IMAGE MINING TECHNIQUES FOR CLASSIFICATION AND SEGMENTATION OF BRAIN MRI DATA.

[1] **L.Jaba Sheela,** [2] **Dr.V.Shanthi**

[1] *Department of MCA, Panimalar Engineering College, Chennai-602103, India.*
E-mail: sujitha14@hotmail.com

[2] *Department of MCA, St.Joseph's Engineering College, Chennai-600 096, India.*

E-mail: drvshanthi@yahoo.com

ABSTRACT

Image segmentation plays a crucial role in many medical imaging applications by automating or facilitating the delineation of anatomical structures and other regions of interest. Automated detection of tumors in different medical images is motivated by the necessity of high accuracy when we dealing with a human life. Also, the computer assistance is demanded in medical institutions due to the fact that it could improve the results of humans in such a domain where the false negative cases must be at a very low rate. It has been proven that double reading of medical images could lead to better tumor detection. But the cost implied in double reading is very high, that's why good software to assist humans in medical institutions is of great interest nowadays. In this paper we propose a system which uses image mining techniques to classify the images either as normal or abnormal and then segment the tissues of the abnormal Brain MRI to identify brain related diseases.

Keywords: *Image Mining, MRI , Segmentation ,Association rule mining, Frequent items*

1. INTRODUCTION

Radiology departments are at the center of a massive change in technology. The ubiquitous radiographic film that has been the basis of image management for almost 100 years is being displaced by new digital imaging modalities such as 1. computed tomography (CT); 2. magnetic resonance (MR); 3. nuclear medicine (NM); 4. ultrasound (US); 5. digital radiography (DF); 6. computed radiography (CR) using storage phosphor imaging plates or film digitizers, 7. digital angiography (DA); 8. MR spectroscopy (MRS);9. electron emission radiography (EMR). These digital modalities are continuously refined and new digital applications are being developed. Advances in image acquisition and storage technology have led to tremendous growth in very large and detailed image databases. A vast amount of image data is generated in our daily life. The use of image technology in medicine has changed enormously with advances in technology developments. Of these imaging technologies magnetic resonance imaging (MRI), x-ray computer tomography (CT), positron emission tomography (PET), and ultrasound has given physicians a non-invasive means to visualize internal anatomical structures and diagnose a wide variety of diseases. MRI is a particularly powerful and versatile modality. Compared to other such techniques, MRI has superior soft tissue differentiation, high spatial resolution and contrast, and does not use ionizing radiation which may be harmful to patients. Such characteristics have shown MRI to be a valuable tool in the clinical and surgical environment

2. BACKGROUND

Medical images are a fundamental part of medical diagnosis and treatment. These images are different from typical photographic images primarily because they reveal internal anatomy as opposed to an image of surfaces. They include both projection x-ray images and cross-sectional images, such as those acquired by means of computed tomography (CT) or magnetic resonance imaging (MRI), or one of the other tomographic modalities (SPECT, PET, or ultrasound, for example). Medical image processing is a branch of image processing that deals with such images. It is driven both by the peculiar nature of the images and by the medical applications that make them useful. The applications of Medical Image Processing primarily focus on the areas of image segmentation, image registration, and image-guided surgery. Image segmentation is the process that permits the automatic extraction of structures of interest from the images. Image registration is the determination of a point-to-point mapping that aligns the anatomy in one image with that in another. The images may have been acquired from the same patient or from different patients and may involve the same or different imaging modalities. Image-guided surgery is any surgical procedure in which a surgeon's approach is guided in part by the tracking of instruments relative to images of the patient.

2.1 A MRI system

A MRI system consists of the following components: 1) a large magnet to generate the magnetic field, 2) shim coils to make the magnetic field as homogeneous as possible, 3) a radio frequency (RF) coil to transmit a radio signal into the body part being imaged, 4) a receiver coil to detect the returning radio signals, 5) gradient coils to provide spatial localization of the signals, and 6) a computer to reconstruct the radio signals into the final image.

2.2 Image Formation in MRI

MRI exploits the inherent magnetic moment of certain atomic nuclei. The nucleus of the hydrogen atom (proton) is used in biologic tissue imaging due to its abundance in the human body and its large magnetic moment. When the subject is positioned in the core of the imaging magnet, protons in the tissues experience a strong static magnetic field and presses at a characteristic frequency that is a function solely of the magnetic field strength, and does not depend, for instance, on the tissue to which the proton belongs. An excitation magnetic field is applied at this characteristic frequency to alter the orientation of precession of the protons. The protons relax to their steady state after the excitation field is stopped. The reason MRI is useful is because protons in different tissues relax to their steady state at different rates. MRI essentially measures the components of the magnitude vector of the precession orientation at different times and thus differentiates tissues. These measures are encoded in 3D using methods for slice selection, frequency encoding and phase encoding. Slice selection is performed by exciting thin cross-sections of tissue one at a time. Frequency encoding is achieved by varying the returned frequency of the measured signal, and phase encoding is done by spatially varying the returned phase of the measured signal.

The MR images normally can be obtained in three planes: the sagittal plane, the horizontal plane, and coronal plane (shown in Fig. 1.)

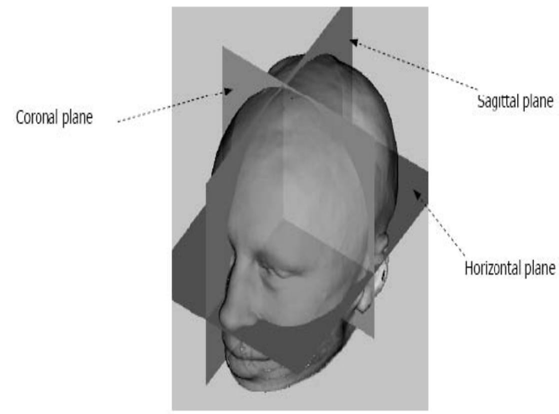

Fig 1. The spatial demonstration of three planes for MRI

3. IMAGE MINING

Image mining is more than just an extension of data mining to image domain but an interdisciplinary endeavor. Very few people have systematically investigated this field. Images are rich in information content. Since the discovery of X-rays in 1895 , medical images have provided significant assistance in medical diagnosis. Medical images contain a wealth of hidden information that can be exploited by physicians in making reasoned decisions about a patient. However, extracting this relevant hidden information is a critical initial step to their use. This motivates the use of data mining techniques for efficient knowledge extraction.

Mining medical images involves many processes. The process to be used depends on the type and complexity of image to be mined. For instance, it is simpler to mine 2-dimentional x-rays as compared to 3- dimensional *CT* scans of the brain. However, some processes are fundamental to the task of medical image mining, regardless of the complexity of image. We briefly discuss these processes below.

Data Preprocessing: 'This stage consists of several processes. These processes include data normalization, data preparation, data transformation, data cleaning, and data formatting. Normalization techniques are required to integrate the different image formats to a common format. Data preparation alters images to present them in a format suitable for transformation techniques, Next, the image is transformed in order to obtain a compressed (lossless) representation of it, e.g., using wavelet transforms. Segmentation is done to identify regions of interest (ROI) for the mining task, usually achieved using classifier systems. The segmentation step finds corresponding regions within an image, since item sets are extremely large. Images are usually represented in pixels. A pixel is a dot of light on an evenly spaced rectangular grid, which has attributes such as color and texture. Collections of neighboring pixels with similar attributes make up a region. *Chu et al.* [11] proposed a temporal evolutionary data model, which enables the temporal and evolutionary descriptions of images, so that their image features and content can be modeled.

Feature Extraction: Images have a large number of features. It is important to identify and extract interesting features for a particular task in order to reduce the complexity of processing. These are attributes or portion of the image being analyzed that is most likely to give interesting rules for that problem. Not all the attributes of an image are useful for knowledge extraction. This stage increases the overall efficiency of the system. Image processing algorithms are used, which automatically extract image attributes such as local color, global color, texture, and structure. Texture is the most useful description property of an image and it specifies attributes, such as resolution, which can be used in image mining. An image can be adequately represented using the attributes of its features. The extraction of the features from an image can be done using a variety of image processing techniques. We localize the extraction process to very small regions in order to ensure that we capture all areas.

Rule Generation: Since this is a highly knowledge based domain, associated domain knowledge can be used to improve the data-mining task. This data integration is an important concept because medical images are not self contained, and are often used in conjunction with other patient data in the process of diagnosis. We expect association rules of two forms: (i) Image contents unrelated to spatial relationships, e.g., if an image has a texture X, it is likely to contain protrusion Y and (ii) Image contents related to spatial relationships, e.g., if X is between Y and Z it is likely there is a T beneath. A low minimum support and high minimum confidence is desirable, since few image data sets have high support .

Interpretation of patterns and knowledge extraction: Not all the interesting rules are medically important. To make **our** technique relevant, the rule presented must be significant and meaningful.

4. LITERATURE REVIEW

Medical Images are generally mined using techniques such as clustering, classifier trees, or regression [5]. Lee [6] developed an intelligent decision making system for breast cancer diagnosis based on segmentation and classification using neural network. Inductive logic programming systems have also been used in image classification for glaucoma diagnosis [7]. Other machine learning methods, such as the Bayesian classifier, have also used in the diagnosis and prognosis of first cerebral paroxysm [8]. Furthermore, several classifiers have been used in the prognosis of the femoral neck fracture [5], in computer aided diagnosis in chest radiography [9, 10] and also in the automated diagnosis and prognosis of breast cancer based on histological images [2]. Xie et al used associative rule mining to detect SARS from radiographs [15]. D.Kontos et al proposed a dynamic recursive partitioning approach for discovering discriminative patterns in fMRI [18] .A neural network based classification of cognitively normal, demented Alzheimer disease and vascular dementia from SPECT image data was performed by Rui et al [19]. Statistical technique involved training of a discriminant analysis classifier (DAC). Kippenhan et al used this type of classifier in the evaluation of neural network classifier applied to perfusion profiles extracted from PET scans.The analysis was carried out using SAS statistical package.The use of pattern recognition methods to segment MR image data sets has been widely described in literature [6] [7] [8] [9]. Nearest neighbor, maximum likelihood and Parzen window classifiers [10] are among the supervised classification algorithms found in literature.

5. RESEARCH OBJECTIVES

The objective of this paper is to propose a new image mining process model to classify normal or abnormal images and develop efficient algorithms to achieve high quality segmentation of WM, GM, and CSF from MR brain images/volumes which in turn can be used to diagnose brain related diseases.

This study is significant for the following reasons:

- The proposed algorithm can assist

physicians improve the accuracy and speed of decision-making based on medical images in medicine, thereby leading to dramatic improvements in health care.

• Experiences gained in applying the proposed algorithm to medical image data can provide the

catalyst to promote [efficient] knowledge extraction from image data sets in other domains,

e.g., geology, astronomy and biology.

6. SYSTEM OVERVIEW

In this research work it is proposed to automate the segmentation of brain magnetic resonance images by using some prior knowledge like pixel intensity and some anatomical features. Currently there are no methods widely accepted therefore automatic and reliable methods for tumor detection are of great need and interest. Magnetic Resonance Imaging proved to provide high quality medical images and became widely used especially for brain.

The advantages of Magnetic Resonance Imaging are that the spatial resolution is high and provides detailed images. Magnetic Resonance Images are used in detecting brain tumors or in tracking them. The tracking of the tumors is important especially when a patient is under medication in order to observe the changes that appear.

Segmentation is an important step in many applications, being also important in those that deal with medical images. When a magnetic resonance image of the brain is segmented to detect a tumor and also its size, it is very important that the segmentation to give results as accurate as possible, because the life of a person could depend on it. The figure below describes the proposed system that is used in order to segment and detect the tumors in MRI images.

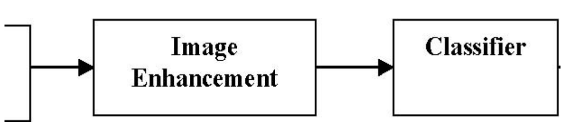

Fig 2. System overview

The image acquisition module represents the first step in an image processing system. In this work, it is initially proposed to obtain the images either by simulation or collect them from online atlases. In this step is very likely that noise could be introduced, thus the quality of the image could be decreased. That's why the next step is dealing with the enhancement of the image in order to improve the quality or to perform a denoising (Data Cleaning).

The denoising methods that can be used are:

- Wavelet-based method or
- Histogram Equalization method

Classifier
Segmentation
Image Enhancement
Image Acquisition

After the denoising has been performed the image is entering the classifier module that will distinguish, according to features of interest of the images in the medical image database. The feature extraction phase is needed in order to create a transactional database to be mined.

The features of interest are:

• Large Object Size (where an object is a pixel in the segment of interest)
• Low Noise Level
• High Contrast
• Coarse Texture
• Abnormal Average Grey Level

where Abnormal Average Grey Level and Coarse Texture are present in all known abnormal cases. In this step the classifier intended to be used is an associative classifier to distinguish between a normal and abnormal slice of brain. . Our goal is to obtain associations between the various values of important features. It is interesting to see if the presence of a particular value in one feature tends to imply another value.

The idea of distinguishing between normal and abnormal images is useful, in order to prune the normal images and to segment only those that have the possibility of having a tumor.

The last step is representing the segmentation of the image in order to detect the tumor. This step uses some prior knowledge from the classification step in order to use an adaptive histogram threshold for a better tumor segmentation. After a MRI image has passed through the system, the result is the segmentation of the existing tumor.

7. PROPOSED ALGORITHM FOR MINING

Classification is an important task in many applications. A classifier is a system that assigns, or predicts, one or more class labels for a given object. One way to create a classifier is to use a learning algorithm to construct a classifier from a training set of objects whose classification(s) is known.

To classify a given object, an associative classifier proceeds in three steps. First, it determines which of its rules apply to the object. Then it selects a subset of the applicable rules (possibly all of them) based on some measure of their strength or precedence. Finally, if it chooses more than one rule, it combines the class predictions of all the selected rules to produce a final classification.

Association rule mining is a popular data mining technique, which aims at discovering strong interesting patterns (associations) between items in vast data sets. For instance, X Y, read as "X implies Y **'**, is an association rule, which is interpreted as "if X occurs it is most likely that Y also will occur". The task of mining association rules is simply searching for items which occur frequently (large items), defined by a user-defined minimum frequency (minimum support), and then finding patterns in their occurrences. Various algorithms exist to efficiently search for and count large item sets. $\Rightarrow$

The Frequent-Pattern (FP) Growth algorithm, however, is a standard ARM algorithm, which is efficient for mining large datasets with frequent patterns (i.e., a large number of "large" items). Its efficiency lies in the compact and complete way it represents the entire set of transactions and patterns in a tree-like form, which eliminates the need to generate candidate items sets. However, for extremely huge datasets, the frequent-tree structure can also exceed the main memory, causing I/O overheads.

We propose the ***Partitioned Frequent-Pattern (FP) Growth algorithm,*** which is based on partitioning the transaction database, and processing each of the partitions in parallel. This approach enables the efficient processing of large datasets with long patterns. This approach, in addition to having a compact data representation scheme, is amenable to parallel implementation and has an efficient input / output structure

8. CONCLUSION

Medical images contain a lot of information. It is desirable to mine them in order to use such knowledge for diagnosis support. The partitioned FPGrowth algorithm we proposed can mine medical image data sets efficiently. There remains a need to apply the proposed algorithm to actual medical image data sets. This constitutes part of our research work, It will also be interesting to investigate the applicability of our proposed algorithm in other domains. Furthermore, a performance analysis of this algorithm is necessary to validate our conclusions Optimizing the construction and mining processes of the FP-tree may further enhance the efficiency of the image mining process.

REFERENCES

[1] R. De La Paz, E. Herskovits, V. Gesu, et al. "Cluster analysis of medical magnetic resonance images MRI data: Diagnostic application and evaluation," *SPIE Extracting Meaning from Complex Data: Processing, Display, Interaction,* Vol. 1259, 1990, pp. 176 –181.

[2] C. Ordonez, C. Santana and L. Brad, "Discovering Interesting Association Rules in MedicaI Data", *ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery,* 2000, pp. 78-85**.**

[3] V. Megalooikonomou, J. For, L. Shen, F Makedon. Data Mining in Brain Imaging. *Statistical Methods in Medical Research* 2000, pp. 359-394.

[4] K. Lee, ''Intelligent Shape based Image Analysis", *Department of Computer Science, University of Iowa,* 55:247 Project, 2001.

[5] N. Lavrac, "Data Mining in Medicine: Selected Techniques and Applications", *In: Proc. of the Second Intemational Conference on Practical Applications of Knowledge Discovery and Data Mining,* pp 11-31**.**

[6] N. Lavac, E. Keraunou and B. Zupan. "Intelligent Data Analysis in Medicine and Pharmacology." *(IDAM4P-97),*

Nagoya, Japonska, Kluwer Acedemic Publishers, 1997**,** pp 61-67

[7] Ginnekan. *Computer Aided Diagnosis in Chest Radiography*. PhD thesis, Utrecht University, March, 2001 pp. 14-17**.**

[8] M. Antonie, 0. Zaiane, A. Coman, "Application of Data Mining Techniques for Medical Image Classification" *In Proceedings of the Second i'nternational Workshop on Multimedia Data Mining,* with *ACM SIGKDD Conference (MDM/KDD'2001),* San Francisco, USA, Aupst 26,2001.

[9] Rui J. P. Defigueiredo, W. Rodman Shanklet, Andrea Maccato, Malcolm B. Dickt, Prashanth Mundkur, "Neural-network-based classification of cognitively normal, demented, Alzheimer disease and vascular dementia from single photon emission with computed tomography image data from brain

[10] Clarke, L.P. et al. (1992): 'Comparison of Supervised Pattern Recognition Techniques and Unsupervised Methods for MRI Segmentation," Proc. SPIE Medical Imaging VI Conference, Vol. 1652 , SPIE Press, Bellingham, WA, 668-677, 1992.

[11] W. Chu, I. Leong, R. Taira, C. Breant, "A Temporal Evolutionary Object Oriented Data Model and its Query Language for Medical Image Management", Proceedings of the 18th KLDB conference, Vancouver, Canada, 1992, pp.53-64

The Algorithm:

Algorithm Partitioned Frequent Pattern (FP) Growth
Divide Transaction Database into n partitions for n number of processors
for i = 1 to n do begin
count 1-item sets in partition i
end for
Read Min supp // User-defined minimum support for large item sets.
Sum local counts for each item X, in each partition to obtain Global count of item, S^G
XIf S^G > = Min supp // Prune all items that are not large
XEnd if
for i = 1 to n do begin
Store large items in Header table, H in descending order of S_X^G
Construct FP-tree in each Partition i
iMine the FP-tree to generate FP conditional pattern bases for each item x
Merge all conditional pattern bases for each item x in partition pi to form local pattern $\boldsymbol{\beta}_x^i$
End for
Merge local patterns to form global pattern, $\boldsymbol{\beta}_X^G$ for item x
if support in global pattern.$\boldsymbol{\beta}_X^G$ < Minsupp
else
Prune X
Mine the global FP-Pattern base, $\boldsymbol{\beta}_X^G$ to obtain the global conditional FP-tree
End if
if tree $\boldsymbol{\beta}_X^G \neq \varphi$ then
Minetree, $\boldsymbol{\beta}_X^G$
End if;
Mine association rules from global FP-tree;
end;.