Hybrid Techniques and Methods for Analysis of Medical Images

[1]Y. Pavan Kumar Reddy & [2]Dr. K. Fayaz
[1]HOD, CSE, Bangalore College of Engineering and Technology, Bangalore, India
[2]Sri Krishnadevaraya University, CSE, Ananthapur, India

Abstract: The medical image processing and analysis methods have been used to help physicians to make important medical decision through physician-computer interaction. Intelligent or model-based quantitative medical image analysis approaches have been explored for computer-aided diagnosis to improve the sensitivity and specificity of radiological tests involving medical images. A critical step in improving accuracy of interpretation of medical images by a computer aided diagnostic system is to focus on hybrid (both image processing and data mining) techniques. The main objective this work is to develop an interactive system for analyzing medical images (such as images MRI, PET, SPECT) by using hybrid methods.

Keywords – Medical Image Processing, Segmentation, Clustering, MAP, K-Means Algorithm, Watershed Segmentation

1. INTRODUCTION

Data mining is the exploration of large quantities of data in order to discover meaningful pattern and rules**.** Clustering is a data mining method used to cluster similar images and, useful in restoration of medical images. The purpose of segmentation is to provide richer information than that which exists in the original medical images alone. Noisy, incomplete, uncertain, nonlinearities in the medical images can be eliminated in the proposed segmentation and data mining methods. Segmentation partitions an image into non-overlapping regions. Segmentation of medical images has significant ramification in clinical medicine [1].Segmentation widely used in detection of cancerous cells in medical images, detection of roads from satellite images and, in many other computer vision based applications [4]. In The segmentation of medical images, the goal is it to make a distinction between normal pathological tissues. For instance, in cerebral medical images some of the tissue classes of interest are white matter, gray matter, cerebrospinal fluid and abnormal tissues. Furthermore, in MR images of breast tissue, the tissue classes of interest are fatty, glandular tissues, dense glandular tissue and cancerous tissue [3]. Most Image Segmentation techniques are categorized as Characteristic feature thresholding or clustering (Feature Domain), Boundary detection (Spatial Domain), and Region growing (Spatial Domain). Characteristic Thresholding or Clustering is Ineffective by itself in image segmentation because it does not take spatial information into consideration. Boundary/Edge detection on noisy, complex images will of the produce missing edges or extra edges. Region Growing Is not clear what at what point the region growing process should be terminated, resulting in under or over image segmentation[6].

i) Watershed Segmentation: The watershed transform is region based image segmentation approach initially proposed by H.Digabel and C.Lantuejoul. It was later refined by lantuejoul and beucher. The concept of watershed is derived from the field of geography and can be explained as: interpret the image as a topographic surface, gray scale images are visualized in 3-dimensions by considering that there is a hole in each minima and the surface is immersed and flood the surface. To avoid the water coming from two different minima to meet, a dam is build. After the final step the only thing visible of the surface would be the dams. The walls between the dams are called the watershed lines [4].

ii) Vincent-soille approach: The technique described in the here is based on the algorithm developed by L.Vincent and P.Soille for implementing the watershed transform. It is based o extraction of catchment basins. Then its complement is computed to get the watershed [10]**.** 1) Sorting of pixels: Watershed segmentation is a simulation of the immersion analogy in a discrete space. It starts with sorting image pixels according to their intensity levels and stores them in a FIFO structure. In order to speed up computations, direct access the pixels are required .this is essential to easily access the pixels of a particular grey value. 2) The immersion process: Beginning with the FIFO containing pixel coordinates with the lowest intensity, it performs a queue –based simulation of immersion. Each minimum is assigned a unique label, and the other nodes are assigned a default label. A pixel is marked to be a part of the watershed if it is beside two or more distinct catchments' basins. If the pixel is surrounded by pixels that belong to a common catchments' basin, then it is marked by the label of those catchments' basins. Lastly, if the pixel still remained with the default label after a particular level, it will belong to the new set of minima. Therefore, it will acquire a unique label.
iii) H.S.Sheshadri and A. Kandaswamy developed a lesion segmentation technique based on the extraction of catchments basins through topographic representation of the mammography image [4]. Their program segmented the digital mammogram taken from mini-MIAS database [4] such that each segment pertains to one suspicious area. The method consists of different modules. Before carrying out segmentation; the mammogram must undergo a preprocessing step to avoid over segmentation. The pre- processing consists in removing or attenuating the curvilinear structures (CLS) by applying an adapted filter mask. The CLS correspond to the blood vessels, the veins, the milk ducts, speculations and fibrous tissue. The second module is to calculate the gradient of the image. The third module is the segmentation method based on the concept of catchments basins. It deals with the immersion principles, applied to a topographic image representation.

The use of the conventional watershed segmentation algorithm for medical image analysis has two drawbacks that include over-segmentation and sensitivity to false edges. The present work addresses the drawbacks of the conventional watershed algorithm when it is applied to medical images by using k-means clustering to produce a primary segmentation of the image. The K-Means clustering is an unsupervised learning algorithm, while the improved watershed segmentation algorithm makes use of automated thresholding on the gradient magnitude map and post-segmentation merging on the initial partitions to reduce the number of false edges and over-segmentation.

All methods (thresholding or clustering, boundary detection, and region growing) in each of the three categories have their own limitations. By applying segmentation and data mining methods hierarchically, a new hybrid method can be proposed that will improve on segmentation. The main motivation of this research work is to investigate on hybrid methods by combining clustering and region merging methods, that leads to develop an interactive system for medical image analysis.

The existing method of medical image analysis (the inspection of two-dimensional grayscale images on a light box) is not sufficient for many applications. When detailed or quantitative information about the appearance, size, or shape of patient anatomy is desired, image segmentation is often the crucial. Applications of interest that depend on image segmentation include three-dimensional visualization, volumetric measurement, research into shape representation of anatomy, image-guided surgery, and detection of anatomical changes over time[3].

2. OBJECTIVES

The hybrid methods have to be applied in design optimum protocol for raw data acquisition then, it is possible to reconstruct the image from raw data. Noise and artifact reduction in image space should be achieved by enhancing images in Regions of Interest. The segmentation and data mining techniques that are applied on medical images will enable the Computer Aided Diagnosis (CAD), Multimodality Image Fusion and Virtual Reality (Virtual Surgery). The hybrid methods have to be applied and validated on various medical images such as X-Ray (CT/Radiography), – MRI (Magnetic Resonance Imaging), PET (Positron Emission Tomography), US (Ultra Sound), SPECT (Single Photon Emission CT), and EIT (Electrical Impedance Tomography).When evaluating segmentation methods for a particular application, the following factors should be tested [2].

1. Segmentation Time**:** The overall time to complete segmentation includes the running time of the algorithm and/or the time spent performing interactive segmentation. All the segmentation algorithms fall into four categories: automatic, automatic after initialization, semi-automatic, and manual. The semi-automatic algorithms are expected to take less time than a manual segmentation, which may take from minutes to days depending on the number of structures and slices. The running time of the automatic algorithms listed is on the order of seconds to hours, depending on the algorithm and the machine on which it is run [8].
2. Accuracy**:** The accuracy of a segmentation algorithm is generally evaluated by comparison with a manual segmentation, as there is no gold standard. Ideally, all methods should be evaluated for performance on data from a phantom or cadaver, but this is not practical. So the expert manual segmentation is compared with the output of the segmentation method, often with volumetric measures that do not address the main question of surface differences on the boundary of the segmentation [8].
3. Reproducibility**:** Reproducibility refers to the ability of an algorithm or operator to produce the same results more than once, or for different operators to produce the same result. This can be evaluated using the same measures as for accuracy, except that instead of comparing with a manual segmentation, comparison is done between segmentations that have been redone, or reproduced.
4. Generality and Applicability: Generality is considered useful in a segmentation method, and the introduction of more Knowledge may limit generality by reducing the applicability of a

method to various types of data. On the other hand, if an algorithm exists with beneficial knowledge of the specific problem at hand, it will be preferred for the application.

3. METHODOLOGY

. In this work the concentration is in Specific data methods to optimize, to identify patterns in medical images. Various phases in medical image analysis can be classified as data collection, data preprocessing and feature extraction. Following fig. shows process of medical image mining and analysis.

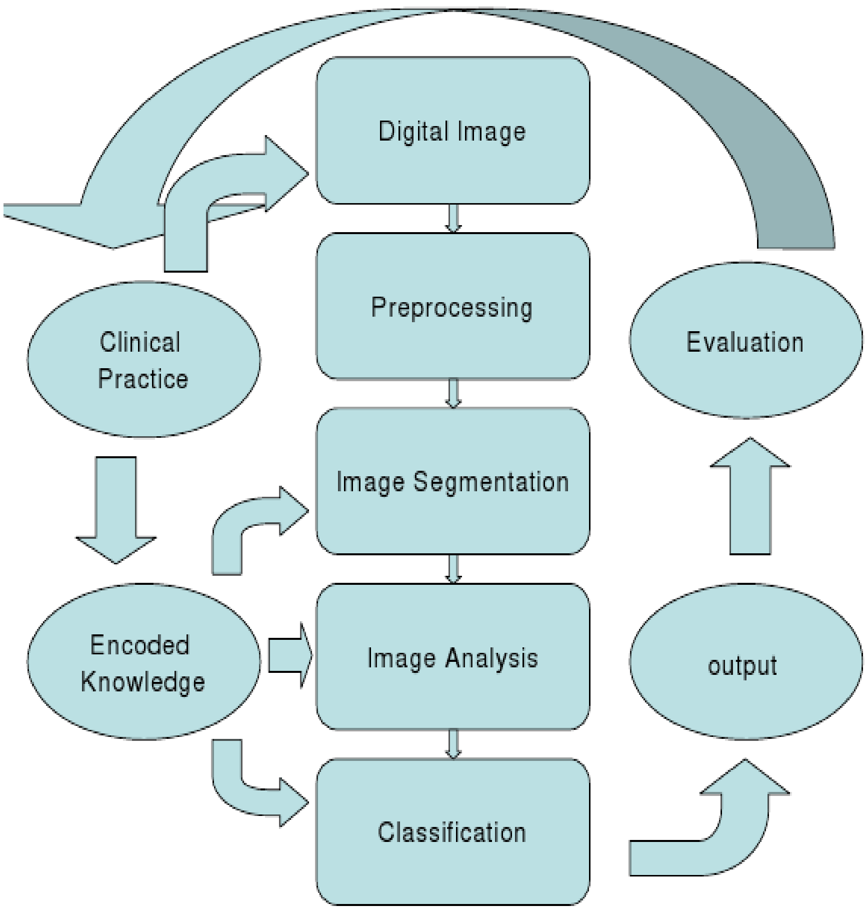

Fig. Process of Medical Image Mining and Analysis

Data Collection:

To have access to real medical images for experimentation is a very difficult undertaking due to privacy issues and heavy bureaucratic hurdles. The data collection that can be used in this work can be taken from the Mammography Image Analysis Society (MIAS) [11]. All the images also include the locations of any abnormalities that may be present. The data in the collection consists of the location of the abnormality (like the centre of a circle surrounding the tumor), its radius, position (left or right), type of t tissues (fatty, fatty glandular and dense) and tumor type if exists (benign or malign). All the mammograms are medico-lateral oblique view.

Pre-processing Phase**:**

Medical images are difficult to interpret, and a preprocessing phase of the images is necessary to improve the quality of the images and make the feature extraction phase more reliable. Pre-processing is always a necessity whenever the data to be mined in noisy, inconsistent or incomplete and pre-processing significantly improves the effectiveness of the data mining techniques [5]. In the digitization process, noise could be introduced that needs to be reduced by applying some image processing techniques. In addition, at the time that the mammograms were taken, the conditions of illumination are generally different. There are two techniques in image preprocessing: a cropping operation and an image enhancement. The first one was employed in order to cut the black parts of the image as well as the existing artifacts such as written labels etc. For most of the images in the dataset, almost 50% of the whole image comprised of a black background with significant noise. Cropping removed the unwanted parts of the image usually peripheral to the area of interest. Image enhancement helps in qualitative improvement of the image with respect to a specific application [10]. In order to diminish the effect of over brightness or over darkness in the images and accentuate the image features, we applied a widely used technique in image processing to improve visual appearance of images known as Histogram Equalization. Histogram equalization increases the contrast range in an image by increasing the dynamic range of grey levels (or colors) [10]. This improves the distinction of features in the image. The method proceeds by widening the peaks in the image histogram and compressing the valleys. This process equalizes the illumination of the image and accentuates the features to be extracted. That is how the different illumination conditions at the scanning phase are reduced.

Feature Extraction:

After cropping and enhancing the images, which represent the data cleaning phase, features relevant to the classification are extracted from the cleaned images. The extracted features are organized in a database in the form of transactions, which in turn constitute the input for both classification algorithms used. The transactions are of the form*f*ImageID, Class Label, $F1; F2; :::; Fng$ where $F1:::Fn$ are ne *n* features extracted for a given image. This database is constructed by merging some already existing features in the original database with some new visual content features that we extracted from the medical images using image processing techniques. The existing features are: _ the type of the tissue (example: dense, fatty and fatty glandular); *the* position of the organ: left or right. The type of tissue is an important feature to be added to the feature database, being well known the fact that for some types of tissue the recognition is more difficult than for others. Training the classification systems with these features incorporated could increase the accuracy rate. The extracted features are four statistical parameters (mean, variance, skewers and kurtosis).

Data mining can be optimized to identify specific patterns in data by using specific methods. These methods can be classified as Association Data, Clustering, Classification and Case-based reasoning methods [5, 7].

Classification: Assignment of an object unknown to a known class is known as assignment of data. The main aim of the classification is to find a model for class attribute as a function of the values of other (predictor) attributes, such that previously unseen records can be assigned a class as accurately as possible.

Clustering: The main aim of clustering is for a given set of data points, a set of attributes is to measure similarities and make cluster such that, data points in one cluster are more similar to one another and data points in separate clusters are less similar to one another.

- Similarity measures
- Euclidean distance if attributes are continuous
- Problem specific measures

Both clustering and classification aim at partitioning a dataset into subsets that bear similar characteristics. Different to classification clustering does not assume any prior knowledge, which are the classes/clusters to be searched for. There exist no class label attributes that would tell which classes exist. Thus clustering serves in particular for exploratory data analysis with little or no prior knowledge. One important application of clustering we have in fact already introduced in information retrieval. The basic problem of information retrieval, i.e. find a set of documents matching a query, can be interpreted as a clustering problem, where the goal is to find two clusters of documents, namely the cluster of relevant ones and the cluster of non-relevant ones. In the tf-idf scheme in fact the tf-measure served to measure intra-cluster similarity for the two document clusters, whereas the id measure served to measure inter-cluster dissimilarity of the document clusters. Clustering has important applications on the Web in order to extract information from large data collections, both document collections and transactional data. Clustering is also an important tool in scientific data analysis and has, for example, a long tradition in image processing and related areas. Data mining frequently adopts techniques from these areas and extends them to make them applicable for analyzing large data sets.

4. FUTURE WORK & CONCLUSION

In this paper, k-means algorithm and water shed algorithms are chosen for medical image analysis. The hybrid methods mentioned for medical image modalities and even can be applied to non-medical images, by this way; it supports a remarkable level of the universality. Unlike to other similar methods, the method is not a problem-oriented solution for the image segmentation. The experimental results are presented by visualization and comparison with the provided medical images by using three well-known measures, namely, sensitivity, precision rate, and overlapping ratio. Obtained numerical results are promising. As a future work, we will enrich the image processing by using more algorithms in both pre- and post-processing phases; furthermore, other optimization method will be investigated to solve the combinatorial part of the optimization problem.

References

[1] Alan Faulkner, Suchendra Bhndarkar, ”Interative Tool for Segmentation, Visualization, and Navigation of Magnetic Resonance Images”, IEEE Transaction on Image Processing and Data mining,16(2):85-91, 2014

[2] X. Falcao, J. K. Udupa, and F. K. Miyazawa, “An Ultra-Fast User-Steered Image

Segmentation Paradigm: Live Wire on the Fly", IEEE Transactions on Medical Imaging, 19(1):55–61, 2000.

[3] H. Li et al. Markov, "Random field for tumor detection in digital mammography for Medical Diagnosis". IEEE Trans. Medical Imaging, 14(3):565–576, 1995.

[4] Jaya Sharma, Paramjeet Kaur, "Performance Analysis of Watershed Segmentation Techniques for breast cancer detection" ICCVGIP, 14(2):58-60,2009

[5] Jiawei Han and Micheline Kamber, "*Data* Mining,Concepts and Techniques. Morgan Kaufmann", IEEE Trans. Data Mining, 4(1):15-20, 2001.

[6] Rafael C. Gonzalez and Richard E. Woods. Digital Image Processing, *2nd edition*. Addison-Wesley, 1993.

[7] Timothy Liao, Dr Sid Ray, "Image Segmentation and Hybrid Methods for Clustering and Region Merging", IEEE Transaction on Computer Science and Software Engineering, 3(2):60-67, 2003.

[8] T. McInerney, D. Terzopoulos. "Deformable Models in Medical Image Analysis": A Survey. Medical Image Analysis, 1(2):91–108, 1996.

[9] Turi, R. and Ray, S. (1998). K-means clustering for colour image segmentation with automatic detection of k, Proceedings of International Conference on Signal and image Processing, Las Vegas, Nevada, USA, pp. 345-349, 1998.

[10] L. Vincent, P.Solle, "Watersheds in digital spaces: an Efficient algorithm on immersion simulations", IEEE Transactions on pattern Analysis and Machine Intelligence, 13, ,583-598**,**1991

[11] http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html.