolution based feature selection algorithm. Then the selected features are obtained using the proposed algorithm. The selected features are given to the classifiers in order to develop the classification model. Finally, the classification model is evaluated on the test datasets.

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Vol. 3, Issue 5, May 2016

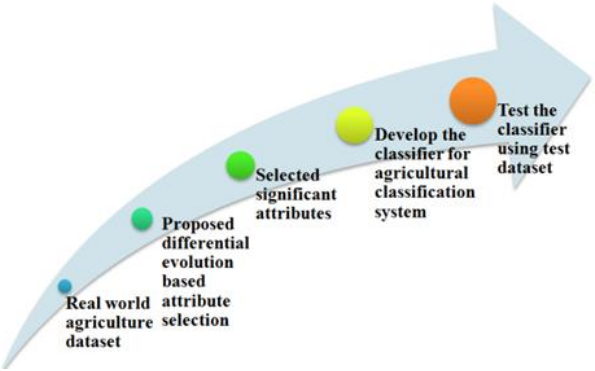

Fig.1. Workflow of proposed system

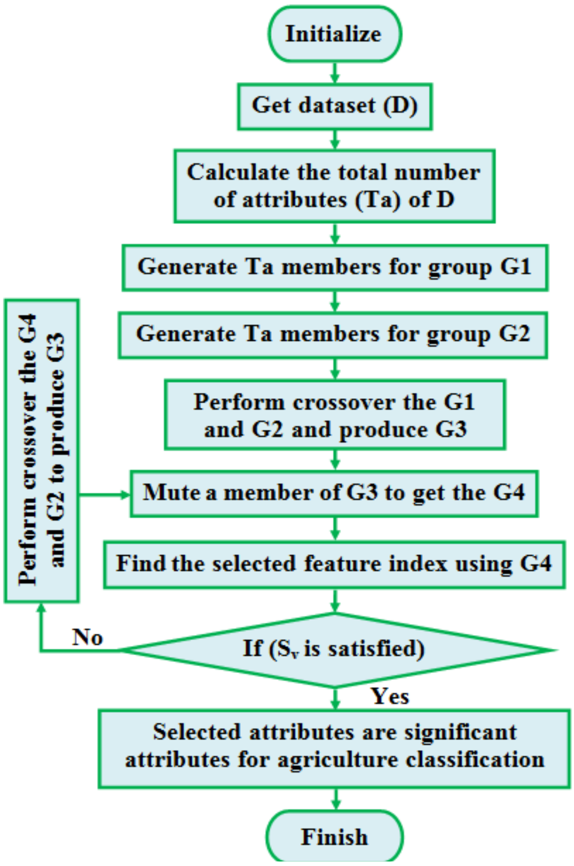

Fig.2. Flowchart representation of proposed algorithm

Algorithm 1: Differential evolution based attribute selection
Input: Dataset D
Output: Selected set of attributes

Step 1: Initialize
Step 2: Get dataset (D)
Step 3: Calculate the total number of attributes (T_a) of D
Step 4: Generate T_a members for group G_1
Step 5: Generate T_a members for group G_2
Step 6: Perform crossover the G_1 and G_2 and produce G_3
Step 7: Mute a member of G_3 to get the G_4
Step 8: Find the selected feature index using G_4
Step 9: Calculate the saturated value S_v
Step10: If Sv is not satisfied, then go to Step 11. Otherwise the selected attributes are significant attribute for agriculture classification
Step11: Perform crossover the G_4 and G_2 to produce new G_3 and go to Step 7
Step12: Finish

The flowchart representation of the proposed algorithm is depicted in Figure 2 and the algorithm is stated in Algorithm 1. Initially the dataset are read and the total number of attributes (T_a) is calculated. Then T_a members for group G_1 of D are generated. T_a members for group G_2 are also generated. Crossover is performed between the G_1, G_2 and G_3 is produced. Then a member of G_3 is muted to get the G_4. Selected feature index are found using G_4. The saturated value S_v is calculated. If the S_v is satisfied the selected attributes are considers as significant attribute for agriculture classification otherwise the crossover is performed between the G_4 and G_2 to produce G_3 and go to Step 7 and the iteration is continued.

IV. IMPLEMENTATION AND EXPERIMENTAL SETUP

The proposed system is implemented using Java programming language with Netbeans IDE8.0 environment. In order to conduct the experiment different agriculture dataset are collected from the Weka dataset repository. Then the proposed system is evaluated with the three classifiers function based Logistic, lazy based K-star, and tree based logistic model tree (LMT)
The experimental results on various datasets are shown with and without proposed feature selection method. The results are compared and analysed. The performance of classifier is analysed in terms of classification accuracy.

V. RESULTS AND DISCUSSION

The experimental results are shown with and without proposed feature selection method on different datasets. The results are compared and analysed. The performance of the classifier is analysed in terms of classification accuracy. Table I shows the accuracy of the classifier with and without proposed system. Figure 3 shows the average accuracy in percentage with respect to the classifiers and the proposed and without proposed system. Figure 4 shows that the accuracy in percentage against the dataset. Figure 3 shows the number of attribute selected using proposed method. Figure 4 shows the number of selected features by the proposed method with respect to the datasets.

From Table I, it is observed that the proposed method reduces the features as much as possible and produces the better accuracy for the classifiers namely Logistic, K-star, and LMT. From the Figure 3, it is observed that the proposed method produces better average accuracy with the classifiers Logistic, K-star, and LMT compared to the normal mode of operation. Figure 4 also appreciates this fact. From Figure 2, it is evident that the proposed method produces better classification accuracy for all the datasets. From Figure 5, it is obvious that the proposed method reduces the attributes almost fifty percentage from the total attributes and also the Figure 6 depicts the same.

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Vol. 3, Issue 5, May 2016

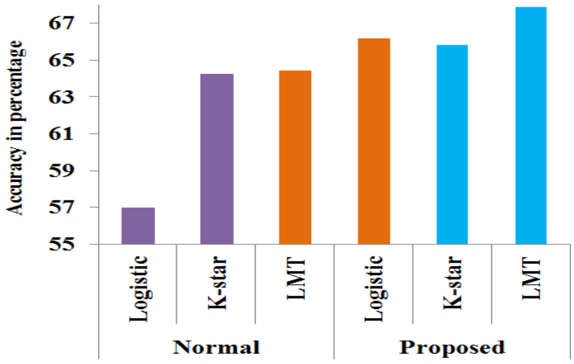

Fig.3 Average accuracy in percentage against classifiers

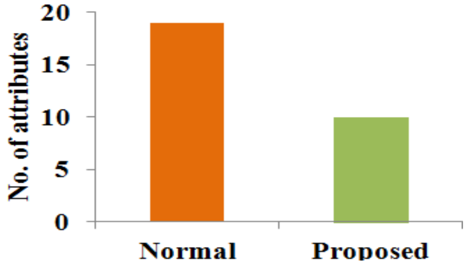

Fig.5 Number of attribute selected using proposed method

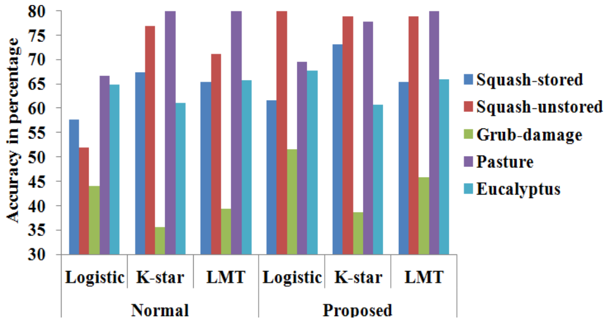

Fig. 4 Accuracy in percentage against the dataset

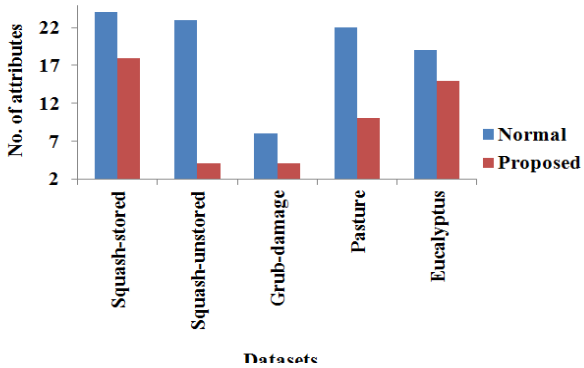

Fig. 6 Number of selected features by the proposed method with respective to the datasets

TABLE I ACCURACY OF THE CLASSIFIER WITH AND WITHOUT PROPOSED SYSTEM

Dataset	Without proposed system (Normal)				Proposed system			
	Total no. of attributes	Logistic	K-star	LMT	Selected attributes	Logistic	K-star	LMT
Squash-stored	24	57.69	67.30	65.38	18	61.53	73.07	65.38
Squash-unstored	23	51.92	76.92	71.15	04	80.76	78.84	78.84
Grub -damage	08	43.93	35.48	39.35	04	51.61	38.70	45.80
Pasture	22	66.66	80.55	80.55	10	69.44	77.77	83.44
Eucalyptus	19	64.80	61.00	65.76	15	67.66	60.73	65.89

VI. CONCLUSION

This paper proposed a differential evolution based feature selection method. This proposed method is implemented using the Java programming language. The performance of the proposed method is tested on the various real world datasets in terms of classification accuracy. In order to evaluate its performance the classifiers namely Logistic, K-star, and LMT are used. From the experimental results it is observed that the proposed method produces better performance compared to others.

REFERENCES

[1] Rami N.Khushaba, Ahmed Al-Ani, and Adel Al-Jumaily, "Differential Evolution based Feature Subset Selection" IEEE, ICPR 2008.
[2] C.Kumar, T.Alwarsamy "Solution of Economic Dispatch Problem using Differential Evolution algorithm", ISSN: 2231-2307, Volume-1, Issue-6, January 2012
[3] Xu, Xing and Li, Yuanxiang, "Comparison between Particle Swarm Optimization, Differential Evolution and Multi-Parents Crossover" International Conference on Computational Intelligence and Security, (2007), pp. 124–127.
[4] Donate, Juan Peralta, et al. "Time series forecasting by evolving artificial neural networks with genetic algorithms, differential evolution and estimation of distribution algorithm." Neural Computing and Applications 22.1 (2013): 11-20.
[5] Sentinella, M. R., "Comparison and integrated use of differential evolution and genetic algorithms for space trajectory optimization," IEEE Congress on Evolutionary Computation, (2007), pp. 973–978.
[6] H. Liu, E. R. Dougherty, J. G. Dy, K. A. Torkkola, E. Tub, H. A. Pang, C. A. Ding, F. A. Long, M. A. Behrens, L. A. Parsons, Z. A. Zhao, L. A. Yu, and G Forman, "Evolving feature selection" IEEE Intelligent Systems, vol. 20, pp. 64-76, 2005.
[7] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning" Artificial Intelligence, vol. 97, pp. 245-271, 2004.
[8] Raja Haroon Ajaz, Lal Hussain "Seed Classification using Machine Learning Techniques, ISSN: 3159-0040 Vol. 2 Issue 5, May - 2015
[9] Ghamari, S. Classification of chickpea seeds using supervised and unsupervised artificial neural networks. African Journal of Agricultural Research, 7(21), 3193-3201, 2012.
[10] Pazoki, A., & Pazoki, Z, Classification system for rain fed wheat grain cultivars using artificial neural network. African Journal of Biotechnology, 10(41), 8031-8038, 2011.
[11] Chen, X., Xun, Y., Li, W., & Zhang, J. "Combining discriminant analysis and neural networks for corn variety identification", "Computers and electronics in agriculture", 71, S48-S53, 2010.
[12] Pun, M., & Ballad, N, "Classification of Wheat Grains Using Machine Algorithms. International Journal",(IJSR), 2319-7064, 2013.
[13] KS, J, "Classification of Seed Cotton Yield Based on the Growth Stages of Cotton Crop Using Machine Learning Techniques". In

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
Vol. 3, Issue 5, May 2016

IEEE International Conference on Advances in Computer Engineering (ACE), (2010) pp. 312-315.

[14] Jefferson R. Souza, Ciao C. T. Mendes, Victor Guizilini, Keen C. T. Vivaldini, Admire Colturato, Fabio Ramos and Denis F. Wolf "Automatic Detection of Ceratocystis Wilt in Eucalyptus Crops from Aerial Images".

[15] ABRAF, "Asocial Brasilia de predators de Florists plantadas", Anurio Statistic ad ABRAF, 2012," Tech. Rep.

[16] D. R. Negro, T. A. F. S. Junior, J. R. S. Passes, C. A. Sangallo, M. T. A.Minhoni, and E. L. Forted, "Bio. of Eucalyptus Urograndis Wood by Fungi", Int. Bio. & Bio., vol. 89, pp. 95–102, 2014.

[17] M. Ghiasi and R. Amirfattahi, "Fast Semantic Segmentation of AerialImages based on Color and Texture" in IEEE ICRA, 2013.

[18] Rizzi, R.; Rudorff, B.F.T. Soybean crop area estimation in Rio Grande do Sul through Landsat images. Revista Brasileira de Cartografia 2005, 57, 226–234.

[19] Gesso, A.; Formaggio, A.R.; Rizzio, R.; Adam, M.; Rudolf, B.T.F. "Soybean area estimation by MODIS/EVI data". Pesquisa Agropecuária Brasilia , 47, 425–435,2012.

[20] Lobell, D.B.; Saner, G.P. "Cropland distributions from temporal unfixing of MODIS data". Remote Sens. Environ. 93, 412–422,2004

[21] Xiao-Fang Wang amebic, De-Shang Huang a, Ji-Xiang Daub, Huan Xu a,b, "Laurent Hutted d Classification of plant leaf images with complicated background Applied Mathematics and Computation", 205, 916–926,2008.

[22] C.L. Lee, S.Y. Chen, Classification for leaf images, Proc. 16th IPPR Conf. Compute. Vision Graphics Image Process. (2003) 355–362.

[23] M. Oide, S. Ninomiya, "Discrimination of soybean leaflet shape by neural networks with image input", Comput. Electron. Agric. 29,59–72,2000.

[24] Q. Wu, C. Zhou, C. Wang, "Feature extraction and XML representation of plant leaf for image retrieval", Lecture Note. Computer Sci. 3842, 127–131, 2006.

[25] Sumitha Thankachan1, Dr. S.Kirubakaran2 E-Agriculture Information Management System, IJCSMC, Vol. 3, Issue. 5, pg.599 – 607, May 2014.

[26] Peter Namisiko and Moses Aballo "Current Status of e-Agriculture and Global Trends: A Survey Conducted in TransNzoia County, Kenya" in International Journal of Science and Research Volume 2 Issue 7, 2013.

[27] Marcel Fafchamps and Bart Minton "Impact of SMS-Based Agricultural Information on Indian Farmers" in Oxford journals VOL. 26, NO. 3, pp. 383–414, 2012.

[28] D. Asir Antony Gnana Singh, E. Jebamalar Leavline, K. Valliyappan and M. Srinivasan, Enhancing the Performance of Classifier Using Particle Swarm Optimization (PSO) - based Dimensionality Reduction, "International Journal of Energy, Information and Communications", Vol.6, Issue 5 (2015), pp.19-26

[29] Danasingh Asir Antony Gnana Singh, Subramanian Appavu Alias Balamurugan, and Epiphany Jebamalar Leavline. "An unsupervised feature selection algorithm with feature ranking for maximizing performance of the classifiers." International Journal of Automation and Computing 12, no. 5 (2015): 511-517.

[30] S.Vidhya, D.Asir Antony Gnana Singh, E.Jebamalar Leavline, Feature Extraction for Document Classification," International Journal of Innovative Research in Science, Engineering and Technology" Vol. 4, Special Issue 6, May 2015, 50-56

[31] D. Asir Antony Gnana Singh, E. Jebamalar Leavline 'Decision Making In Enterprise Computing: A Data Mining Approach', International Journal Of Core Engineering & Management (IJCEM) Volume 1, Issue 11, February 2015, 103-113

[32] D. Asir Antony Gnana Singh, P.Surenther, E. Jebamalar Leavline, "Ant Colony Optimization Based Attribute Reduction for Disease Diagnostic System" International Journal of Applied Engineering Research, Vol. 10 No.55 (2015).