International Journal of Computer Applications (0975 – 8887)
Volume 31– No.8, October 2011

A Novel SVM based CSSFFS Feature Selection Algorithm for Detecting Breast Cancer

S. Aruna
Research Scholar
Dr M.G.R Educational and Research Institute
University, Chennai-95

Dr S.P. Rajagopalan
Professor Emeritus
Dr M.G.R Educational and Research Institute
University, Chennai-95

ABSTRACT

In this paper a feature selection algorithm CSSFFS (Constrained search sequential floating forward search) based on SVM is proposed for detecting breast cancer. It is a greedy algorithm with search strategy of constrained search. The aim of this algorithm is to achieve a feature subset with minimal BER (Balanced error rate). This is a hybrid algorithm with the combination of filters and wrappers. Feature ranking with SVM acts as filters for removing irrelevant features. Then SFFS acts as wrapper which further removes the redundant features yielding the optimal subset of features. WDBC dataset from UCI machine learning depository is used for the experiment. The experiments are conducted in WEKA. After feature selection the accuracy and BER for WDBC dataset is 98.2425 and 0.0226 respectively with 15 features.

General Terms

Pattern Recognition, Machine learning, Algorithms.

Keywords

SVM, breast cancer diagnosis, feature selection, SFFS, BER.

1. INTRODUCTION

Breast cancer is one of the most common cancers among women. Although breast cancer is a potentially fatal condition, early diagnosis of disease can lead to successful treatment [1]. One of the important steps to diagnose the breast cancer is classification of tumor. Tumors can be either benign or malignant but only the latter is cancer. So, malignant tumors generally are more serious than benign tumors. Early diagnosis needs a precise and reliable diagnosis procedure that allows physicians to distinguish between benign breast tumors and malignant ones [2]. Unfortunately not all the physicians are experts in cross domain. Hence automation of diagnostic system is needed. There may be irrelevant, redundant, noisy data in real world data and not all the attributes are useful for classification. Hence feature selection is needed while dealing with real world data sets.

Feature selection, as a preprocessing step to machine learning especially in real world data, has been very effective in reducing dimensionality, removing irrelevant data, and noise from data to improve result comprehensibility [3]. Therefore these features increase the cost of retain and management of data and cause of confusing the algorithm of classification. Generally, they lead to a low learning precision [4, 5, 6]. Feature selection task is to choose a subset of the original features present in a given dataset that provides most of the useful information [7]. Feature selection directly reduces the number of original features by selecting a subset of them that still retains sufficient information for classification. Feature selection has many advantages; some benefits include facilitating data visualization and data understanding, reducing the measurement and storage requirements, reducing training and utilization times, defying the curse of dimensionality to improve prediction performance [8].

In this paper a novel SVM based CSSFFS feature selection algorithm is proposed for detecting breast cancer. This paper is structured as follows. Section 2 gives a brief introduction to Support vector machines, Sequential forward selection method and WDBC dataset used for the experiment. The CSSFFS feature selection algorithm is explained in Section 3. Section 4 discusses the results obtained and concluding remarks are given in Section 5 to address further research issues.

2. MATERIALS AND METHODS

2.1 Support vector Machines

The support vector machine (SVM) [9] is a very powerful pattern recognition technique useful for recognizing subtle patterns in complex datasets. SVMs have been applied in numerous bioinformatics domains including recognition of translation start sites [10], protein fold recognition [11], microarray gene expression analysis [12,13], prediction of protein–protein interactions [14] and peptide identification from mass spectrometry data [15].

SVMs are supervised learning technique used for classification and regression. The algorithm performs discriminative classification, learning by example to predict the classifications of previously unseen data. Their aim is to devise a computationally efficient way of identifying separating hyper planes in a high dimensional feature space. In particular, the method seeks separating hyper planes maximizing the margin between sets of data. This should ensure a good generalization ability of the method, under the hypothesis of consistent target function between training and testing data. To calculate the margin between data belonging to two different classes, two parallel hyper planes are constructed, one on each side of the separating hyper plane, which is “pushed up against” the two data sets. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the neighboring data points of both classes, since in general the larger the margin the lower the generalization error of the classifier. The parameters of the maximum-margin hyper plane are derived by solving large quadratic programming (QP) optimization problems.

There exist several specialized algorithms for quickly solving these problems that arise from SVMs, mostly reliant on heuristics for breaking the problem down into smaller, more manageable chunks. In this work the implementation of John Platt's [16] sequential minimal optimization (SMO) algorithm for training the support vector classifier is used. SMO works by breaking the large QP problem into a series of smaller 2-dimensional sub-problems that may be solved analytically, eliminating the need for numerical optimization algorithms such as conjugate gradient methods. The implementation we used is the one contained in the WEKA public domain software [17]. This implementation globally replaces all missing values and transforms nominal attributes into binary ones. It also normalizes all attributes by default (in that case the coefficients in the output are based on the normalized data, not the original data and this is important for interpreting the classifier). The main parameter values used in this work are reported in Table 1. All these parameter values correspond to the standard values offered by the WEKA software and they are defined for instance in [16].

Table 1. Parameters for kernel

C	Epsilon	Exponent	Cache size	T	CV
1.0	1.0E-12	1.0	250007	0.0010	10

2.2 Sequential forward floating selection

Sequential forward floating selection (SFFS) and sequential backward floating selection (SBFS) methods proposed by Pudil *et al.* [18] efficiently overcome the nesting problem by dynamically backtracking after each sequential step to locate a better subset. The SFFS algorithm begins the search with an empty feature set and uses the SFS algorithm to add one feature at a time to the selected feature subset. Every time a new feature is added to the current feature subset, the algorithm attempts to backtrack by using the SBS algorithm to remove one feature at a time to locate a better subset. The search terminates when the size of the current feature set is larger than the number of features *d* we want. This is necessary to allow backtracking. SFFS is renowned for presenting an excellent cost-benefit in terms of the computational complexity and the quality of the returned solution. There are some variants of this algorithm (adaptive and generalized floating search methods) that try to improve the SFFS solutions at the expense of an increase on the computational cost. However, they cannot avoid the nesting effect completely [19].

2.3 Wisconsin Diagnostic Breast Cancer

This breast cancer database was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg available in UCI Machine learning depository [20]. Features are computed from a digitized image of a Fine Needle Aspiration (FNA). They describe characteristics of the cell nuclei present in the image. Number of instances: 569, Number of attributes: 32 (ID, diagnosis, 30 real-valued input features).

Attribute Information :

1) ID number

2) Diagnosis (M = malignant, B = benign)

3-32) ten real-valued features are computed for each cell nucleus:

a) Radius (mean of distances from center to points on the perimeter)

b) Texture (standard deviation of gray-scale values)

c) Perimeter

d) Area

e) Smoothness (local variation in radius lengths)

f) Compactness (perimeter^2 / area - 1.0)

g) Concavity (severity of concave portions of the contour)

h) Concave points (number of concave portions of the contour)

i) Symmetry

j) Fractal dimension ("coastline approximation" -1)

The mean, standard error, and "worst" or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. For instance, field 3 is Mean Radius, field 13 is Radius SE, and field 23 is Worst Radius. All feature values are recoded with four significant digits. Class distribution: 357 benign, 212 malignant.

3. CSSFFS ALGORITHM

CSSFFS is a greedy algorithm based on constrained search. The aim of this algorithm is to achieve a feature subset with minimal Balanced Error Rate (BER). This is a hybrid algorithm with the combination of filters and wrappers. The ranking acts as filters for removing irrelevant features. Then SFFS acts as wrapper which further removes the redundant features if any, yielding the optimal subset of features. This algorithm uses BER as the main criterion for selecting feature subset. BER is the average of the error rate on positive class examples and the error rate on negative class examples. If both class have uneven examples then overall accuracy is different from BER. For example, assume a dataset contains 90 positive and 10 negative instances if all the instances are predicted positive then accuracy is 90% but BER us 50%. Hence the feature subset selected by this algorithm will have features that help in improving the classification accuracy of both classes. At the end of the algorithm the dimension d of the resultant set is d<=D/3, where D is the dimension of the original feature set. The worth of an attribute is evaluated by using an SVM classifier. Parameters used for attribute evaluation are C=1.0, epsilon= 1.0E-25, tolerance parameter= 1.0E-10 The weight of every feature $w(t)$ is calculated from equation 1.

$$w(t)=\sum_{SVs} yiaixi\,(t) \qquad (1)$$

where $xi(t)$ is the value of the tth feature of the ith sample; Attributes are ranked by the square of the weight assigned by the SVM. Fig 1 shows the schematic flowchart of CSSFFS algorithm.

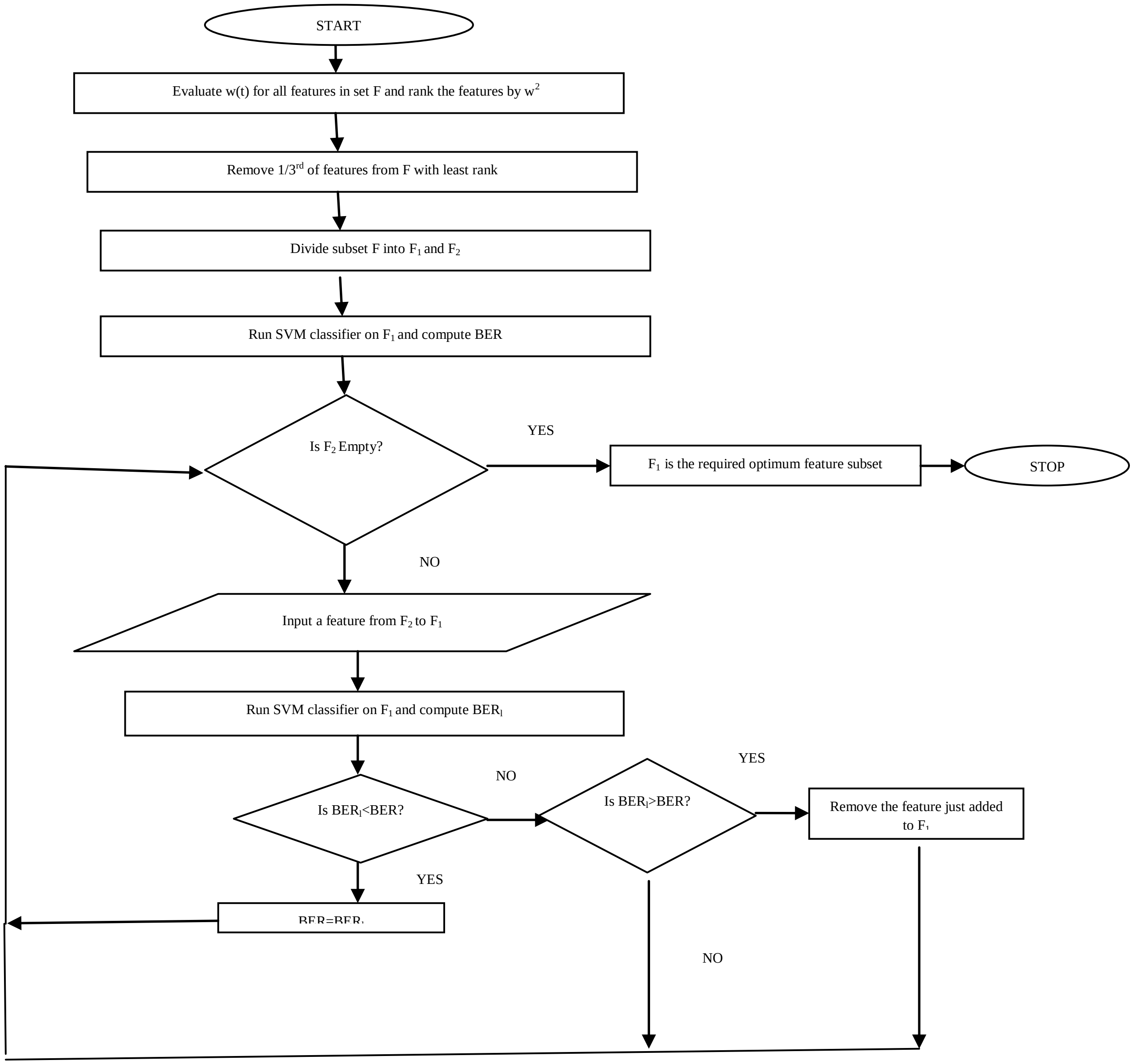

Fig 1: Flow chart for CSSFFS algorithm

Agorithm:

Input: Data set F with all features.

Output: Optimum feature subset F_1

Step 1: Rank the features in F using $w^{2.}$

Step 2: Remove $1/3^{rd}$ of the features from F which has least ranks.

Step 3: Divide the set F into 2 subsets $F_1=\{x_{1,}x_2........x_{n/2}\}$ and $F_2=\{x_{n/2+1},x_{n/2+2},....x_n\}$.

Step 4: Run the classifier on set F_1 and obtain BER.

Step 5: Do steps 6-9 until set F_2 is empty.

Step 6: Add a feature from F_2 to $F_{1.}$

Step 7: Run the classifier on F_1 and compute BER_l.

Step 8: If $BER_l <$ BER update BER as BER= BER_l

Step 9: If BER_l>BER remove the feature just added to F_1

Step 10: The set F_1 is the required optimum feature subset.

4. RESULTS

The Experiments were conducted in WEKA with 10 fold cross validation. Ten fold cross validation has been proved to be statistically good enough in evaluating the performance of the classifier [21]. The attributes are ranked according to the square of weights calculated by SVM and divided into 3 subsets. Table 2 show the most relevant, relevant and irrelevant attributes ranked by SVM classifier. After removing irrelevant features SVM classifier is trained over most relevant features and the BER obtained is 0.03955. Then constrained search for optimum feature set is done for most relevant and relevant features with the help of SVM poly kernel of degree 1. Table 3 shows the steps carried out during the process of subset selection.

Table 2. Most relevant, relevant and irrelevant features ranked by SVM attribute selector

Most relevant features	Relevant features	Irrelevant features
21,28,23,22,8,24,29,1,25,4	11,2,3,7,16,13,10,27,14,9	6,5,15,20,30,id,12,17,18,19,26

Table 3. Steps carried out in each iteration.

Iteration	Feature added	Feature eliminated	Balanced error rate	Comments
1	11		0.03485	BER$_l$ <BER-update GER
2	2		0.02730	BER$_l$ <BER-update BER
3	3		0.02870	BER$_l$ >BER-Backtrack
4	7	3	0.02730	BER$_l$ =BER
5	16		0.02495	BER$_l$ <BER-update BER
6	13		0.02260	BER$_l$ <BER-update BER
7	10		0.02730	BER$_l$ >BER-Backtrack
8	27	10	0.02495	BER$_l$ >BER-Backtrack
9	14	27	0.02495	BER$_l$ >BER-Backtrack
10	9	14	0.02635	BER$_l$ >BER-Backtrack

Table 4 summarizes the Memory utilized by the data set (in KB), Model build time, Accuracy, Mean absolute and Root mean squared error, AUC for SMO poly kernel before and after feature selection. It can be seen from the table 4 results that the Memory utilized in KB, Model build time, Mean absolute error and Root mean squared error has been decreased and the Accuracy and Area under ROC(AUC) has been increased. The data mining algorithms such as Simple Cart, RBF Network, Naïve bayes and J48 are used to classify the WDBC dataset with all features(32) and with optimum features (15) selected by CSSFFS algorithm. The results are shown in Table 5 and 6. Table 5 shows the results for various data mining algorithms with all 32 features and Table 6 shows the results for various data mining algorithms with the 15 features selected by CSSFFS algorithm. The empirical comparison shows that the optimum features selected by CSSFFS algorithm also improved the accuracy of all the datamining algorithms. Classification accuracies of CSSFFS algorithm and other methods for WDBC dataset from literature are summarized in Table 7.

Table 4. Memory(in KB),Model build time, Accuracy, mean absolute and root mean squared error, AUC for SMO before and after feature selection.

No of features	Memory(KB)	Modelbuild time(sec)	Accuracy(%)	Mean absolute error	Root mean squared error	AUC
All	121	0.59	97.8910	0.0211	0.1452	0.973
15	55.8	0.05	98.2425	0.0176	0.1326	0.977

Table 5. Accuracy, mean absolute and root mean squared error for various datamining algorithms without feature selection

Algorithm	Modelbuild time(sec)	Accuracy(%)	Mean absolute error	Root mean squared error
Simple Cart	0.72	92.9701	0.0878	0.2558
RBF Network	0.30	93.6731	0.0948	0.2338
Naïve bayes	0.03	92.6186	0.0732	0.2648
J48	0.06	92.9701	0.0758	0.2608

Table 6. Accuracy, mean absolute and root mean squared error for various datamining algorithms with feature selection

Algorithm	Modelbuild time(sec)	Accuracy(%)	Mean absolute error	Root mean squared error
Simple Cart	0.27	93.6731	0.0823	0.2415
RBF Network	0.25	95.0791	0.0781	0.2027
Naïve bayes	0.02	94.3761	0.0557	0.2246
J48	0.03	94.0246	0.0675	0.2406

Table 7. Accuracy rate comparisons of CSSFFS with other approaches from previous researches on WDBC medical database. (*Mean value)

Author (year) Reference	**Prediction Method**	**Accuracy (%)**
Quinlan (1996) [22]	C4.5	94.74
Ster and Dobnikar (1996) [23]	Multi layer perceptron/backpropag. LVQ kNN Euclidean/Manhattan NB—naıve Bayes Linear discriminant analysis Assistant I tree—ASI Assistant R tree—ASR Lookahead feature construction binary tree	96.7 96.6 96.6 96.4 96.0 95.6 94.7 94.4
Shang etal(1996) [24]	DB-CART	96.2
Hamilton et al(1996) [25]	C4.5-decision tree RIAC	96.0 94.99
Adamczak and Duch (1997) [26]	Feature space mapping	98.30
Bennett and Blue (1997) [27]	Memetic pareto artificial neural Support vector machine OC1 DT GTO DT C4.5	97.2 95.9 95.7 93.4
Jankowski et al(1997)[28]	IncNet	97.1
Nauck and Kruse (1999) [29]	NEFCLASS	95.06
Pena-Reyes and Sipper (1999) [30]	Fuzzy-GA1	97.36
Duch and Adamczak (2000) [31]	3-NN standard Manhatan kNN with DVDM distance 21-NN standard Euclidean	97.1 97.1 96.9
Abbas et ak2001)[32]	Back propagation	97.5
Abbas etal(2002) [33]	Network—MPANN	98.1
Goodman et al. (2002) [34]	Optimized-LVQ Big-LVQ AIRS	96.70 96.80 97.20
Abonyi and Szeifert (2003) [35]	Supervised fuzzy Clustering	95.57
Liu etal(2004) [36]	Ant_Miner3 Ant_Miner1	94.3* 92.6*
Ioannis etal(2006) [37]	DiagNN - min. b over AIC/MDL	97.9
A. K. Jain et al (2006)[38]	K-means 1-Nearest Neighbor Bayes Linear Classifier RF algorithm	90.78 95.25 96.66 91.03
Chung-Jui Tu et al (2007)[39]	PSO-SVM	95.61
Hua-Liang Wei et al(2007)[40]	FOS-MOD algorithm 5NN(full) 7NN(subset searched)	97.94 97.04
Maglogiannis et al.(2009)[41]	SVM Gaussian RBF (σ = 0.7) ANN Bayes Net Naïve Bayes	97.54 97.90 92.80 91.39
Gadaras and Mikhailov (2009) [42]	Fuzzy rule classification	96.08
Fernando E. B et al (2009)[43]	Ant-Miner cAnt-Miner cAnt-Miner-MDL cAnt-Miner2 cAnt-Miner2-MDL J48	90.39 * 93.88* 93.30 * 93.94* 93.64* 92.63*
C.-Y. Fan et al(2011)[44]	CBFDT	98.90
Li-Yeh Chuang (2011) [45]	CatfishBPSO	98.17
Mohammad Darzi et al(2011)[46]	CBRGenetic(all) After feature selection	94.74 97.37
Our method CSSFFS algorithm (10-CV)		98.25

5. CONCLUSION

In this paper CSSFFS feature selection algorithm is proposed for detecting breast cancer. This is a greedy algorithm based on constrained search. This is a hybrid algorithm with the combination of filters and wrappers. Attributes are ranked with the square of weights calculated by the SVM classifier. This acts as Filters to remove irrelevant features. From the remaining features SFFS with SVM is used to select the optimum subset of features. This act as a wrapper to remove the redundant features if any yields the required optimum subset. BER is used as the main criterion for selecting features. The objective of this algorithm is to select features with minimal BER. The experiments are conducted in WEKA. WDBC dataset with 32 features is used for the experiment. The CSSFFS algorithm yielded an optimum feature subset of 15 features with Accuracy, BER of 98.2425 and 0.0226 respectively. In this work this algorithm is used for breast cancer domain. In future work this algorithm will be experimented with other domains.

6. REFERENCES

[I]. Harirchi, et al., “Breast cancer in Iran: a review of 903 case records,” Public Health, 2000. 114(2): p. 143-145.

[2] T. Subashini, V. Ramalingam, and S. Palanivel, “Breast mass classification based on cytological patterns using RBFNN and SVM,” Expert Systems with Applications, 2009. 36(3): p. 5284-5290

[3] Y. liu and Y. F. Zheng, (2006) "FS_SFS: A novel feature selection method for support vector machines", Pattern Recognition, vol. 39, pp. 1333–1345.

[4] R. Kohavi, and G.H. John, “Wrappers for feature subset selection”,Artificial intelligence, 1997. 97(1-2): p. 273-324.

[5] Y. Yuling, “A Feature Selection Method for Online Hybrid Data Based on Fuzzy-rough Techniques,” 2009: IEEE.

[6] N. Abe, et al., “A divergence criterion for classifier-independent feature selection,” Advances in Pattern Recognition, 2000: p. 668-676.

[7] R. Jensen, and Q. Shen, “Computational intelligence and feature selection: rough and fuzzy approaches,” IEEE Press Series On Computational Intelligence, 2008: p. 340.

[8] I. Guyon, and A. Elisseeff, “An introduction to variable and feature selection,” The Journal of Machine Learning Research, 2003. 3: p. 1157-1182.

[9] Vapnik,V.N. (1998) Statistical Learning Theory. Adaptive and Learning Systems for Signal Processing, Communications, and Control. Wiley, New York.

[10] Zien,A., Rätch,G., Mika,S., Schölkopf,B., Lengauer,T. and Müller,K.-R. (2000) Engineering support vector machine kernels that recognize translation initiation sites. Bioinformatics, **16**,799–807.

[11] Ding.C and Dubchak;I (2001) Multi class protein fold recognition using support vector machines and neural networks. Bioinformatics, 17, 349-358.

[12] Guyon,I., Weston,J., Barnhill,S. and Vapnik,V. (2001) Gene selection for cancer classification using support vector machines. Machine Learning, **46**, 389–422.

[13] Furey,T.S., Cristianini,N., Duffy,N., Bednarski,D.W., Schummer,M. and Haussler,D. (2001) Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics, 16, 906–914.

[14] Bock,J.R. and Gough,D.A. (2001) Predicting protein–protein interactions from primary structure. Bioinformatics, **17**, 455–460.

[15] Anderson,D.C., Li,W., Payan,D.G. and Noble,W.S. (2003)

A new algorithm for the evaluation of shotgun peptide sequencing in proteomics: support vector machine classification of peptide MS/MS spectra and SEQUEST scores. J. Proteome Res., 2, 137–146

[16] Platt J: Fast Training of Support Vector Machines using

Sequential Minimal Optimization. Advances in Kernel Methods- Support Vector Learning 1998.

[17] Weka: A multi-task machine learning software developed by Waikato University. 2006, [http://www.cs.waikato.ac.nz/ml/weka]. [18] P. Pudil, J. Novovicova, and J. Kittler, “Floating search methods in feature selection,” Pattern Recognition Lett., vol. 15, pp. 1119-1125, 1994.

[19] P. Somol, P. Pudil, J. Novoviˇcová, and P. Paclík, “Adaptive floating search methods in feature selection Pattern Recognition Letters, vol. 20, pp. 1157–1163, 1999.

[20] UCI Machine learning depository, (http://archive.ics.uci.edu/ml).

[21] P. Baldi, S. Brunak, Y. Chauvin, et al. (2000) “Assessing the accuracy of prediction algorithms for classification and overview”, Bioinformatics, 5(5):412–424. [22] J.R. Quinlan, Improved use of continuous attributes in C4.5, Journal of Artificial Intelligence Research 4 (1996 77–90

[23] B. Ster, A. Dobnikar, Neural networks in medical diagnosis: comparison with other methods, in: Proceedings of the International Conference on Engineering Applications of Neural Networks (EANN’96), 1996, pp. 427–430.

[24] Shang N, Breiman L (1996) Distribution based trees are more accurate. In: Proceedings of the International conference on neural information processing, Hong Kong vol 1, pp133–138

[25] H.J. Hamilton, N. Shan, N.Cercone (1996) RIAC: a rule induction algorithm based on approximate classification. Technical report CS 96-06, Regina University

[26] Adamczak R, Duch W (1997) New developments in the feature space mapping model. In: 3rd Conference on neural networks and their applications, October 14–18 [27] K.P. Bennet, J.A. Blue, A Support Vector Machine Approach to Decision Trees, Math Report, Rensselaer Polytechnic Institute, Troy, 1997, pp. 97–100 [28] Jankowski N, Kadirkamanathan V (1997) Statistical control of RBF-like networks for classification. In: Proceedings of the 7th international conference on artificial neural networks, Lausanne, pp 385–390

[29] D. Nauck, R. Kruse, Obtaining interpretable fuzzy classification rules from medical data, Artificial Intelligence in Medicine 16 (1999) 149–169.

[30] C.-A.Pena-Reyes, M. Sipper, A fuzzy-genetic approach to breast cancer diagnosis, Artificial Intelligence in Medicine 17 (1999) 131–155.

[31] Duch W, Adamczak R, Grabczewski K (2000) A new methodology of extraction, optimization and application of crisp and fuzzy logical rules. IEEE Trans Neural Netw11(2):1-31

[32] Abbass HA, Towsey M, Finn GD (2001) C-net, a method for generating non-deterministic and dynamic multivariate decision trees. Knowl Inf Syst 3:184–197

[33] Abbass HA (2002) An evolutionary artificial neural networks approach for breast cancer diagnosis. Artif Intell Med 25(3):265–281

[34] D.E. Goodman, L. Boggess, A. Watkins, Artificial immune system classification of multiple-class problems, in: Proceedings of the Artificial Neural Networks in Engineering ANNIE, 2002, pp. 179–183.

[35] János Abonyi, Ferenc Szeifert, Supervised fuzzy clustering for the identification of fuzzy classifiers, Pattern Recognition Letters 24 (14) (2003) 2195–2207.

[36] Liu B, Abbass HA, McKay B (2004) Classification rule discovery with ant colony optimisation. IEEE Comput Intell Bull3(1): 31-35

[37] Ioannis Anagnostopoulos E, Ilias Maglogiannis(2006), Neural network-based diagnostic and prognostic estimations in breast cancer microscopic instances, Med Bio Eng Comput 44:773–784 DOI 10.1007/s11517-006-0079-

[38] A. K. Jain, Pavan K. Mallapragada, Martin Law, "Bayesian Feedback in Data Clustering," ICPR, vol. 3, pp.374-378 18th International Conference on Pattern Recognition (ICPR'06) Volume 3, 2006

[39] Chung-Jui Tu, Li-Yeh Chuang, Jun-Yang Chang, and Cheng-Hong Yang, Member, IAENG , IAENG International Journal of Computer Science, 33:1, IJCS_33_1_18

[40] Hua-Liang Wei and Stephen A. Billings Ieee Transactions On Pattern Analysis And Machine Intelligence, Vol. 29, No. 1, January 2007 [41] Ilias Maglogiannis · Elias Zafiropoulos Ioannis Anagnostopoulos ,An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers, Appl Intell (2009) 30: 24–36

[42] I. Gadaras, L. Mikhailov, An interpretable fuzzy rule-based classification methodology for medical diagnosis, Artificial Intelligence in Medicine 47 (1) (2009) 25–41 [43] Fernando E. B. Otero, Alex A. Freitas, and Colin G Johnson, Proceedings of the IEEE Symposium Computational Intelligence and Data Mining, CIDM 2009 part of the IEEE Symposium Series on Computational Intelligence 2009, Nashville, TN, USA, March 30, 2009 April 2, 2009. IEEE 2009

[44] Chin-Yuan Fana, Pei-Chann Changb, Jyun-Jie Linb, J.C Hsiehb, A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification, Applied Soft Computing 11 (2011) 632–644

[45] Li-Yeh Chuang, Sheng-Wei Tsai, Cheng-Hong Yang(2011), Catfish Binary Particle Swarm Optimization for Feature Selection, In: Proceedings of the international Conference on Machine Learning and Computing IPCSIT vol.3 (2011)pp 40-44

[46] Mohammad Darzi, Ali AsgharLiaei,Mahdi Hosseini, HabibollahAsghari , Feature Selection for Breast Cancer Diagnosis:A Case-Based Wrapper Approach ,(2011), World academy of Science, Engineering and Technology 77, 2011,pp 1142-1143.