Journal of Theoretical and Applied Information Technology
29th February 2012. Vol. 36 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN EFFICIENT MEDICAL IMAGE DIAGNOSIS SYSTEM USING SOFT COMPUTING TECHNIQUES

[1]V.SIVAKRITHIKA, [2]S.SENGOLE MERLIN [3]K.SUGIRTHA

[1]Asstt Prof., Department of Computer Science and Engineering, PRIST University

[2]Assoc. Prof., Department of CSE, Loyola-ICAM College of Engineering and Technology

[3] Pursuing M.E, CSE, HITS University

E-mail: [1]sivakrithikav@gmail.com, [2]sengolemerlin@gmail.com [3]sgitha@gmail.com

ABSTRACT

In this paper, we aim at presenting the process of building up of an accurate data mining architecture for detecting cancer at very early stages to enhance the probability of successful cure. Existing Computer Aided Diagnosis (CAD) systems make use of the association rule-based classification that suffers from some problems inherited from the Association Rule Mining (ARM) such as handling of continuous data and the support/confidence framework. Such problems lead to impreciseness in its operation and uncertainty about the results. The proposed method is a hybrid system that combines the principles of soft computing techniques such as Fuzzy Discretization, Genetic Network Programming (GNP) and GNP based Classification. Application of Fuzzy Discretization on the extracted features of the image makes it feasible to identify the potential cancer forming cells of tumors even at initial stages of their development. GNP is used for the Rule mining as it optimizes the rule extraction process and make the system autonomous and self-learning. This method of diagnosis of cancer based on Fuzzy Discretization and Rule Mining using GNP provides more accurate results in identifying the images in a tumor to be Malign (cancerous cell), Benign (not cancerous) .

Keywords: *Medical Image Diagnosis, Cancer Diagnosis, Soft Computing Techniques, Fuzzy logic, Genetic Network Programming ,Texture Feature Extraction, Fuzzy Discretization, Genetic Network Programming Based Rule Mining.*

1. INTRODUCTION

The world had witnessed the birth and growth of the computing technology that has and is revolutionizing every sphere of life on our planet. The introduction of automation with help of computing machines, human errors due to various biases and other known and unknown causes have been almost completely eliminated. The field of medicine has its own computer aided simple as well as automated tools for various activities. Diagnosis is the most important aspect of treatment for diseases. Though diagnosis is easy and simple for many diseases there are few diseases including cancer which requires much caution because, the fatal diseases are required to be detected and confirmed in real time or at very early stages, because the available treatment methods call for it. CT, Computer tomography uses scanning methods to give multi-dimensional images. These images are found to be most reliable in early detection of various diseases including tumors, in a system called Computer Aided Diagnosis (CAD). It is an interdisciplinary field that utilizes techniques available in areas of data mining, digital image processing, radiology. It is so because the scan images are the most difficult to read due to their low contrast that displays the tissue texture gradient which the experts in oncology use, to determine the nature of the cells of a tumor. In fact as we are knowledgeable, CAD as on today employs different classification methods available in different gadgets in data mining for cancer detection. However, there is wild search for a high tech methodology in data mining [1] for narrowing the possibility of escaping detection, especially at the early stages of tumors to be malign.

The project in hand proposes cancer detection using scan images, which can assist the medical image diagnosis system. The method proposed here makes use of association rule mining technique to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

classify the scan images. It combines the low-level features extracted from images and high level knowledge from intelligent system. Feature extraction is done using texture analysis [2] which deals with the spatial variations of the pixel intensities. Fuzzy Discretization is used in the partitioning of the continuous attributes and fuzzifies the data into LOW, MID and HIGH using fuzzy membership functions. Association rule-based classification is one of the most important data mining techniques applied to many scientific problems. ARM is one of the most popular data mining methods for wide range of applications and is used to discover association rules or correlations among a set of attributes in a dataset.In this paper, a novel Fuzzy Discretized feature extraction and Rule mining using GNP that can deal with the problems inherited by Association rule-based classifiers is proposed.GNP helps to minr rules efficiently. GNP is one of the evolutionary optimization algorithms that uses directed graph structures as solutions instead of strings (Genetic Algorithms) or trees (Genetic Programming). GNP can deal with more complex problems by using the higher expression ability of graph structures.

The proposed system consists of: a pre-processing phase, feature extraction phase, a phase for mining the resultant transaction database with classification rules, a final phase to build the classifier and generating the suggestion of diagnosis.

The paper is organized as follows: Related work is represented in section 2. Details of the proposed method are described in section 3. Section 4 contains details of experimental procedure and results. Conclusion and future work is presented in section 5

2. RELATED WORKS

Most of the medical image mining method makes use of segmentation method or feature extraction process for mining information from images. Water immersion method for detecting region of interest in brain tumors was presented in . Computational effort increases when we consider segmentation process and labeling objects for scan images. Medical images are analyzed using the differences in the tissue density that could be represented by textural variations .In [3], QURAT-UL-AIN et al, has used co-occurrence matrix for feature extraction purpose. Histogram based features are local in nature. So for this purpose gray-level spatial co-occurrence matrix based texture features that consider spatial information into consideration are extracted. This method is simple and efficient to represent images which generate the single signature for an image. For image analysis, CAD systems involve image classification and segmentation which is not sufficient. The data mining concept proposed in has been widely used in medical information system for processing large volumes of data recently. In [4], Marcela X.riberiro et al proposed association rule mining based medical image diagnosis method. They used two new algorithms PreSAGe for feature discretization and HiCARe for classification purpose. PreSAGe algorithm suffers from sharp boundary problem while handling continuous values of the features. Discretization process involves large computational effort and affects the speed of the system. Moreover classifier also provides less accuracy.

In [5], Hüllermeier and Yi justify the relevance of fuzzy logic being applied to association rule mining in today‟s data-mining setup. Like the crisp version of Apriori, fuzzy Apriori is a very slow and inefficient algorithm for very large datasets (in the order of millions of transactions) In [6], Rajendran et al, has reported that Novel Fuzzy Association Rule Mining (NFARM) can be applied on the image transaction database which contains the features that are extracted from the brain scan images. NFARM is based on fuzzy partitioning method which is a tedious task.

In [7], Karla Taboada et al, proposed a Genetic Network Programming(GNP) based data mining method has been proposed for discovering comprehensible fuzzy association rules, which is potentially useful for classification tasks. GNP-based association rule mining is a tool for extracting a large number of rules into a general pool. The genetic operators executed in GNP such as crossover and mutation allows the GNP-based association rule mining to find new rules according to the progress of evolution.

In this paper we propose a hybrid system, novel Fuzzy - Genetic Network Programming (GNP) based rule mining to classify the images for the user‟s purpose in a short time very accurately without human intervention by allowing machine learning.

Journal of Theoretical and Applied Information Technology
29th February 2012. Vol. 36 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

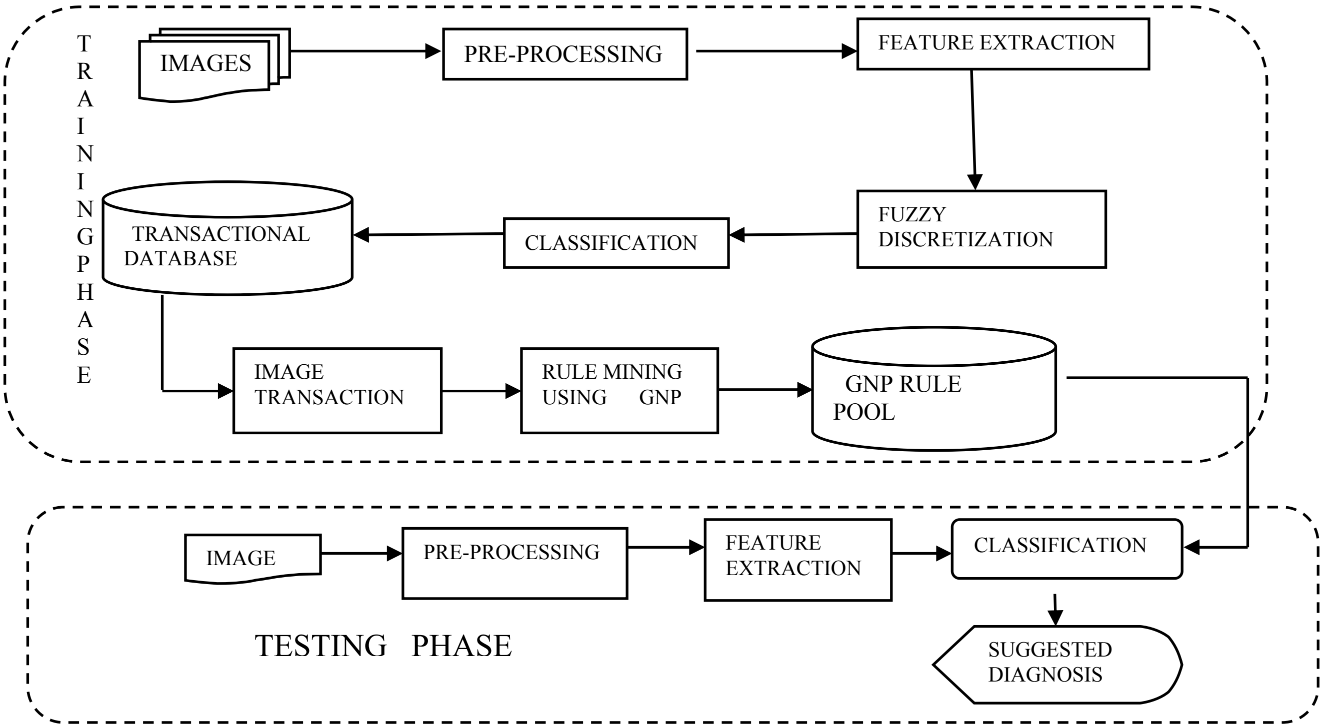

Fig.1 Pipeline of the proposed System

3. PROPOSED SYSTEM

The proposed intelligent system for classifying medical images using soft computing techniques consists of two phases: Training phase and Testing phase (Fig.1). The proposed novel method involves texture feature extraction, fuzzy discretization, rule mining using GNP to classify the images accurately. In training phase, image samples are collected and features are extracted from them. The features are transformed into feature vector and the values are discretized using fuzzy partitioning method. In the previous work [8], we have analyzed the efficiency of fuzzifed feature values in classifying the images. And it is proved that fuzzified feature values improved the accuracy of the classifier.From these transactions association rules are mined using GNP. In testing phase new sample image is given as input by the user to the classifier and the trained system could classify the image as malign or benign accurately.

3.1 Algorithm 1 summarizes the steps of the proposed method.

Input: Training images ,a test image
Output: Keywords Benign, malign

1. Extract features of the training images
2. Discretize the feature values using fuzzy
4. Mine association rules using GNP
5. Build GNP based classifier
6. Extract features of the test image ,discretize the Feature values and feed into classifier
7. Return the keywords

3.2 Feature extraction and feature selection

Texture features are effectively estimated from the co-occurrence matrices, the gray-level spatial dependence matrix with good discriminating power. The proposed work makes use of the gray-level co-occurrence matrix(GLCM) that characterize the texture of an image by considering the distribution of intensities and relative positions of pixels in an image. GLCM matrix is created, where each element (i,j) in the resultant matrix is the number of times that the pixel pairs with intensities i and j occur in the specified spatial

Journal of Theoretical and Applied Information Technology
29th February 2012. Vol. 36 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

relationship position in the input image. Statistical features are calculated using GLCM matrix. These statistics provide information about the texture of an image. Table 1 lists the statistics. The feature values are calculated by applying the statistics formula to the GLCM matrix. Features that distinguish the classes are selected using class separation distance method [9]. Table 2 shows the selected five feature values.

Table 1 Texture Features

Feature	Equation	Meaning
Step	$\sum_i \sum_j P(i, j)$	Distribution
Variance	$\sum_i \sum_j (i-j)^2 P(i, j)$	Contrast
Entropy	$\sum_i \sum_j P(i, j)\log(P(i, j))$	Randomness
Energy	$\sum_i \sum_j P(i, j)^2$	Uniformity
Homogeneity	$\sum_i \sum_j \frac{P(i, j)}{(1+\lvert i-j \rvert)}$	Closeness of the distribution of the elements
3° Moment	$\sum_i \sum_j (i-j)^3 P(i-j)$	Distortion
Inv.Contrast	$\sum_i \sum_j \frac{P(i, j)}{((i-j)^2)}$	Local variations

3.3 FUZZY DISCRETIZATION

The crisp data set leads to uncertainty and loss of information at the boundaries of ranges. The proposed work makes use of fuzzy partition method. For each feature trapezoidal fuzzy membership function [10] is calculated using the formula

$$\alpha = 2\beta - \Upsilon$$

Where α = lower end range

β = average value of the feature f_i in the database where f is the feature and i=1to5

Υ = the largest value of the feature f_i in the database

The values of the continuous feature is discretized into three linguistic terms LOW,MEDIUM and HIGH.The probability of the crisp dataset value with respect to the membership function is known as fuzzy categorical value of the features. Highest membership value is assigned to the feature. Fig.2 shows the fuzzy membership function. More over fuzzy discretization technique [11] provides high accuracy while classifying the test images.

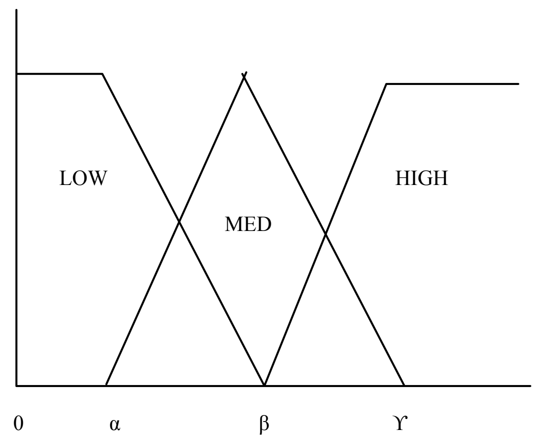

Fig..2 Trapezoidal Fuzzy Membership Functions

3.4 RULE GENERATION

3.4.1 Creation Of GNP Individuals

In the proposed work rules are extracted from the GNP structures.

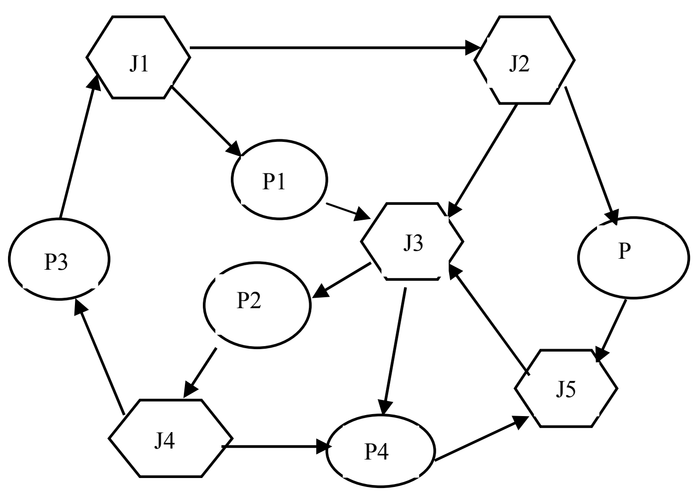

Fig.3 GNP Individual

{J1…J5} Judgement Nodes (Extracted Features)

{P1…P5} Processing Nodes

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

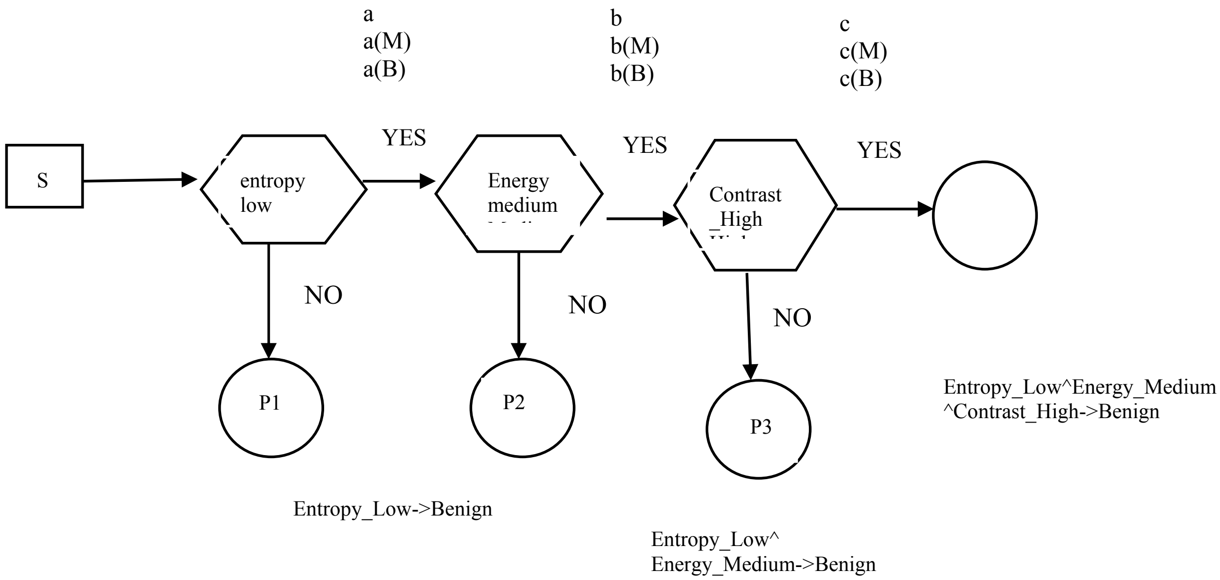

Fig.4 Rule Mining Using GNP Structure

GNP individuals [12] are created with processing nodes and judgement nodes as shown in Fig.3. The number of judgement nodes are proportional to the number of features extracted from the image .the judgement nodes and the processing nodes are inter connected ,so that rules could be framed with different set of features. As initial population four to six GNP structures are generated and rules are mined. To mine rules with various combinations of attributes, GNP individuals are remodelled using Genetic Operators such as mutation and cross over. The structures are reproduced until the individuals are significant measured using fitness.

3.4.2 Association Rule Mining Using GNP

Here association rules are mined from the database using GNP structures [13] that are created with processing nodes and judgement nodes. Each judgement node evaluates the attribute value with the membership function. The 'yes' side moves the control to other judgment node and 'no' side moves to processing node. Rules are extracted in the processing nodes. The connection from the current activated node to the next processing node forms the association rule. Rules are mined with the attribute whose membership value of the fuzzy attribute is greater than or equal to the membership value of all other lingustic terms of the fuzzy attribute. The association rules mined in such a manner is stored in the rule pool, every time rule is created. In order to reduce the number of association rules important rules are selected with the help of chi-square test [14] or support and confidence. The extracted rules are stored under two different classes. Thus the extracted rules would be used when a test image needs to be diagnosed.

3.5 CLASSIFIER

All the above phases are used to build a trained intelligent system. And now in the testing phase a new sample image is given as input to the trained system. The features of the new image are extracted using feature extraction method as in training phase. The extracted features are discretized using fuzzy .Now the features values are substituted in the rules and their summation is obtained with respect to two different classes. For each test image the average distance between new image and rules in the class is computed . The class

Journal of Theoretical and Applied Information Technology
29th February 2012. Vol. 36 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

with lowest distance is predicted as the diagnosis class.Classification of test data is determined as follows

$m_z(d)$: average distance between data d and rules in class Z

R_z : number of rules in class z

$$m_z(d) = \frac{1}{|R_z|} \Sigma D_z(d,i)$$

$$D_z(d,i) = 1 - \frac{r_{di}(Z)}{r_i(Z)})$$

$D_z(d,i)$: distance between data d and rule i in class z

$r_{di}(z)$: sum of membership values in antecedent of rule in class z calculated by data d

$r_i(z)$: number of attributes in the antecedent of rule i in Class Z

4. EXPERIMENTAL RESULTS

Image database consists of about 15 images collected from the hospitals. The images are split into two groups one group is used for training phase and other group is used for testing. As in the first phase the image texture features are extracted using GLCM from the training image database as shown in Table 2.Extracted features are fuzzified as shown in the Table 3. From this fuzzzified data set the rules are mined. In constructing GNP individuals five judgement nodes and four processing nodes are used with respect to number features extracted. GNP individuals are created. As initial population four GNP structures are created.To reproduce GNP individuals for extracting different combination of rules GNP operations such as cross over and mutation are done with the 0.8 probability. Mutation is done by changing the connections of the judgement nodes.Cross over is done by producing the off springs from the parents. About six significant individuals were reproduced. From these individuals rules are extracted. Significance of the rules is measured using support, confidence, chiSquare tests. Unique and significant rules are stored in the rule pool as shown in Table 4. The test images are fed into the classifier and the classes are determined.

Table 2 Selected Texture Features

Energy v	Entropy	Contrast	Moment	Homegenity	Image Type
815	255	1547	105	32567	M
1236	382	1259	870	23451	B
1160	307	1123	745	34563	B

Table 3 Fuzzified Features

Energy			Entropy		
Low	Medium	High	Low	Medium	High
1	0	0	0	1	0
0	0	1	1	0	0
0	.5	.5	.4	.5	.1

Table 4 Mined Rules Form GNP Structure

RULES	Support	Confidence
entropy_low=>benign	a(B)/N	a(B)/a
entropy_high ∧ energy_med=>malign	b(M)/N	b(M)/b
entropy_low ∧ energy_med ∧ contrast_high => malign	c(M)/N	c(M)/c

N : Number of tuples in data base.
B: Benign Image Type
M:Malign Image Type

5. PERFORMANCE ANALYSIS

Performance of classifier is measured by analyzing how it recognize the different classes. Confusion matrix is a tool that is used to measure the performance of the classifier. Confusion matrix represents how far the classifier label correctly and how far it mislabel the classes. It is a table with two rows and two columns in binary classification that reports the number of True Negative (TN), False Positive (FP), False Negative (FN), and True Positive (TP).In the proposed work, system recognizing the malign images correctly is considered as true positive. Confusion matrix is created using the experimental results as shown in Table 5.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 5 Confusion Matrix

	PREDICTED CLASS		
ACTUAL CLASS		Malign	Benign
	Malign(pos)	6(TP)	0(FN)
	Benign(neg)	3(FP)	6(TN)

Sensitivity measures the proportion of malign images correctly identified as malign. Specificity measures the proportion of negatives which are correctly identified that is correct identification of benign images. Accuracy measures the probability or proportion of images correctly classified. Accuracy, Sensitivity, Specificity are calculated using the formula stated in the Table 6, taking values from the confusion matrix. N(TN) is the number of True negatives and N(TP) is the number of Tue Positives. Positives (pos) are the total number of malign images. Negatives (neg) are the total number of benign images in the database. Performance of the classifier is analyzed in terms of accuracy, sensitivity and specificity The accuracy of the proposed system is 83%.

Table 6 Performance Measures For Classifier

Performance Measure	**Formula**
Specificity	$\frac{N(TN)}{N(TP)+N(TN)}$
Sensitivity	$\frac{N(TP)}{N(TP)+N(TN)}$
Accuracy	$sensitivity\frac{pos}{pos+neg}+specficity\frac{pos}{pos+ne}$

6. CONCLUSION

The existing cancer diagnosis systems have been found to be inadequate in one way or the other in detection. Different methods that are used for this purpose have been studied and a novel framework has been proposed in this project. The proposed method use Soft Computing techniques. A hybrid method that involves fuzzy and GNP is used for discretizing the feature values extracted from images and for mining association rules. Features of the images are extracted by texture analysis. Proposed cancer diagnosing system classifies the new unknown images more accurately. The accuracy of the proposed system is 83% .When the data set size is increased surely the accuracy rate could be increased.

7. REFERENCES

[1] J.Han,M.Kamber:"Data Mining : Concepts and Techniques", Morgan Kaufmann Publishers, San Francisco ,2000.

[2] Robert .M, Haralick .K "Textural Features For Image Classification"IEEE Trans. On Systems, Man and Cybernetics, VO L.SMC-3, No.6, November 1973.

[3] Qurat-Ul-Ain, Ghazanfar Latif, Sidra Batool Kazmi, M. Arfan Jaffar, Anwar M. Mirza "Classification And Segmentation Of Brain Tumor Using Texture Analysis", Recent Advances In Artificial Intelligence, Knowledge Engineering And Data Bases

[4] Marcela X.Riberio "An Association Rule-Based Method to Support Medical Image Diagnosis With efficiency",IEEE Transactions On Multimedia,feb 2008

[5] Hüllermeier, E., Yi, Y. 2007 "In Defense of Fuzzy Association Analysis" IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, 37

[6] P. Rajendran, M. Madheswaran 2010 "Novel Fuzzy Association Rule Image Mining Algorithm for Medical Decision Support System" International Journal of Computer Applications (0975 - 8887) Volume 1 – No. 20

[7] Karla Taboada, Shingo Mabu, Eloy Gonzales, Kaoru Shimada, Kotaro Hirasawa 2009 "Fuzzy Classification Rule Mining Based on Genetic Network Programming Algorithm" IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TX, USA

[8] V.Sivakrithika ,B.Shanti ," Comparative Study on Cancer Image Diagnosis using Soft Computing Techniques", International Journal Of Computer Application, April 11.

[9] M.Vasantha,R.Dhamodharan, "Medical Image Feature, Extraction, Selection And

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Classification", International Journal Of Engineering Science And Technology Vol. 2(6), 2010, 2071-2076.

[10] Kaoru Shimada, Kotaro Hirasawa, and Jinglu Hu 2006 "Genetic Network Programming with Acquisition Mechanisms of Association Rules" Journal of Advanced Computational Intelligence and Intelligent Informatics Vol.10 No.1

[11]Ashish Mangalampalli, Vikram Pudi 2009 "Fuzzy Association Rule Mining Algorithm for Fast and Efficient Performance on Very Large Datasets" IEEE International Conference on Fuzzy Systems.

[12] K. Hirasawa, M. Okubo, H. Katagiri, J. Hu and J. Murata 2001 "Comparison between Genetic Network Programming and Genetic Programming" Proceedings of the Congress of Evolutionary Computation

[13] Karla Taboada, Shingo Mabu, Eloy Gonzales, Kaoru Shimada, Kotaro Hirasawa 2009 "Fuzzy Classification Rule Mining Based on Genetic Network Programming Algorithm" IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TX, USA

[14] K. Hirasawa , Kaoru Shimada, Jinglu Hu 2006 "Class Association Rule Mining with Chi-Squared Test Using Genetic Network Programming " IEEE International Conference on Systems, Man, and Cybernetics.

Journal of Theoretical and Applied Information Technology
29th February 2012. Vol. 36 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AUTHOR PROFILES:

V. **Sivakrithika** received the M.Tech degree in computer science and engineering from SASTRA University, 2011. Currently, she is an Assistant Professor at PRIST University. Her interests are in data mining and cloud computing.

S. Sengole Merlin received the M.E degree in computer science and engineering from the Anna University, 2011. Currently, she is a Assistant Professor at Loyala Engineering college. Her research interests include image mining and data mining

K. Sugirtha pursuing the M.Tech degree in computer science and engineering in Hindustan University .Her research interests include image mining and soft computing techniques.

.