A STUDY OF MEDICAL DATA CLASSIFICATION USING ARTIFICIAL NEURAL NETWORK

G.Purusothaman [1] and P Krishnakumari P [2]

Abstract

Medical area produces gradually more voluminous amounts of electronic data which are becoming intricate. The produced medical data have a certain characteristics that make their analysis very challenging and striking. In this study we present an overview of medical domain data mining from different objectives including nature of medical data, collection of requirements dealing with such data and the different techniques used for medical data mining. Among the different methods we underline on the use of Artificial Neural Network Algorithms which is one of the effective and efficient evaluation method. To support our argument, empirical comparison of ANN versus with other methods on different medical data sets, shows that ANN is well suited for medical application and has high performance in most of the examined medical problems.

Key words: Data mining, Artificial Neural Network Classification and Medical data

I. Introduction

Nowadays modern hospitals are well equipped with monitoring and other data collection devices resulting in enormous data which are collected continuously through health examination and medical treatment. All this led to the fact that medical area produces increasingly voluminous amounts of electronic data which are becoming more complicated.

In the past, various statistical methods have been used for modeling in the area of disease diagnosis. These methods require prior assumptions and are less capable of dealing with massive and complicated non linear and dependant data [37] . However , data mining has proven to be more powerful and effective and it provides processes for discovering useful patterns from large data sets. [53] . These data mining techniques are generally classified, into supervised and unsupervised models. Clustering techniques which are unsupervised learning, have emerged as popular techniques for pattern recognition and image processing [1,55] and have also been applied to problems with medical data [25]. However in this paper, we are concerned with predictive methods [i.e. supervised] methods which require the data to include a special response attribute, known as the class attribute and therefore known as classification models.

The importance of Medical Data Mining [MDM] is to assist the physician to make the final decision without hesitation, minimizing diagnosis errors [especially from inexperienced physicians] , improving diagnostic speed

[1]Asst, Professor, Dept. of MCA, RVS Arts and Science College, Sulur, Coimbatore.

[2]Director, Dept. of MCA, RVS Arts and Science College, Sulur, Coimbatore.

and increasing the quality of medical treatment [39,5, 37,51].

II. Background of Medicaldata Mining

A. Characteristics of medical data

The data gathered in medicine is generally collected as a result of patient-care activity to benefit the individual patient and research is only a secondary consideration. As a result, medical data contain many features that create problems for the data mining techniques and they might be in a format which is not suitable for the direct application of those techniques [13, 53].

In general, medical collections, diagnoses and treatments are subject to error rates, imprecision and uncertainty [43]. As with any large databases and due to the collection method, medical databases may contain missing values and can introduce noisy, redundant, incomplete or inconsistent data [13].

B. Requirements for systems dealing with medical data:

For a data mining system to be useful in solving medical problems, the following features are desired:

Handling missing values and noisy data: In real medical data sets, missing values are frequently present and most patients' records lack certain data. This can be a result of certain tests not performed or certain questions that were not asked [27,10,6]. Therefore, medical mining systems have to be able to appropriately deal with such incompleteness of the data. Some data mining approaches are robust to missing values while other approaches deal with this requirement through preprocessing of the data. In addition to missing values, medical data are characterized by their incorrectness, inconsistency, redundancy, sparseness and inexactness.

For this reason, in most cases, a robust data preprocessing system is required in order to draw any kind of knowledge from even medium-sized medical data sets [35,50]

High performance and efficiency of the produced model: For a medical diagnostic system to be accepted by the user, its accuracy must be as high as possible. In most cases several approaches are tested on the available data and the one with best performance is considered. However, for small differences in predictive performance it might be necessary to take into account other features for selecting the appropriate method [27;42]. Efficiency of the data mining method used is also important, because the final application is a user interactive and for many optimal solutions they are usually time consuming [36]

Transparency of the model: Data mining techniques differ in their degree of transparency, i.e., the users' ability to analyze and understand how the patterns were generated. For some techniques which are considered as "black boxes", their results may not be accepted by the end user, especially when producing unexpected solution [35]. In medical applications the user should be able to use the model's logic to explain how the conclusion was reached which may significantly increase a physician's confidence in the model [6]

Interpretability and understandability of results: Interpretability and acceptability by the medical community intervene in favour of a method that may not have the highest predictive performance [42]. In general, users do not care how sophisticated a data mining method is but they do care how understandable its results are [36]. It is crucial for a medical diagnosis system to be able to explain and justify its decisions when diagnosing a new patient [35]

Reduction of the number of tests and generalization: Since the collection of medical data is sometimes expensive and harmful for the patients, it is desirable to have a system that is able to reliably diagnose with a small amount of data [27]. However this should not result in overfitting situations and the produced model must be able to perform well with unseen cases [6]

Protecting the privacy of data: When dealing with medical data it is important to protect the privacy and sensitive information from disclosure and to identify possible ways to have secure channels for transfering medical data [8,41]

III. Techniques And Methods Uded Medical Data Mining

Soft computing methods have been widely used for medical data mining and proven to be well suited to cope with the special characteristics of medical data such as imprecision and uncertainty. For example: Fuzzy Logic [24, 7], Rough Sets [54,21], Genetic Algorithms [34, 18] and Neural Networks [57,38].

Statistical methods have been considered, by many researchers, less capable of dealing with massive, non-linear and dependent data [such as the health care data]. However some predictive statistical approaches such as the proposed model by Cong and Tsokos [12], the k-Nearest Neighbour [k-NN] [16], Logistic Regression [LR] [29] and Bayesian Classifiers [28], have been successfully applied to medical data.

Decision Tree [DT] algorithm is one of the most popular classification algorithms used for data mining. It has been applied to medical data providing competitive performance as compared to other approaches as discussed by Delen et al. [13] and Kuo et al. [30].

Agent-based systems and artificial immune systems [AISs] have been also applied to medical problems. Examples of their use for medical applications are given by Lanzola et al. [32], Hudson and Cohen [22], Polat et al. [46] and Latifoglu et al. [33].

Recently, the need of a hybrid data mining approach is widely recognized by the data mining community and much current work in data mining tends to hybridize diverse methods [20]. In Medical domain there are a lot of hybrid models which have been proposed, such as Evolutionary decision tree [44], Polynomial Fuzzy DT [40], ANN with MARS [Multivariate Adaptive Regression Splines] [9] and Fuzzy AIS with k-NN [48]. Most of the above mentioned methods combine two or three methods, while some researchers have proposed combining more models, such as Hassan and Verma [2007] which combines self-organizing map [SOM], k-means and naïve Bayes with a neural network based classifier.

Apart from improving [or hybridizing] existing data mining techniques, other attempts to enhance the final predicted output are based on improving the quality of the data itself. Approaches which fall under this category aim to study the medical data itself and apply different techniques to the data such as Decomposition using structured rule-feature matrix [31], discretization [2], filtering outliers [45] and filtering with over-sampling [53].

However, among the different approaches and techniques used for medical applications, in this paper we are concerned with the use of ANN Algorithms for medical classification. In the following we discuss its basic features and how it suits for this domain.

IV. GENETICALGORITHMS

The GA is a type of structured random search algorithm so-called by most of researchers who used GAs that mimics the process of biological evolution. The algorithm begins with a collection of parameter estimates [called a chromosome or individual] and each is evaluated for its fitness in solving the given minimization or maximization task. At each generation [algorithm time-step], the most fit chromosomes are allowed to mate and bear offspring. The biological analogy suggests that such a procedure will be likely to lead to workable solution for complex non-linear problems.

A GA traditionally contains three types of operators: selection, crossover and mutation. A simple GA executes as follows:

a] Start with a randomly generated population of n k-bit chromosomes.

These are the candidate solutions to the problem.

b] Calculate the fitness F[x] of each chromosome x in the population.

c] Repeat the following steps until n offspring have been created.

i. Select a pair of chromosome playing the role as parents. The probability of an individual been selected is usually a function of fitness. The fitter the individual is, the more likely it will be selected to reproduce.

ii. With a probability Pc [the crossover rate], crossover the pair at a randomly chosen point to form two offspring. If no crossover takes place, form two offspring that are exact copies of their respective parents.

iii. Mutate the some worse individuals in the population at each locus with probability Pm [the mutation rate] and place the resulting chromosomes in the new population.

d] Replace the current population with the new population.

e] Go to step b.

Random search algorithms have achieved increasing popularity as researchers recognize the shortcomings of calculus-based and enumerative schemes. Random walks and random schemes that search and save the best must be discounted because of efficiency requirements. Random searches, in the long run, can be expected to do no better than enumerative schemes. Random search methods are distinct from randomized techniques.

A GA is an example of a search procedure that uses random choice as a tool to guide a highly exploitative search through a coding of a parameter space. Many search techniques require auxiliary information in order to work properly. GAs has no need for all this auxiliary information; they are blind. They only require payoff values associated with individual strings in performing an effective search for better and better structures. This characteristic makes a GA a more canonical method than many search schemes. Many researchers have tried to improve the GAs performance by handling some modifications on the genetic operators and analyzing chromosomes space properties: dealing with genotype-phenotype mapping [3, 11], analyzing schema theory at aim of catching some idea for improving GAs performance [14, 59]. Some researchers investigated the effects of GAs operators and tried to modify GAs operators [23,25,55]. Some people defined the new version of GAs [19, 26, 59-62]. In noisy environment, fitness of an individual cannot be evaluated precisely, but its fitness has to be estimated [25]. Most of researchers used GAs applied modifications on the GA operators to improve the performance of GAs. However, there are some problems related to GAs such as to be trapped in local solution/solutions or diverging from best or sub-best solution. These are important points for improving the performance of GAs.

V. Artifical Neural Network

ANN Structure i.e., the graph representing the network. The vertices of the graph represent the neurons, while the edges of the graph represent the connections. In the most general form any graph can be evolved. From a global perspective learning in ANNs is equivalent to adjusting its internal parameters, e.g., weights and biases. Evolution of network parameters, e.g., weights, biases, and activation functions, is the most direct approach to ANN learning, hence it could be termed Evolutionary Training of ANNs [63]. In order to improve ANN teaching, we could generate data sets for training, validation, and testing by simply evolving data subsets of all available data. We could also evolve data "from scratch" within restrictions given by the specific task the ANN is used for[64].

VI. Empirical Comparision

Here we present an empirical comparison of Artificial Neural Network with popular algorithms on different medical data sets (Table 1). The selected algorithms are: Decision Tree [DT], Genetic Algorithm and a simple rule-based algorithm [ZeroR]. These algorithms were chosen because they represent quite different approaches to learning and they have been used in medical data mining applications as discussed earlier.

TABLE I: Medical data sets used for the experiment

		No. of Attributes		
Medical Problem	**Instances**	**Numeric**	**Nominal**	**Classes**
Dermatology	*366*	*1*	*33*	*6*
Echocardiogram	*132*	*8*	*8*	*2*
Lever Disorders	*345*	*6*	*0*	*2*
Pima Diabetes[Indians]	*768*	*8*	*0*	*2*
Haberman	*306*	*2*	*1*	*2*
Heart – c [Cleveland]	*303*	*6*	*7*	*2*
Heart Stalog	*270*	*5*	*8*	*2*
Hepatitis	*155*	*6*	*13*	*2*
Lung Cancer	*32*	*0*	*56*	*3*

TABLE II: Comparative analysis based on predictive accuracy

Problem	ANN	DT	GA	ZeroR
Dermatology	*97.43*	*94.10**	*96.45*	*30.60**
Echocardiogram	*95.77*	*96.41*	*93.64*	*67.86**
Lever Disorders	*54.89*	*65.84 !*	*68.73!*	*57.98*
Pima Diabetes [Indians]	*75.75*	*74.49*	*74.75*	*65.11**
Haberman	*75.36*	*72.16*	*70.32 **	*73.53*
Heart – c [Cleveland]	*83.34*	*77.13**	*80.99*	*54.45**
Heart Stalog	*84.85*	*75.59**	*81.78*	*55.56**
Hepatitis	*83.81*	*79.22*	*80.78*	*79.38*
Lung Cancer	*53.25*	*40.83*	*44.08*	*40.00*

VII. Conclusion

This study reviewed the current state of medical data mining from different perspectives. ANN with other Algorithms classification approach has been discussed and its main features are highlighted based on the medical mining requirements. Based on various experiments study we prove empirically its suitability to the medical domain problems as compared to other approaches. The experimental results show that ANN is better than the compared approaches on most of the used medical data sets.

References

[1] Abed, H. and L. Zaoui, 2011. *"Partitioning an image database"* by K_means algorithm. J. Applied Sci., 11: 16-25.

[2] Abraham, R., J.B. Simha and S.S. Iyengar, 2006. *"A comparative analysis of discretization methods for medical data mining with Naive Bayesian classifier"*. Proceeding of the 9th International Conference on Information Technology [ICIT'06], Dcember 18-21, 2006, Bhubaneswar, pp: 235-236.

[3] R.E. Keller, W. Banzhaf," *Evolution of Genetic Code on a Hard Problem*", Proc. Of the Genetic and Evolutionary Computation Conference, San Fransisco, California, pp:50-56, July 7-11, 2001.

[4] Asuncion, A. and D. Newman, 2007. *UCI machine learning repository of machine learning databases.* University of California, School of Information and Computer Science, Irvine, CA.

[5] Bai, V.T. and S.K. Srivatsa, 2006. *"Wireless tele care system for intensive care unit of hospitals using bluetooth and embedded technology."* Inform. Technol. J., 5: 1106-1112.

[6] Bellazzi, R. and B. Zupan, 2008. *"Predictive data mining in clinical medicine":* Current issues and guidelines. Int. J. Med. Inform., 77: 81-97.

[7] Blessia, T.F., S. Singh, A. Kumar and J.J. Vennila, 2011. " *Application of knowledge based system for diagnosis of osteoarthritis*". J. Artif. Intell., 4: 269-278.

[8] Chen, T.S., J. Chen and Y.H. Kao, 2010. *"A novel hybrid protection technique of privacy-preserving data mining and anti-data mining.* " Inform. Technol. J., 9: 500-505.

[9] Chou, S.M., T.S. Lee, Y.E. Shao and I.F. Chen, 2004. " *Mining the breast cancer using artificial neural networks and multivariate adaptive regression splines*". Expert Syst. Appli., 27: 133-142.

[10] Cios, K.J. and G.W. Moore, 2002. " *Uniqueness of medical data mining*". Artif. Intell. Med., 26: 1-24.

[11] S. Luke, "When Short Runs Beat Long Runs", *"Proc. Of the Genetic and Evolutionary Computation Conference"*, San Fransisco, California, pp:74-80, July 7-11, 2001.

[12] Cong, C. and C.P. Tsokos, 2010. *"Statistical modeling of breast cancer relapse time with different treatments"*. J. Applied Sci., 10: 37-44.

[13] Delen, D., G. Walker and A. Kadam, 2005. *"Predicting breast cancer survivability: A comparison of three data mining methods"*. Artificial Intell. Med., 34: 113-127.

[14] R. Poli, N. F. McPhee, « Exact Schema *"Theory for GP and Variable-Length Gas w Homologous Crossover"*, Proc. Of the Genetic and Evolutionary Computation Conference, San Fransisco, California, pp:104-111, July 7-11, 2001.

[15] J.A. Vasconcelos, J.A. Ramirez, R.H.C. Takahashi, R.R. Saldanha, *"Improvement in Genetic Algorithms"*, IEEE Trans. On Magnetics, vol:37, no:5, pp:3414-3417, 2001.

[16] Dzeroski, S. and N. Lavrac, 1996. *"Rule induction and instance-based learning applied in medical diagnosis"*. Technol. Health Care, 4: 203-221.

[17] N.Y. Nikolaev, H. Iba, *"Regularization Approach to Inductive Genetic Programming"*, IEEE Trans. On Evolutionary Computation, vol:5, no:4, pp:359-375, 2001.

[18] Goletsis, Y., C. Papaloukas, D.I. Fotiadis, A. Likas, and L.K. Michalis, 2004. *"Automated ischemic beat classification using genetic algorithms and multicriteria decision analysis. IEEE Trans. Biomed"*. Eng., 51: 1717-1725.

[19] Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I.H. Witten, 2009. The WEKA data mining software: An update. SIGKDD Explorations Newslett., 11: 10-18.

[20] Hassan, S.Z. and B. Verma, 2007. A hybrid data mining approach for knowledge extraction and classification in medical databases. Proceeding of the IEEE Seventh International Conference on Intelligent Systems Design and Applications, Octuber 20-24, 2007, Rio de Janeiro, pp: 503-508.

[21] Hassanien, A.-E., M.E. Abdelhafez and H.S. Own, 2008. Rough sets data analysis in knowledge discovery: A case of kuwaiti diabetic children patients. Adv. fuzzy syst., Vol. 2008. 10.1155/2008/528461

[22] Hudson, D.L. and M.E. Cohen, 2002. Use of intelligent agents in the diagnosis of cardiac disorders. Disorders. Compu. Cardiol., 29: 633-636.

[23] H.E. Aguirre, K. Tanaka, S. Oshita, *"Increasing the Robustness of Distributed Genetic Algorithms by Parallel Cooperative- Competitive Genetic Operators"*, Proc. Of the Genetic and Evolutionary Computation Conference, San Fransisco, California, pp:195-202, July 7-11, 2001.

[24] John, R.I. and P.R. Innocent, 2005. Modeling uncertainty in clinical diagnosis using fuzzy logic. IEEE Trans. Syst. Man Cybern., 35: 1340-1350.

[25] J. Branke, *"Reducing the Sampling Variance When Searching for Robust Solutions"*, Proc. Of the Genetic and Evolutionary Computation Conference, San Fransisco, California, pp:235-242, July 7-11, 2001.

[26] Y.-W. Leung, Y. Ewang, éAn Orthogonal Genetic *Algorithm with Quantization for Global Numerical Optimization"*, IEEE Trans. On Evolutionary Computation, vol:5, no:1, pp:41-53, 2001.

[27] Kononenko, I., 2001. Machine learning for medical diagnosis: History, state of the art and perspective. Artif. Intell. Med., 23: 89-109.

[28] Kononenko, I., I. Bratkoa nd M. Kukar, 1998. Application of Machine Learning to Medical Diagnosis. In: Machine Learning and Data Mining: Methods and Applications, Michalski, R.S., R.S. Michalski, I. Bratko and M. Kubat [Eds.]. J. Wiley, New York.

[29] Kuhnert, P.M., K.A. Do and R. McClure, 2000. Combining non-parametric models with logistic regression: An application to motor vehicle injury data. Comput. Statist. Data Analys., 34: 371-386.

[30] Kuo, W.J., R.F. Chang, D.R. Chen and C.C. Lee, 2001. Data mining with decision trees for diagnosis of breast tumor in medical ultrasonic images. Breast Cancer Res. Treat., 66: 51-57.

[31] Kusiak, A., 2001. Decomposition in data mining: A medical case study. Proceedings of the SPIE Conference on Data Mining and Knowledge Discovery: Theory, March 27, 2001, Tools and Technology III, pp: 267-278.

[32] Lanzola, G., L. Gatti, S. Falasconi and M. Stefanelli, 1999. A framework for building cooperative software agents in medical applications. Artif. Intell. Med., 16: 223-249.

[33] Latifoglu, F., K. Polat, S. Kara and S. Gunes, 2008. Medical diagnosis of atherosclerosis from carotid artery doppler signals using principal component analysis [PCA], k-NN based weighting pre-processing and Artificial Immune Recognition System [AIRS]. J. Biomed. Inform., 41: 15-23.

[34] Laurikkala, J., M. Juhola, S. Lammi, and K. Viikki, 1999. Comparison of genetic algorithms and other classification methods in the diagnosis of female urinary incontinence. Method. Inform. Med., 38: 125-131.

[35] Lavrac, N., 1999. Selected techniques for data mining in medicine. Artificial Intell. Med., 16: 3-23.

[36] Li, J., A. Wai-Chee Fu, H. He, J. Chen and H. Jin, 2005. Mining risk patterns in medical data. Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug. 21-24, Chicago, Illinois, USA., pp: 770-775.

[37] Lin, R.H., 2009. An intelligent model for liver disease diagnosis. Artif. Intell. Med., 47: 53-62.

[38] Lisboa, P.J.G., 2002. A review of evidence of health benefit from artificial neural networks in medical intervention. Neural Netw., 15: 11-39.

[39] Maria, P.R.F., 2002. An intelligent, interactive, web-based platform for effective clinical education through VR and distance learning modules. Inform. Technol. J., 1: 126-131.

[40] Mugambi, E.M., A. Hunter, G. Oatley and L. Kennedy, 2004. Polynomial-fuzzy decision tree structures for classifying medical data. Knowledge-Based Syst., 17: 81-87.

[41] Nabi, M.S.A., M.L.M. Kiah, B.B. Zaidan, A.A. Zaidan and G.M. Alam, 2010. Suitability of using SOAP protocol to secure electronic medical record database transmission. Int. J. Pharmacol., 6: 959-964.

[42] Ohno-Mochado, L., 2001. Modeling medical prognosis: Survival analysis techniques. J. Biomed. Inform., 34: 428-439

[43] Pattaraintakorn, P., N. Cercone and K. Naruedomkul, 2005. Hybrid intelligent systems: Selecting attributes for soft-computing analysis. Procceding of the 29th Annual International Computer Software and Applications Conference, July 26-28, 2005, Department of Math., Mahidol University, Thailand, pp: 319-325.

[44] Podgorelec, V. and P. Kokol, 2001. Towards more optimal medical diagnosing with evolutionary algorithms. J. Med. Syst., 25: 195-219.

[45] Podgorelec, V., M. Hericko and I. Rozman, 2005. Improving mining of medical data by outliers prediction. Proceedings of 18th IEEE Symposium on Computer-Based Medical Systems, June 23-24, 2005, Dublin, Ireland, pp: 91-96.

[46] Polat, K., S. Sahan, H. Kodaz and S. Gunes, 2005. A new classification method for breast cancer diagnosis: Feature selection artificial immune recognition system [FS-AIRS]. Lect. Notes Comput. Sci., 3611: 830-838.

[47] M. Elhadef, D.A. Coley, "*Adaptive Mutation for Semi-Separable Problems*", Proc. Of the Genetic and Evolutionary Computation Conference, San Fransisco, California, pp:306-312, July 7-11, 2001.

[48] Sahan, S., K. Polat, H. Kodaz and S. Gunes, 2007. A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis. Comput. Biol. Med., 37: 415-423.

[50] Sumathi, S. and S.N. Sivanandam, 2006. Data mining in biomedicine and science. Stud. Comput. Intell., 29 : 499-543.

[51] Temurtas, H., N. Yumusak and F. Temurtas, 2009. A comparative study on diabetes disease diagnosis using neural networks. Expert Syst. Appl., 36: 8610-8615.

[52] Thongkam, J., G. Xu, Y. Zhang and F. Huang, 2008. Breast cancer survivability via AdaBoost algorithms. Proc. 2nd Aust. Workshop on Health Data Knowledge Manage., 80: 55-64.

[53] Thongkam, J., G. Xu, Y. Zhang, and F. Huang, 2009. Toward breast cancer survivability prediction models through improving training space. Expert Syst. Appl., 36: 12200-12209.

[54] Tsumoto, S., 1999. Knowledge discovery in medical Multi-databases: A rough set approach. Principles Data Min. Knowl. Dis., 1704: 147-155

[55] Velmurugan, T. and T. Santhanam, 2011. A survey of partition based clustering algorithms in data mining: An experimental approach. Inform. Technol. J., 10: 478-484.

[56] Witten, I.H. and E. Frank, 2005. Data Mining Practical Machine Learning Tools and Techniques. 2nd Edn., Morgan Kaufman, San Francisco, CA.

[57] Yao, X. and Y. Liu, 1999. Neural networks for breast cancer diagnosis. Proceedings of the 1999 Congress on Evolutionary Computation, July 06-09, 1999, Washington, DC, USA -.

[58] Zelic, I., I. Kononenko, N. Lavrac and V. Vuga, 1997. Induction of decision trees and bayesian classification applied to diagnosis of sport injuries. J. Med. Syst., 21: 429-444.

[59] J.E. Rowe, N.F. McPhee, "The Effect of Crossover and Mutation Operators on Variable Lengrh Linear Structures", Proc. Of the Genetic and Evolutionary Computation Conference, San Fransisco, California, pp:535-542, July 7-11, 2001.

[60] A. Simoes, E. Costa, *"On Biologically Inspired Genetic Operators: Transformation in the Standard Genetic Algorithm"*, Proc. Of the Genetic and Evolutionary Computation Conference, San Fransisco, California, pp:584-591, July 7-11, 2001.

[61] J. Smith, *"Modelling GAs with Self Adaptive Mutation Rates"*, Proc. Of the Genetic and Evolutionary Computation Conference, San Fransisco, California, pp:599-606, July 7-11, 2001.

[62] R. P. Srivastava, D.E. Goldberg, *"Verification of the Theory of Genetic Algorithm Continuation"*, Proc. Of the Genetic and Evolutionary Computation Conference, San Fransisco, California, pp:623-630, July 7-11, 2001.

[63] Karthik Balakrishnan and Vasant Honavar. Evolutionary design of neural architectures – a preliminary taxonomy and guide to literature. Technical Report CS TR #95-01, Iowa State University, Department of Computer Science, Ames, Iowa 50011–1040, U.S.A., January 1995.

[64] J¨urgen Branke. Evolutionary Algorithms in Neural Network Design and Training – A Review. In Jarmo T. Alander, editor, Proceedings of the First Nordic Workshop on Genetic Algorithms and their Applications, pages 145–163, 1995.

AUTHOR BIOGRAPHY

G. Purusothaman, received the M.Sc. degree from GRI Deemed University and M.Phil. degree from Bharathiar University. He has more than seven years of teaching experience and four years of research experience. His research interests include Data mining and soft computing. At present, he is pursuing Ph.D., Computer Science at Bharathiar University, Coimbatore, Tamil Nadu.

Dr P Krishnakumari, She has more than 18 years of teaching experience. Her area of Specialization is in Data mining and Networking. She has published research articles in International Journals and presented research papers in National and International Conferences. She has more than 10 years of research experience. She acts as reviewer in major research journals & book Publishers. She acted as Convener, Chairperson and Technical review committee member for various conferences. She is a member in various professional bodies. She acted as chairman and member of BoS of various institutions.