Published in 2016 International Conference on Systems in Medicine and Biology (ICSMB), IIT Kgp, 2016

Knowledge Abstraction from Textural Features of Brain MRI Images for Diagnosing Brain Tumor using Statistical Techniques and Associative Classification

Abhishek Singh Sambyal
Computer Sc. and Engineering
Bangalore Institute of Technology
Bangalore, India
abhishek.sambyal@gmail.com

Dr. Asha T.
Professor, Computer Sc. and Engineering
Bangalore Institute of Technology
Bangalore, India
asha.masthi@gmail.com

***Abstract* – This paper presents a methodology for finding the association rules using associative classification which can be used to abstract knowledge from brain MRI images. Reducing the size of images using different thresholds help to reduce the complexity of the proposed system without affecting the correctness of these images. Textural features are taken into consideration because when there is a wide variation of features of discrete gray tone, the texture dominates more. Gray-Tone Spatial-Dependence matrices are calculated from images in which textural information is contained. The system uses a supervised learning approach for selecting the important features from different textural features. Using associative classification, the rules are generated from selected textural features which abstract the knowledge from the images.**

Keywords - Interpolation, image thresholding, novel supervised learning, associative classification, knowledge abstraction

I. Introduction

Brain tumor is one of the major leading causes of cancer-related deaths [1]. Brain tumors are of mainly two types: malignant or cancerous tumors and benign tumors. Diagnosing vast number of MRI images by radiologists is a tiresome task and susceptible to errors. An expert system can really be helpful in assisting the radiologists

Scaling of an image is used very frequently in image processing. In this paper, we have used Nearest Neighbor Interpolation technique for scaling an image which interpolates each pixel value of its nearest sampling point.

Image thresholding is a method of image segmentation. In multi modal systems, multi-level thresholding technique [2] finds multiple thresholds from an image and a segment on the basis of multiple thresholds.

Co-occurrence matrix [3] is defined over an image to be the distribution of co-occurring values at a given offset. It is calculated for each feature like color, texture, edge [4]. Feature extraction is a technique of dimensionality reduction. The large image data is transformed into a reduced representation of sets of features which is called feature vector. The features selected are discretized as per the cut points, which is a statistical technique for finding the relevant features. Selected features are taken as data set of items and frequent patterns are searched. Frequent pattern shows the relationship between attribute-value pairs [5] that occur frequently in a given data set.

Associative classification first mines the data for frequent itemsets that are generated by association rule mining [9] by satisfying the minimum *support* and rules are generated by satisfying minimum *confidence* specified by the user. Each rule is organized by a rule based classifier. Associative classification rule contains antecedent and consequent where antecedent can be the conjunction of only items and consequent can have only class. Every value in selected features is discretized, which is an item and each image is associated with a class. So the rule structure will be like:

$$\text{item1} \wedge \text{item2} \wedge \ldots \text{itemp} \rightarrow \text{class}$$

In this paper, image processing and statistical techniques are used to organize data, which is fed to the Associative classification algorithm. This methodology helps in detecting the tumors in brain MRI images.

Rest of the paper is organized as follows: Section II contains the literature survey. In Section III, we have used different methods and techniques and a framework is proposed for diagnosing brain tumors. In Section IV, we have discussed our results. Section V includes algorithms designed for this study and Section VI is the conclusion of this paper.

II. Literature Survey

K. Beyer, Jonathan Goldstein et al. [6] have shown the effect of dimensionality on the nearest neighbor problem and the tradeoff between the discrimination power and the feature vector size. In this paper, we have considered the dependency of the features and selected the relevant ones without losing information.

P. G. Foschi et al. [4] have shown the relevance of different features of an image and how to identify the unique feature of a particular domain. In this paper, we have taken textural feature, where variation of each pixel with respect to its neighboring pixel defines texture.

Ping-Sung Liao et al. [2] have proposed a thresholding technique for finding thresholds in multi modal systems. In this paper, we have used Multi-Thresholding for finding multiple thresholds of an image.

R.M. Haralick et al. [7] have proposed the Gray-Tone Spatial-Dependency Matrices from which the Textural features are extracted. This paper uses the Gray-Tone Spatial-Dependency Matrices to extract relevant textural features from medical images.

Pan Haiwei, Jianzhong Li et al. [8] have shown how preprocessed data is stored by transactions containing the image ID and different features of each image. In our study, we have taken the transactions as an input for finding frequent patterns from which the rules are generated.

R. Agrawal et al. [9] have provided association rule algorithms on the basis of support and confidence measures. This paper uses these techniques to generate association rules from large transactional database.

Maria-Luiza A. et al. [10] provide classification models for medical images based on association rule mining. In this paper, association rules are generated by frequent patterns combined with class labels which are considered to have more accuracy and scalability.

III. Proposed Framework

Data flow diagram (DFD) depicts the use of image processing and statistical techniques along with various algos at different processes. It is shown in the Figure 1.

A. Reducing the time complexity of the system

The complexity of the higher dimension medical images are scaled down using Nearest Neighbor Interpolation [6]. Reduction in the size of images leads to faster system, but finding the co-occurrence matrix is still a tedious task. Further, the images are quantized among 16 different thresholds which are calculated by Multi Threshold method [2] as shown in Table 1. The dimension of the co-occurrence matrix depends on the maximum value of the pixel in an image which is 256x256 in this case. But the co-occurrence matrix only sees the relationship between the neighboring pixels and the function of distance between them. So, we substitute each threshold value ranging from 12 to 255 with the corresponding value ranging from 1 to 16. By this technique, the dimension of the co-occurrence matrix reduces to 16x16. By substituting the pixel values of the image, there is no effect on the correctness of the calculation of the co-occurrence matrix but it greatly reduces the time complexity of the system.

Table 1: Threshold Table

12	33	55	67	81	.	.	194	215	230	243	255

B. Gray-Tone Spatial Dependency Matrix and Feature extraction

Gray-Tone Spatial-Dependency matrix [7] is calculated for various angular relationships like 0°, 45°, 90° and 135° at distance 1, 2, 3, 4, 5 between neighboring resolution pairs on the image. It is shown in Figure 2.

In Figure 2, the pixel of interest is taken as the center point from where angular relationship of 0° is taken for 1, 2, 3, 4, 5

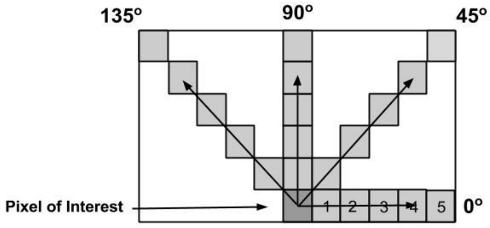

Figure 2: Angular relationships with distances

distances. This is repeated until all the angular relationships are over. For each image, Gray-Tone Spatial-Dependency matrix is created and is normalized for further computation. Total of 20 co-occurrence matrices are created for each image. From this, many textural features are extracted which is used to find the tumor. List of textural features is shown in Table 2.

Table 2: Texture Features

Feature	Equation
Step	$\Sigma_i\Sigma_j$ P(i,j)
Contrast	$\Sigma_i\Sigma_j$ $(i-j)^2$ P(i,j)
Entropy	$\Sigma_i\Sigma_j$ P(i,j) log (P(i,j))
Energy	$\Sigma_i\Sigma_j$ $P^2(i,j)$
Correlation	$(\Sigma_i\Sigma_j (ij)P(i,j) - \mu_x \mu_y)/\sigma_x \sigma_y$
Homogeneity	$\Sigma_i\Sigma_j$ P(i,j)/(1+|i-j|)
Angular Third Moment	$\Sigma_i\Sigma_j$ $\{P(i,j)\}^3$

The variance feature gives the level of contrast of an image, entropy gives the suavity of an image, energy gives how uniform the image is, homogeneity feature tells the homogeneity of the pixel distribution and angular third moment is the level of distortion in an image. All the features in Table 2 are computed for each co-occurrence matrix. There are seven features and a total of 140 i.e. 7 for each co-occurrence matrix which are stored in the feature vector.

Feature vector [8] is created which constitutes ImageID, features and its class (type of the brain tumor) where each transaction is one image with its features and class. The structure of the feature vector is shown in Table 3.

Table 3: Feature vector with Classes

ImageID	f1	f2	f3	…	…	f139	f140	Class
IMG1	1	.77	3.0	.	.	7.89	8.3	Benign
IMG2	.98	1.6	2.6	.	.	8.93	7.9	Malignant
IMG3	..	..	..			..	..	Malignant
…								…

C. Novel Supervised learning algorithm for feature selection and discretization using cut points

Each of 140 features has its own importance in feature vector and we have to select the most relevant ones. The selection of relevant features is done using cut-points, where

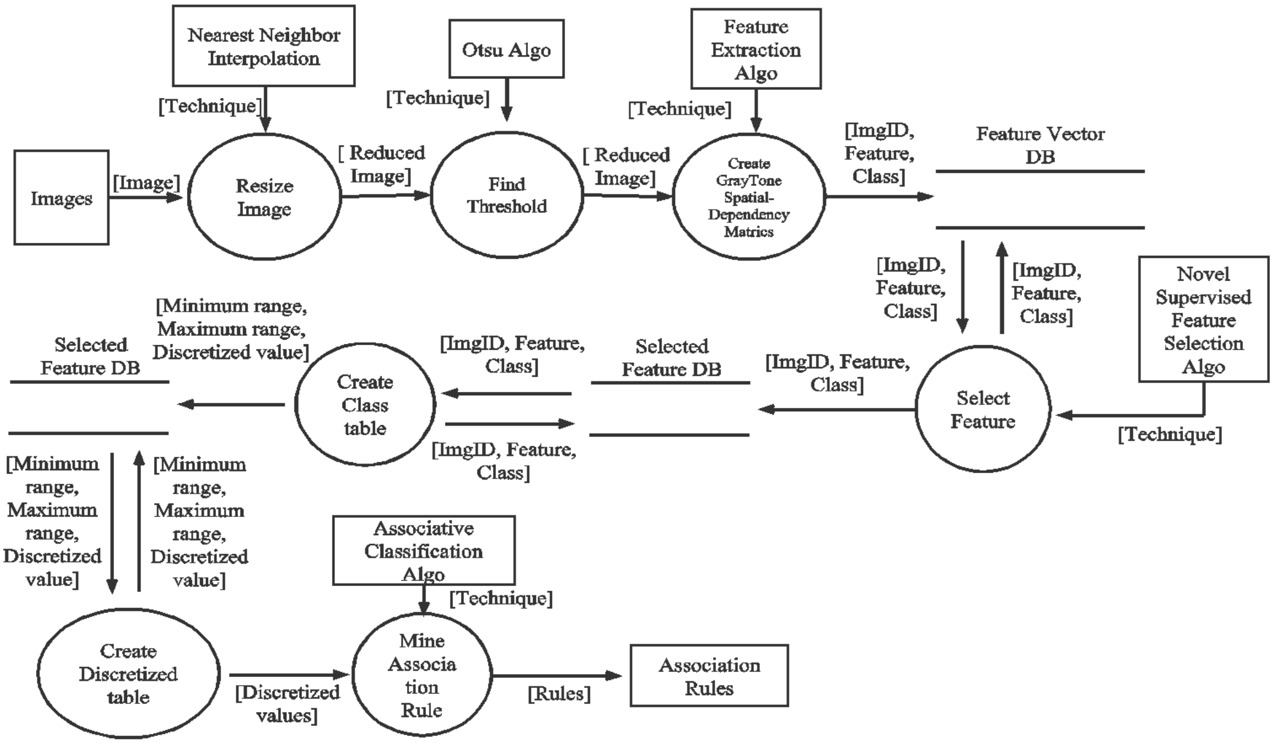

Figure 1: Data Flow Diagram

each feature is divided into intervals called cut-points which is determined by the class of each image

Let I be the set of images transactions, f be the feature of feature vector F where f_i is the value of feature in transaction i. c_i is the class of transaction i where $i \in I$. For each feature f, if the class label of instance is different from class label of previous instance, the cut point is introduced. With this method, lot of cut points are generated especially with noisy data. Each interval represents an item in the process of mining association rules. Use of many items generate rules with low confidence which are unnecessary and increases the time complexity of the system. So to overcome this problem, constraints are applied to reduce the number of cut points.

1) Constraint 1: The number of occurrence in each interval should be greater than or equal to minimum threshold.

2) Constraint 2: The cut point in two consecutive intervals can be removed if majority class of current interval is equal to the majority class of the next interval and ratio of magnitude of majority class to the interval class is greater or equal to join_threshold and ratio of magnitude of next majority class to the next interval class is greater or equal to join_threshold. It means two consecutive intervals can be combined if both have same majority class and the occupancy of majority class is equal or greater than the join_threshold.

Features with minimum number of cut points are selected and then discretized. For discretization, the range of every interval of feature in *class_table* is stored. The attributes of *class_table* are minimum range, maximum range and discretized value. Minimum and maximum range is the minimum and maximum value which is calculated for each image and its feature for every interval.

All the values of selected features are substituted by their corresponding discretized values as shown in Table 4. It is considered as the pre-processing of data, which is given to the associative classifier to find the useful patterns which are necessary for extracting the knowledge from the database.

Table 4: Discretized table

f1	f2	f3	f4					f22	f23	f24	Class
2	4	8	14	. .	. .	. .	. .	114	126	130	135
2	6	9	12	. .	. .	. .	. .	120	124	134	136
1	6	9	13	. .	. .	. .	. .	120	124	134	136

D. Associative classification for rule generation using Classification based on Predictive Association Rules (CPAR)

Association classification [10] consists of three steps, first is to find the frequent sets, second is to analyze the frequent sets and generate rules satisfying support and confidence, and third is the ordering of the rule.

We have used CPAR algorithm [11] for finding rules and rule ordering based on Laplace accuracy. CPAR chooses the number of attributes which have similar best gain. For calculating the bestGain, we have chosen *Information Gain* [5] as an attribute selection measure. The Local Gain Threshold

(LGT) is calculated by multiplying attribute bestGain with GAIN_SIMILARITY_RATIO.

$$LGT = bestGain * GAIN\_SIMILARITY\_RATIO \quad \text{..... (1)}$$

Attributes with gain better than eq. (1) are selected. All the attributes whose gain is greater than LGT are added to the rule list. This approach minimizes the expected number of tests needed to classify a given tuple. Suppose, if it generates two or more rules with identical antecedents and consequents, the rule with lower accuracy is removed from the rule list.

IV. RESULTS AND DISCUSSIONS

Laplace accuracy is used to find the accuracy of the generated rules. Every generated rule is tested on multiple brain MRI images which are being used for training the system. The accuracy of each rule is given by eq. (2).

$$Accuracy_{RULE} = \frac{Success}{Total\ number\ of\ images\ tested} \quad \text{….. (2)}$$

where, Success is the number of times the rule predicted the correct behavior.

As per the formula for Rule 1, we get 94% correctly classified instances and 6% incorrectly classified instances. The diagnosis results of the data used in our study are very encouraging as discussed below:

Rule	% Accuracy
Rule 1: {93} → {136}	94
Rule 2: {14 48} → {136}	90
Rule 3: {14 88} → {136}	90
Rule 4: {52} → {135}	80
Rule 5: {9} → {135}	75
Rule 6: {69} → {135}	67
Rule 7: {40} → {135}	67
Rule 8: {13} → {135}	50

Rule ordering is done based on Laplace accuracy and eight rules are obtained from our brain MRI data. First column represents the rule and the second column represents the accuracy of the rule. The antecedent of rule is a discretized value of the feature with some range as discussed in Section III-C and consequent is the class of that feature.

Considering Rule 2, {14 48} are the discretized values of the 5^{th} and 30^{th} feature of an image i.e. Correlation and Contrast respectively. Discretized value {14} is the correlation feature of distance 1 at angle 0° and discretized value {48} is the contrast feature of distance 2 at angle 0°. The range of {14} is 0.8593 to 0.9055 and the range of {48} is 1.0123 to 2.6875. The rule says, if feature value of {58} and {113} exists then {136} exists which means malignant behavior. In our study, values 135 and 136 represent benign and malignant nature of the tumor respectively.

V. ALGORITHMS

We have designed and implemented algorithms to create feature vector and select the most relevant features and their discretization. These are shown in Algorithm 1 and 2 respectively.

Algorithm 1: Build feature vector

```
Input: Images, image classes, distance, angular relation
Output: Feature vector
for each image do
    Resize image using nearest neighbor interpolation
    Finding threshold values for an image
    Quantize image using threshold values
    for each distance do
        for each angular relation do
            Find Gray-Tone Spatial Dependency matrix
            Find features from Gray-Tone Spatial Dependency matrix
            Store feature values in feature vector with classes
        end for
    end for
end for
Return V
```

Algorithm 2: Feature Selection and Discretization

```
Input: Feature vector F, image classes, minimum threshold, join_threshold, feature reduction
Output: Selected feature vector V
for each f in feature vector F
    sort f values
    Create an instance Ii of the form ci, fi for each transaction i.
    Create a cut point vector which store the information about the intervals
end for
for each cut point in cut point vector
    Determine cut points
    Remove cut points using Constraint 1
    Remove cut points using Constraint 2
    Store the remaining cut points in Final Cut point vector
end for
Sort features according to the maximum number of cut points
Select features with minimum number of cut points
Create a class table with minimum and maximum range of
    each interval calculated above with its corresponding
    discretized value
Discretized the selected features in V
Return V
```

VI. CONCLUSION

This paper describes the usage of image processing techniques, statistical concepts and data mining methods for detecting the tumor in brain MRI images. The proposed framework identifies malignant and benign brain tumors so that the patient is diagnosed accordingly. Thus giving an expert like opinion as the radiologist.

The source code generated for this study is freely available under GNU General Public License (GPLv3).

References

[1] Brain Tumor Statistics, http://www.abta.org/about-us/news/brain-tumor-statistics/.

[2] Ping-Sung Liao, Tse-Sheng Chen and Pau-Choo Chung, "A Fast Algorithm for Multilevel Thresholding". Journal of Information Science and Engg. 17 (5): 713–727(2001).

[3] Co-occurrence matrix http://en.wikipedia.org/wiki/Co-occurrence_matrix.

[4] P. G. Foschi, D. Kolippakkam, H. Liu, and A. Mandvikar, "Feature extraction for image mining," in *Proc. 8th Int. Workshop Multimedia Inf. Syst.*, Tempe, AZ, 2002, pp. 103–109.

[5] Jiawei Han, Micheline Kamber, Jian Pei, "DATA MINING: Concepts and Techniques", Third Edition, Morgan Kaufmann Publications, ISBN 978-93-80931-91-3, p.no. 336-340, 415-418.

[6] K. Beyer, J. Godstein,R. Ramakrishnan, and U. Shaft, "When is nearest neighbor meaningful?" in *Proc. Int. Conf. Database Theor. (ICDT)*, 1999, pp. 217–235.

[7] R. M. Haralick, K. Shanmugam, and I. Distein, "Textural features for image classification", *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, pp. 610–621, 1973.

[8] P. Haiwei, J. Li, and Z. Wei, "Medical image clustering for intelligent decision support" in *Proc. 2005 27th Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, Shanghai, China, 2005, pp. 3308–3311.

[9] R. Agrawal, T. Imielinski, and A. N. Swami, "Mining association rules between sets of items in large databases", in *Proc. 1993 ACMSIGMOD Int. Conf. Manage. Data - SIGMOD 93 SIGMOD 93*, Washington, DC, 1993, pp. 207–216.

[10] M.-L. Antonie, O. R. Zaiane, and A. Coman, "Associative classifiers for medical images" in *LNAI 2797, MMCD*. New York: Springer-Verlag, 2003, pp. 68–83.

[11] The University of Liverpool https://cgi.csc.liv.ac.uk/~frans/KDD/Software/FOIL_PRM_CPAR/cpar.html.