A Novel Method for Brain Tumor Detection in MRI Images

Abdelmonim Mohammed Yahya Naway

B.SC in computer science
Faculty of computer studies
International University of Africa (2004)

A Dissertation
Submitted to the University of Gezira in Partial Fulfillment of the Requirements for the Award of the Degree of Master of Science in Computer Science
Department Of Computer Science
Faculty of Mathematical sciences and computer

August/2015
A Novel Method for Brain Tumor Detection in MRI Images

Abdelmonim Mohammed Yahya Naway

Examination Committee:

<table>
<thead>
<tr>
<th>Name</th>
<th>Position</th>
<th>Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Magdi Bakr Amien</td>
<td>Chair Person</td>
<td></td>
</tr>
<tr>
<td>Dr. Yasir Abdelgadir Mohammed</td>
<td>External Examiner</td>
<td></td>
</tr>
<tr>
<td>Dr. Mohammed Abbas elameen</td>
<td>Internal Examiner</td>
<td></td>
</tr>
</tbody>
</table>

Date of Examination: 8/8/2015
A Novel Method for Brain Tumor Detection in MRI Images

Abdelmonim Mohammed Yahya Naway

Supervision Committee:

<table>
<thead>
<tr>
<th>Name</th>
<th>Position</th>
<th>Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Magdi Bakr Amien</td>
<td>Main Supervisor</td>
<td>..................</td>
</tr>
</tbody>
</table>
| Dr. Abdalla Akode Osman| Co-supervisor | ..................

Date: 4/5/2015
Dedication

To My Family.
A Novel Method for Brain Tumor Detection in MRI Images

Abdelmonim Mohammed Yahya Naway

Abstract

Brain tumor, is serious illness that affect the human most vital organ. Identification of the tumor is precious for treatment. Radiotherapists use medical images to locate the tumor. Magnetic resonance imaging (MRI) scans the brain to affirm the presence of brain tumor and to identify its location for selected specialist treatment. However, identification of brain tumor from medical images is still a critical and complicated job for Radiotherapists. Brain tumor identification from magnetic resonance imaging (MRI) consists of several phases. Segmentation is known to be an essential step in medical image analysis. Performing the brain MR images segmentation manually is an onerous task as there are several challenges associated with it. Radiotherapists and medical experts spend lengthy of time for manually segmenting brain MR images. In addition, images sometimes are corrupted with noise when performing MRI procedures. This problem make the radiotherapist face the difficulties in identifying tumor or specifying the location of the tumor precisely. This dissertation deals with the task of automatically segmenting the brain tumors. The presented methodology for brain tumor detection consists of three phases: firstly, pre-processing MRI images captured usually susceptible to speckle noise, salt and pepper noise. This phase perform denoising for the MRI image by applying the median filter. Secondly, the image is segmented into various regions. The segmentation performed with the aid of fuzzy c-means algorithm (FCM). Thirdly, features extraction: texture is an important characteristic used in identifying regions of interest in the image. Grey Level Co-occurrence Matrices (GLCM) is used for feature extraction. Cluster analysis has been performed for the extracted features. The results show the success of images segmentation and features extraction for both normal and abnormal images. The system performance analysis result indicates that there are improvement in the measurements of sensitivity (1.0), specificity (0.93), and accuracy (0.89). The study recommends cooperation of personnel in medical centers, build website or blog for information exchange, and translation materials related to medical image mining from different languages.
طريقة جديدة لتحديد الورم في الدماغ في صور الرنين المغناطيسي

عبد المنعم محمد يحي نواي

ملخص الدراسة

الأورام في المخ، هي واحدة من أمراض الدماغ الأكثر شيوعا، وقد أثرت ودمرت العديد من الأرواح. الكشف الدقيق عن الورم هو مسألة جوهرية للغاية لتحديد العلاج. من أجل تقليل الأخطاء التشخيصية يستخدم الأطباء الصور الطبية لتشخيص الأمراض بدقة. صور الرنين المغناطيسي (MR) تقوم بعمل مسح للدماغ للتأكد من وجود ورم في المخ وتحديد موقعه لتلقي العلاج. ومع ذلك، تحديد ورم في المخ من الصور الطبية لا يزال مهمة صعبة ومعقدة للأطباء. عملية تحديد ورم في المخ من خلال التصوير بالرنين المغناطيسي (MRI) تتكون من عدة مراحل. تجزئة صورة الورم هي خطوة أساسية في تحليل الصور الطبية. القيام بتجزئة الصورة بشكل يدوي مهمة شاقة كما أن هناك العديد من التحديات المرتبطة بها. الأطباء واختصاصيو المعالجة بالأشعة يقضون فترات طويلة من الزمن لتجزئة صورة MR بدلاً، وهذا مهمة غير قابلة للتكرار. بالإضافة إلى ذلك، هناك مشكلة عند القيام بإجراء التصوير بالرنين المغناطيسي، وهي أن الصور أحيانا تكون متأثرة بالتشويش (noise). مما يؤدي لصعوبات تلقيي إخضائي العلاج بالإشعاع في تحديد الورم ويمكن أن يؤدي إلى تحديد مكان الورم بشكل غير سليم. تتناول هذه الأطروحة مهمة تجزئة صورة الورم المغناطيسي بشكل أتوماتيكي. وتتبع منهجية تقوم على نموذج نظام لكشف الورم في المخ يتكون من ثلاث مراحل: أولاً المرحلة التمهيدية، حيث يتم تخليص الصور من التشويش بنطاق أحد تقنيات إزالة التشويش الأقتصادي (median filter) وهو الفلتر الوسيط، ثانياً، تجزئات الصورة إلى عدة أجزاء مختلفة وذلك من خلال تطبيق خوارزمية التحليل (FCM). ثالثاً، استخراج خصائص الصور (texture features) وهي عملية هامة تستخدم في تحليل المناطق ذات الاهتمام في صورة ما. ومن أجل هذه الغاية تم استخدام الـ (GLCM). النتائج أظهرت نجاح عملية تجزئة الصور وتحديد خصائص الصور الطبيعية والصور غير الطبيعية.
# Table of contents

Abstract ............................................................................................................................... iv  
Table of Contents ............................................................................................................ vi 
List of Tables ...................................................................................................................... vii 
List of Figures .................................................................................................................... ix 

**Chapter 1** 1.1 Introduction .......................................................................................... 1 
1.2 Theoretical background .......................................................................................... 3 
1.2.1 Digital Image ...................................................................................................... 3 
1.2.2 Digital Image Processing ................................................................................... 7 
1.2.3 Data Mining ....................................................................................................... 11 
1.2.4 Image mining ..................................................................................................... 14 
1.2.5 Image Mining Topics ......................................................................................... 14 
1.2.6 Image Mining Process ....................................................................................... 14 
1.2.7 Image Mining frameworks ............................................................................... 15 
1.2.8 Image Mining Techniques ............................................................................... 16 
1.3 Organization of the dissertation ............................................................................ 19 

**Chapter 2** 2.1 Introduction ......................................................................................... 20 
2.2 Related Studies ...................................................................................................... 20 
2.3 Problem statement ............................................................................................... 22 
2.4 Objectives .............................................................................................................. 23 

**Chapter 3** 3.1 Introduction .......................................................................................... 25 
3.2 Brain Anatomy ..................................................................................................... 26 
3.3 Brain Tumors ........................................................................................................ 27 
3.4 MR imaging (MRI) ................................................................................................. 27 
3.5 The dataset ............................................................................................................ 27 
3.6 Preprocessing ....................................................................................................... 28 
3.6.1 Noise types ....................................................................................................... 28 
3.6.2 Noise Model ..................................................................................................... 28
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7 Denoising techniques</td>
<td>28</td>
</tr>
<tr>
<td>3.7.1 Filtering</td>
<td>29</td>
</tr>
<tr>
<td>3.7.2 Median Filter</td>
<td>29</td>
</tr>
<tr>
<td>3.7.2.1 Median Filter Advantages</td>
<td>28</td>
</tr>
<tr>
<td>3.7.2.2 Median Filter Disadvantages</td>
<td>29</td>
</tr>
<tr>
<td>3.7.2.3 Evaluation Metrics</td>
<td>30</td>
</tr>
<tr>
<td>3.8 Segmentation</td>
<td>31</td>
</tr>
<tr>
<td>3.8.1 Fuzzy sets and fuzzy logic</td>
<td>31</td>
</tr>
<tr>
<td>3.8.2 Fuzzy Clustering</td>
<td>31</td>
</tr>
<tr>
<td>3.8.3 Fuzzy C-means (FCM)</td>
<td>31</td>
</tr>
<tr>
<td>3.8.4 Algorithmic steps for the FCM</td>
<td>32</td>
</tr>
<tr>
<td>3.9 Feature extraction</td>
<td>33</td>
</tr>
<tr>
<td>3.9.1 GLCM</td>
<td>33</td>
</tr>
<tr>
<td>3.9.2 GLCM features</td>
<td>34</td>
</tr>
<tr>
<td><strong>Chapter 4</strong></td>
<td></td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>36</td>
</tr>
<tr>
<td>4.2 preprocessing</td>
<td>37</td>
</tr>
<tr>
<td>4.3 Segmentation</td>
<td>38</td>
</tr>
<tr>
<td>4.4 Feature extraction</td>
<td>39</td>
</tr>
<tr>
<td>4.5 Cluster Analysis</td>
<td>43</td>
</tr>
<tr>
<td>4.6 Results</td>
<td>45</td>
</tr>
<tr>
<td>4.7 Performance Analysis</td>
<td>45</td>
</tr>
<tr>
<td>4.8 Comparing results</td>
<td>47</td>
</tr>
<tr>
<td><strong>Chapter 5</strong></td>
<td></td>
</tr>
<tr>
<td>Conclusion</td>
<td>48</td>
</tr>
<tr>
<td>Recommendations</td>
<td>49</td>
</tr>
<tr>
<td><strong>References</strong></td>
<td>50</td>
</tr>
</tbody>
</table>
List of Tables

Table 3-1 GLCM calculation .................................................................34
Table 4-1 Features extracted from abnormal images .................................39
Table 4-2 Features extracted from normal images ....................................41
Table 4-3 Initial Cluster Centers for abnormal images ...............................43
Table 4-4 Number of Cases in each Cluster - abnormal images ..................43
Table 4-5 Final Cluster Centers - abnormal images ...................................43
Table 4-6 Initial Cluster Centers – normal images ....................................44
Table 4-7 Number of Cases in each Cluster -normal images ......................44
Table 4-8 Final Cluster Centers normal images .......................................44
Table 4-9 GLCM Features For abnormal images .....................................45
Table 4-10 GLCM Features For normal images ......................................45
Table 4-11 performance Analysis Results .............................................46
Table 4-12 comparing results ...............................................................47

List of Figures

Figure 1-1 MRI scanner .........................................................................6
Figure 1-2 Image Mining Process ..........................................................15
Figure 3-1 phases of brain tumor detection ..........................................25
Figure 3-2 human brain .....................................................................26
Figure 4-1 system interface .................................................................36
Figure 4-2 metrics before median filter .................................................37
Figure 4-3 metrics after median filter ..................................................37
Figure 4-4 Image after applying FCM
Chapter One

Introduction

1.1 Introduction:

Image mining is a critical technique, which is used to mine knowledge straightforwardly from image. Image mining deals with the hidden knowledge extraction, image data association and additional patterns, which are not apparently accumulated in the images. The most important function of the mining is to generate all significant patterns without prior information of the patterns.

Due to recent technology progresses, large volumes of medical data is prevailed. Medical images include vast amount of unobserved information that exploited by physicians in making reasoned decisions about a patient. However, extracting this relevant concealed information is a critical first step to their use. For this reason, image-mining techniques are used for efficient knowledge extraction.

Brain tumor, which is one of the most common brain diseases, has affected and devastated many lives.

Magnetic resonance (MR) imaging scan the brain to confirm the presence of brain tumor and to identify its location for selected specialist treatment options.

Radiotherapists use medical images to diagnose diseases precisely. However, identification of brain tumor from medical images is still a critical and complicated job for Radiotherapists [1]. Brain tumor identification form magnetic resonance imaging (MRI) consists of several stages. Segmentation is known to be an essential step in medical image analysis. Performing the brain MR images segmentation manually is an arduous task as there are several challenges associated with it. Radiotherapists and medical experts spend lengthy of time for manually segmenting brain MR images, and this is a non-repeatable task. In addition, a problem when performing MRI procedures is images sometimes are corrupted with noise. This problem make the radiotherapist face difficulties in identifying tumor and may lead to discommode [2].

In view of this, an automatic segmentation of brain MR images is needed to perform correctly and in a shorter span of time.
The accurate segmentation is all-important as otherwise the wrong identification of disease can lead to severe consequences.

The accuracy can be improved by using computer aided diagnosis (CAD) systems.

1.2 Theoretical background

1.2.1 Digital Image

“An image may be defined as a two dimensional function, f(x, y), where x and y are spatial (plane) coordinates, and amplitude of f at any pair of coordinates(x, y) is called the intensity or gray level of the image at that point. When x and y, and intensity values of f are all finite, discrete quantities, we call the image a digital image”. [3]

1.2.1.1 Image Information

a. Pixel
An element in the display on a monitor or data projector.

b. Size
It is the rectangular pixel dimensions of the 2D image – for example, 512 x 512 might describe a single image [16].

c. Scale
d. Resolution
1.2.1.2 Pixel Information

An image represents the state of a subject at some time in the past (e.g. an MR image), then the image data represents a discrete sampling of some physical property of the subject. An MR image, for example, will have been acquired with a specific field of view and matrix size. A pixel in the raw MR image data represents the average MR signal intensity in a specific volume of space inside the MR scanner [16].

1.2.1.3 Image Metadata

Significant information such as who or what was imaged, and how and when the imaging was carried out would have to be registered somewhere and faithfully connected with the
pixel intensity data. All the non-intensity data is called the image metadata, or image file header [16].

1.2.1.4 Image File Formats

a. Bitmaps and BMP Files
It can be visualized as a table in which each entry represents the intensity of a pixel.

For a gray scale image with 256 possible intensities, we will need m x n x 8 bits to store the image data.

b. Vector Graphics
A vector graphics image describes the line and tonal detail as a collection of vectors – lists of points that describe the geometry of objects in an image.

c. JPEG
JPEG is a compression method, not a file format, and it may be used within file formats other than JFIF.

d. Graphic Interchange Format (GIF)
GIF provides for multiple layers, including transparent layers. Transparency permits an image to be displayed on a background such that pixels designated as transparent in the image are displayed with the background color [16].

e. Portable Network Graphics (PNG)
File format was developed as a lossless storage format that would still provide efficient compression. PNG provides for variable precision (8–16 bits) and variable transparency, but does not allow multiple layers. Because of its high precision and lossless compression, PNG could safely be used for storage and transmission of individual medical images. The PNG file will, however, lack the extensive and standardized metadata capability of the DICOM format [16].

f. Tagged Image File Format (TIFF or TIF)
Was designed by developers of color printers, monitors, and scanners. It focuses on the quality of the image rather than the size of the image file; however, several different
compression methods are supported. A useful feature of the TIF format is that it can store multiple images, or layers, in a single file [16].

g. DICOM
Most medical imaging systems archive and transmit image data in DICOM (Digital Imaging and Communications in Medicine) format. The DICOM standard is designed to enable efficient exchange of radiological (images, patient information, scheduling information, treatment planning, etc.) independent of modality and device manufacturer. A DICOM image file comprises a header of image metadata and the raw image data within a single file. The header contains information about the imaging system, the acquisition parameters, and some information about the patient (or the object that was imaged).

1.2.1.5 Medical Images
The purpose of medical imaging is to reveal and record the structural or functional state of the body. Mostly we want to see what is going on inside the body – to check that all is well, or to find out why all is not well. Sometimes we want a record of the current state of the body to be referred to at some future time – to monitor the progress, or absence of progress, of a disease or a treatment [16].

1.2.1.6 Medical Imaging Methods
Images are produced in several ways such as:

a. Visible Light Imaging interacts strongly with tissue and thus does not penetrate more than a few millimeters into the body. Visible light imaging is therefore used for characterization of the body surface, the internal surfaces of the eye, and the body cavities accessible with an endoscopic camera.[16]

b. X-Ray Imaging: Differential transmission of X-ray photons by body tissue is fundamental to image formation – it is the mechanism of contrast generation.

c. Computed Tomography: is the creation of sectional images of an object. X-ray computed tomography (X-ray CT, or just CT) uses multiple projection images to construct a cross sectional image of the body that represents a 2D map of the X-ray attenuation coefficients of the tissue.
d. Positron Emission Tomography (PET): a nuclear imaging technique that provides physicians with information about how tissues and organs are functioning [33].

e. Magnetic Resonance Imaging:
MRI is a medical imaging technique, and radiotherapists use it for visualization of the internal structure of the body. MRI can provide plentiful of information about human soft tissues anatomy as well as helps diagnosis of brain tumor. MR images are used to analyze and study behavior of the brain.

A powerful magnetic field is used to align the nuclear magnetization of hydrogen atoms (or protons) of water in the body. In the presence of RF (radio frequency) electromagnetic fields, hydrogen nuclei produce a rotating magnetic field, which is detectable by the scanner. Since protons can absorb energy at specific frequency and have the ability to reemit that energy; therefore, a transmitter coil is normally fitted around the human skull to measure the net magnetization. The transmitter coil functions in the following way: first, it produces electromagnetic waves and transmits these waves inside the brain, and then a receiver coil measures the intensity of the emitted electromagnetic waves. Moreover, an additional gradient coil is used for spatial localization of the signal. Lastly, the recorded signals (or electromagnetic waves) are reconstructed into an image by a specialized computer program [1].

Figure 1-1 MRI scanner [17]
1.2.2 Digital Image Processing

The field of Digital Image Processing refers to processing digital images by means of digital computer. [3]

The major topics within the field of image processing include [18]:

a. Image restoration

Image restoration is often used in the field of photography or publishing where an image was somehow degraded but needs to be improved before it can be printed.

b. Image Enhancement

Involves taking an image and improving it visually.

c. Image Compression

Involves reducing the typically massive amount of data needed to represent an image.

d. Image Representation

The digital image I (r, c) is represented as a two-dimensional array of data, where each pixel value corresponds to the brightness of the image at the point (r, c). In linear algebra terms, a two-dimensional array like our image model I(r, c) is referred to as a matrix, and one row (or column) is called a vector [18].

1.2.2.1 Image types

a. Binary images are the simplest type of images and can take on two values, typically black and white, or ‘0’ and ‘1’.

b. Gray scale images are referred to as monochrome, or one-color image. They contain brightness information only, no color information. The number of different brightness level available. The typical image contains 8 bit/ pixel.

c. Color Image can be modeled as three band monochrome image data, where each band of the data corresponds to a different color. When the image is displayed, the corresponding brightness information is displayed on the screen by picture elements that emit light energy corresponding to that particular color.
1.2.2.2 Image analysis

Image analysis involves manipulating the image data to determine exactly the information necessary to help solve a computer-imaging problem [18]. The image analysis process can be broken down into three primary stages:

a. Preprocessing

Is used to remove noise and eliminate irrelevant, visually unnecessary information.

Other preprocessing steps might include:

- Gray–level or spatial quantization (reducing the number of bits per pixel or the image size).
- Finding regions of interest for further processing.

b. Data Reduction

Involves either reducing the data in the spatial domain or transforming it into another domain called the frequency domain, and then extraction of features for the analysis process.

c. Features Analysis

The features extracted by the data reduction process are examined and evaluated for their use in the application. After preprocessing, we can perform segmentation on the image in the spatial domain or convert it into the frequency domain via a mathematical transform. After these processes, we may choose to filter the image. This filtering process further reduces the data and allows us to extract the feature that we may require for analysis.

1.2.2.3 Noise

Noise is any undesired information that contaminates an image. Noise appears in image from a variety of source. The digital image acquisition process, which converts an optical image into a continuous electrical signal that is then sampled, is the primary process by which noise appears in digital images. Noise can be removed with the help of filtering techniques.
1.2.2.4 Image segmentation

Image segmentation is a broad and active field, not only in medical imaging, but also in computer vision and satellite imagery. Its purpose is to divide an image into regions, which are meaningful for a particular task [20].

1.2.2.5 Segmentation approaches

a. Segmentation Based on Edge Detection
This method endeavor to unravel an image segmentation by detecting the edges or pixels between different regions that have rapid transition in intensity are extracted and linked to form closed object boundaries [26].

b. Segmentation Based on Thresholding
This approach is for images with different intensities. Using this approach, the image is partitioned directly into different regions based on the intensity values [50].

c. Region based segmentation
Region based approach, partition an image into regions that are similar according to a set of predefined criteria [50].

d. Segmentation Based on Clustering
Clustering classifies pixels into classes, without knowing previous information or training. It classifies pixels with highest probability into the same class. Clustering approach training is done by using pixel features with properties of each class [50].

1.2.2.6 Fuzzy C-means (FCM)

Bezdek introduced Fuzzy C-Means clustering method in 1981, extend from Hard C-Mean clustering method. FCM is an unsupervised clustering algorithm that is applied to wide range of problems connected with feature analysis, clustering and classifier design. FCM is widely applied in agricultural engineering, astronomy, chemistry, geology, image analysis, medical diagnosis, and shape analysis and target recognition. This algorithm is used for analysis based on distance between various input data points. The clusters are formed according to the distance between data points and the cluster centers are formed for each cluster. In fact, FCM is a data clustering technique in which a data set is grouped into n
clusters with every data point in the dataset related to every cluster and it will have a high degree of belonging (connection) to that cluster and another data point that lies far away from the center of a cluster which will have a low degree of belonging to that cluster [32].

1.2.3 Data Mining

Data mining come along in the middle of 1990's and come out as powerful tool that is suited for bringing in antecedently unknown pattern and useful information from enormous dataset [12].

Data mining is the process of discovering interesting patterns and knowledge from large amounts of data [4].

1.2.3.1 Data mining tasks

Task in data mining can be classified into:

a. Classification

Classification divides data samples into target classes. The classification technique predicts the target class for each data points [12].

There are diverse classification algorithm used in the medical field such as K-Nearest Neighbor (K-NN), Decision Tree (DT), Support Vector Machine (SVM), Neural Network (NN) and Bayesian Methods.

b. Regression

Regression is used to find out functions that explain the correlation among different variables [12].

c. Clustering

In clustering large database are separated into the form of small different subgroups or clusters. Clustering partitioned the data points based on the similarity measure [12]. Clustering approach is used to identify similarities between data points. Each data points within the same cluster are having greater similarity as compare to the data points belongs to other cluster.
There are many clustering methods

- Partitioned Clustering In this clustering method the datasets having ‘n’ data points partitioned into ‘k’ groups or clusters. Each cluster has at least one data point and each data point must belong to only one cluster [12].

- Hierarchical Clustering algorithm decomposes the data points in hierarchical way. It decompose the data points either using bottom up approach or top down approach [12].

- Density Based Clustering The problem with partition and hierarchical clustering method is that they can handle only spherical shaped cluster and are not suitable for discovering cluster of arbitrary shapes [12]. Density clustering methods remove this drawback and efficiently handle outliers and arbitrary shaped cluster.

d. Association

Association is one of the most vital approach of data mining that is used to find out the frequent patterns, interesting relationships among a set of data items in the data repository [12].

1.2.3.2 Data mining Techniques

As multi-disciplinary arena, data mining espouse its techniques from many research areas [13].

a. Statistics

Many statistical tools have been used for data mining including correlation analysis, Cluster analysis, and Bayesian network.

b. Machine Learning

Machine learning produces comparable (and often better) predictive accuracy. Its good performance as compared to statistical methods [14]. This can be ascribed to the fact that it is free from parametric and structural assumptions that underlie statistical methods. Some of the machine learning techniques are Neural Networks, Genetic Algorithms, Support Vector Machine, and Decision Tree Induction.
c. Fuzzy Logic

Fuzzy logic, which may be viewed as an extension of classical logical systems, provides an effective conceptual framework for dealing with the problem of knowledge representation in an environment of uncertainty and imprecision.

1.2.3.3 Data mining application in medical sciences:

In medical science, there is large scope for application of data mining. Radiotherapists face lot of difficulties in detection of tumors that is why CAM (Computer Aided Methods) could help to the medical staff. So that they can produce the good quality of the result detection [15].

1.2.4 Image mining

Digitization in every sectors accompanied with advances in in image acquisition and storage technology have led to enormous growth in significantly large and detailed image databases(satellite images, medical images). These images, if analyzed can reveal useful information to the human users.

Regrettably it is unmanageable or even inconceivable for human to discover the underlying knowledge and patterns in the image when handling a large collection of images.

“Image mining deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the images. It is an interdisciplinary attempt that draws upon expertise in computer vision, image processing, image retrieval, data mining, machine learning, database, and artificial intelligence”. [5]

1.2.5 Image Mining Topics

Research in image mining can be generically classified into two main topics:

- The first topic involves domain-specific applications where the focus is in the process of extracting the most relevant image features into a form suitable for data mining [6].
The second topic involves general applications where the focus is on the process of generating image patterns that may be helpful in the understanding of the interaction between high-level human perceptions of images and low-level image features [6].

1.2.6 Image Mining Process

The images from an image database are first preprocessed to improve their quality. These images then undergo various transformations and feature extraction to generate the important features from the images [6]. With the generated features, mining can be carried out using data mining techniques to discover significant patterns [6]. The resulting patterns are evaluated and interpreted to obtain the final knowledge, which can be applied to applications.

![Image Mining Process](image.png)

Figure 1-2 Image Mining Process [7].

1.2.7 Image Mining frameworks

Early work in image mining has concentrated on developing a suitable framework to perform the process of image mining.
The image database containing crude image data cannot be directly used for mining aims[8]. Crude image data need to be processed to generate the information that is usable for high-level mining modules.

A good image mining system is expected to provide users with an effective access into the image repository and generation of knowledge and patterns underneath the images. Such a system typically encompasses the following functions: image storage, image processing, feature extraction, image indexing and retrieval, and patterns and knowledge discovery.

It can be discerned that there are three kinds of frameworks used to characterize image-mining systems:

- Function-driven frameworks:
  Focuses on the functionalities of different component modules to organize image-mining systems [5].

- Information-driven frameworks:
  Designed as a hierarchical structure with special emphasis on the information needs at various levels in the hierarchy [9].

- Knowledge Driven Framework: Focus on how to mine the maximum knowledge from the mining course [34].

1.2.8 Image Mining Techniques:

a. Features extraction

A feature is a characteristic that can capture a certain visual property of an image either globally (whole image) or locally (object or regions) [10].

There are three types of feature vectors for an image [11]

- Pixel level features: store spectral and textural information about each pixel of the image.
- Region level features describe groups of pixels.
- Tile level present information about whole images.
Color, edges, shape, and texture are the common image attributes that are used to extract features for mining.

a. Object Recognition
An object recognition system finds objects in the real world from an image. This is one of the major tasks in the domain of image mining.

b. Image classification
Image classification is the task of assigning objects to one of several predefined categories. There are three processes involved in image classification, which are Feature extraction, training and Classification [19].

c. Image Clustering
Image clustering is an unsupervised learning method, which groups a given set of unlabeled images into meaningful clusters according to the image content.

The process normally comprises of four steps [19]:

- Image preprocessing, feature extraction and selection.
- Set up similarity metrics suitable for special application.
- Image clustering
- Form cluster.

d. Image retrieval
Image mining requires that images be retrieved according to some requirement description. The requirement description can be classified into three levels [19]:

- Level 1 comprises image retrieval by primary features such as color, texture and shape.
- Level 2 comprises image retrieval by derived or logical features like objects of a given type or individual objects or persons.
- Level 3 comprises image retrieval by obtaining attributes, involving a significant amount of high level reasoning about the meaning or purpose of the objects or scenes depicted.
e. Image Indexing
The objective of image indexing is to retrieve similar images from an image database for a given query image (i.e., a pattern image). Each image has its unique feature. Hence, image indexing can be implemented by comparing their features, which are derived from the images [19].

f. Association rule
Association rule mining finds interesting associations and correlation relationships among large set of image data. Association rules show attribute value conditions that occur frequently together in a given dataset [20].

g. Neural Networks
Neural Networks is one of the technique used in image processing and image retrieval.

1.2.8.1 Gray Level Co-occurrence Matrix (GLCM)

Gray Level Co-occurrence Matrix by Haralick contains the information about the gray level intensities of pixels and their neighbors, at fixed distance and orientation. GLCM depicts how often-different combinations of gray level co-occur in an image. The idea is to scan the image and keep track of gray levels of each of two pixels separated within a fixed distance d and direction θ. This spatial relationship can be specified in different ways, the default one is between a pixel and its immediate neighbor to its right [21].

1.2.9 Image mining and data mining

Image mining considered as applying existing data mining algorithms on images. This is unquestionably not true, because there are significant factors differentiate relational databases versus image databases, the following are some of these differences [19]:

- Absolute versus relative value. In relational databases, the data value are semantically meaningful. However, in image databases, the data value themselves may not be significant unless the context supports them.
- Spatial information: implicit spatial information is critical for interpretation of image contents but there is no such requirement in relational databases.
Unique versus multiple interpretations: In images, there is multiple interpretations for the same visual patterns, but there is nothing like that in data.

1.3 Organization of the dissertation

- Chapter 1 establish the theoretical background in image processing, data mining, image mining.
- Chapter 2 include related studies concerning image mining, medical image segmentation, and fuzzy clustering.
- Chapter 3 present the modules for the proposed system for the automatic detection of brain tumor from MRI images.
- Chapter 4 present the results of the proposed system.
- Chapter 5 Recommendations and conclusion
Chapter Two
Related Studies

2.1 Introduction

A review of tens of studies was performed to support the attempt of this dissertation. A general survey was first performed to past research efforts in brain tumor detection. This survey encompasses the whole process of the detection, which is generally include three stages, pre-processing, segmentation, and feature extraction. In addition, to the different techniques used in every stages is previewed.

2.2 Related Studies

P. Tamije Selvy et al (2011) [29], point out that Complex medical process cannot be done without image processing techniques. Structures like tumor, brain tissue and skull cannot be identified without image segmentation. They also call attention to that Clustering is suitable for biomedical image segmentation. The performance of four clustering algorithms (K-means, Self-Organizing Maps, Hierarchical Clustering, and Fuzzy C-Means) were analyzed. The results show that the execution time for K-means Clustering and Self-Organizing Maps were less compared to the other clustering methods and regarding the number of tumor pixels, K-means clustering and Hierarchical clustering gave a better result than the other methods.

Rajesh war Dass et al (2012) [27], investigates and compiles some of the technologies used for image segmentation. The authors discuss the problem of classification of the pixel in an image correctly, when there is no crisp boundaries between objects in an image, they provided that Fuzzy clustering technique classify pixel values with great extent of accuracy and it is basically suitable for decision oriented applications like tissue classification and tumor detection. A major finding from their work is that there is no ubiquitously recognized method for image segmentation, as the result of image segmentation is affected by many factors, such as: spatial characteristics of the image, texture, image content. Hence, there is no single method, which can be considered good for neither all type of images nor all methods equally good for a particular type of image.

Rachana Rana et al (2013) [22], presents a review of the various methods used in brain MRI image segmentation. This review covers imaging modalities (with their Types, and Characteristics), magnetic resonance imaging and methods for brain MRI segmentation (Seed-based region growing, Level-set segmentation, Split and Merge-based segmentation, and Edge Based Segmentation) with brief demonstration for each approach. They conclude their work with the emphasis on the needs for more research in this domain to ameliorate precision and accuracy of segmentation methods.

Priyanka, Balwinder Singh (2013) [25], in their review they first established theoretical background on brain tumor, diagnose, MRI, and segmentation. They concisely introduced segmentation techniques such as edge based segmentation, k-means algorithm, and region
growing. They present some of related studies, they also provide an evaluation for the outcome of some techniques that can be used to detect tumor from scanned MRI images of brain.

Prof. Dinesh D. patil, Ms. Sonal G Decore (2013) [26], in their review article they categorize image segmentation approaches into detecting discontinuities (to partition an image based on abrupt changes in intensity), and detecting similarities (to partition an image into regions that are similar according to a set of predefined criterion). The authors perform a comparison of image segmentation techniques with description, advantages, and disadvantages.

B.Jegathees kumar et al (2014) [23], they propose an approach for automated brain tumor detection in four phases (pre-processing, segmentation, feature extraction, and classification). They claim that this approach improve precision by applying classifications.

Nooshin Nabizadeh et al (2014) [24], they highlight the significance of feature extraction techniques, because automatic tumor detection depend on it. They also refers to that although there are many techniques utilized, it is still is not obvious which of feature extraction methods should be favored. They compare the efficiency of using Gabor wavelet features and statistical features that are two chief groups of textured based feature in tumor segmentation. Their results show that statistical features offer higher accuracy than Gabor wavelet.

SivaSankari.S et al (2014) [28], address the concept of brain tumor segmentation and feature extraction, they proposed a system for that consist of three modules pre-processing, segmentation, and feature extraction. The author select k-means clustering for the segmentation and justify that k-means has less computation time. The authors apply the methodology on thirty MRI images and supply results for their work.

Sanjaya nag et al (2014) [30], take a holistic view for the topic. Firstly, they provide adequate descriptions for Brain Morphology, Brain Abnormalities, CAD System and their utility, and MRI Technique. Secondly, move to preprocessing and enhancement. Thirdly, cite some segmentation techniques. Finally perform comparative analysis for various techniques for image segmentation with image modality, benefits, and estimation for their accuracy.

V.Velusamy et al (2014) [31], reviewed different methods of diagnosing brain tumor through MRI used in preprocessing and segmentation techniques. In the first place, the authors depict the removal of the artifacts or label on MRI images, and then highlight the need for pre-processing. They provide general overview for pre-processing methods. They carry on a survey for segmentation methods and present remarks for each methods.
### 2.3 Summary

<table>
<thead>
<tr>
<th>No.</th>
<th>Study title</th>
<th>Techniques used</th>
<th>Result and Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Performance Analysis of Clustering Algorithms in Brain Tumor Detection of MR Images</td>
<td>K-means, Self-Organizing Maps, Hierarchical Clustering, and Fuzzy C-Means</td>
<td>Execution time for K-means Clustering and Self-Organizing Maps were less compared to the other clustering methods</td>
</tr>
<tr>
<td>2</td>
<td>Image Segmentation Techniques</td>
<td>Edge Detection, Thresholding Method, Region Based Segmentation Methods, Segmentation Based on Clustering</td>
<td>Fuzzy clustering technique classify pixel values with great extent of accuracy and it is basically suitable for decision oriented applications like tissue classification and tumor detection</td>
</tr>
<tr>
<td>3</td>
<td>Study of Various Methods for Brain Tumor Segmentation from MRI Images</td>
<td>Seed-based region growing, Level-set segmentation, Split and Merge-based segmentation, and Edge Based Segmentation</td>
<td>There is a need for more research in this domain to ameliorate precision and accuracy of segmentation methods</td>
</tr>
<tr>
<td>4</td>
<td>A REVIEW ON BRAIN TUMOR DETECTION USING SEGMENTATION</td>
<td>Edge based segmentation, Clustering Method, Region growing method</td>
<td>Methods and techniques that are being used to detect the brain tumor from scanned MRI images of brain are evaluated</td>
</tr>
<tr>
<td>5</td>
<td>Medical Image Segmentation: A Review</td>
<td>Detecting discontinuities, and detecting similarities</td>
<td>Comparison of image segmentation techniques with description, advantages, and disadvantages</td>
</tr>
<tr>
<td>6</td>
<td>VARIOUS TECHNIQUES FOR BRAIN TUMOR IDENTIFICATION AND SEGMENTATION APPROACH IN MRI IMAGES</td>
<td>Median Filtering, region growing, GLCM, SVM and ANFIS classifiers</td>
<td>Enhance the precision by applying classification and by using SVM, ANFIS is used to discover the tumor</td>
</tr>
<tr>
<td>7</td>
<td>Efficacy of Gabor-Wavelet versus Statistical Features for Brain Tumor Classification in MRI</td>
<td>GLCM, GLRLM, HOG, HA- SVM, KNN, K-Means</td>
<td>Statistical features offer higher accuracy than Gabor-wavelet</td>
</tr>
<tr>
<td>8</td>
<td>Feature Extraction of Brain Tumor Using MRI</td>
<td>Median Filtering, K-means, GLCM and Gabor</td>
<td>The proposed Method Has a good performance</td>
</tr>
<tr>
<td>9</td>
<td>A Review of Image Segmentation Methods on Brain MRI for Detection of Tumor and Related Abnormalities</td>
<td>Preprocessing techniques, segmentation, hybrid methods for segmentation</td>
<td>Comparative analysis of the discussed methods</td>
</tr>
<tr>
<td>10</td>
<td>Enhancement Techniques and Methods for MRI A Review</td>
<td>Preprocessing and segmentation techniques-median filter</td>
<td>Review of preprocessing techniques, a survey of segmentation techniques</td>
</tr>
</tbody>
</table>
The previous studies uses different techniques for the task of MRI segmentation and feature extraction, sometimes combine more than one techniques. In this dissertation, the presented methodology, which consist of three, phases preprocessing, segmentation and features extraction. In the preprocessing, the median filter is used to denoise the images while preserving the edges, which is very important for the next stage. In the segmentation, the fuzzy C-means is used to segment the image for its competency in working with group of pixels. In features extraction texture features are extracted by using the gray level co-occurrence matrix.

It is believed that this methodology and the techniques as described is new and it is an effort to the endeavors for the brain tumor detection from MRI Images.

2.4 Problem statement

Brain tumor, which is one of the most common brain diseases, has affected and devastated many lives.

Manual segmentation and analysis of MR brain tumor images by radiotherapists is carried through in almost all Hospitals at present. The boundary of the tumor in an image is usually traced by hand, which is time consuming, and difficult to detect and localize, detection becomes infeasible with large set of data sets. The dependability of the segmentation hinges on the knowledge and skill of the radiotherapists. However, the process is draggy, and impractical.

Computer aided detection of abnormal growth of tissues is primarily motivated by the necessity of achieving maximum possible accuracy. The accuracy can be improved by using computer aided diagnosis (CAD) systems. This improves the accuracy and consistency of radiological diagnosis. However, segmentation of the image of brain tumors is a very difficult task. In the first Place, there are a large class of tumor types, which have a variety of shapes and sizes. Appearance of brain tumors at different locations in the brain with different image intensities is another factor that makes automated brain tumor image detection and segmentation difficult.

Accurate detection of brain tumor is highly essential for treatment planning in order to minimize diagnostic errors

2.5 Objectives:

General objective:

To automate the tumor detection and segmentation process.

Specific Objective:

Specifically, this dissertation intend to build a system for the automation of brain tumor detection. This system will include three phases. First phase pre-processing deal with the removal of the noise that may corrupt the image and prepare it for the next phase. Second phase, Segmentation of the MRI image to specify the tumor in the image. Last phase, features extraction for the segmented image.
Chapter Three
The Methodology

3.1 Introduction

In order to achieve the goal of automatic brain tumor detection, a system consisted of three phases is proposed. In the first place, there is a need to eliminate the noise by using median filter. Then the segmentation of the image to locate the tumor is performed with Fuzzy C-means algorithm. Finally, features will be extracted using Gray level co-occurrence matrices.

![Diagram of phases of brain tumor detection](image)

Figure 3-1 phases of brain tumor detection
3.2 Brain Anatomy

The human brain, which functions as the center for the control of all the parts of human body, is a highly specialized organ that allows a human being to accommodate and abide varying environmental conditions. The human brain enables a human to enunciate words, perform actions, and share thoughts and feelings [17].

The brain is composed of two tissue types, namely gray matter (GM) and white matter (WM). Gray matter is made of neuronal and glial cells, also known as neuroglia or glia that controls brain activity and the basal nuclei, which are the gray matter nuclei, located deep within the white matter. The basal nuclei include caudate nucleus, putamen, palladium and clastrum. White matter fibers consist of many alienated axons, which connect the cerebral cortex with other brain regions. The left and the right hemispheres of the brain are connected by corpus callosum, which is a thick band of white matter fibers [17].

The brain also contains cerebrospinal fluid (CSF) which consists of glucose, salts, enzymes, and white blood cells. This fluid circulates through channels (ventricles) around the brain and the spinal cord to protect them from injury. There is also another tissue called meninges which are the membrane covering the brain and spinal cord.

![Human Brain Diagram](image)

Figure 3-2 human brain [17]

Figure 3-1 shows the anatomy of the brain. It composed of the cerebrum and the brain stem. The cerebrum occupies the largest part of the brain. It is connected with the conscious thoughts, movement and sensations. It further consists of two halves, the right and the left hemispheres. Each controls the opposite side of the body. Moreover, each hemisphere is divided into four lobes: the frontal, temporal, parietal and occipital lobes. The cerebellum is the second largest structure of brain. It is connected with controlling motor functions of body such as walking, balance, posture and the general motor coordination. It is situated toward the backside of the brain and is linked to brain stems. Both, cerebellum and cerebrum have a very thin outer cortex of gray matter, internal white matter and small but deeply situated masses of the gray matter. The spinal cord is connected to the brainstem. It is located toward the bottom of the brain. Brainstem
controls vital functions in human body such as motor, sensory pathways, cardiac, repository and reflexes. It has three structures: the midbrain, pons and medulla oblongata [17].

3.3 Brain Tumors

Under certain conditions, brain cells grow and multiply uncontrollably because for some reasons, the mechanism that control normal cells is unable to regulate the growth of the brain cells. The abnormal mass of brain tissue is the brain tumor that occupies space in the skull, interrupts the normal functions of brain, and creates an increasing pressure in the brain. Due to increased pressure on the brain, some brain tissues are shifted, pushed against the skull or are responsible for the damage of the nerves of the other healthy brain tissue [17].

3.4 MR imaging (MRI)

Raymond V. Damadian invented MRI in 1969 and was the first person to use MRI to investigate the human body.

MRI became the most preferred imaging technique in radiology because MRI enabled internal structures be visualized in some detail. With MRI, good contrast between different soft tissues of the body can be observed. This makes MRI suitable for providing better quality images for the brain, the muscles, the heart and cancerous tissues compared with other medical imaging techniques, such as computed tomography (CT) or X-rays [17].

Brain images in MRI scan can be normal or abnormal. An image from MRI scan is composed of gray level intensity values in the pixel spaces. The gray level intensity values depend on the cell concentration in the volume scanned.

3.5 The dataset

The image dataset obtained from DICOM sample image sets [35].

3.6 Preprocessing

3.6.1 Noise types

The main sources of noise in digital images are imperfect instruments, problem with data acquisition process, interference natural phenomena, transmission and compression. Image denoising forms the pre-processing step in the field of photography, research, technology and medical science, where somehow image has been degraded and needs to be restored before further processing[36].

3.6.2 Noise Model

Noise is present in image either in additive or multiplicative form [36]

a. Additive Noise Model Noise signal that is additive in nature gets added to the original signal to produce a corrupted noisy signal
b. Multiplicative Noise Model In this model, noise signal gets multiplied to the original signal.

3.6.3 Types of noise

a. Gaussian noise is evenly distributed over the signal. This means that each pixel in the noisy image is the sum of the true pixel value and a random Gaussian distributed noise value [36].

\[ f(g) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(g-m)^2/2\sigma^2} \]  (3-1)

Where \( g \) represents the gray level, \( m \) is the mean or average of the function and \( \sigma \) is the standard deviation of the noise.

b. Salt and Pepper is an impulse type of noise and is referred to as intensity spikes. It is generally caused due to errors in transmission. This is caused generally due to errors in data transmission. It has only two possible values, \( a \) and \( b \). The probability of each is typically less than 0.1. The corrupted pixels are set alternatively to the minimum or to the maximum value, giving the image a “salt and pepper” like appearance. Unaffected pixels remain unchanged [36].

MRI images captured usually are susceptible to speckle noise, and salt and pepper noise. Image filtering algorithms are applied over the noisy images to carry off the noise and conserve the image details [37].

3.7 Denoising techniques

Image Denoising is the process of removing the noise from the digital images using some prior knowledge about the noise while retaining as much as possible important image features. Basically, there are two approaches to image denoising based on the domain in which the denoising taken place. These approaches are named as spatial domain and transform domain-filtering approaches. Spatial Filtering approaches remove the noise by manipulating the image in the spatial domain itself, whereas Transform Filtering approaches manipulate the image in transform domain [41].

3.7.1 Filtering

The word filter comes from frequency-domain processing, where “filtering” refers to the process of accepting or rejecting certain frequency components [38].

Spatial filtering of images is an important aspect of image processing as it provides means for removing noise and sharpening blurred images. There are many types of spatial filters, which can be classified into linear and non-linear filters. The simplest linear spatial filter is mean filter, which works by passing a mask over the image calculating the mean intensity and setting the central pixel to this value. They tend to remove the fine details in the image and fail to remove high-level noise effectively. Among spatial filters, the famous non-linear filter is standard median filter [41].
3.7.2 Median Filter

Median filtering is extensively used in de-noising and image smoothing applications. Median filters exhibit edge-preserving characteristics, which is very desirable for many image-processing applications as edges contain important information for segmenting, and preserving detail in images [39].

This filter can be represented by the following equation:

\[ G(u,v) = \text{median}\{I(x,y), (x,y) \in wF}\) \] (3-2)

Where:

wF=\(wxw\) Filter window with pixel \((u,v)\) as its middle

The median filter considers each pixel in the image in turn and looks at its nearby neighbors to decide whether it is representative of its surroundings. Instead of simply replacing the pixel value with the mean of neighboring pixel values, it replaces it with the median of those values. Median filter controls the pepper and Gaussian noises. [39].

Most scanned images contain noise caused by the scanning method (sensor and its calibration, electrical components, and radio frequency spikes) this noise may look like dots of black and white. Median filter helps us by erasing the black dots, called the Pepper, and it fills in white holes in the image, called Salt "impulse noise" [42].

3.7.2.1 Median Filter Advantages

- No reduction in contrast across steps, since output values available consist only of those present in the neighborhood (no averages).
- Median filtering does not shift boundaries, as can happen with conventional smoothing filters (a contrast dependent problem).
- Since the median is less sensitive than the mean to extreme values (outliers), those extreme values are more effectively removed [40].

3.7.2.2 Median Filter Disadvantages:

- Less effective in removing Gaussian or random-intensity noise.
- Repeating will remove noise but at the expense of detail (posterization occurs) where pixel brightness values are leveled across regions "group of pixels having similar brightness values" [42].

3.7.2.3 Evaluation Metrics

To evaluate the effects of a filter the following metrics are used:

Root Mean Squared Error (RMSE), Peak Signal-to-Noise Ratio (PSNR), Mean Absolute Error (MAE) [54].
\[ PSNR = 10\log_{10}\left(\frac{R}{MSE}\right) \quad (3-3) \]

Where \( R \) is maximum value of the pixel present in an image and MSE is mean square error between the original and de-noised image with size \( M \times N \). Mean square error is defined as:

\[ MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} [x(i, j) - y(i, j)]^2 \quad (3-4) \]

Where, \( x(i, j) \) is original image and \( y(i, j) \) is denoised image. Root mean square error is defined as:

\[ \text{RMSE} = \sqrt{\text{MSE}} \]

**Root Mean Squared Error (RMSE):**

Root-Mean-Square Error (RMSE) is used to measure of the differences between the original and reconstructed images.

**Peak Signal-to-Noise Ratio (PSNR):**

The PSNR is used as a measure of quality of reconstruction in image compression.

**Mean Absolute Error (MAE):**

Mean Absolute Error (MAE) is used to measure how the reconstructed image close to the original image.

\[ MAE = \frac{1}{n} \sum_{i=1}^{n} |f_i - y_i| = \frac{1}{n} \sum_{i=1}^{n} |e_i| \quad (3-5) \]

the mean absolute error is an average of the absolute errors \( e_i = |f_i - y_i| \), where \( f_i \) is the prediction and \( y_i \) the true value.

The higher PSNR value, lower RMSE value and the lower MAE indicates the better noise removal [43].

### 3.8 Segmentation

Image segmentation is indispensable step in a series of operations with the final cause at overall image understanding. The image is segmented into various regions and the aim of segmentation is to partition an image into significant regions in connection with a particular application. The segmentation is based on measurements taken from the image and might be grey level, color, texture, and depth [49].
3.8.1 Fuzzy sets and fuzzy logic

Lotfi A. Zadeh, a professor at University of California at Berkeley was the first to propose a theory of fuzzy sets and an associated logic, namely fuzzy logic. Essentially, a fuzzy set is a set whose members may have degrees of membership between 0 and 1, as on contrary to classical sets where each element must have either 0 or 1 as the membership degree - if 0, the element is entirely outside the set; if 1, the element is entirely in the set. As classical logic is based on classical set theory, fuzzy logic is based on fuzzy set theory [51].

Fuzzy set theory provide exact representations of concepts and relations that are obscure, that is, with no acute yes-no borderline between cases covered, and cases not covered, by the concept or relation. Fuzzy sets and fuzzy logic extend membership degrees and truth-values from zero and one to the real interval from 0 to 1, the definition of the fuzzy logic formalism still rely on the classical logic [51].

3.8.2 Fuzzy Clustering

Clustering is the process of dividing data elements into different groups (known as clusters) in such a way that the elements within a group possess high similarity while they differ from the elements in a different group. In fuzzy clustering, fuzzy sets are used to cluster data, so that each point may belong to two or more clusters with different degrees of membership. In this case, data will be associated to an appropriate membership value [52].

3.8.3 Fuzzy C-means (FCM)

The algorithm furnish a method for dividing the n data-points into c fuzzy clusters (where c<n), while simultaneously determining the locations of the clusters in the suitable space [53].

It is an iterative algorithm that finds clusters in data and which uses the concept of fuzzy membership. Instead of assigning a pixel to a single cluster, each pixel will have different membership values on each cluster. The Fuzzy C-Means attempts to find clusters in the data by minimizing an objective function shown in the equation below:

\[ J = \sum_{i=1}^{N} \sum_{j=1}^{C} \mu_{ij}^m |x_i - c_j|^2 \]  

(3-6)

J is the objective function. After one iteration of the algorithm, the value of J is smaller than before. It means the algorithm is converging or getting closer to a good separation of pixels into clusters. N is the number of pixels in the image, C is the number of clusters used in the algorithm, and must be decided before execution, \( \mu \) is the membership table -- a table of NxC entries which contains the membership values of each data point and each cluster, \( m \) is a fuzziness factor (a value larger than 1), \( x_i \) is the ith pixel in N, \( c_j \) is j th cluster in C and \( |x_i - c_j| \) is the Euclidean distance between \( x_i \) and \( c_j \) [29].
3.8.4 Algorithmic steps for the FCM

The input to the algorithm is the N pixels on the image and m, the fuzziness value.

**Step 1**: Initialize $\mu$ with random values between zero and one; but with the sum of all fuzzy membership table elements for a particular pixel being equal to 1 -- in other words, the sum of the memberships of a pixel for all clusters must be one.

**Step 2**: Calculate an initial value for $J$ using

$$J = \sum_{i=1}^{N} \sum_{j=1}^{C} \mu_{ij}^m \left| x_i - y_j \right|^2$$  \hspace{1cm} (3-7)

**Step 3**: Calculate the centroids of the clusters $c_j$ using,

$$c_j = \frac{\sum_{i=1}^{N} \mu_{ij}^m x_i}{\sum_{i=1}^{N} \mu_{ij}^m}$$  \hspace{1cm} (3-8)

**Step 4**: Calculate the fuzzy membership table using

$$M_{ik} = \frac{1}{\sum_{k=1}^{C} \left( \frac{|x_i - c_j|}{2} \right)^{2/m-1}}$$  \hspace{1cm} (3-9)

**Step 5**: Recalculate $J$.

**Step 6**: Go to step 3 until a stopping condition was reached.

Some possible stopping conditions of the algorithm are:

1. When a number of iterations were executed, we can consider that the algorithm achieved a "good enough" clustering of the data.

2. The difference between the values of $J$ in consecutive iterations is small (smaller than a user-specified parameter $\varepsilon$), therefore the algorithm has converged [29].

3.9 Feature extraction

Texture is a very useful characterization for a wide range of image. It is generally believed that human visual systems use texture for recognition and interpretation. In general, color is usually a pixel property while texture can only be measured from a group of pixels. A large number of techniques have been proposed to extract texture features. Based on the domain from which the texture feature is extracted, they can be broadly classified into spatial texture feature extraction methods and spectral texture feature extraction methods. For the former approach, texture
features are extracted by computing the pixel statistics or finding the local pixel structures in original image domain, whereas the latter transforms an image into frequency domain and then calculates feature from the transformed image[44].

3.9.1 GLCM

Texture is an important characteristic used in identifying regions of interest in an image. Grey Level Co-occurrence Matrices (GLCM) is one of the earliest methods for texture feature extraction proposed by Haralick back in 1973. Since then it has been widely used in many texture analysis applications and remained to be an important feature extraction method in the domain of texture analysis. Fourteen features were extracted by Haralick from the GLCMs to characterize texture [45].

The GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image [46].

The GLCM described here is used for a series of "second order" texture calculations [46].

- First order texture measures are statistics calculated from the original image values, like variance, and do not consider pixel neighbor relationships.
- Second order measures consider the relationship between groups of two (usually neighboring) pixels in the original image.
- Third and higher order textures (considering the relationships among three or more pixels) are theoretically possible but not commonly implemented due to calculation time and interpretation difficulty.

A GLCM is a matrix where the number of rows and columns is up to the number of gray levels, G, in the image. GLCM matrix formulation can be demonstrated with the example depicted in Table 3-1 for four different gray levels. Here one pixel offset is used (a reference pixel and its immediate neighbor). If the window is large enough, using a larger offset is possible. The top left cell will be filled with the number of times the combination 0,0 occurs, i.e. how many times within the image area a pixel with grey level 0 (neighbor pixel) falls to the right of another pixel with grey level 0 (reference pixel) [47].

<table>
<thead>
<tr>
<th>Neighbor pixel value</th>
<th>ref pixel value</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0,0</td>
<td>0,1</td>
<td>0,2</td>
<td>0,3</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1,0</td>
<td>1,1</td>
<td>1,2</td>
<td>1,3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2,0</td>
<td>2,1</td>
<td>2,2</td>
<td>2,3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3,0</td>
<td>3,1</td>
<td>3,2</td>
<td>3,3</td>
<td></td>
</tr>
</tbody>
</table>

Table 3-1 GLCM calculation [47]

3.9.2 GLCM features [48]

In the following, we will use \( I(x, y), 0 \leq x \leq Nx - 1, 0 \leq y \leq Ny - 1 \) to denote an image with \( G \) gray levels. The \( G \times G \) gray level co-occurrence matrix \( Pd\theta \) for a displacement vector \( d = (dx, dy) \) and direction \( \theta \) is defined as follows. The element \((i, j)\) of \( Pd\theta \) is the number of occurrences of the pair of gray levels \( i \) and \( j \), which the distance between \( i \) and \( j \) following direction \( \theta \) is:
\[ d.P\theta \ d (i, j) = \# \{ (r, s), (t, v) : I (r, s) = i, I (t, v) = j \} \]

Where \((r, s), (t, v) \in N_x \times N_y; (r, v) = (r + dx, s + dy)\).

(1) Angular second moment (ASM) feature:

The ASM is known as uniformity or energy. It measures the uniformity of an image. When pixels are very similar; the ASM value will be large.

\[
f_1 = \sum_{i=0}^{n_o-1} \sum_{j=0}^{n_o-1} pd \cdot \theta(i, j)^2 \quad (3-10)\]

(2) Contrast feature:

Contrast is a measure of intensity or gray-level variations between the reference pixel and its neighbor. In the visual perception of the real world, contrast is determined by the difference in the color and brightness of the object and other objects within the same field of view.

\[
f_2 = \sum_{n=0}^{N_g-1} \left( \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} pd \cdot \theta(i, j) \right)^{\frac{1}{2}}, \text{ where } n = |i - j| \quad (3-11)\]

(3) Entropy Feature:

The concept comes from thermodynamics; it refers to the quantity of energy that is permanently lost to heat every time a reaction or a physical transformation occurs. Entropy cannot be recovered to do useful work. Because of this, the term can be understood as amount of irreparable chaos or disorder.

(4) Variance Feature:

Variance is a measure of the dispersion of the values around the mean of combinations of reference and neighbor pixels. It is similar to entropy, answers the question ‘What is the dispersion of the difference between the reference and the neighbor pixels in this window.

(5) Correlation Feature:

Correlation feature shows the linear dependency of gray level values in the co-occurrence matrix. It presents how a reference pixel is related to its neighbor, 0 is uncorrelated, 1 is perfectly correlated.

\[
f_5 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} pd \cdot \theta(i, j) \cdot (i - \mu_x) \cdot (j - \mu_y) \frac{(j - \mu_y)}{\sigma_x \sigma_y} \quad (3-12)\]

Where \(\mu_x, \mu_y\) and \(\sigma_x, \sigma_y\) are the means and standard deviations of \(px\) and \(py\).
\begin{equation}
\mu_x = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} i \cdot pd_{i,j}, \theta(i, j), \mu_y = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} j \cdot pd_{i,j}, \theta(i, j) \end{equation}

(3-13)

\begin{equation}
\sigma_x = \sqrt{\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (i - \mu)^2 pd_{i,j}, \theta(i, j)}, \sigma_y = \sqrt{\sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (j - \mu)^2 pd_{i,j}, \theta(i, j)} \end{equation}

(3-14)

(6) Inverse Difference Moment (IDM) Feature

IDM is usually called homogeneity that measures the local homogeneity of an image. IDM feature obtains the measures of the closeness of the distribution of the GLCM elements to the GLCM diagonal.

(7) Sum Average Feature

(8) Sum Variance Feature

(9) Sum Entropy Feature

(10) Difference Variance Feature

(11) Difference Entropy Feature

(12) Information Measures of Correlation Feature

(13) Information Measures of Correlation Feature

To reduce the computational complexity, only some of these features were selected (Contrast, Correlation, Energy, and Homogeneity).
Chapter Four

Results

4.1 Introduction

The methodology described in the previous chapter based on the following discussed techniques: Gray-Level Co-occurrence matrix, median filter, and Fuzzy c-means segmentation. This methodology consist of three phases: pre-processing, segmentation, and feature extraction. Preprocessing for noise removal; segmentation to divide image into segments; feature extraction to extract four texture features.

The matlab software is used to implement the system.

Figure 4-1 system interface
4.2 preprocessing

Median filter metrics

<table>
<thead>
<tr>
<th></th>
<th>PSNR</th>
<th>RMSE</th>
<th>MEA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35.3404</td>
<td>3.53949</td>
<td>0.97918</td>
</tr>
</tbody>
</table>

Figure 4-2 metrics before median filter

<table>
<thead>
<tr>
<th></th>
<th>PSNR</th>
<th>RMSE</th>
<th>MEA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>37.1178</td>
<td>3.53949</td>
<td>0.97918</td>
</tr>
</tbody>
</table>

Figure 4-3 metrics after median filter

The higher PSNR value, lower RMSE value and the lower MAE indicates the better noise removal
4.3 Segmentation

Figure 4-4 Image after applying FCM
### 4.4 Feature extraction

Table 4-1 Features extracted from abnormal images

<table>
<thead>
<tr>
<th></th>
<th>Contrast</th>
<th>Correlation</th>
<th>Energy</th>
<th>Homogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>image1</td>
<td>0.8781</td>
<td>0.8922</td>
<td>0.7499</td>
<td>0.9791</td>
</tr>
<tr>
<td>image2</td>
<td>0.8947</td>
<td>0.8983</td>
<td>0.7313</td>
<td>0.9787</td>
</tr>
<tr>
<td>image3</td>
<td>1.0054</td>
<td>0.8924</td>
<td>0.7133</td>
<td>0.9761</td>
</tr>
<tr>
<td>image4</td>
<td>1.0519</td>
<td>0.8938</td>
<td>0.6965</td>
<td>0.975</td>
</tr>
<tr>
<td>image5</td>
<td>1.1128</td>
<td>0.8928</td>
<td>0.6817</td>
<td>0.9735</td>
</tr>
<tr>
<td>image6</td>
<td>1.197</td>
<td>0.8875</td>
<td>0.6722</td>
<td>0.8781</td>
</tr>
<tr>
<td>image7</td>
<td>1.311</td>
<td>0.8799</td>
<td>0.6616</td>
<td>0.9688</td>
</tr>
<tr>
<td>image8</td>
<td>1.4417</td>
<td>0.8726</td>
<td>0.6472</td>
<td>0.9657</td>
</tr>
<tr>
<td>image9</td>
<td>1.5934</td>
<td>0.8634</td>
<td>0.6337</td>
<td>0.9621</td>
</tr>
<tr>
<td>image10</td>
<td>1.7628</td>
<td>0.8601</td>
<td>0.6035</td>
<td>0.958</td>
</tr>
<tr>
<td>image11</td>
<td>1.8497</td>
<td>0.8602</td>
<td>0.5836</td>
<td>0.956</td>
</tr>
<tr>
<td>image12</td>
<td>1.8613</td>
<td>0.8636</td>
<td>0.5719</td>
<td>0.9557</td>
</tr>
<tr>
<td>image13</td>
<td>1.8658</td>
<td>0.8664</td>
<td>0.5629</td>
<td>0.9556</td>
</tr>
<tr>
<td>image14</td>
<td>1.8303</td>
<td>0.8722</td>
<td>0.5541</td>
<td>0.9564</td>
</tr>
<tr>
<td>image15</td>
<td>1.7916</td>
<td>0.8773</td>
<td>0.5473</td>
<td>0.9573</td>
</tr>
<tr>
<td>image16</td>
<td>1.7318</td>
<td>0.884</td>
<td>0.5397</td>
<td>0.9588</td>
</tr>
<tr>
<td>image17</td>
<td>1.7047</td>
<td>0.8875</td>
<td>0.5342</td>
<td>0.9594</td>
</tr>
<tr>
<td>image18</td>
<td>1.7512</td>
<td>0.8866</td>
<td>0.5247</td>
<td>0.9583</td>
</tr>
<tr>
<td>image19</td>
<td>1.6964</td>
<td>0.8914</td>
<td>0.5211</td>
<td>0.9596</td>
</tr>
<tr>
<td>image20</td>
<td>1.7229</td>
<td>0.8906</td>
<td>0.5169</td>
<td>0.959</td>
</tr>
<tr>
<td>image21</td>
<td>1.7661</td>
<td>0.8889</td>
<td>0.5118</td>
<td>0.9579</td>
</tr>
<tr>
<td>image22</td>
<td>1.7683</td>
<td>0.8896</td>
<td>0.5085</td>
<td>0.9579</td>
</tr>
<tr>
<td>image23</td>
<td>1.8171</td>
<td>0.8871</td>
<td>0.5051</td>
<td>0.9567</td>
</tr>
<tr>
<td>image24</td>
<td>1.8791</td>
<td>0.8841</td>
<td>0.5002</td>
<td>0.9553</td>
</tr>
<tr>
<td>image25</td>
<td>1.8198</td>
<td>0.8886</td>
<td>0.4981</td>
<td>0.9567</td>
</tr>
<tr>
<td>image26</td>
<td>1.9278</td>
<td>0.8827</td>
<td>0.4928</td>
<td>0.9541</td>
</tr>
<tr>
<td>image27</td>
<td>1.9095</td>
<td>0.8846</td>
<td>0.4903</td>
<td>0.9545</td>
</tr>
<tr>
<td>image28</td>
<td>1.9466</td>
<td>0.8826</td>
<td>0.4882</td>
<td>0.9537</td>
</tr>
<tr>
<td>image29</td>
<td>1.951</td>
<td>0.883</td>
<td>0.4856</td>
<td>0.9535</td>
</tr>
<tr>
<td>image30</td>
<td>1.9975</td>
<td>0.8808</td>
<td>0.4823</td>
<td>0.9524</td>
</tr>
<tr>
<td>image31</td>
<td>2.0241</td>
<td>0.8795</td>
<td>0.4804</td>
<td>0.9518</td>
</tr>
<tr>
<td>image32</td>
<td>2.0679</td>
<td>0.8769</td>
<td>0.4792</td>
<td>0.9508</td>
</tr>
<tr>
<td>image33</td>
<td>2.0557</td>
<td>0.8782</td>
<td>0.4775</td>
<td>0.9511</td>
</tr>
<tr>
<td>image34</td>
<td>2.0667</td>
<td>0.8783</td>
<td>0.4743</td>
<td>0.9508</td>
</tr>
<tr>
<td>image35</td>
<td>2.1791</td>
<td>0.8716</td>
<td>0.4716</td>
<td>0.9481</td>
</tr>
<tr>
<td>image36</td>
<td>2.2633</td>
<td>0.8667</td>
<td>0.4696</td>
<td>0.9461</td>
</tr>
<tr>
<td>image37</td>
<td>2.302</td>
<td>0.8648</td>
<td>0.4672</td>
<td>0.9452</td>
</tr>
<tr>
<td>image38</td>
<td>2.3203</td>
<td>0.864</td>
<td>0.4659</td>
<td>0.9448</td>
</tr>
<tr>
<td>image39</td>
<td>2.3773</td>
<td>0.8608</td>
<td>0.4639</td>
<td>0.9434</td>
</tr>
<tr>
<td>Image</td>
<td>Contrast</td>
<td>Correlation</td>
<td>Energy</td>
<td>Homogeneity</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>-------------</td>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>image1</td>
<td>1.9278</td>
<td>0.8635</td>
<td>0.5569</td>
<td>0.9541</td>
</tr>
<tr>
<td>image2</td>
<td>1.9389</td>
<td>0.8691</td>
<td>0.5376</td>
<td>0.9538</td>
</tr>
<tr>
<td>image3</td>
<td>1.7451</td>
<td>0.8797</td>
<td>0.5509</td>
<td>0.9585</td>
</tr>
<tr>
<td>image4</td>
<td>1.7395</td>
<td>0.8827</td>
<td>0.542</td>
<td>0.9586</td>
</tr>
<tr>
<td>image5</td>
<td>2.0009</td>
<td>0.8616</td>
<td>0.546</td>
<td>0.9524</td>
</tr>
<tr>
<td>image6</td>
<td>1.8038</td>
<td>0.8786</td>
<td>0.5398</td>
<td>0.9571</td>
</tr>
<tr>
<td>image7</td>
<td>1.8115</td>
<td>0.8783</td>
<td>0.5387</td>
<td>0.9569</td>
</tr>
<tr>
<td>image8</td>
<td>2.0662</td>
<td>0.8588</td>
<td>0.5395</td>
<td>0.9508</td>
</tr>
<tr>
<td>image9</td>
<td>1.94</td>
<td>0.8696</td>
<td>0.5359</td>
<td>0.9538</td>
</tr>
<tr>
<td>image10</td>
<td>1.9278</td>
<td>0.8635</td>
<td>0.5569</td>
<td>0.9541</td>
</tr>
<tr>
<td>image11</td>
<td>3.1967</td>
<td>0.7874</td>
<td>0.5014</td>
<td>0.9239</td>
</tr>
<tr>
<td>image12</td>
<td>1.9145</td>
<td>0.8733</td>
<td>0.5298</td>
<td>0.9544</td>
</tr>
<tr>
<td>image13</td>
<td>2.1581</td>
<td>0.8549</td>
<td>0.5304</td>
<td>0.9486</td>
</tr>
<tr>
<td>image14</td>
<td>2.0031</td>
<td>0.8657</td>
<td>0.533</td>
<td>0.9523</td>
</tr>
<tr>
<td>image15</td>
<td>1.909</td>
<td>0.8728</td>
<td>0.5331</td>
<td>0.9545</td>
</tr>
<tr>
<td>image16</td>
<td>2.0961</td>
<td>0.8583</td>
<td>0.5342</td>
<td>0.9501</td>
</tr>
<tr>
<td>image17</td>
<td>2.0064</td>
<td>0.8656</td>
<td>0.5327</td>
<td>0.9522</td>
</tr>
<tr>
<td>image18</td>
<td>1.9865</td>
<td>0.8681</td>
<td>0.5294</td>
<td>0.9527</td>
</tr>
<tr>
<td>image19</td>
<td>2.2212</td>
<td>0.8511</td>
<td>0.5277</td>
<td>0.9471</td>
</tr>
<tr>
<td>image20</td>
<td>1.9278</td>
<td>0.8635</td>
<td>0.5569</td>
<td>0.9541</td>
</tr>
<tr>
<td>image21</td>
<td>3.2078</td>
<td>0.7862</td>
<td>0.5021</td>
<td>0.9236</td>
</tr>
<tr>
<td>image22</td>
<td>2.188</td>
<td>0.8482</td>
<td>0.5425</td>
<td>0.9479</td>
</tr>
<tr>
<td>image23</td>
<td>3.2554</td>
<td>0.7864</td>
<td>0.4943</td>
<td>0.9225</td>
</tr>
<tr>
<td>image24</td>
<td>1.7395</td>
<td>0.8827</td>
<td>0.542</td>
<td>0.9586</td>
</tr>
<tr>
<td>image25</td>
<td>3.4115</td>
<td>0.7763</td>
<td>0.4905</td>
<td>0.9188</td>
</tr>
<tr>
<td>image26</td>
<td>3.4492</td>
<td>0.7753</td>
<td>0.487</td>
<td>0.9179</td>
</tr>
<tr>
<td>image27</td>
<td>3.5389</td>
<td>0.765</td>
<td>0.4931</td>
<td>0.9157</td>
</tr>
<tr>
<td>image28</td>
<td>1.8038</td>
<td>0.8786</td>
<td>0.5398</td>
<td>0.9571</td>
</tr>
<tr>
<td>image29</td>
<td>3.4636</td>
<td>0.777</td>
<td>0.4817</td>
<td>0.9175</td>
</tr>
<tr>
<td>image30</td>
<td>1.9344</td>
<td>0.8699</td>
<td>0.536</td>
<td>0.9539</td>
</tr>
<tr>
<td>image31</td>
<td>3.3318</td>
<td>0.7877</td>
<td>0.48</td>
<td>0.9207</td>
</tr>
<tr>
<td>image</td>
<td>Feature 1</td>
<td>Feature 2</td>
<td>Feature 3</td>
<td>Feature 4</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>image32</td>
<td>2.1105</td>
<td>0.8567</td>
<td>0.5356</td>
<td>0.9498</td>
</tr>
<tr>
<td>image33</td>
<td>1.9278</td>
<td>0.8635</td>
<td>0.5569</td>
<td>0.9541</td>
</tr>
<tr>
<td>image34</td>
<td>3.6452</td>
<td>0.7677</td>
<td>0.4731</td>
<td>0.9132</td>
</tr>
<tr>
<td>image35</td>
<td>2.4737</td>
<td>0.8343</td>
<td>0.5212</td>
<td>0.9411</td>
</tr>
<tr>
<td>image36</td>
<td>2.5822</td>
<td>0.8253</td>
<td>0.5229</td>
<td>0.9385</td>
</tr>
<tr>
<td>image37</td>
<td>3.4813</td>
<td>0.779</td>
<td>0.4751</td>
<td>0.9171</td>
</tr>
<tr>
<td>image38</td>
<td>2.56</td>
<td>0.8299</td>
<td>0.5159</td>
<td>0.9390 0</td>
</tr>
<tr>
<td>image39</td>
<td>3.8921</td>
<td>0.75</td>
<td>0.4711</td>
<td>0.9073</td>
</tr>
<tr>
<td>image40</td>
<td>2.6984</td>
<td>0.824</td>
<td>0.5048</td>
<td>0.9358</td>
</tr>
<tr>
<td>image41</td>
<td>3.7592</td>
<td>0.7601</td>
<td>0.4712</td>
<td>0.9105</td>
</tr>
<tr>
<td>image42</td>
<td>1.909</td>
<td>0.8728</td>
<td>0.5331</td>
<td>0.9545</td>
</tr>
<tr>
<td>image43</td>
<td>3.4414</td>
<td>0.78</td>
<td>0.4791</td>
<td>0.9181</td>
</tr>
<tr>
<td>image44</td>
<td>2.6243</td>
<td>0.83</td>
<td>0.5036</td>
<td>0.9375</td>
</tr>
<tr>
<td>image45</td>
<td>3.5965</td>
<td>0.7705</td>
<td>0.4747</td>
<td>0.9144</td>
</tr>
<tr>
<td>image46</td>
<td>2.0396</td>
<td>0.8663</td>
<td>0.5229 0</td>
<td>0.9514</td>
</tr>
<tr>
<td>image47</td>
<td>3.6452</td>
<td>0.7677</td>
<td>0.4731</td>
<td>0.9132</td>
</tr>
<tr>
<td>image48</td>
<td>2.1825</td>
<td>0.854</td>
<td>0.5279</td>
<td>0.948</td>
</tr>
<tr>
<td>image49</td>
<td>3.7681 3</td>
<td>0.7599</td>
<td>0.4703</td>
<td>0.9103</td>
</tr>
<tr>
<td>image50</td>
<td>3.0993</td>
<td>0.7972</td>
<td>0.4968</td>
<td>0.9262</td>
</tr>
</tbody>
</table>

Table 4-2 Features extracted from normal images
4.5 Cluster Analysis

4.5.1 Cluster analysis for abnormal images

Table 4-3 Initial Cluster Centers for abnormal images

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Contrast</th>
<th>Correlation</th>
<th>Energy</th>
<th>Homogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.8781</td>
<td>.8922</td>
<td>.7499</td>
<td>.9791</td>
</tr>
<tr>
<td></td>
<td>2.8380</td>
<td>.8280</td>
<td>.4691</td>
<td>.9324</td>
</tr>
</tbody>
</table>

Table 4-4 Number of Cases in each Cluster- abnormal images

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Number of Cases in each Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29.000</td>
</tr>
<tr>
<td>2</td>
<td>21.000</td>
</tr>
<tr>
<td>Valid</td>
<td>50.000</td>
</tr>
<tr>
<td>Missing</td>
<td>.000</td>
</tr>
</tbody>
</table>

Table 4-5 Final Cluster Centers - abnormal images

<table>
<thead>
<tr>
<th>Final Cluster Centers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast</td>
</tr>
<tr>
<td>Correlation</td>
</tr>
<tr>
<td>Energy</td>
</tr>
<tr>
<td>Homogeneity</td>
</tr>
</tbody>
</table>

4.5.2 Cluster analysis for normal images

Table 4-6 Initial Cluster Centers – normal images

<table>
<thead>
<tr>
<th>Initial Cluster Centers</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast</td>
</tr>
<tr>
<td>correlation</td>
</tr>
<tr>
<td>energy</td>
</tr>
<tr>
<td>homogeneity</td>
</tr>
</tbody>
</table>
Table 4-7 Number of Cases in each Cluster - normal images

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Number of Cases in each Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33.000</td>
</tr>
<tr>
<td>2</td>
<td>17.000</td>
</tr>
</tbody>
</table>

Table 4-8 Final Cluster Centers normal images

<table>
<thead>
<tr>
<th>Final Cluster Centers</th>
<th>contrast</th>
<th>2.0575</th>
<th>3.4267</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>correlation</td>
<td>.8611</td>
<td>.7749</td>
</tr>
<tr>
<td></td>
<td>energy</td>
<td>.5336</td>
<td>.4832</td>
</tr>
<tr>
<td></td>
<td>homogeneity</td>
<td>.9512</td>
<td>.9171</td>
</tr>
</tbody>
</table>
Results

After analysis, it has been concluded that values of GLCM features for abnormal images is as in table 4-9

Table 4-9 GLCM Features For abnormal images

<table>
<thead>
<tr>
<th>Feature</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>0.8822 - 0.8581</td>
</tr>
<tr>
<td>Energy</td>
<td>0.4716 - 0.5768</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>0.9433 - 0.9583</td>
</tr>
</tbody>
</table>

Similarly, the values of GLCM features for normal images is as in table 4-10

Table 4-10 GLCM Features For normal images

<table>
<thead>
<tr>
<th>Feature</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast</td>
<td>2.0575 - 3.4267</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.7749 - 0.8611</td>
</tr>
<tr>
<td>Energy</td>
<td>0.4832 - 0.5336</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>0.9171 - 0.9512</td>
</tr>
</tbody>
</table>
4.7 Performance Analysis

**Performance measures:**

- True positive: Sick people correctly diagnosed as sick (TP)
- False positive: Healthy people incorrectly identified as sick (FP)
- True negative: Healthy people correctly identified as healthy (TN)
- False negative: Sick people incorrectly identified as healthy (FN)

**Sensitivity:**

Relates to the test's ability to identify a condition correctly.

\[ \text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \]

**Specificity:**

Specificity relates to the test's ability to exclude a condition correctly.

\[ \text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \]

**Accuracy (ACC):**

\[ \text{ACC} = \frac{\Sigma \text{True positive} + \Sigma \text{True negative}}{\Sigma \text{Total population}} \]

<table>
<thead>
<tr>
<th>TP</th>
<th>FN</th>
<th>FP</th>
<th>TN</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>0</td>
<td>3</td>
<td>45</td>
<td>1.0</td>
<td>0.93</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Table 4.11 performance Analysis Results

4.8 Comparing results

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>This Study</td>
<td>100</td>
<td>93</td>
<td>89</td>
</tr>
<tr>
<td>Mubashir Ahmad et al(2012)(^{57})</td>
<td>98.46</td>
<td>70</td>
<td>94.7</td>
</tr>
<tr>
<td>Prof. P. Tamije Selvy et al(2013)(^{56})-results for FCM</td>
<td>81</td>
<td>93.7</td>
<td>87.6</td>
</tr>
<tr>
<td>Minakshi Sharma, Dr. Sourabh Mukherjee(2013)(^{60})</td>
<td>96</td>
<td>93</td>
<td>86</td>
</tr>
<tr>
<td>Anamika Ahirwar(2013)(^{61})</td>
<td>72</td>
<td>88</td>
<td>75.5</td>
</tr>
<tr>
<td>SAID CHARFI et al(2014)(^{55})</td>
<td>100</td>
<td>82</td>
<td>90</td>
</tr>
<tr>
<td>A.Prabin and J.Veerappan(2014)(^{58})</td>
<td>92</td>
<td>73</td>
<td>90</td>
</tr>
<tr>
<td>G.THAMARAI SELVI, K.DURAISAMY(2014)(^{59})</td>
<td>66</td>
<td>100</td>
<td>80</td>
</tr>
</tbody>
</table>

Table 4.12 comparing proposed work with some previous work..... From table 4-12 it is obvious that the proposed work has an excellent sensitivity, very good specificity and a good accuracy.
Chapter Five

Conclusion

- Brain tumor, which is one of the most frequent brain diseases, has influenced and destroyed many lives. Accurate detection of the tumor is extremely intrinsical for treatment planning in order to minimize diagnostic errors.
- Radiotherapists use medical images to diagnose diseases precisely. Magnetic resonance (MR) imaging scan the brain to affirm the presence of brain tumor and to identify its location for selected specialist treatment. However, identification of brain tumor from medical images is still a critical and complicated job for Radiotherapists.
- Brain tumor identification form magnetic resonance imaging (MRI) consist of several phases. Segmentation is known to be an essential step in medical image analysis.
- Performing the brain MR images segmentation manually is an onerous task as there are several challenges associated with it. Radiotherapists and medical experts spend lengthy of time for manually segmenting brain MR images. In addition, a problem when performing MRI procedures is that images sometimes are corrupted with noise. This problem make the radiotherapist face difficulties in identifying tumor and may lead to inconvenience in locating the tumor.
- Chapter 1 lay out the foundation of the dissertation as the subject is interdisciplinary and involve many knowledge domains.
- Chapter 2 Survey the tens of recent Previous studies in order to get acquaintance with the subject and develop better understanding of the problem.
- Chapter 3 present a methodology consists of three phases: firstly, pre-processing to perform denoising for the MRI image by applying the median filter. Secondly, the image is segmented into various regions using GLCM. Finally, Texture features are an important characteristic used in identifying regions of interest in an image. Grey Level Co-occurrence Matrices (GLCM) is used for feature extraction.
- Chapter 4 present the implementation of the methodology, and show the results of the segmentation. Cluster analysis is performed for the extracted features and the results show the characteristics for the normal images and abnormal images.
Recommendations

- There is a need for the cooperation of the personnel in the medical centers to provide the data sets for the researchers, and this does not entail the revelation of the personal data of the patient, just the images in a way that provide data for the researchers and keep the privacy of the patient. It is better to get data from local medical center than to probe the web for datasets.

- The study recommend that To build a website or blog to share information and exchange experiences among the researchers in the image mining and medical image mining fields, and to upload image datasets to the blog/website. Furthermore, the website may take advantage of server virtualization technology to provide application over the web for the interested users.

- The study recommend that to call for translation from different languages materials related to image mining and medical image mining, so as to broaden and enrich the knowledge of the researchers.
Chapter Five

Conclusion

- Brain tumor, which is one of the most frequent brain diseases, has influenced and destroyed many lives. Accurate detection of the tumor is extremely intrinsic for treatment planning in order to minimize diagnostic errors.

- Radiotherapists use medical images to diagnose diseases precisely. Magnetic resonance (MR) imaging scan the brain to affirm the presence of brain tumor and to identify its location for selected specialist treatment. However, identification of brain tumor from medical images is still a critical and complicated job for Radiotherapists.

- Brain tumor identification from magnetic resonance imaging (MRI) consist of several phases. Segmentation is known to be an essential step in medical image analysis.

- Performing the brain MR images segmentation manually is an onerous task as there are several challenges associated with it. Radiotherapists and medical experts spend lengthy of time for manually segmenting brain MR images. In addition, a problem when performing MRI procedures is that images sometimes are corrupted with noise. This problem make the radiotherapist face difficulties in identifying tumor and may lead to inconvenience in locating the tumor

- Chapter 1 lay out the foundation of the dissertation as the subject is interdisciplinary and involve many knowledge domains.

- Chapter 2 Survey the tens of recent Previous studies in order to get acquaintance with the subject and develop better understanding of the problem.

- Chapter 3 present a methodology consists of three phases: firstly, pre-processing to perform denoising for the MRI image by applying the median filter. Secondly, the image is segmented into various regions using GLCM. Finally, Texture features are an important characteristic used in identifying regions of interest in an image. Grey Level Co-occurrence Matrices (GLCM) is used for feature extraction.

- Chapter 4 present the implementation of the methodology, and show the results of the segmentation. Cluster analysis is performed for the extracted features and the results show the characteristics for the normal images and abnormal images.
Recommendations

- There is a need for the cooperation of the personnel in the medical centers to provide the data sets for the researchers, and this does not entail the revelation of the personal data of the patient, just the images in a way that provide data for the researchers and keep the privacy of the patient. It is better to get data from local medical center than to probe the web for datasets.

- The study recommend that To build a website or blog to share information and exchange experiences among the researchers in the image mining and medical image mining fields, and to upload image datasets to the blog/website. Furthermore, the website may take advantage of server virtualization technology to provide application over the web for the interested users.

- The study recommend that to call for translation from different languages materials related to image mining and medical image mining, so as to broaden and enrich the knowledge of the researchers.
References


[4] Jiawei Han, Micheline kamber, Jang Pei, (2012). Data mining concepts and techniques. 3rd ed. USA: Elsevier Inc.


