Global Journal
of Technology
& Optimization
GJTO
Volume 3, 2012

ADAPTIVE ENSEMBLE AND HYBRID MODELS FOR CLASSIFICATION OF BIOINFORMATICS DATASETS

Tarek Helmy[2,3], Mosleh M. Al-Harthi[1], Mohamed T. Faheem[1,3]

[1]College of Engineering, Taif University, Saudi Arabia
[2]College of Computer Science & Engineering, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia
[3]College of Engineering, Department of Computers Engineering & Automatic Control, Tanta University, Egypt
Emails: m_t_faheem@yahoo.com, mosleh2k@hotmail.com, helmy@kfupm.edu.sa

Abstract
Clinical databases have accumulated large quantities of information about patients and their clinical histories. Data mining is the search for relationships and patterns within this data that could provide useful knowledge for effective decision-making. Classification analysis is one of the widely adopted data mining techniques for healthcare applications to support and improving the quality of medical diagnosis. This paper presents individual, ensembles and hybrid of computational intelligence techniques such as Support Vector Machine (SVM), Neural Networks (NN), Function Network (FN) and Fuzzy Logic (FL) to classify real bioinformatics datasets. The performance of the proposed computational techniques measured using well known bioinformatics datasets. As expected, the performance of the proposed ensembles and hybrid computational intelligence models is better compared to the monolithic models and overcome the weaknesses of existing classifiers particularly in the classification accuracy.

1. Introduction

In recent years, rapid developments in bioinformatics have generated a large amount of data. Often, drawing conclusions from these data require sophisticated computational analyses. Bioinformatics, or computational biology, is the interdisciplinary science of interpreting biological data using information technology and computer science. The importance of this new field of inquiry will grow as we continue to generate and integrate large quantities of data. A particularly active area of research in bioinformatics is the application and development of machine learning techniques to biological problems. Analyzing large biological datasets requires making sense of the data by inferring structure or generalizations from the data. Examples of these types of dataset analysis include cancer classification, diabetes classification, uncertainty manipulation, etc [15, 20]. Each of these problems can be framed as a problem in machine learning [1, 2, 3]. Therefore there is a great potential to increase the interaction between machine learning and bioinformatics. The types of learning algorithms fall along several classifications but we can summarize them as supervised and unsupervised learning [4]. Supervised learning used in medical diagnoses adopts the general paradigm of pattern recognition where objects are described by a collection of features that form a multidimensional space in which all discrimination activities take place. Various classifiers, both linear and nonlinear can be used, including SVM, Fisher's linear discriminates, polynomial classifiers, NN, fuzzy rule-based systems etc. [6, 7, 8, 9, 33]. The classifier is developed through training on a dataset of examples that are representative of the objects to be classified. During the training process, the classifier learns the feature patterns that distinguish between the different object classes. For instance, given a set of attributes about potential cancer patients, and whether those patients actually had cancer, the classifier could learn how to distinguish between likely cancer patients and possible false alarms. For example, breast cancer dataset from the Wisconsin Hospital University contains 699 samples with 683 complete data and 16 samples with missing attributes. The attributes contained by the dataset are Lump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli and Mitoses. There are two values in the class variable of breast cancer which are benign (non-cancerous) and malignant (cancerous). This sort of learning could take place with FL systems, NN or SVM. Recently, SVM and NN have emerged as a powerful tool in pattern recognition [2], classification and forecasting in many areas. They have featured in a wide range of medical and business journals, often with promising results [11, 12, 13, 19]. Inspired by promising results obtained in other fields [26, 27], we explored the use of these intelligence techniques for classifying disease types from datasets related to bioinformatics. The main focus of this paper resides in the classification of bioinformatics datasets using individual, ensemble and hybrid of computational intelligence techniques such as SVM, NN, FN, and FL.

The rest of this paper is organized as following. A review of related earlier work is presented in Section 2. Section 3 introduces the ensemble models used in this work. Section 4 introduces the hybrid model used in this work. Section 5 describes the experimental set up. Section 6 presents models

development steps of the proposed models. Section 7 describes the datasets used in the experiments. Section 8 discusses the experimental results of using the developed computational intelligence techniques. Finally, Section 9 concludes the paper and highlights the future work.

2. Literature Survey

Recently, many classifiers were developed exploring various fields with the help of computer science. In fact, most of the research work found in the literature related to disease classification either makes use of statistical models or artificial neural networks. Statistical methods such as linear discriminate analysis, generalized linear regression such as logistic regression, and nearest neighbor classification are widely used. There are many methods and algorithms used to mine biomedical datasets for hidden information. They include NNs, Decision Trees (DT), FL Systems, Naive Bayes, SVM, cauterization, logistic regression and so on. Studying the literature it turns out that the most frequent choices for the medical decision support systems are the DT (C4.5 algorithm), NNs and the Naive Bayes [1, 2, 8, 9, 10, 17, 22, 42, 43, 44]. These algorithms are very useful in medicine because they can decrease the time spent for processing symptoms and producing diagnoses, making them more precise at the same time. Also, many of the research assessed the algorithms on a narrow set of medical databases [16]. However and to the best or our knowledge ensembles of these techniques have not been used in the bioinformatics dataset classification.

NNs are networks of units, called neurons that exchange information in the form of numerical values with each other via synaptic interconnections, inspired by the biological neural networks of the human brain. They become very powerful and flexible approaches to function approximation. NNs are mainly refer to the feed forward networks such as multilayer perceptrons and radial basis function neural networks, which have been widely used to develop diagnostic models. In order to improve the costs benefit ratio of breast cancer screenings, authors of [14] evaluated the performance of a back-propagation NN to predict an outcome (cancer/not cancer) to be used as classifier. NNs were trained on data from family history of cancer, and socio demographic, gyneco obstetric and dietary variables. Research is going on in capitalizing the use of NNs in medical diagnosis of breast cancer. This work indicates that statistical NNs can be effectively used for breast cancer diagnosis to help oncologists [12] in which classification is based on a feed forward NN rule extraction algorithm. General regression NN, or probabilistic NN was used in order to get the suitable result. The problem with NNs is that they usually adopt gradient-based learning methods which are susceptible to local minima and long training times especially when the number of classes/categories is high. The authors of [17] introduce artificial NNs with back propagation for classification of heart disease cases. This solution is implemented in a medical system to support the classification of the Doppler signals in cardiology. The predictions yielded by the method were more accurate than similar presented in [25]. The NNs' major disadvantage is complexity, which makes classification process difficult to interpret. Nevertheless, the authors prove that they produce effective classifications in case of medical data. As far as NN is concerned, the influence of the noisy inputs on the output variable together with the transfer functions, implicit in the

values of the weights. Hence an unattractive feature of such networks is that the number of weights and complexity increase greatly as the network grows. Also the weights may not always be easy to interpret if the data is imprecise and uncertain which leads to the problem of under fitting or over fitting and the problem becomes difficult to visualize from an examination of the weights.

SVM has been proposed as a very effective method for pattern recognition, machine learning and data mining [2, 8]. The general idea is to map non-linearly D-dimensional input space into a high dimensional feature space. A linear classifier (separating hyper plane) is constructed in this high dimensional space to classify the data. The use of the kernel trick allows constructing the classifier without explicitly knowing the feature space. It is considered to be a good candidate because of its high generalization performance. Intuitively given a set of points which belong to either one of the two classes, a SVM can find a hyper plane having the largest possible fraction of points of the same class on the same plane. This hyper plane called the optimal separating hyper plane (OSH) can minimize the risk of misclassifying examples of the test set. SVM, when using One-Versus-All (OVA) approach to make binary classifiers applicable to multi category problems, it can be seen that, when the number of classes increases, the complexity of the overall classifier also increases. So the system becomes more complex and requires extra computations. In SVM classifiers, problems with corrupted inputs are more difficult than problems with no input uncertainty. Even if there is a large margin separator for the original uncorrupted inputs, the observed noisy data may become non-separable. For example by using a kernel function in SVM, the input vector is mapped to in a usually high dimensional feature space and the uncertainty in the input data introduces uncertainties in the feature space. To overcome this problem, researchers used total least square regression methods with SVM but could not achieve promising results.

FNs are extensions of NN which consist of different layers of neurons connected by links. Each computing unit or neuron performs a simple calculation: a scalar typically monotone function f of a weighted sum of inputs. The function f, associated with the neurons, is fixed and the weights are learned from data using some well-known algorithms such as the least-square fitting. A FN consists of a layer of containing the input data; a layer of output units containing the output data; one or several layers of neurons or computing units which evaluate a set of input values, coming from the input units, and which give a set of output values to the output units. The computing units are connected to each other, in the sense that the output from one unit can serve as part of the input to another neuron. Once the input values are given, the output is determined by the neuron type, which can be defined by a function [26, 27, 28, 29].

Type-2 FLS was introduced as an extension of the concept of Type-1 FLS [37]. Type-2 FLS has membership grades that are themselves fuzzy. For each value of a primary variable (e.g., pressure and temperature), the membership is a function (not just a point value). The secondary Membership Function (MF) has its domain in the interval (0, 1), and its range may also be in (0, 1). Hence, the MF of a Type-2 FLS is three dimensional, and it is the newly introduced third dimension that provides new degrees of design freedom for handling uncertainties. Type-2 FLS does not obtain good performance when the number of training data is small, but it

can perform better when the number of training prototypes is large [30, 31, 32].

3. The Ensemble Models

Ensemble learning is an effective technique that has increasingly been adopted to combine multiple learning algorithms to improve overall prediction/classification accuracy. An ensemble model is constructed using a set of machine learning techniques to train for a portion of a given problem and then to integrate these machine learning techniques to solve of the entire problem. This method is also called committee of learning machines. For classification problem it is also called multiple classifier systems, classier fusion, etc. One approach of ensemble system is built by a group of base learners to solve a problem which is similar to divide-and-conquer method. Instead of solving the same problem, the base learners are trained for different sub-problems. This approach is classified as mixture of experts. However, in ensemble, the base learners are used to solve the same problem. Ensemble methods are basically used to improve the generalization capability of a single machine learning technique. Recent work in computational biology has seen an increasing use of ensemble learning methods due to their unique advantages in dealing with small sample size, high-dimensionality, and complex data structures [5, 23, 24]. In this paper we have used three NN for the ensemble model. Training data of those models are obtained through bagging sampling principle. The next task is to combine these member models using an appropriate strategy which include linear and nonlinear ensemble. In this paper we have adopted two methods for combining the three NN. For the first model we have combined them using weighted average (Figure 1) and for the second model we have combined them using another NN (Figure 2) to produce the final output.
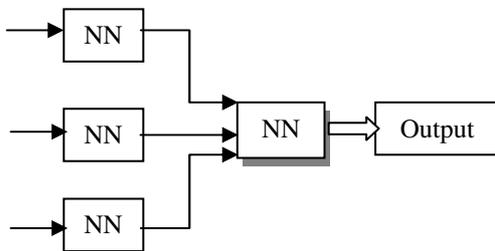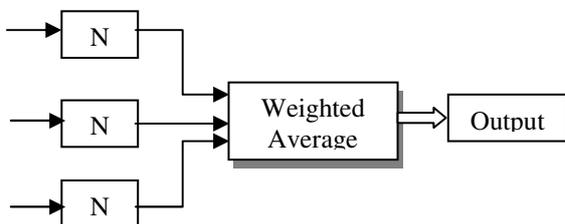

Figure 1: Ensemble of NN Combined by Weighted Average

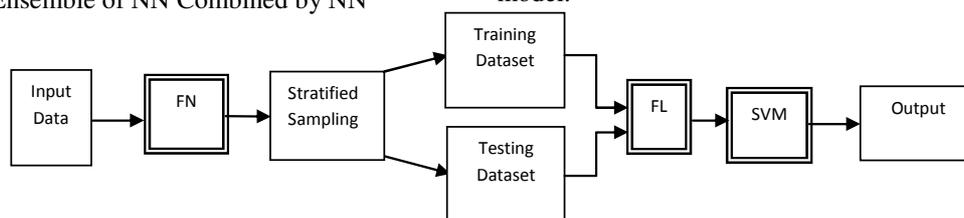Figure 2: Ensemble of NN Combined by NN

4. The Hybrid Model

Multiple Kernels Learning (MKL) has been an attractive topic in machine learning [34]. Multiple kernel learning searches for a combination of base kernel functions/matrices that maximizes a generalized performance measure. Typical measures studied for multiple kernel learning, include maximum margin classification errors [35, 36]. It has been regarded as a promising technique for identifying the combination of multiple data sources or feature subsets and been applied in a number of domains, such as genome fusion, splice site detection, image annotation and so on [40, 41]

To achieve our aim, a hybrid models was built: FN-FL-SVM (FFS). FN was used as the base for each of the models. This is due to its functional approximation capability and its ability to select the best variables for the system directly from the training data. Next we describe the framework design of the hybrid model and the optimized parameters of each of the techniques before using in the hybrid. The model is composed of three major blocks containing respectively: Functional Networks (FN), Type-2 Fuzzy Logic and Support Vector Machines (SVM). The FN block, using its least-squares fitting algorithm, is used to select the best variables from the input data. The dimensionality of the input data can be ignored by the user as it is automatically handled by the FN block that plays the role of a best-variable selector in the model. The best variables are extracted from the input data and then divided into training and test sets using the Stratified Sampling approach. The training set is passed to the Fuzzy Logic block where uncertainties are removed, if any exists. Already, Type-2 Fuzzy Logic System has been shown in several works such as in [14, 26, 27, 31] to have the ability to remove uncertainties using its extension to a third dimension. It would be futile to attempt to revalidate this already established fact in the literature. Hence, the focus of this work is the successful combination of the individual techniques to implement hybrid models that demonstrate the combined capabilities of each technique. The training data, with uncertainties removed, is then used to train the SVM block in readiness for prediction with the test data. Finally, the test data is passed to the trained SVM block to perform the regression task in order to evaluate the performance of the model. The role performed by the Fuzzy Logic block in this model is to ensure that in case an input data containing uncertainties is used, such uncertainties would have been removed before the data is passed to the SVM block for training. In this way, only "clean" data is allowed to enter the SVM block which performs the prediction task after the training process. This is an attempt to complement the performance of the hybrid with the ability of Type-2 Fuzzy Logic to handle uncertainties. However, in the absence of uncertainties in the input data, then the performance of the Type-2 FLS block is reduced to that of Type-1 FLS [26, 27, 37, 38, 39]. Figure 3 shows the design framework of this model.

Figure 3: Framework of FFS Hybrid Model

5. Experimental Setup

Various computational models for classification have been developed for the classification of real bioinformatics datasets; these include monolithic, hybrid and ensemble models. We conducted our experiments on Matlab R2007a. The datasets are stored in MS Excel documents and read directly from Matlab. All the graphs are generated by using the same Matlab R2007a. We used Naïve Bayes classifier and the SVM attribute evaluation feature selection technique available in the machine learning library with Java implementation [21]. We adopted the same splitting used by earlier published work using the dataset to allow direct comparison of results, i.e. same number of cases for training dataset and same number of cases for evaluation dataset. The diagnostic performance of the developed models is evaluated using Receiver Operating Characteristic (ROC) curve [10]. In ROC curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points. Each point on the ROC plot represents a sensitivity/specificity pair corresponding to a particular decision threshold. A model with perfect discrimination (no overlap in the two distributions) has a ROC plot that passes through the upper left corner (100% sensitivity, 100% specificity). Therefore the closer the ROC plot is to the upper left corner, the higher the overall accuracy of the model. Figures 5~10 show the ROC of NN and SVM models with the datasets.

6. Models Development Steps

Combining the output of several classifiers is useful only if they disagree on some inputs. Theoretical and empirical work showed that an effective ensemble should consist of a set of networks that are not only highly correct, but ones that make their errors on different parts of the input space as well. Diverse individuals can be obtained by adopting different model structure. In case of NN, different types of models can be obtained by having different network types, number of neuron in hidden layer, learning algorithm and initial state in weight space. Diversity can be supported by training the ensemble/hybrid members on different training datasets which can be achieved by bagging, boosting or cross validation. We divide the datasets randomly into training and testing set. We have used 80% of the datasets for training and 20% for testing. Homogeneous ensemble models with the same kind of computational techniques with different fixed parameters are chosen in each run. Performing optimization by other computational techniques with different fixed parameters results into a completely different architecture of the models in each run. Furthermore, the computational techniques of the models are trained by different portion of the training datasets and thus the ensemble and hybrid models are enforced to be diverse enough in order to substantiate better generalization. The algorithm can be continued to N runs so as to have an ensemble of N members. Figure 4 shows the models development steps.

7. Datasets Description

Datasets from the machine learning repository [18], at the University of California Irvine have been used. A brief description of these datasets is shown in Table 1.
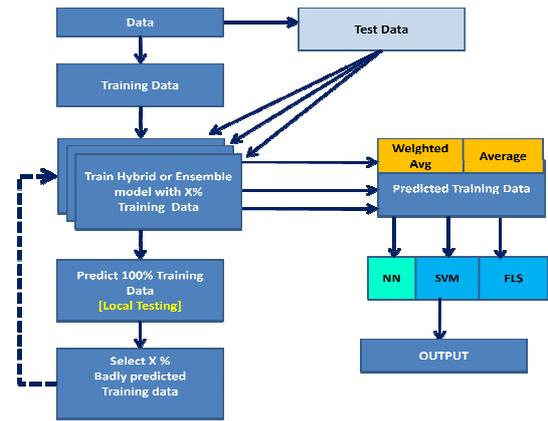

Figure 4: Models Development Steps

Table 1: Dataset Description

Dataset	Size	Number of Attributes	Missing	Class
WBC	699	9	Yes	2
Hepatitis	155	19	Yes	2
Hypothyroid	3772	30	Yes	3
Heart	303	14	Yes	5
Pima	768	8	Yes	2

For these datasets, the size represents the number of instances/entries in a dataset, the number of attributes represents how many values is contained in one instance, missing shows whether there is any incomplete entry, and class represents the number of categories to be classified in a dataset. For example, the Wisconsin Breast Cancer (WBC) dataset contains 699 samples with 683 complete data and 16 samples with missing attributes. Te attributes contained by the dataset are Lump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli and Mitoses. There are two values in the class variable of WBC which are benign (non-cancerous) and malignant (cancerous). The patient's problem is then diagnosed as either benign or malignant. For the datasets having more than two classes, such as hypothyroid datasets, the problem is then diagnosed as normal, hyper-thyroid or hypo-thyroid.

8. Results and Discussion

It is evident from Figures (5~16) that the performance of the ensemble models is better compared to the monolithic models because the ROC plot passes through the upper left corner (100% sensitivity, 100% specificity) more. It is also shown in Tables 2, 3 and Figures (17~26) that the performance of the hybrid model is better compared to the monolithic models in terms of the training and testing classification accuracy. As we have used MATLAB toolboxes of NN, and both of Type-1 and Type-2 FLS, it is noticed that the classification accuracy of different datasets with different sizes is satisfactory. The reason is that, for small and simple datasets, Type-1 has better generalization property as it keeps the overall process simple. However if we look at the ROC graphs, we can say that hybrid model with Type-2 FLS outclassed the other models. Therefore our intuition for

handling the uncertainty in a classification framework by using Type-2 FLS is justified.

Table 2: Training Classification Accuracy

	Training Classification Accuracy			
	SVM	FN	FL	Hybrid
WBC	0.93157	0.887252	0.94682	0.98565
Heart Stat log	0.863244	0.84015	0.882545	0.89659
Hepatitis	0.87156	0.825688	0.97985	0.984862
Hypothyroid	0.782143	0.807143	0.814286	0.978571
Pima Diabetes	0.742546	0.71645	0.765799	0.788674

Table 3: Testing Classification Accuracy

	Testing Classification Accuracy			
	SVM	FN	FL	Hybrid
WBC	0.93746	0.873255	0.928726	0.96852
Heart Stat log	0.851852	0.839506	0.864198	0.876543
Hepatitis	0.804348	0.76087	0.847826	0.934783
Hypothyroid	0.731092	0.722689	0.806723	0.932773
Pima Diabetes	0.726087	0.656522	0.734783	0.756522
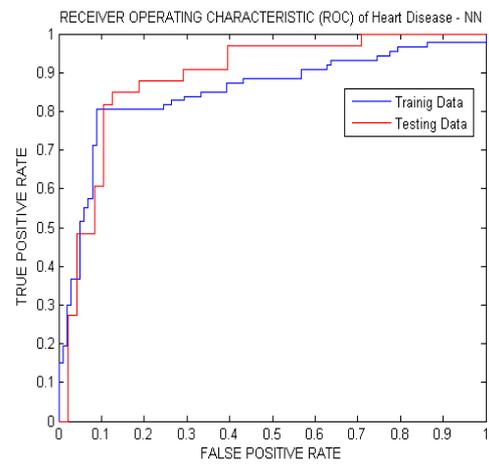
Fig. 7: ROC of Neural Networks using Heart Disease Dataset

Fig. 8: ROC of SVM using Hepatitis Dataset

Fig. 5: ROC of Neural Networks using WBC Dataset
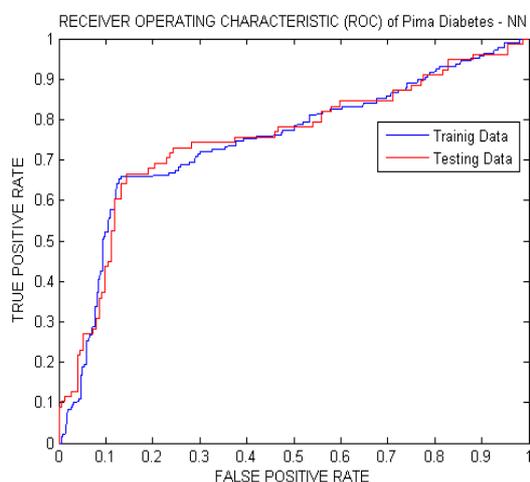
Fig. 6: ROC of Neural Networks using Pima Dataset
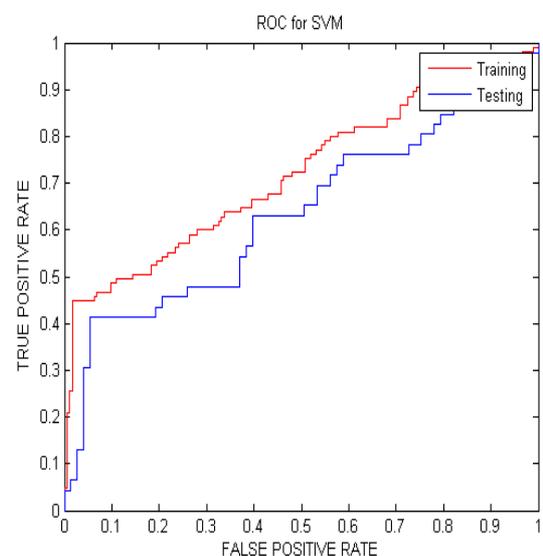
Fig. 9: ROC of SVM using Hypothyroid Dataset

Copyright @ 2012/gjto
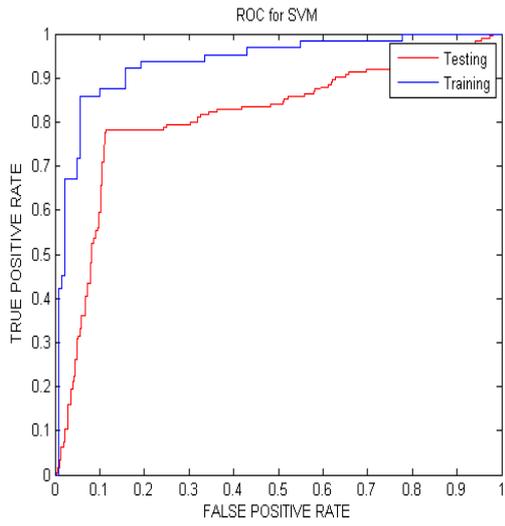
Fig. 10: ROC of SVM using WBC Dataset
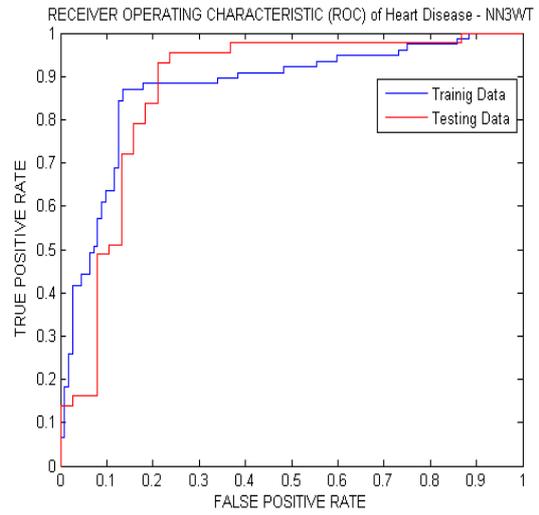
Fig. 13: ROC of Ensemble of 3 Neural Networks: Weighted Average Method using Heart Disease Dataset
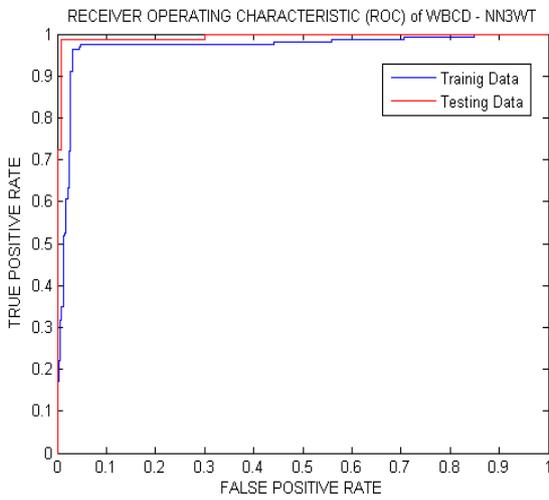

Fig. 11: ROC of Ensemble of 3 Neural Networks: Weighted Average Method using WBC Dataset
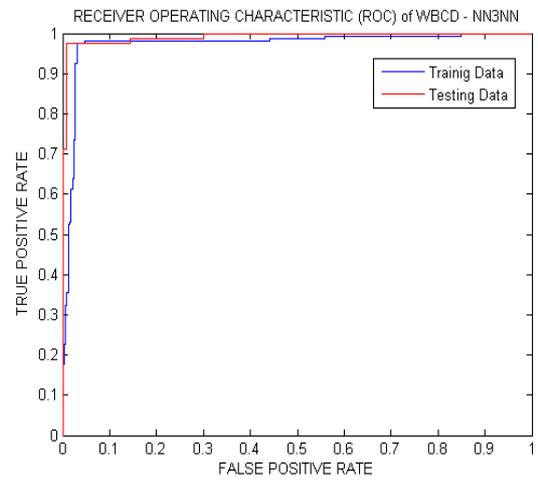

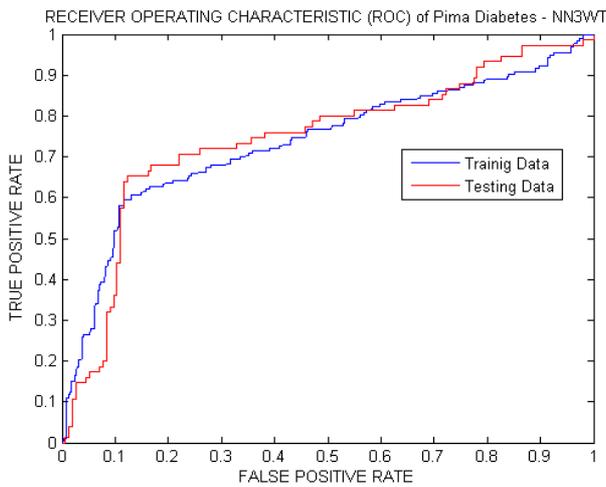Fig. 14: ROC of Ensemble of 3 Neural Networks Combined by NN using WBC Dataset

Fig. 12: ROC of Ensemble of 3 Neural Networks: Weighted Average Method using Pima Diabetes Dataset
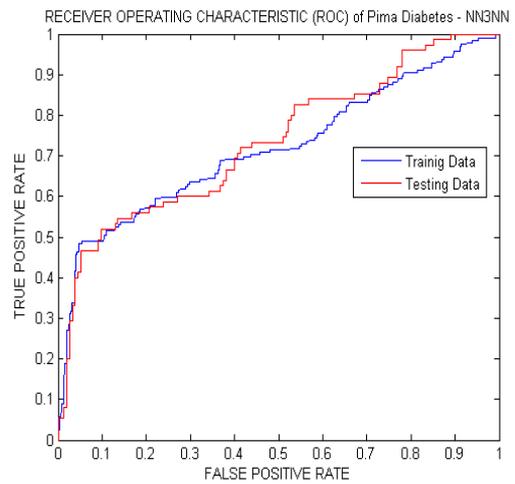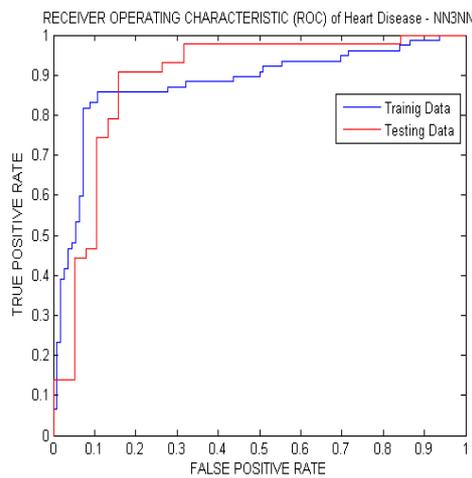

Fig. 15: ROC of Ensemble of 3 Neural Networks Combined by NN using Pima Diabetes Dataset

Fig. 16: ROC of Ensemble of 3 Neural Networks Combined by NN using Heart Disease Dataset
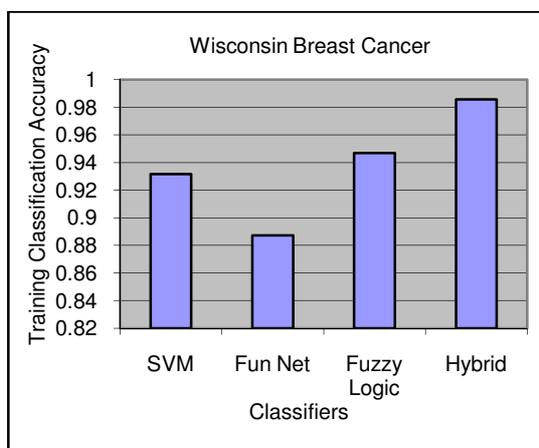
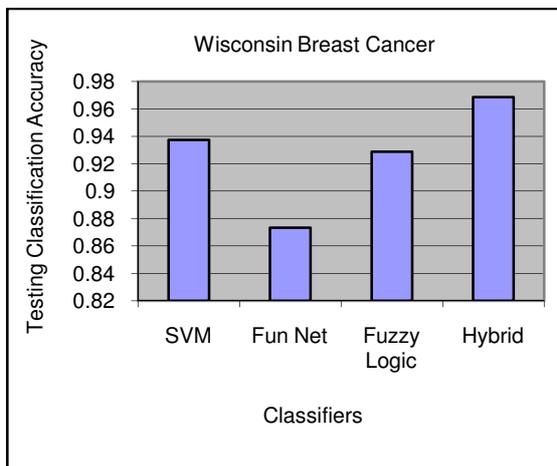Fig. 17: Training Classification Accuracy of WBC Dataset

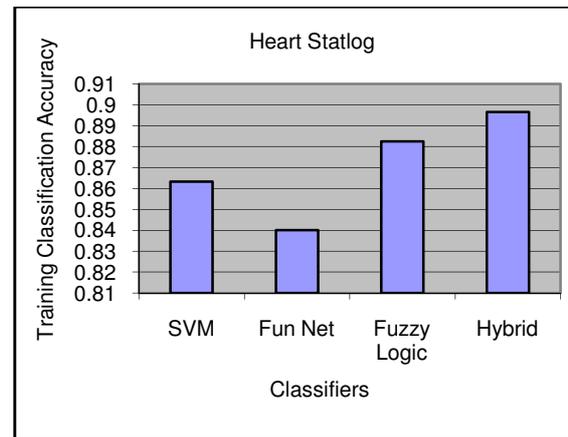Fig. 18: Testing Classification Accuracy of WBC Dataset
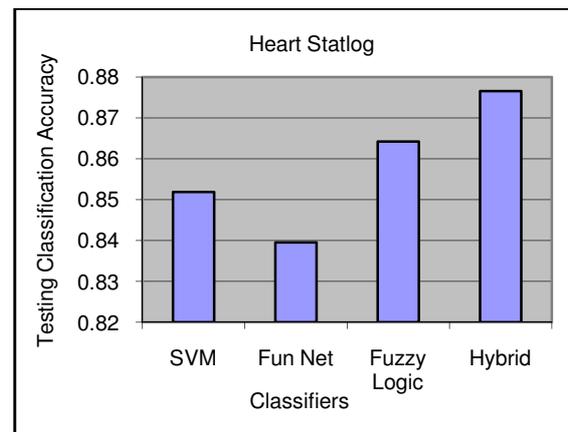

Fig. 19: Training Classification Accuracy of Heart Statlog Dataset

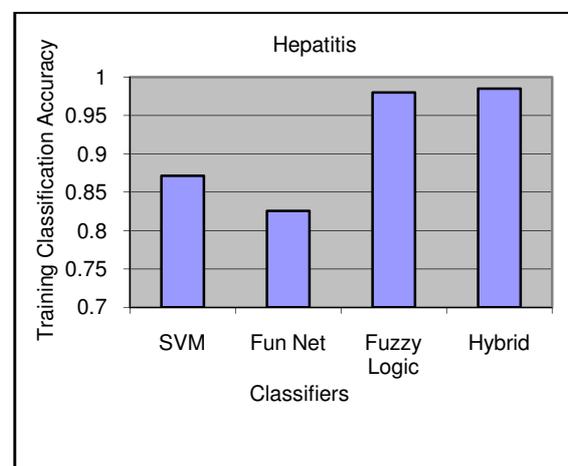Fig. 20: Testing Classification Accuracy of Heart Statlog Dataset

Fig. 21: Training Classification Accuracy of Hepatitis Dataset

9. Conclusion and Future Work

This paper demonstrates the use of individual computational intelligence techniques, ensemble and hybrid of computational models for classifying of real bioinformatics datasets. Individual, ensemble and hybrid computational models were developed and their performance checked under different modeling conditions. The results obtained indicate that the diagnostic performance of the hybrid and ensemble models is much better compared to the performance of the individual models in terms of the classification accuracy. As a future work, more research will be done with the aim of developing more heterogeneous ensemble and hybrid models

Fig. 22: Testing Classification Accuracy of Hepatitis Dataset

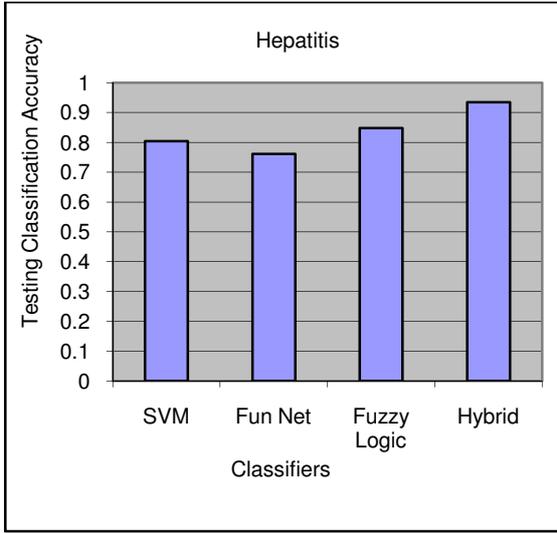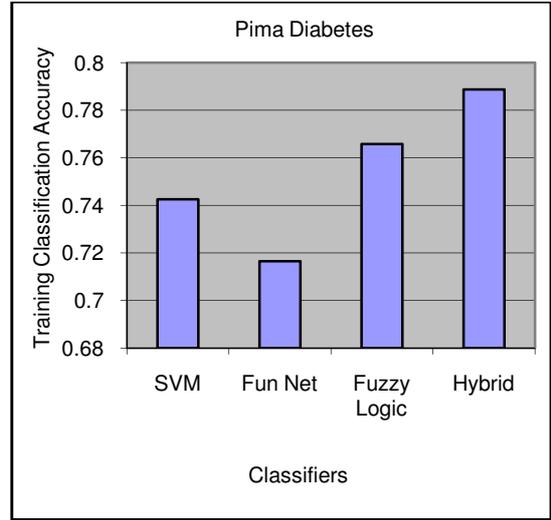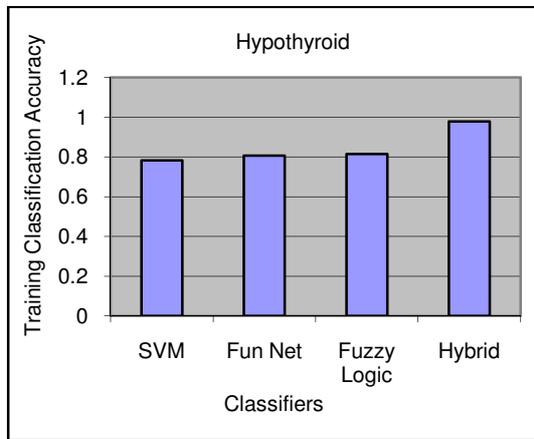Fig. 23: Training Classification Accuracy of Hypothyroid Dataset

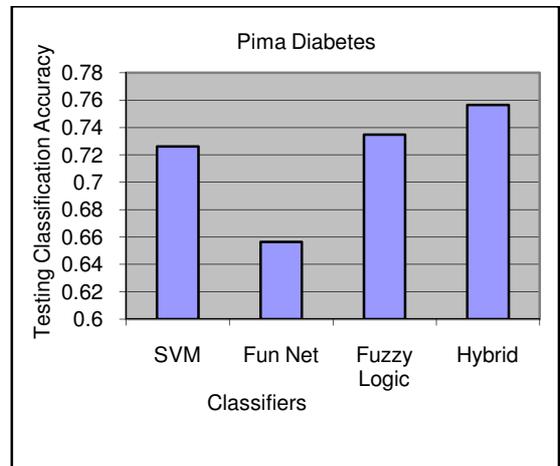Fig. 24: Testing Classification Accuracy of Hypothyroid Dataset

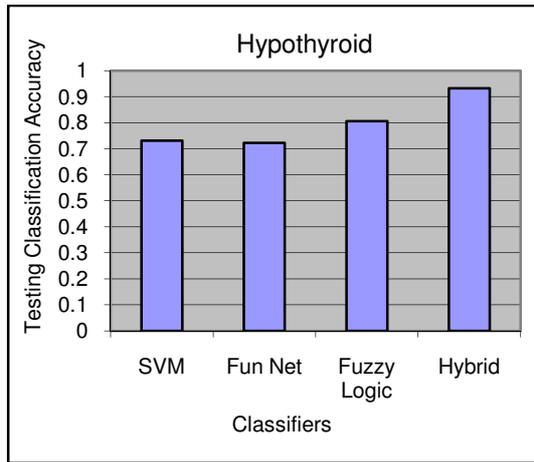Fig. 25: Training Classification Accuracy of Pima Diabetes Dataset

Fig. 26: Testing Classification Accuracy of Pima Diabetes Dataset

References

1. Bharathi A., M. Natarajan, "Cancer Classification of Bioinformatics data using ANOVA", Inter. Journal of Computer Theory and Engineering, Vol. 2, No. 3, June, 2010, pp. 1793-8201.

2. Balakrishnan Sarojini, Narayanasamy Ramaraj and Savarimuthu Nickolas, "Enhancing the Performance of LibSVM Classifier by Kernel F-Score Feature Selection", Contemporary Computing, 2009, Volume 40, Part 10, pp. 533-543.

3. Comak E., Arslan A., Turkoglu I., A decision support system based on support vector machines for diagnosis of the heart valve diseases. Elsevier, 2007, vol. 37, 21-27.

4. Charles X. Ling, William Stafford Noble, and Qiang Yang, "Guest Editors' Introduction to the Special Issue: Machine Learning for Bioinformatics", IEEE/ACM Transactions computational biology and bioinformatics, Vol. 2, NO. 2, April-June 2005.

5. Chitra and S. Uma, An Ensemble Model of Multiple Classifiers for Time Series Prediction, International Journal of Computer Theory and Engineering, Vol. 2, No. 3, pp. 1793-8201, 2010.

6. Lihua Li, Hong Tang, Zuobao Wu, Jianli Gong, Michael Gruidl, Jun Zou, Melvyn Tockman, Robert A. Clark, "Data mining techniques for cancer detection using serum proteomic profiling, Artificial Intelligence in Medicine, 2004, vol. 32, 71-83.

7. Kiyana, T., Yildirim, T., "Breast cancer diagnosis using statistical neural networks. In International XII, Turkish symposium on artificial intelligence and neural networks, 2003.

8. Cortes, C., Vapnik, V., "Support-vector networks", Machine Learning, 20(2), pp. 273–297, 1995.

9. Herron P., Machine Learning for Medical Decision Support: Evaluating Diagnostic Performance of Machine Learning Classification Algorithms, INLS 110, Data Mining, 2004.

10. Daya S., "Diagnostic test-Receiver Operating Characteristic Curve", Evidence-based Obstetrics and Gynecology, 2006, vol. 8, no. 1-2, 3-4.

11. Jonathan L. Lustgarten, Vanathi Gopalakrishnan, Himanshu Grover, and Shyam Visweswaran, "Improving Classification Performance with Discretization on Biomedical Datasets", AMIA annual symposium Proc. 2008, pp. 445–449.

12. Linder R., Dew D., H. Sudhoff, D. Theegarten, K. Remberger, S.J. Poppl, and M. Wagner, "The Subsequent Artificial Neural Network Approach Might Bring More Classificatory Power to ANN-Based DNA micro Array analyses", Bioinformatics 20 (2004), pp. 3544–3552

13. Tarek Helmy, Zeehasham Rasheed, "Multi-Category Bioinformatics Dataset Classification using Extreme Learning Machine", Proceedings of IEEE Congress on Evolutionary Computation, pp. 3234-3240, 18-21 May, 2009, Norway.

14. Tarek Helmy, Mohamed Al-Mulhim, Zeehasham Rasheed, "Type-2 Fuzzy Logic Based System for Classification of Bioinformatics Datasets" Proc. of IEEE 2009 International Conference on Bioinformatics & Computational Biology, pp. 812-816, USA, July 13-16, 2009.

15. Tsymbal A., Bolshakova N., Guest Editorial Introduction to the Special Section on Mining Biomedical Data, IEEE Transactions On Information Technology In Biomedicine, 2006, vol. 10, no. 3, 425-428.

16. Bittencourt V. G., M. C. C. Abreu, M. C. P. d. Souto, and A. M. d. P. Canuto, "An empirical comparison of individual machine learning techniques and ensemble approaches in protein structural class prediction", Inter. Joint Conference on Neural Networks, 2005, pp. 527 - 531.

17. Van Gerven M. A.J., Jurgelenaite R., Taal B. G., Heskes T., Lucas P. J.F., Predicting carcinoid heart disease with the noisy-threshold classifier. Artificial Intelligence in Medicine, 2007, vol. 40, 45-55.

18. UCI repository of bioinformatics Databases, Website: http://www.ics.uci.edu/~mlearn/MLRepository.html

19. Vlahou A., Schorge J. O., Gregory B. W., Coleman R. L., Diagnosis of ovarian cancer using decision tree classification of mass spectral data. Journal of Biomedicine and Biotec, 2003 vol. 5 308-314.

20. Witten I. H., Frank E., Data Mining, Practical Machine Learning Tools and Techniques, 2nd Elsevier, 2005

21. Witten, I., Frank E. "Data Mining: Practical Machine Learning Tools with Java Implementations", Morgan Kaufmann, San Francisco, 2000.

22. Wutao Chen, Huijuan Lu, Mingyi Wang, Cheng Fang, "Gene Expression Data Classification Using Artificial Neural Network Ensembles Based on Samples Filtering", IEEE proceedings of 2009 International Conference on Artificial Intelligence and Computational Intelligence, pp. 626-628.

23. Wangshu Zhang, Feng Zeng, Xuebing Wu, Xuegong Zhang, Rui Jiang, "A comparative study of ensemble learning approaches in the classification of breast cancer metastasis", Proceedings of the IEEE 2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, pp. 242-245.

24. Yang Pengyi, H. Yang Yee, B. Zhou Bing, Y. Zomaya Albert, "A Review of Ensemble Methods in Bioinformatics", Bioinformatics Volume 5, Number 4, December 2010, pp. 296-308(13).

25. Turkoglu I., Arslan A., Ilkay E., An expert system for diagnosis of the heart valve diseases. Expert Systems with Applications, 2002, vol. 23, no.3, 229–236.

26. Tarek Helmy, Anifowose Fatai, Kanaan Faisal, "Hybrid Computational Models for the Characterization of Oil and Gas Reservoirs", Elsevier Journal of Expert Systems with Applications 37 (2010), pp. 5353-5363.

27. Tarek Helmy, Anifowose Fatai, "Hybrid Computational Intelligence Models for Porosity and Permeability Prediction of Petroleum Reservoirs". International Journal Computational Intelligence and Applications, Vol. 9, No. 4, pp. 1-25, 2010.

28. Castillo E., "Functional Networks", Neural Processing Letters, Vol. 7, 1998, pp. 151-159.

29. Castillo E., Gutiérrez J., Hadi S., and Lacruz B., "Some Applications of Functional Networks in Statistics and Engineering", Technometrics, Vol. 43, 2001, pp. 10–24.

30. Karnik N. and Mendel J., "Type-2 Fuzzy Logic Systems", IEEE Transactions on Fuzzy Systems, Vol. 7, No. 6 (1999).

31. Mendel J., "Type-2 Fuzzy Sets: Some Questions and Answers," IEEE Connections, Newsletter of the IEEE Neural Networks Society, Vol. 1, 2003, pp. 10-13.

32. Zarandi M. F., Rezaee B., Turksen I. B. and Neshat E., "A Type-2 Fuzzy Rule-Based Expert System Model for Stock Price Analysis", Expert Systems with Applications, 2007, pp. 9-34.

33. Süreyya Özöğür-Akyüz, Devrim Ünay and Alex Smola, "model selection and optimization in machine learning", Volume 85, Numbers 1-2, (2011), DOI: 10.1007/s10994-011-5261-8.

34. Zenglin Xu, Rong Jin, Haiqin Yang, Irwin King, Michael R. Lyu, "Simple and Efficient Multiple Kernel Learning by Group Lasso", *Proceedings of the* 27th *International Conference on Machine Learning*, Haifa, 2010.

35. Argyriou, Andreas, Hauser, Raphael, Micchelli, Charles A., and Pontil, Massimiliano. A dc-programming algorithm for kernel selection. In *ICML '06: Proc. of the 23rd international conference on Machine learning*, pp. 41–48, 2006.

36. Zien, Alexander and Ong, Cheng Soon. Multiclass multiple kernel learning. In *ICML '07: Proc. of the* 24th international conference on Machine learning, pp. 1191–1198, 2007.

37. Mendel, J.M. and Liang, Q.: "Pictorial Comparison of Type-1 and Type-2 Fuzzy Logic Systems", In Proceedings of IASTED International Conference on Intelligent Systems & Control, Santa Barbara, CA, Oct., 1999.

38. Mendel, J. M.: "Uncertain Rule-Based Fuzzy Logic Systems", Prentice-Hall, Upper Saddle River, NJ 07458, 2001.

39. Tarek Helmy, Zeehasham Rasheed, Mohamed Al-Mulhem, "Adaptive Fuzzy Logic-Based Framework for Handling Imprecision and Uncertainty in Classification of Bioinformatics Datasets", International Journal of Computational Methods, Vol. 8, Issue 3, Sep., 2011, pp. 513-534.

40. Sonnenburg Soren, Ratsch Gunnar, Schafer Christin, and Scholkopf Bernhard, "Large scale multiple kernel

learning", Journal of Machine Learning Research, 7:1531–1565, 2006.

41. Harchaoui Zaid and Bach Francis, "Image classification with segmentation graph kernels", Proc. Computer Vision and Pattern Recognition (CVPR), pp. 1–8, 2007.

42. Ganesan T., P. Vasant, I. Elamvazuthi, "Optimization of non linear geological structure mapping using hybrid Neuro-Genetic techniques", Mathematical and Computer Modeling, 54 (2011), pp. 2913-2922.

43. Vasant P., Bhattacharya A., Sarkar B., Mukherjee S.K., "Detection of level of satisfaction and fuzziness patterns for MCDM model with modified flexible S-curve MF", Applied Soft Computing Journal, volume 7, issue 3, year 2007, pp. 1044 – 1054.

44. Vasant P., I. Elamvazuthi, T. Ganesan, "Iterative fuzzy optimization approach for crude oil refinery industry", Scientific Annals of Computer Science, Vol. 8, No. 2, pp. 262-280, 2010.