Automated Detection and Grading of

Brain Tumors using Particle Swarm Optimization

by

SATISH CHANDRA

A thesis submitted in fulfillment of the requirement for the degree of

DOCTOR OF PHILOSOPHY

IN

COMPUTER SCIENCE & ENGINEERING

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
WAKNAGHAT-173 215, H.P. (INDIA)

December 2010

Certificate

This is to certify that the thesis entitled "***Automated Detection and Grading of Brain Tumors using Particle Swarm Optimization***" and submitted by Mr. Satish Chandra to Jaypee University of Information Technology, Waknaghat, HP, India, as a requirement for award of degree of ***Doctor of Philosophy in Computer Science***, is a record of bona fide research work carried out by him under our guidance and supervision and no part of this work has been submitted for any other degree or diploma.

Dr. Rajesh Bhat
Indian Institute of Technology
New Delhi
India

Prof. D. S. Chauhan
Vice Chancellor
Uttarakhand Technical University
Dehradun, Uttarakhand, India

Prof. Harinder Singh
Jaypee University of Information Technology
Waknaghat, Solan, Himachal Pradesh
India

Dated: December 07, 2010.

Acknowledgments

First of all, my abeyance and sincere gratitude to *The Holy Mother Mahamaya* who is always protecting. May your name be exalted, honoured and glorified. Next, I would like to show my gratitude to all people who have helped and inspired me during my doctoral study.

My deepest gratitude to my supervisor Dr. Rajesh Bhat who, despite his extremely hectic schedule at Indian Institute of Technology, Delhi, was always available to me with his continuous guidance, valuable suggestions, and encouragement. I enjoyed working with him, making every moment spent in the research work, as enjoyable as can be imagined. I would like to express my warm thanks to my supervisor Professor Harinder Singh who spared no effort in supporting me with all care and patience. His perpetual energy and enthusiasm always motivated me to work hard. In addition, he was always accessible and willing to help me in the preparation and presentation of all the reports during my research. I would also like to thank Prof. Durg Singh Chauhan, Vice Chancellor, Uttarakhand Technical University for making the medical information and contacts available.

I am highly indebted to Dr.Yajulu Medury, Hon'ble Vice Chancellor, Jaypee University of Information Technology for his support and blessings since the day I joined the University. I would also like to place on record my hearty thanks to the Registrar of Jaypee University of Information Technology Brigadier (Retd.) Balbir Singh for his enormous cooperation. Next, my sincere thanks to Brig.(Retd.) S. P. Ghrera, Head, Department of Computer Science & Engineering and Information Technology for inspiring me in research through our interactions over long discussions.

I am indebted to all my colleagues at JUIT for their love and encouragement. In particular, I would like to thank my friend Dr. Nitin for his friendship and help in the past six years. All my lab staff at Jaypee University of Information Technology made it a convivial place to work.

This thesis would never have been complete without the blessings of my parents. I know it could take a million ages for me to give back what they have done for me. Whatever my goals and ambitions in my life I always want them to feel happy and proud about me being their son. I would also like to thank my in laws for their good wishes and kind motivation. My children Akshat and Mauli need special mention here for sacrificing a lot as they could not get the attention that they deserved from me during all these years. Last but not the least; I thank my wife Priyadarshini for her support and care, who stood by me in all thick and thin.

To all acknowledged I solemnly owe this work.

Place: Waknaghat, HP, India Satish Chandra

Date:

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, the objective of the research work is set and a brief description of anatomy of Human Brain with characteristic features of brain tumor is presented. Later, Magnetic Resonance Imaging (MRI) technique and motivations behind this thesis is discussed briefly. Finally the contribution and outline of the entire thesis is presented.

1.1 Problem Definition

There has been great impact on the field of medical imaging by the advancements in the computer technology as new and improved techniques of data acquisition, analysis, processing and visualization have evolved. Magnetic resonance images (MRI) provides information about potential abnormal tissues necessary for medical follow up. Brain MRI gets additional importance in medical science as it is the only preliminary method of diagnosing a brain tumor. The available and upcoming techniques of Artificial Intelligence can assist the medical practitioners in deciding whether an evident anomaly in a human Brain MRI is a tumor or not. This thesis addresses the problem of automated detection and grading of brain tumors and tries to find a novel yet efficient solution for it. The formal problem definition is enlisted below:

a) To explore possibilities of developing a robust algorithm for classification of Brain MRIs as normal or abnormal based on the absence or presence of tumor of any grade.

b) To determine the grade of tumor present in abnormal Brain MRIs.

c) To study comparatively the applicability and implementation of various medical image segmentation and classification techniques and finally apply the best one.

1.2 Human Brain: Anatomy and Anomalies

1.2.1 Brain Anatomy

Before dealing with the brain tumors and MR Images, the basic structure of human brain must be well understood. Following are the main parts of human brain [Martin 2003, Duvernoy and Vannson 1999]:

a) Brainstem – It is the lower extension of the brain where it connects to the spinal cord. Neurological functions located in the brainstem include those necessary for survival (breathing, digestion, heart rate, blood pressure) and for arousal (being awake and alert). The brainstem is the pathway for all fiber tracts passing up and down from peripheral nerves and spinal cord to the highest parts of the brain.

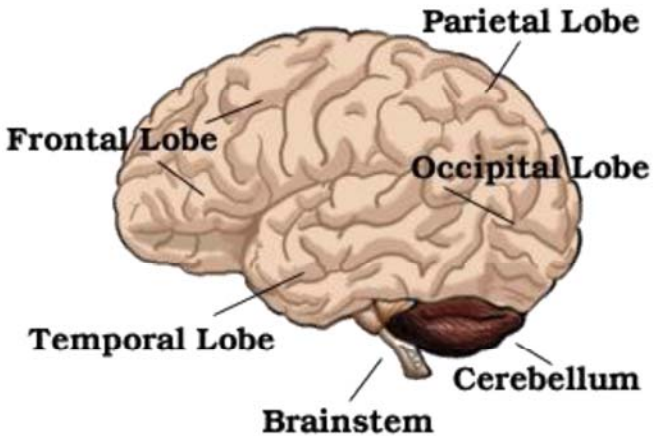

Fig.1.1. Parts of human brain

b) Cerebellum - The portion of the brain (located at the back) which helps coordinate movement (balance and muscle coordination).

c) Frontal Lobe – It is the front part of the brain and is involved in planning, organizing, problem solving, selective attention, personality and a variety of "higher cognitive functions" including behavior and emotions.

d) Occipital Lobe – It is the region in the back of the brain which processes visual information. It also contains association areas that help in the visual recognition of shapes and colors.

e) Parietal Lobe - One of the two parietal lobes of the brain located behind the frontal lobe at the top of the brain.

f) Temporal Lobe - There are two temporal lobes, one on each side of the brain located at about the level of the ears. These lobes allow a person to tell one smell from another and one sound from another. They also help in sorting new information and are believed to be responsible for short-term memory.

Most of times in adults tumors develop in the cerebral cortex, area of the brain that has a role in memory, thought, and more, whereas in case of children tumors often develop in the brain stem and cerebellum, which is located near the brain stem and affects movement and coordination.

1.2.2 Brain Tumors: Features and classification

A brain tumor is an abnormal growth of tissue in the brain. Unlike other tumors spread by local extension and rarely metastasize (spread) outside the brain. A brain tumor

develops when abnormal cells multiply for unknown reasons. Glioma is the term used to refer to the most prevalent primary brain tumors. Gliomas arise from glial tissue, which supports and nourishes cells that send messages from the brain to other parts of the body. These tumors may be either malignant or benign [Greenberg and Chandler 1999]. Astrocytomas, ependymomas, and mixed gliomas are three of the most common gliomas.

a) Benign and Malignant brain tumors

Benign tumors are composed of harmless cells and have clearly defined borders. They can usually be completely removed, and are unlikely to recur. Benign brain tumors do not infiltrate nearby tissues but can cause severe pain, permanent brain damage, and death. Malignant brain tumors do not have distinct borders. They tend to grow rapidly, increasing pressure within the brain and can spread in the brain or spinal cord beyond the point where they originate. It is highly unusual for malignant brain tumors to spread beyond the Central Nervous System (CNS).

b) Primary and Secondary brain tumors

Primary brain tumors originate in the brain. They represent about 1% of all cancers and 2.5% of all cancer deaths. Approximately 25% of all cancer patients develop secondary or *metastatic brain tumors* when cancer cells spread from another part of the body to the brain.

c) Naming and grading brain tumors

The name of a brain tumor describes where it originates, how it grows, and what kind of cells it contains. A tumor in an adult is graded or staged according to:

- how malignant it is
- how rapidly it is growing and how likely it is to invade other tissues

- how closely its cells resemble normal cells. (The more abnormal a tumor cell looks, the faster it is likely to grow)

Low-grade brain tumors usually have well-defined borders. Some low-grade brain tumors form or are enclosed (encapsulated) in cysts. Low-grade brain tumors grow slowly, if at all. They may spread throughout the brain, but rarely metastasize to other parts of the body.

Mid-grade and high-grade tumors grow more rapidly than low-grade tumors. Described as "truly malignant," these tumors usually infiltrate healthy tissue. The growth pattern makes it difficult to remove the entire tumor, and these tumors recur more often than low-grade tumors. A single brain tumor can contain several different types of cells. The tumor's grade is determined by the highest-grade (most malignant) cell detected under a microscope, even if most of the cells in the tumor are less malignant. An infiltrating tumor is a tumor of any grade that grows into surrounding tissue. World Health Organization grading system classifies the brain tumors on the basis of rate of growth into four categories, grade I, II, III and IV [Chaa 2006]. Grade I tumors are the least malignant and grow slowly. But even a grade I tumor may be life-threatening if it is inaccessible for surgery. Grade II tumors grow slightly faster than grade I tumors and have a little abnormal microscopic appearance. These tumors may invade surrounding normal tissue, and may recur as a grade III or higher tumor. Grade III tumors are malignant. The chances of recurrences of these tumors are quite high. Grade IV tumors are the most malignant and invade wide areas of surrounding normal tissue.

1.3 Magnetic Resonance Imaging

Great convenience has been brought by modern clinical imaging techniques in studying anatomy and in making screening, prognosis, diagnosis and examination of disorders or diseases present in human body. Clinical imaging techniques can be categorized on two basis:

a) According to the target position, medical imaging can be classified into cerebral imaging, cardiac imaging, thoracic/lung imaging, liver imaging, bone / arthral imaging, vascular imaging and so on; and

b) According to imaging modality, medical imaging can be categorized as one of the following: Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), Functional Magnetic Resonance Imaging (fMRI), Ultrasound, Diffusion Tensor Imaging (DTI) etc.

1.3.1 Working of MRI Machines

Due to the importance of brain in human body, brain imaging has always been one of the focuses of attention for researchers in medical imaging. MRI machines use magnetetic and radio waves to look inside the human body. It provides an unparalleled view inside the human body. The level of detail we can see is extraordinary compared with any other imaging modality. MRI is a popular method for the diagnosis of many types of injuries and conditions because of the incredible ability to tailor the examination to the particular medical question being asked. The basic idea of MRI is to calculate the signal changes of protons caused by a strong external magnetic field and low-energy radio frequency signal.

In MRI, only the proton in hydrogen interacts with the magnetic field. So, the hydrogen is also termed as the proton which moves in a wobbling fashion in a static magnetic field. This movement makes the proton like a tiny magnet; therefore it aligns to the direction of the magnetic field in a high external static magnet. The direction of the precession is parallel to the magnetic field and the frequency of the precession is determined by Larmor equation [Bargmann *et al.* , 1959]. The units for measuring the strength of magnetic field are Tesla and gauss. Currently, 1.5-Tesla is the gold standard for the clinic diagnosis. Magnetic fields greater than 2 Tesla have not been approved for use in medical imaging.

1.3.2 Directions in MRI

In MR imaging domain, the plane that parallels to the z axis is called longitudinal plane, and the one consists of x-y axis is called transverse plane. The orientation of x, y, and z axes are defined as:

z axis - from patients' head to feet.

x axis - from patients' back to front.

y axis - from patients' left arm to right arm.

When a patient is placed in a strong magnetic field, the protons in the body precess along z direction with a precession frequency determined by the Larmor equation. To generate a magnetic resonance signal, a RF pulse with the same precession frequency is transmitted to the protons. The protons absorb the transmitted energy and are forced to

precess in the direction perpendicular to z direction. After that, the RF pulse is turned off and the protons begin to realign to the z direction (longitudinal plane). During this realignment period, the protons will release the absorbed energy in the form of radio frequency. These radio frequencies, or signals, are received by a RF receiver and sent to a computer to reconstruct the final images. The period during which the protons release the energy is called "relaxation". Generally, there are two kinds of relaxation times can be measured, *i.e.* T1 relaxation time and T2 relaxation time.

1.3.3 T1 and T2 relaxation time

The time that the protons release the absorbed energy to their adjacent tissues (lattice), and realign to z-direction (longitudinal plane), is called T1 relaxation time, or spin-lattice/longitudinal relaxation time. Typically, this time ranges from 0.1 to 2 second.

T2 relaxation time measures the time over which the protons release the absorbed energy to their surrounding protons (spins) in the x-y plane, is called T2 relaxation time, or spin-spin/transverse relaxation time. This time is usually shorter than or equal to T1 relaxation time. Since different tissues in the human body have different T1/T2 relaxation times, the records of these times can reconstruct the image of the corresponding tissue. Most of MRI has three kinds of image contrasts, *i.e.* T1, T2, and Proton Density (PD)-weighted images. PD-weighted images are generated by calculating the number of protons per unit of tissue.

1.3.4 Planes in MRI imaging

In MR imaging, the images are normally acquired in three planes, known as the sagittal plane, the horizontal (axial) plane, and coronal plane as shown in Figure 1.2.

Figure 1.3, 1.4 and 1.5 shows the T1-weighted brain images generated in these three planes, respectively.

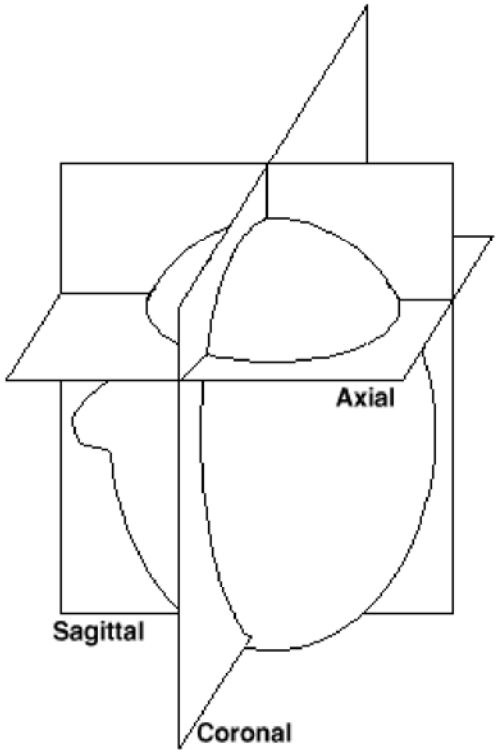

Fig.1.2. The spatial demonstration of three planes for MRI (a) Coronal plane (b) Axial plane (c) Saggital Plane

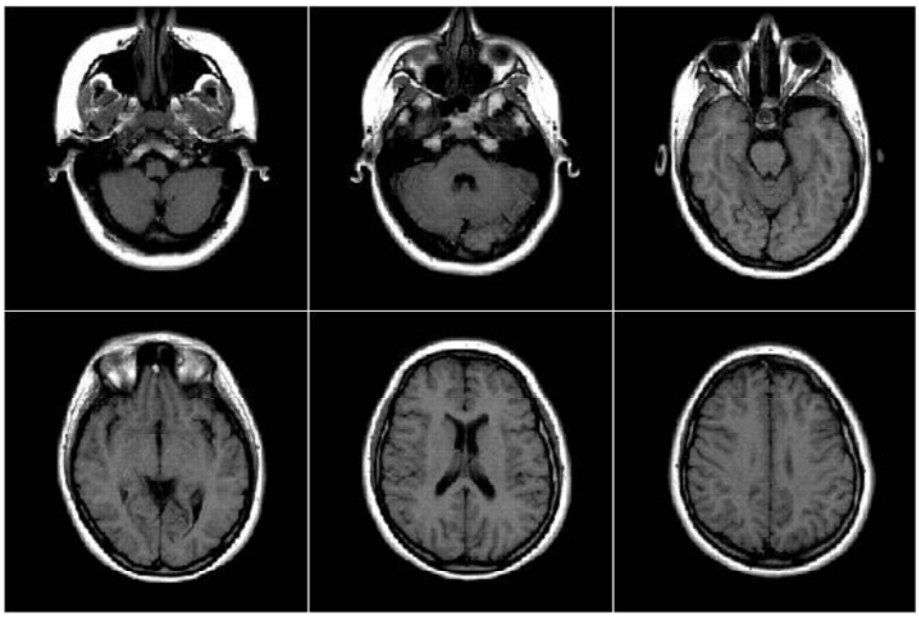

Fig.1.3. Axial T1 MRI of Normal Brain

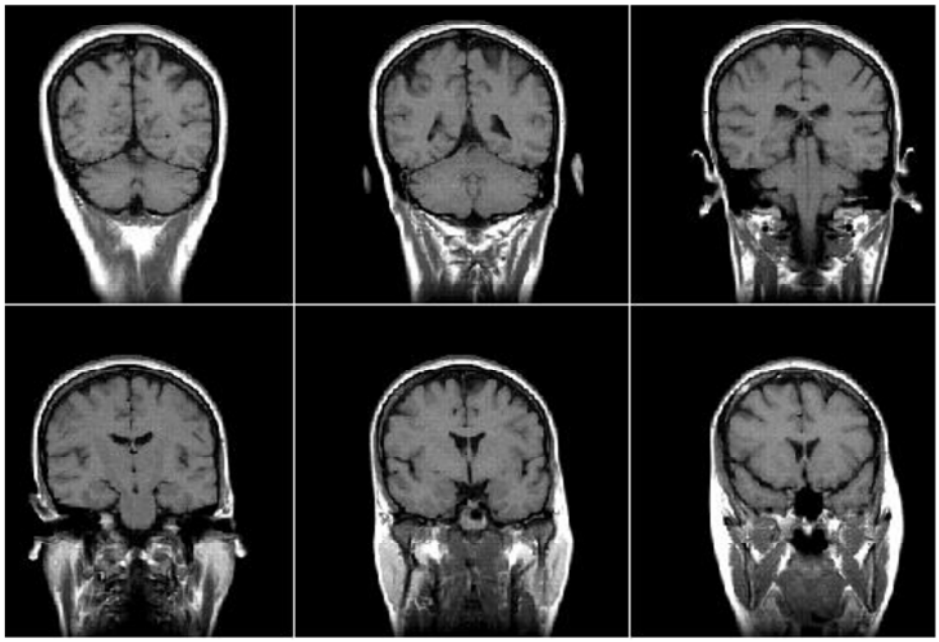

Fig.1.4. Coronal T1 MRI of Normal Brain

Figure 1.6 shows a slide of MRI with tumors. The tumor can be seen as the small white region on the lower right side of the slices.

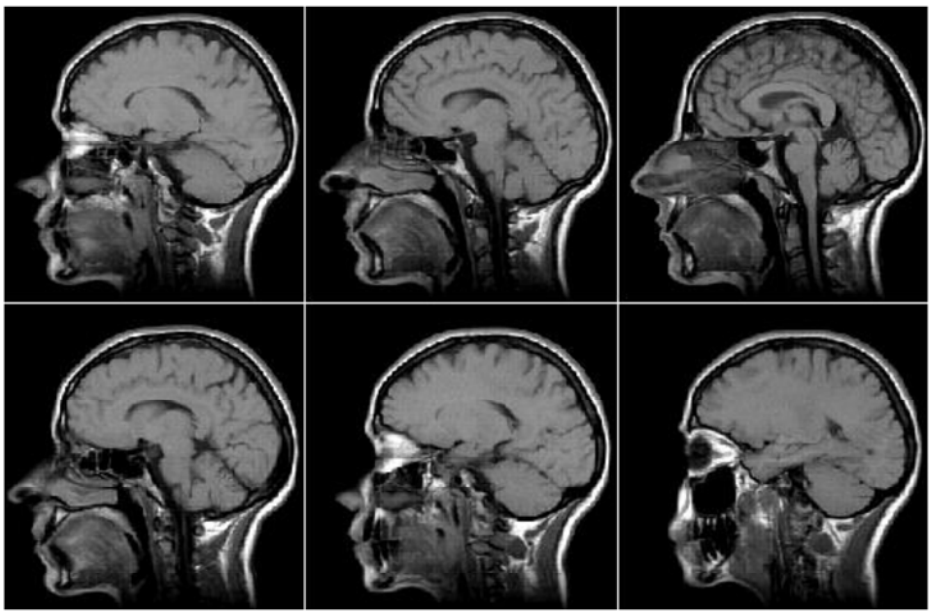

Fig.1.5. Saggital T1 MRI of Normal Brain

1.3.5 The advantages and disadvantages of MRI

Following are the advantages of MRI that make it the most preferable image modality:

1.) High spatial resolution

2.) Exceptional description of soft tissues.

3.) Functional brain measurement

4.) No hazard for human body

5) By changing examination parameters, the MRI system can cause tissues in the body to take on different appearances. This is very helpful to the in determining if something seen is normal or not. We know that when we do "A," normal tissue will look like "B" -- if it doesn't, there might be an abnormality.

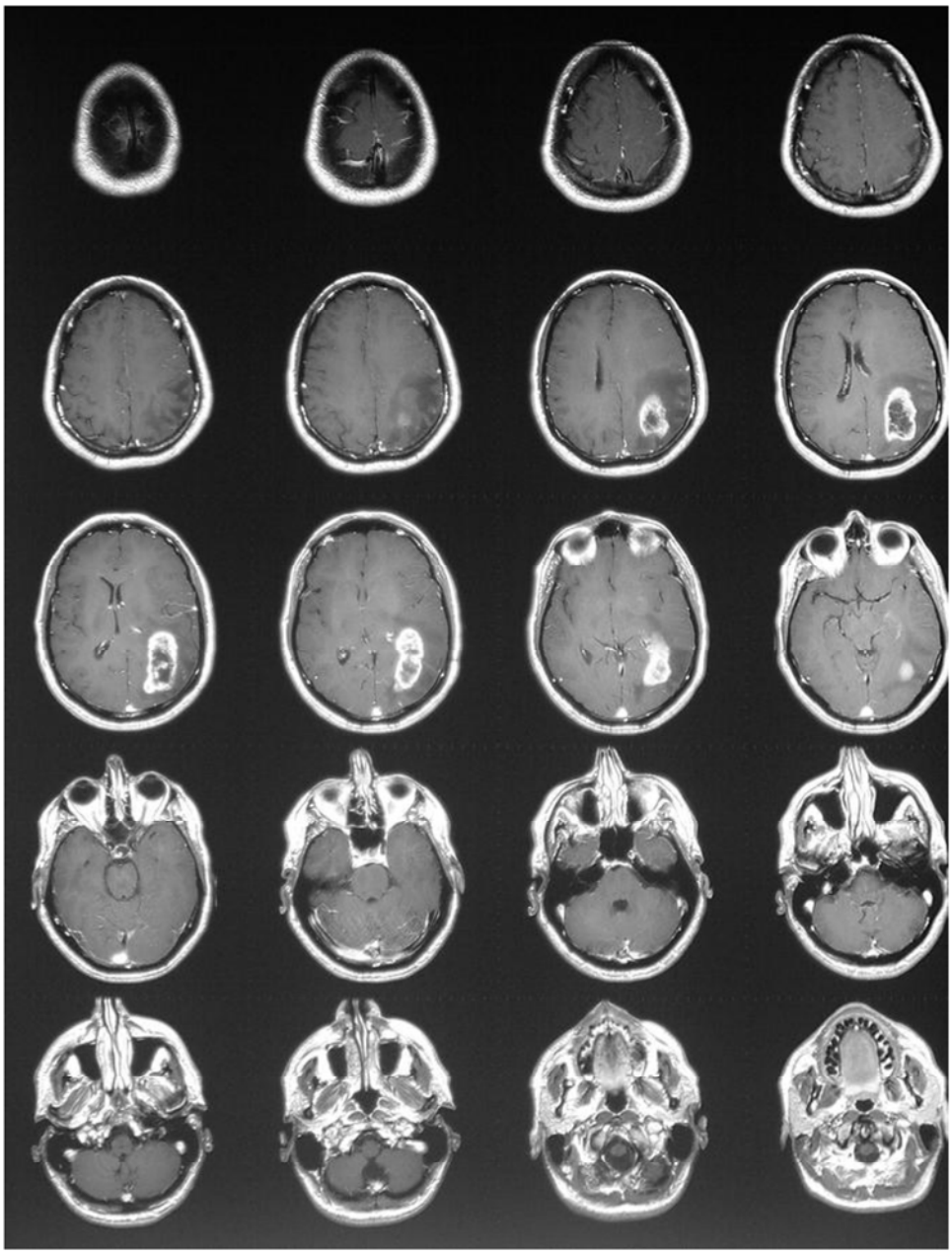

Fig.1.6. A slide of MRI of Brain with tumor

1.4 Motivations

The properties of the problem of automatic tumor detection make it an excellent research challenge in the fields of medical image analysis and Pattern Recognition, in general. The main motivations for this research work are as follows:

i) The manual segmentation by qualified professionals has two major drawbacks. The first drawback is that producing manual segmentations or semi-automatic segmentations is extremely time consuming, with higher accuracies on more finely detailed volumes demanding increased time from medical experts. The second problem with manual and semiautomatic methods is that results are subject to variations. This is due to the fact that several anomalies in brain appear similar in the MRI. A recent study quantified an average of 28% ± 12% variation in quantified volume between individuals performing the same brain tumor segmentation task (the variation ranged from 11% to 69%), and quantified a 20% ± 15% variation within individuals repeating the task three times at 1 month intervals [Mazzara *et al.* 2004].

ii. Accurate automatic segmentation methods could also lead to new applications, including effective content based image retrieval in large medical databases. This could allow clinicians to find similar images in historical data based on tumor location, grade, size, enhancement, extent of edema, similar patterns of growth, or a variety of other factors.

1.5 Methodology

There are various forms of medical image analysis. Well-known examples include segmentation, registration, reconstruction, classification, validation, visualization, interaction, simulation, etc. The methodologies used for the detection of brain tumors in this thesis are image segmentation and classification. Medical image segmentation means to divide an image into non-overlapping regions that belong to meaningful objects such as tissues, organs, anatomical structures, cell colonies and so on. Image segmentation is a fundamental process in several image processing and computer vision applications. It can be considered as the first low-level processing step in image processing and pattern recognition [Cheng *et al.* 2001].

Image segmentation divides an image into disjoint homogenous regions. These homogenous regions should represent objects or parts of them [Lucchese and Mitra 2001]. The homogeneity of the regions is measured using some image property (e.g. pixel intensity) [Jain *et al.* 1999].

Brain image segmentation can be difficult because the target object may be indistinctive and is hard to extract from the surroundings. Some examples of obstacles to segmentation are noise corruption, obscure boundaries, interference from irrelevant objects or background clutter, fusion of target object with other objects, imaging artifacts, etc. Automatic brain segmentation is further compromised by variation of normal and pathologic anatomy, different image qualities and characteristics resulting from various acquiring protocols and machine specifications. Therefore, we need sophisticated

segmentation methods to tackle the problem of cerebral segmentation. There are many techniques for image segmentation in the literature. In general, these techniques can be categorized into thresholding, edge-based, region growing and clustering techniques. These methods are detailed in the next chapter.

Models proposed in this thesis are first presented and discussed. Experimental results were then obtained using various MR images with known characteristics in order to show the accuracy and efficiency of the proposed algorithms. The results of *state-of-the-art* algorithms when applied to the same test images were also reported to show the relative performance of the proposed approaches when compared to other well-known approaches. Due to the stochastic nature of the proposed algorithms, all the presented results are averages and standard deviations over several simulations. However, due to the computational expensive nature of the simulations, results were generally taken over 10 or 20 runs.

1.6 Challenges in Automatic Detection

In this section we try to explore why brain tumor detection is more challenging than other cancers such as mammograms. The automated detection of abnormalities in MR images is a complicated task. This is due to the following problems With respect to the MR imaging modality; this thesis will focus on five problems that can complicate the segmentation task:

1. Local Noise

2. Partial Volume Averaging

3. Intensity Inhomogeneity

4. Inter-slice Intensity Variatoins

5. Intensity Non-Standardization

The differences in the intensities of MR images are present even in controlled settings. MRIs lack standard interpretation of image intensities. Unlike in other modalities, MR images taken for the same patient on the same scanner at different times may appear different from each other due to a variety of scanner-dependent variations and, therefore, the absolute intensity values do not have a fixed meaning. [Nyul *et al.* 2000]. The acquisition of MR images is therefore not a calibrated measure, and the intensities represented in the image do not have an exact meaning with respect to the underlying tissue [Clatz *et al.* 2004]. This variation can cause major problems in intensity based segmentation methods, since differences in a wide variety of factors can lead to different observed intensity distributions.

Local noise, partial volume averaging, intensity inhomogeneity, and inter-slice intensity variations can be taken care of by preprocessing or postprocessing steps. However, intensity nonstandardization represents a major problem in the quantitative analysis of MR images. Despite the presence of the above challenges for automatic segmentation of MR images, effective methods for the automatic segmentation of normal brains into different normal brain tissue classes exist. The design and selection of algorithms in the classification or segmentation is one of the major challenges. In next chapter an exhaustive literature survey has been done to find out the algorithms which have been used so far, for the purpose of Brain MRI Segmentation.

1.7 Thesis Outline:

This thesis comprises of the following chapters:

Chapter 2 presents an extensive report on historical perspective of this problem.

Chapter 3 details the outline of automated detection and highlights the approach followed for each step in the process.

Chapter 4 of the thesis presents few standard approaches for the MR image classification.

Chapter 5 introduces Particle Warm Optimization (PSO). It also presents a survey of a few versions of PSO and experiments with some of them.

Chapter 6 presents three methods based on Particle Warm Optimization (PSO) for MR image segmentation. Finally, an approach for grading of tumors is also presented in this chapter.

1.8 Contributions

The main contributions of this study are:

a) Possibilities of developing a robust algorithm have been explored, for classification of Brain MRIs as normal or abnormal based on the absence or presence of tumor of any grade.

b) Efficient algorithm for image segmentation and classification algorithm based on the PSO has been developed for brain tumor detection from MRI, which can be used for other diseases as well, with little or no modification.

Chapter 2

Related Works

2.1 Introduction

The important forms of medical image analysis are segmentation, registration, reconstruction, classification, validation, visualization, interaction, simulation, etc. Out of these, image segmentation is the most popular technique. The methodologies used for the detection of brain tumors in this thesis are image segmentation and classification. Medical image segmentation means to divide an image into non-overlapping regions that belong to meaningful objects such as tissues, organs, anatomical structures, cell colonies and so on. Image segmentation is a fundamental process in several image processing and computer vision applications. Brain tumor detection has been a challenge in the field of brain computer-aided diagnosis. A large amount of research effort has been focused on the segmentation of images of the brain in MR images. Segmentation is an important process in most medical image classification and analysis for radiological evaluation or computer-aided [Dhawan 1990] diagnosis. It is the process of separating out mutually exclusive homogeneous regions of interest. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze [Shapiro and Stockman 2001]. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images.

The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s). This chapter provides a survey of many of the proposed approaches for automatic brain tumor segmentation in MR images. The first two types of methods we examine are unsupervised and supervised methods. These methods do not incorporate spatial registration. The difference between these two is that supervised methods make use of training data that has been manually labeled, while unsupervised methods do not. Figure 2.1 summarizes the main methods of image segmentations under study in this chapter.

2.2 Unsupervised segmentation

In unsupervised segmentation the computer selects natural groupings of pixels based on their spectral properties. An unsupervised classification algorithm still requires user interaction; however, these occur after the classification has been performed. In unsupervised classification the user attempts to assign information classes to the spectral classes the computer has created. Following are the main categories of unsupervised image segmentation:

2.2.1 Thresholding Techniques :

Thresholding [Gonzalez and Woods 1992; Jain et al. 1995] is the simplest image segmentation technique. In its simplest version an image is divided into two segments:

object and background by specifying a threshold. A pixel above the threshold is assigned to one segment and a pixel below the threshold is assigned to the other segment. For more sophisticated images multiple thresholds can be used.

2.3.2 Edge-based Techniques

In edge-based techniques [Gonzalez and Woods 1992; Jain et al. 1995; Kwok and Constantinides 1997], segmentation is achieved by finding the edges of the regions.

2.2.3 Region growing Techniques

In region growing [Gonzalez and Woods 1992; Jain et al. 1995; Fuh et al. 2000], a set of seed pixels are chosen. Neighboring pixels of a seed are agglomerated if they satisfy a homogeneity criterion. This is repeated until no more pixels can be added to the region. This approach has following problems [Turi 2001]:

- The selection of the seed pixels which is not a straightforward task.
- The selection of the homogeneity criterion.

Region splitting and merging divide the image into regions. A region is then split if it does not satisfy a homogeneity condition. Regions can also be merged if their merging results in a region that satisfies some condition. This is repeated until no more splitting and merging can occur [Gonzalez and Woods 1992].

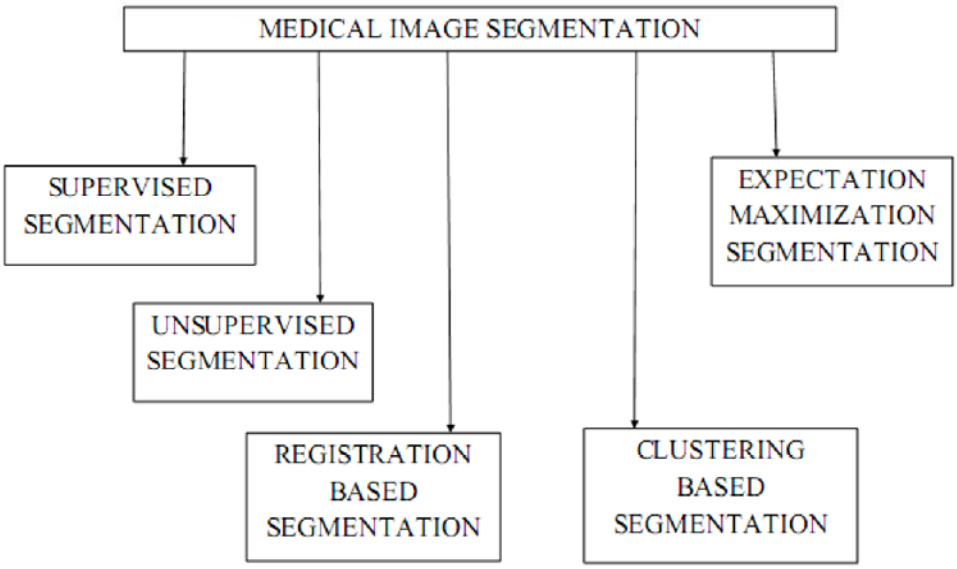

Fig.2.1. Important methods of Medical Image Segmentation

An unsupervised approach for the segmentation of enhancing tumor pixels from T1-weighted post-contrast images was proposed by Gibbs *et al.* [1996]. This system first applied an intensity threshold to a manually selected region of interest and then used a region growing algorithm to expand the thresholded regions up to the edges defined by a Sobel edge detection filter. A similar approach was proposed by Zhu and Yan [1997] for the segmentation element of their enhancing tumor boundary detection approach. These methods represent a clearly justified approach for segmenting image objects that are different in intensity than their surroundings. A fully unsupervised approach for tumor segmentation was proposed by Ho *et al.* [2002]. This system used both the T1- weighted pre-contrast and the T1-weighted post-contrast images as input, and the first step in this system was the coregistration of these two volumes. Coregistration refers to the spatial alignment of two volumes that may not be of the same modality, but that represent a

(potentially unaligned) measurement of the same underlying object. After this alignment step, an image was computed that represented the difference between the T1-weighted images before and after the injection of the contrast agent. A Mixture Model was then applied to the histogram of this difference image. This method was not subject to many of the disadvantages of the earlier methods.

The advantages of this system are a) use of a Mixture Model allows the technique to adaptively find the enhancing area and is thus more robust to differences in intensity between images and b) non-enhancing areas surrounded by enhancing areas will be included in the segmentation through the use of the active contour. Some more systems based on the similar approach were presented by Yoon *et al.* [1999], Gosche *et al.* [1999], Fletcher-Heath *et al.* [2001] etc.

2.3 Supervised Segmentation

The classification problem formulation is a popular method to perform image segmentation using a supervised approach. The task in classification is to assign a class, from a finite set of classes, to an entity based on a set of features. Supervised classification involves a training phase that uses labeled data to learn a model that maps from features to labels, and a testing phase that is used to assign labels to unlabeled data based on the measured features. While many unsupervised approaches also use these two phases, the use of labeled data in the training phase of supervised approaches forces the model to focus on making discriminations in the feature space that correspond to the desired semantic discriminations.

Labels, as normal or tumor are used as classes in the segmentation task. The training phase under this formulation would consist of learning a model that uses the MR image intensities to discriminate between normal and tumor pixels. The testing phase would consist of the use of this model to classify unlabeled pixels into one of the two classes based on their intensities. A major advantage of using a supervised formulation is that supervised methods can perform different tasks simply by changing the training set. The supervised approaches that have been used in medical imaging include kNN, Artificial Neural Networks (ANN), Maximum Likelihood classifier (ML), Decision Tree Classifier, Support Vector Machines (SVM) etc. Figure 2.2 shows the basic outline of a supervised learning framework.

A comparison between Maximum Likelihood (ML) and Artificial Neural Network (ANN) was done by Clarke [1991] for brain tumor segmentation in MR images. It was found that the ANN performed better than the ML approach. Training ML classifiers consists of optimizing the parameters of an assumed model of the features (often assuming a parametric model such a univariate or multivariate Gaussian) and assigning pixels to the class that they are statistically most likely to belong to, based on these models. On the other hand, ANN approaches 'feed' the features through a series of nodes, where mathematical operations are applied to the input values at each node and a classification is made at the final output nodes.

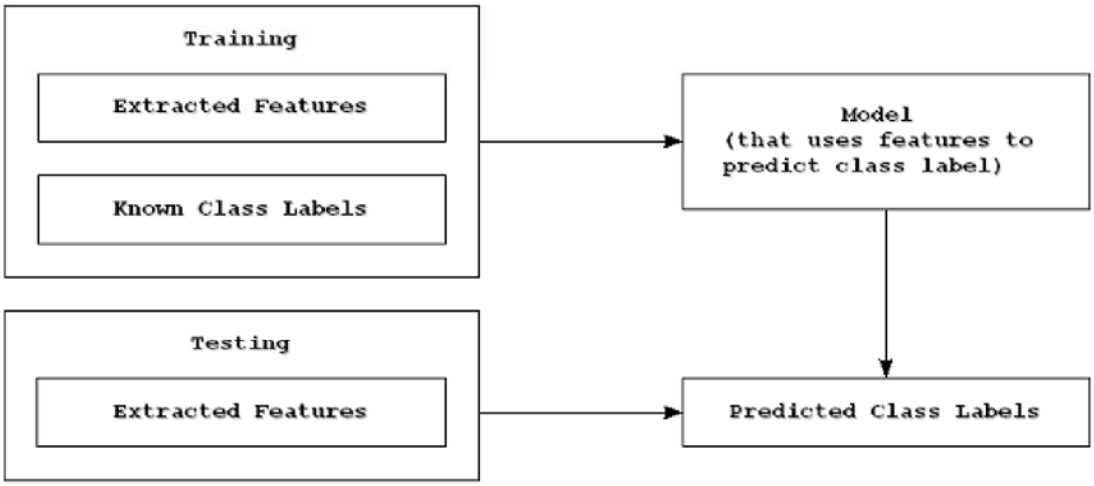

Fig.2.2. Supervised Learning Framework: In the training phase labeled data is used to make a model for classification, whereas in testing phase this model is used to predict labels of unlabelled images.

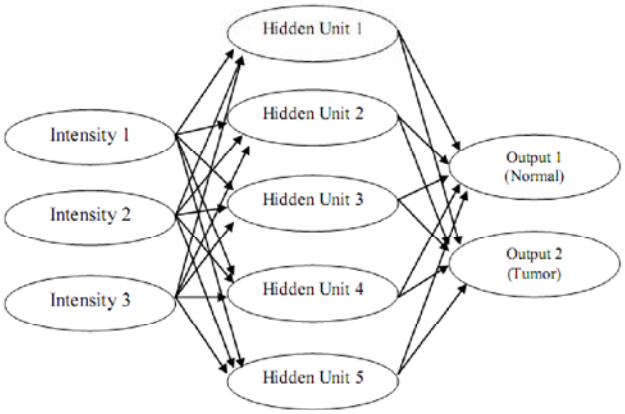

Fig.2.3. Architecture of a Neural Network: Multi-spectral intensities represent the input to this network, linear combinations of the intensities; the output node values are formed from linear combinations of the results of the hidden layer transformations.

Figure 2.3 shows the overall architecture of ANN. Training for these models consists of determining the values of the parameters for the mathematical operations such

that the error in the predictions made by the output nodes is minimized. ANN approaches are non-parametric techniques and, with the use of 'hidden' layers of nodes, allow the modeling of non-linear dependencies in the features. In another system developed by Ozkan *et al.* [1993] pixel intensities were used in the different channels and patient-specific training was done. This work also confirmed that Neural Networks outperformed Maximum Likelihood methods.

In a different work by Schad *et al.* [1993] a Decision Tree classifier was used based on first-order and second-order texture features, that is statistical moments and spatial co-occurrence features, respectively. Decision Trees are a popular classification technique due to their ability to model non-linear dependencies in the features, and their intuitive graphical representation of the learned model (as opposed to, for example, ANNs). Decision Trees perform classification by making a set of decisions based on the features, beginning from a root 'node' and following decision made to other nodes in the tree where new decisions are made, leading finally to a 'leaf" where a classification is made. There are many methods of automatic Decision Tree learning, including the popular C4.5 classifier [Quinlan, 1993].

Vinitski *et al.* [1997] also presented a supervised method that addressed several issues previously ignored in most automatic systems for tumor segmentation. This method used several preprocessing steps before the classification in order to improve results. These steps were step coregistration, anisotropic diffusion filter and intensity inhomogeneity correction. The classifier used was a k-Nearest Neighbors (kNN) classifier, that assigns labels to pixels based on the most frequent label among the k closest training points under

a distance metric applied to the features (referred to as 'lazy' learning, since no explicit model is learned). The kNN algorithm is a simple and effective method for multi-class classification that is able to model non-linear distributions. Disadvantages of the kNN algorithm include the dependence on the parameter k, large storage requirements (the model consists of all training points), sensitivity to noise in the training data, and the undesirable behavior that can occur in cases where a class is underrepresented in the training data.

To overcome the need of patient - specific training in supervised methods, a method was proposed by Dickson and Thomas [1997] which used a set of 50 hand-labeled MR slices from the same area of the head of different patients with brain tumors, and learned to automatically label this without patient specific training. The features used in this system included not only the pixel intensities, but the intensities of neighboring pixels and the pixel's location within the image. This work compared the use of a kNN classifier, a Learning Vector Quantization (LVQ) classifier, and an ANN. The comparative studies done in this work have provided valuable insights into the problem. These results indicated that a) the ANN outperformed the other two methods, b) pixel neighborhood intensities increase classification performance, c) the combination of intensity and texture information performed better than either individually, and d) 1 hidden layer in the network topology outperformed 0 or 2 hidden layers. After pixel classification with the ANN, this system performed an unsupervised segmentation to divide the image into homogeneous regions. These regions were assigned a label based on the results of the classifier, and were

processed with morphological operations. A second ANN was used to determine whether the resulting regions represented tumors based on a feature set.

One more method was presented by [Busch, 1997] that did not require patient-specific training. This work focused on the segmentation of a specific type of non-enhancing homogeneous tumor (low-grade astrocytomas) from T1-weighted, T2-weighted, and coregistered CT (X-ray) images. Recently, a method was proposed by Zhang *et al.* [2004] for automatic tumor segmentation in MR images. This approach used Support Vector Machines (SVMs), which are currently an extremely popular method for performing binary classification.

In addition to the supervised learning methods discussed above many methods have been proposed that use Supervised Segmentation with Advanced Image Modalities. The advantages of these methods are that they may facilitate an easier automatic segmentation task and that they may more appropriately characterize the extent of the tumor infiltration. The disadvantages of these approaches are that they require additional acquisition time and that the additional modalities are not available for historical data. A method to segment tumors was evaluated by Soltanian-Zadeh *et al.* [1998]. The system presented to segment this large combination of images used patient specific training, and consisted of coregistration, brain masking, anisotropic filtering, intensity non-uniformity correction, and finally an eigenimage analysis.

2.4 Registration-Based Segmentation

Registration-based methods, also known as atlas based method make use of a segmented image (deformable atlas) which is elastically warped to a new image and tissue labels are simply transferred. A deformable atlas is usually obtained using one or several manual segmentations. The main advantage of these methods is possibility to propagate any brain structure available in the atlas without any additional cost.

The aim of registration of image X to image Y is to find a transformation which maps any point x in image X to its corresponding point y in image Y. First, the global 12-parameter affine transformation $A(X \rightarrow Y)$ is calculated to perform translation, rotation, scaling and skewing so that the best possible alignment of the images is achieved [Hajnal *et al.* 2001]. Affine registration is often not sufficient due to natural variability in shape and size of normal healthy brains. One of the ways to model the local component is using uniform tensor-product 3D cubic B-splines [Rueckert *et al.* 1999]. During the iteration step in both affine and non-rigid registration, Normalized Mutual Information NMI [Studholme *et al.* 1998] is used as the similarity metric of the two images, given as follows:

$$NMI(X,Y) = \frac{H(X)+H(Y)}{H(X,Y)}$$

where, H( ) denotes the entropy of normalized image histogram. A cost function consisting of similarity term and in case of B-spline registration also of regularization term to ensure smoothness is then minimized using gradient descent. The B-spline registration is performed in multi-resolution framework. The resolution is refined by halving the spacing

between the B-spline control points and consequently inserting new B-spline control points. Affine registration is always performed before B-spline registration so we can write

$$N_{X\to Y}(X) = A_{X\to Y}(X) + B_{X\to Y}(X) \tag{2.1}$$

Let R be the reference image and M corresponding manual segmentation of image R. A new image I can be segmented by transferring the manual segmentation using non-rigid registration. Let $Seg_{\mathrm{nr}}(\mathrm{I})$ denote registration-based segmentation of I. Then,

$$Seg_{nr}(I) = N_{R\to I}(M) \tag{2.2}$$

where, transformation of the image means that the transformation is applied to every pixel of the image. Segmentation approaches based on this idea typically first align a labeled template (or atlas) image with the image to be segmented, and infer the labels for the new image by assuming that they correspond to the labels of the aligned template.

The advantage of this type of method is that spatial information is encoded through the use of the template, as opposed to pixel classification based methods that encode limited spatial information. The major disadvantage of this method is that the registration may not be perfect, and that there may be anatomical differences between the template and the image to be segmented. These disadvantages make template registration methods inappropriate to apply directly for tumor segmentation, since the template does not have a tumor, nor is its anatomy affected by the presence of a tumor. However, the ability to use spatial information derived from the spatial alignment of a template is appealing, and there has been considerable recent effort focusing on the incorporation of template registration into methods for tumor segmentation.

The major work using this approach include those of Kaus *et al.*(2001]. This model is shown in Figure 2.3. They employed a kNN classification algorithm with patient-specific training, used a label-based registration algorithm based on principles of optical flow, and preprocessed images with an anisotropic diffusion filter before analysis. After preprocessing, the segmentation consisted of performing kNN classifications.

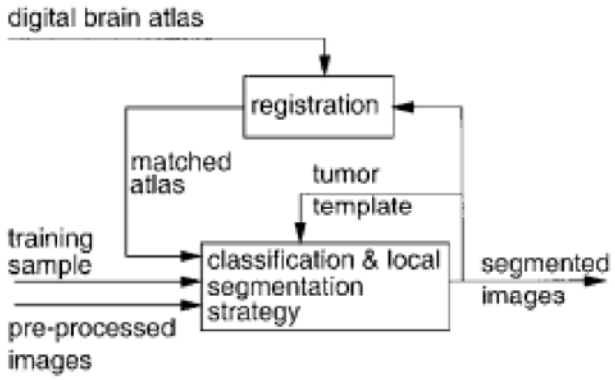

Fig.2.4. Diagram of the tumor segmentation scheme proposed by [Kaus *et al.* 2001]

Gering [2003] proposed a system that used template registration in segmentation. This method used a database of normal brains as training data. Each normal brain would be registered with the image to be segmented, and a simple multi-resolution statistic would be computed for each pixel to determine how significantly it differed from the most similar normal brain at that location. One of its limitations is that it does not account for the intensity non-standardization effects that would be present in a large database of normal brains, while another weakness is the lack of availability of a database of completely normal brains. In a hybrid approach, Murgasova *et al.* [2006] combined intensity based

and registration based approaches to obtain a robust segmentation method which is successfully used on healthy and diseased brain MRI of 2-year-old children.

2.5 EM-based segmentation

The expectation-maximization algorithm (EM) [Dempster *et al.* 1977] is a general technique for finding missing data based on observed data and maximum likelihood parameters estimates. [Leemput *et al.* 1999a] presented a model in which observed data are the image intensities, the missing data are the labels and the parameters are the means and variances of the Gaussian distribution which is assumed for the intensity distribution of each tissue class. This is an iterative process which interleaves the calculation of posterior probabilities of each voxel belonging to each tissue class (usually white matter, gray matter, cerebrospinal fluid, other) - the expectation step, with maximum likelihood estimation of the Gaussian distribution parameters the maximization step.

For the task of segmenting head MR images into the three normal brain classes (gray matter, white matter, and cerebrospinal fluid), Expectation Maximization approaches have become a popular framework, since they have shown to be robust to both intensity inhomogeneity and intensity non-standardization. Wells *et al.* [1996] was the first group to formulate the task of normal brain segmentation as an Expectation Maximization problem. Some other works using this approach are those by Leemput *et al.* [1999a, 1999b], Evans and Collins [1993] etc.

2.6 Image Segmentation using Clustering

In one more category of unsupervised methods, rather than dividing the image along anatomically meaningful distinctions, images are divided into homogeneous regions using image-based features such as intensities and/or textures. Clustering is one method to this. The main disadvantages of this approach are a) the number of regions often needs to be pre-specified, b) tumors can be divided into multiple regions, and c) tumors may not have clearly defined intensity or textural boundaries. These disadvantages limit the use of this approach. Some significant works done using these methods are by Capelle *et al.* [2000] and Sammouda *et al.* [1996].

Image segmentation can be treated as a clustering problem where features describing each pixel correspond to a pattern and an image region (i.e. segment) corresponds to a cluster [Jain *et al.* 1999]. It can be inferred from the definitions of the clustering problem and the image segmentation problem that both the problems are similar in nature. Therefore, clustering algorithms have been widely used to solve the problem of image segmentation (e.g. K-Means [Tou and Gonzalez 1974], FCM [Trivedi and Bezdek 1986], ISODATA [Tou and Gonzalez 1974] etc. But since the number of clusters is usually not known a priori in image segmentation, clustering algorithms that do not require the user to specify the number of clusters are usually preferred.

In this thesis, the clustering problem and the image segmentation problem are considered to be similar. Thus, algorithms are proposed for both problems interchangeably. Image segmentation is a fundamental process in several image processing and computer

vision applications. It can be considered as the first low-level processing step in image processing and pattern recognition [Cheng *et al.* 2001]. Image segmentation is defined as the process of dividing an image into disjoint homogenous regions. These homogenous regions should represent objects or parts of them [Lucchese and Mitra 2001]. The homogeneity of the regions is measured using some image property (e.g. pixel intensity) [Jain *et al.* 1999]. Image segmentation can be formally defined as follows:

Given an image I and a homogeneity predicate P. The segmentation of image I is the partitioning of I into K regions, $\{R_1, R_2,\ldots,R_K\}$, satisfying the following conditions:

- Each pixel in the image should be assigned to a region.
- Each pixel is assigned to one and only one region.
- Each region satisfies homogeneity predicate P.
- Two different regions cannot satisfy the same predicate P.

In this work, the various classification techniques used for the detection of anomalies in brain MRI are a) Hybrid of Genetic Algorithms and Artificial Neural Networks b) Adaboost c) Support Vector Machine. As the research interest of this work is to explore the applicability of Particle Swarm Optimization, in chapter 6 the same has been used for a) segmentation using clustering and b) optimization of feature sets to be input to existing classifier such as kNN and clustering algorithm *K-Means*. The supervised classification stage has two components, a training phase and a testing phase. In the training phase, pixel features and their corresponding manual labels represent the input, and the output is a model that uses the features to predict the corresponding label.

2.7 Conclusion

Several methods of segmentation of brain tumors were discussed. These included supervised and unsupervised methods, registration and EM based methods. Advantages and limitations of all the approaches were presented. It can be concluded that the current state of the art methods for automatic brain tumor segmentation are Expectation Maximization approaches that use outlier detection due to the robustness to intensity non-standardization, along with the registration-based method of [Kaus *et al.*, 2001], due to the use of non-linear registration and the more extensive use of spatial information to enhance discrimination and the neural network based methods of [Dickson and Thomas, 1997, Busch, 1997], due to the use of textural information and more powerful classification techniques. Many methods make use of two or more of the previously discussed approaches. These are termed as Hybrid methods, which has been used in the upcoming chapters. In this work, the use of textural and spatial information has been integrated with some classification techniques which have not yet been used in brain tumor segmentation. Particle Swarm Optimization is one such method. Other methods include Support Vector Machine, kNN, *K-Means*, Genetic Algorithms etc.

Chapter 3

MRI Classification Outline

3.1 Introduction

The existing work provides important information needed in designing a new system. In the previous chapter a variety of methodologies were presented for the detection of tumors using MRI. This chapter provides a common outline of few new systems for this purpose. The next section points out some of the steps which can enhance the tumor detection process and presents a basic model for image preprocessing and segmentation of MRI. Next section provides brief explanations of all the components of the model. Section 3.4 deals with MRI preprocessing in which each step of the MRI preprocessing has been detailed.

3.2 Model for Tumor Detection

From the previous chapter, it can be concluded that following are the steps which can enhance the tumor detection process:

a) Intensity inhomogeneity correction to reduce the effects of intra-volume intensity inhomogeneity.

b) Inclusion of intensity and texture information to improve discrimination between the classes.

c) Inclusion of information about shape, size and symmetry to improve segmentation results.

As this work is aimed at automating a task which is presently performed by human experts, the methods used by human experts can also be useful in this problem. But human experts do not make use of an internal intensity-based pixel classifier, and are able to incorporate much more complex information. This includes knowledge of the expected appearance, location, and variability of normal anatomy, patient-specific bi-lateral symmetry, knowledge about the expected intensities of different tissues relative to each other in a modality and the evaluation of the appearance of regions of pixels and/or shapes present within the image.

Manual methods consider and combine the above mentioned diverse sources of information and also consider previous experiences in related tasks. Therefore, when discrimination between normal and abnormal areas is not trivial those sources of evidences are used in making a decision. These sources of evidence may be the intensities in different modalities, the textures observed, bi-lateral symmetry, similarity to a normal brain, expected tissues or structures at spatial positions, expected and observed shapes of different structures, assessing normal anatomic variations, assessing variations due to the presence of a tumor, and evaluating pixel regions.

The main issue involved in designing an automated model is of combining the sources of these evidences to achieve acceptable result. The shortcoming of the existing methods lie in the fact that the patterns are complex and involve interactions between the different sources of evidence. These complex interactions are difficult to be represented by

a well defined set of manually determined rules. Therefore use of a supervised method is appropriate as it focuses specifically on finding patterns in large sets of features for optimizing a performance measure, such as the number of misclassified pixels. Intensity and texture information or different textural measures can be combined to increase performance.

According to Egmont-Petersen *et al.* [2002] image classification process uses the following main steps:

a. Preprocessing - These are filtering operations that give as a result a modified image with the same dimensions as the original image (e.g., contrast enhancement and noise reduction).

b. Data reduction (Feature extraction): Any operation that extracts significant components from an image (window). The number of extracted features is generally smaller than the number of pixels in the input window.

c. Segmentation - Any operation that partitions an image into regions that are coherent with respect to some criterion. One example is the segregation of different textures.

d. Object detection and recognition - Determining the position and, possibly, also the orientation and scale of speci1c objects in an image, and classifying these objects.

e. Image understanding - Obtaining high level (semantic) knowledge of what an image shows.

f. Optimization - Minimization of a criterion function which may be used for, e.g., graph matching or object delineation.

The above image processing chain has been described in Figure 3.1.

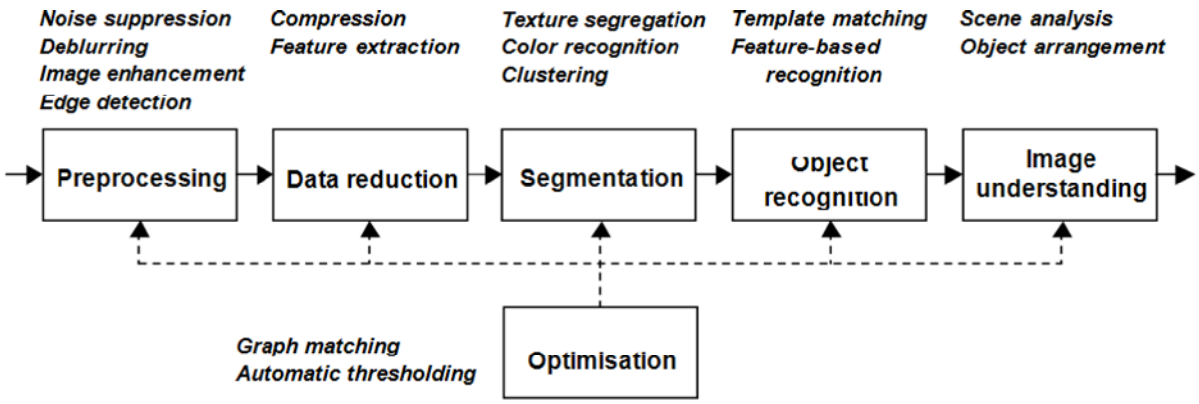

Fig.3.1. The image processing chain containing the five different tasks: preprocessing, data reduction, segmentation, object recognition and image understanding. Optimisation techniques are used as a set of auxiliary tools that are available in all steps of the image processing chain Genetic Algorithms (*Source*: [Egmont-Petersen *et al.* 2002])

As shown in Figure 3.2 the models proposed in this work have following three main phases: MRI preprocessing, Feature Extraction and Classification. Each of these phases is explained in the following sections. The goal of MR image segmentation is to accurately identify the principal tissue structures in these image volumes. In this work studies and experiments have been done with some of the existing and promising techniques of MRI classification and then, better models have been developed.

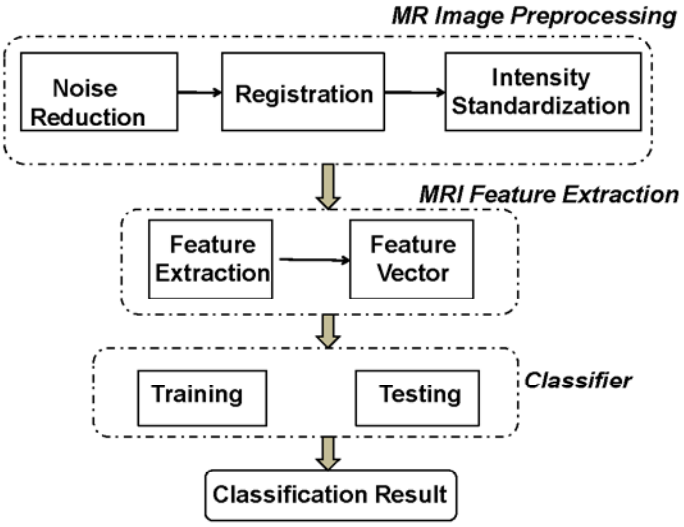

Fig.3.2. General model of image classification

3.3 MRI Preprocessing

It is necessary to process any image before using it for any segmentation or classification task. Usually the preprocessing involves noise reduction, resampling and intensity or contrast standardization. These steps are explained in the following subsections.

3.3.1 Noise Reduction

Noise reduction is done to reduce the effects of local noise, inter-slice intensity variations, and intensity inhomogeneity. But as the MRI machines nowadays are specifically designed to reduce these effects, this stage of preprocessing is not of much importance. This phase yields images with the same or reduced levels of local noise, inter-slice intensity variations, and intensity inhomogeneity.

a) Local Noise Reduction

When the signal recorded at each pixel gets corrupted, it is termed as 'Local Noise'. Local noise is additive and independent of pixel location but it depends on the tissue measured at the location.

b) Inter-Slice Intensity Variation Reduction

When sudden changes occur in intensity that can be observed between adjacent slices, it is termed as Inter-slice intensity variations. The presence of this type of noise depends on the acquisition protocol used. This effect can be corrected by modeling it as part of a three-dimensional inhomogeneity field as in [Leemput *et al.* 1999a], or it can be corrected by using techniques that standardize the intensities of adjacent slices. The method used in the present study for reducing inter-slice intensity variations is weighted least squares estimation method [Schmidt Mark 2005].

c) Intensity Inhomogeneity Reduction

Inhomogeneity of intensities within images can be present due to a variety of factors. Numerous methods have been suggested to reduce intensity inhomogeneity. Most common methods for Intensity Inhomogeneity Reduction are a) Non-parametric Non-uniform intensity Normalization (popularly known as N3) [Sled *et al.* 1998] and b) Bias Field Corrector (BFC) [Shattuck *et al.* 2001]. Some other significant methods have been suggested by Montillo *et al.* [2003] and Lai and Fang [1999]. We shall use N3 is due to its ease of use and no limitation of skull-stripping, as in case of BFC.

3.3.2 Registration

Image registration is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors [Barbara *et al.* 2003]. It geometrically aligns two images—the reference and sensed images. The present differences between images are introduced due to different imaging conditions. This is done by computing a transformation that maps each location in an input volume onto a template volume, then 're-slices' the input image such that pixels align spatially and are the same size as the corresponding pixels in the template image. Registration methods typically aim to compute a transformation that minimizes a measure of dissimilarity between the images, or maximizes a measure of similarity. The input to the Registration phase will be the images produced by the Noise Reduction phase. The method used for image registration in this work is Advanced Normalization Tools (ANTs) [Avants, Tustison and Gang 2010].

3.3.3 Intensity Standardization

The next step of image preprocessing, intensity standardization allows the intensities to be used without the need of patient-specific training. Three popular methods available for this technique are histogram-based method, model-based approach and template based approach. In the first two methods large abnormalities are present due to interference, so template based method [Studholme *et al.* 2004] has been used in this work. Template-based methods perform registration as a preprocessing step, and estimate a transformation between the intensities in the image and the intensities in the template, based on pixels at corresponding locations. The advantage to this type of approach is that

spatial information is used, in addition to the intensity distribution used by the other two approaches.

3.4 Feature Extraction

Shape and texture features have been used for some time for pattern recognition in datasets such as remote sensed imagery, medical imagery, photographs, etc. The fourth stage in the framework is the calculation of the features that will be used in pixel classification. MRI features include intensities and textures, intensities and distances to labels, intensities and spatial tissue prior probabilities. These feature combinations comprise a very limited characterization of a pixel within an image that has a relatively known structure. It would be advantageous to simultaneously use intensities, textures, 'distances to labels', spatial prior probabilities, and symmetry. A variety of other features can also be used, such as features based on histogram analysis, anatomic variability maps and template comparisons.

The main consideration when selecting features is that the features used should reflect properties measured at the pixel-level that can aid in discriminating between normal pixels and tumor pixels. However, there does not necessarily need to be an obvious (or even linear) relationship between the features used and the likelihood of a pixel representing a tumor, since the classification stage will learn an appropriate means of combining the features to perform the task. There are four main sources for the computation of useful features:

1. Image-based Features: Features that can be calculated directly from the image data.

2. Coordinate-based Features: Features that take advantage of a standard coordinate system.

3. Registration-based Features: Features that use one or more registered templates.

4. Feature-based Features: Features formed from subsets or combinations of other features.

Image-based features are the most commonly used feature sets in brain tumor segmentation and can be used to represent various properties of pixels and their neighborhoods. The most obvious pixel-level feature of this type is the pixel intensities. The different image-based features are as follows:

a.. Intensity-based: A pixel's intensity in each channel, the intensities of the pixel's neighbors, and aggregations over the intensities in the pixel's neighborhood.

b. Histogram-based: The intensity percentile of the pixel within the histogram, the multi-spectral distances to the intensities of normal tissues, and the number of pixels close to the pixel's intensities in the multi-channel intensity space.

c. Texture-based: Explicit calculations of texture features based on a pixel's neighborhood. Textural features consist of a set of calculations that can characterize patterns in the intensities of the region containing a pixel, and have been used previously for tumor segmentation. Three main types of textural features are as follows:

a) Gray level co-occurrence matrices

b) Wavelets

c) Gabor filters.

In this work Haralick textural features [Haralick, Shanmugam and Dinstein, 1973] have been used. These feature sets characterize texture using a variety of quantities derived from second order image statistics. Co-occurrence texture features are extracted from an image in two steps. First, the pairwise spatial co-occurrences of pixels separated by a particular angle and/or distance are tabulated using a gray level co-occurrence matrix (GLCM). Second, the GLCM is used to compute a set of scalar quantities that characterize different aspects of the underlying texture. While these quantities can be interpreted using intuitive notions of texture, their main benefit is in providing a quantitative description that can be used for image analysis, such as forming a high-dimensional feature vector to support content based retrieval.

The choice of Haralick features based on GCM was made considering due to their proven applicability to analyze objects with irregular outlines [Russ 1995]. The GCMs are constructed by mapping the gray level co-occurrence probabilities based on spatial relations of pixels in different angular directions. A GCM P(i,j) reflects the distribution of the probability of occurrence of a pair of gray levels (i,j) given the spacing between the pixels is Δx and Δy in the x and y dimensions respectively. Four angles namely 0, 45, 90, 135 as well as a predefined offset distance of one pixel in the formation of symmetric co-occurrence matrices are considered. From these GCM matrices, set of features are computed as follows:

a) Entropy: A measure of nonuniformity in the image based on the probability of co-occurrence values.

$$\text{Entropy} = -\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{P(i,j)}{R}\right)\log\left(\frac{P(i,j)}{R}\right) \tag{3.1}$$

where N is the number of gray levels, equal to 256 for images in the present study. R is equal to the total number of pixel pairs used for the calculation of texture features in the specified angular direction.

b) Difference moment: A measure of contrast.

$$\text{Contrast} = -\sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2\left(\frac{P(i,j)}{R}\right) \tag{3.2}$$

c) Energy: A measure of homogeneity.

$$\text{Energy} = \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{P(i,j)}{R}\right)^2 \tag{3.3}$$

d) Inverse Difference Moment: A measure of local homogeneity.

$$\text{Inverse Difference Moment} = \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{P(i,j)/R}{1+(i-j)^2}\right) \tag{3.4}$$

e) Correlation: A measure of linear dependency of brightness.

$$\text{Correlation} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left((ijP(i,j)/R) - \mu_X\mu_y\right)}{\sigma_X\sigma_y} \tag{3.5}$$

where,

$$\mu_X = \sum_{i=1}^{N} i \sum_{j=1}^{N} P(i,j) / R \tag{3.6}$$

$$\mu_y = \sum_{i=1}^{N} j \sum_{j=1}^{N} P(i,j) / R \tag{3.7}$$

$$\sigma_X{}^2 = \sum_{i=1}^{N} (i - \mu_X)^2 \sum_{j=1}^{N} P(i,j) / R \tag{3.8}$$

$$\sigma_y{}^2 = \sum_{i=1}^{N} (i - \mu_y)^2 \sum_{j=1}^{N} P(i,j) / R \tag{3.9}$$

In the above expressions, $\mu_X, \sigma_X, \mu_{y,} \sigma_y$ are the mean and standard deviation values of GCM values accumulated in the x and y directions, respectively.

3.5 Classification

The computed features are used by a classifier to decide whether each pixel represents a tumor pixel or a normal pixel. The time required for training is an important factor in choosing a classifier, since each image has a large number of pixels and training times can grow large as feature sets and the number of training pixels increases. If the system is to be used in a semi-automatic way, where the user can input additional training pixels where the system has made mistakes, then a classifier that can be trained incrementally should be used to prevent the need to completely retrain. For most classifiers, assigning classes based on a model is computationally efficient, while initially learning the model can be computationally intensive.

For deciding the type of approach to be used the advantages and disadvantages of following three kinds of methods were studied.

Unsupervised Methods: The advantage of these methods is that they do not rely on training data, and therefore are not subject to any degree of variation due to human interpretation. These methods have the disadvantage that they have been limited to simple tasks, where there is an obvious indicator of abnormality such as the presence of a contrast agent. Another disadvantage of these methods is that significant re-engineering is required in order to apply these methods to new tasks or to use different modalities than those the system was designed for.

Supervised Methods: These methods have the advantage that they can be applied to new tasks or can use different modalities without the need for redesign. One more advantage of supervised methods is that learning to meaningfully combine different potential sources of evidence for the presence of tumor can be done automatically. The main disadvantage of this approach is that methods require patient-specific training. Therefore, these methods are not fully automatic, and they are also subject to manual variability.

Registration-Based Methods: These methods use spatial patterns and constraints within a system. This improves results but it incorporating the registration-based information properly is the prime concern. The details of above four methods have been given in chapter 2.

Chapter 4
MRI Classification using Existing Approaches

4.1. Introduction

In this chapter some existing methods of image classification have been used for brain tumor detection from MRI. The approaches that have been considered here are 1. Hybrid of a neural network called Self Organizing Map (SOM) and Genetic algorithms (GA), 2. Support Vector Machine (SVM) and 3. Adaptive Boosting. Although the choice of these algorithms is arbitrary, an attempt has been made to pick algorithms from the set of methods which are popular these days. The registration based methods and expectation-maximization algorithms have not been implemented as sufficient work has already been done with these methods as described in chapter 2.

Section 2 of this chapter presents a model based on Hybrid of Neural Network and Genetic Algorithms. The next section deals with the application of SVM in MRI classification. This section also presents an overview of the various kernels used in SVM. Section 4 presents a model based on yet another classifier termed as Adaptive Boosting. In the next section implementation details of the three models are presented which is followed by the performance measure description. Finally, results and conclusion are presented.

4.2 Image classification using Hybrid of Neural Network and Genetic Algorithm

This method uses a two-step approach to retrieval. In the first step, we propose the use of a Neural Network called Self Organizing Map (SOM) [Kohonen 2001] for clustering the images with respect to their basic characteristics. The Self-Organizing Map (SOM), with its variants, is the most popular artificial neural network algorithm in the unsupervised learning category. Many fields of science have adopted the SOM as a standard analytical tool: in statistics, signal processing, control theory, financial analyses, experimental physics, chemistry and medicine. The SOM is also one of the most realistic models of the biological brain functions.

In the second step, the Genetic Algorithm (GA)[Goldberg 2003] based search will be made on a sub set of images which were having some basic characteristics of the input query image. It was found that this method radically improves the result over the single feature vector approach.

In the medical-imaging context, the ultimate aim of Content Based Image Retrieval (CBIR) is to provide radiologists with a diagnostic aid in the form of display of relevant past cases, along with proven pathology and other suitable information [Ho *et al.* 2002]. Statistical pattern recognition methods like the Bayesian discriminant and the Parzen windows were popular until the beginning of the 1990s. Non-parametric feed-forward ANNs [Bishop 1996] are attractive trainable machines for feature-based segmentation and object recognition. The role of feed-forward ANNs and SOMs has been extended to encompass also low-level image processing tasks such as noise suppression and image

enhancement. Hopfield ANNs were introduced as a tool for finding satisfactory solutions to complex optimization problems. This makes them an interesting alternative to traditional optimization algorithms for image processing tasks that can be formulated as optimization problems.

4.2.1 MR Image Data Set

The sources and types of MR Images obtained are shown in table 4.1. The images under study were acquired using the Siemens 1.5-Tesla MR Systems. A major limitation with medical and biological data is that special cases or "positive" samples are very less compared to control or "negative" samples. On the contrary, the reverse is true for some of the medical imaging data, including brain MRI. Huge difference in the number of positive and negative cases, therefore, creates an imbalance of the data class distribution [Chawla *et al.* 2004]. To control such imbalances, two sets of images were prepared as per table 4.2.

Table 4.1: Image Data Set

Source	Normal Images	Tumor Images	Grades of Tumor , if known				Grade NA
			I	II	III	IV	
Kamayani Hospital, Agra, India	47	143	30	14	28	29	42
Silveroak Hospital, Mohali, India	31	89	14	11	19	11	34
IBSR, Human Brain Project, Harward	60	173	_	_	_	_	173
The Brain Institute, Utah University	72	85	_	_	_	_	85
Total	210	490	44	25	47	40	334

4.2.2. Genetic Algorithm

A genetic algorithm is a stochastic based beam search in which successor states are generated by combining two parent states, rather by modifying a single state [Norvig and Russel, 2002] It takes its inspiration from natural selection and survival of the fittest in the biological world.

Table 4.2. Number of Images in the two sets :I. For Biased training and II. For Unbiased Training

SET	Training Type	No. of Images for Training		No. of Images for Testing	
		Normal	Abnormal	Normal	Abnormal
I	Biased Training	36	65	12	21
II	Unbiased	36	36	12	50

GA differs from more traditional optimization techniques in that they involve a search from a "population" of solutions, not from a single point. Every iteration of a GA involves a competitive selection that weeds out poor solutions. The solutions with high "fitness" are "recombined" with other solutions by swapping parts of a solution with another. Solutions are also "mutated" by making a small change to a single element of the solution [Goldberg 1993]. Recombination and mutation are used to generate new solutions that are biased towards regions of the space for which good solutions have already been seen. Pseudo-code for a genetic algorithm is as follows:

1. Initialize the population
2. Evaluate each chromosome of the initial population
3. Generate new chromosome of the mating current chromosomes, by performing Mutation and Crossover.

4. Remove numbers of population for new chromosomes.
5. Evaluate new population
6. Repeat through Step 3 until some convergence criteria is satisfied.

The block diagram of GA is given in Fig.4.1

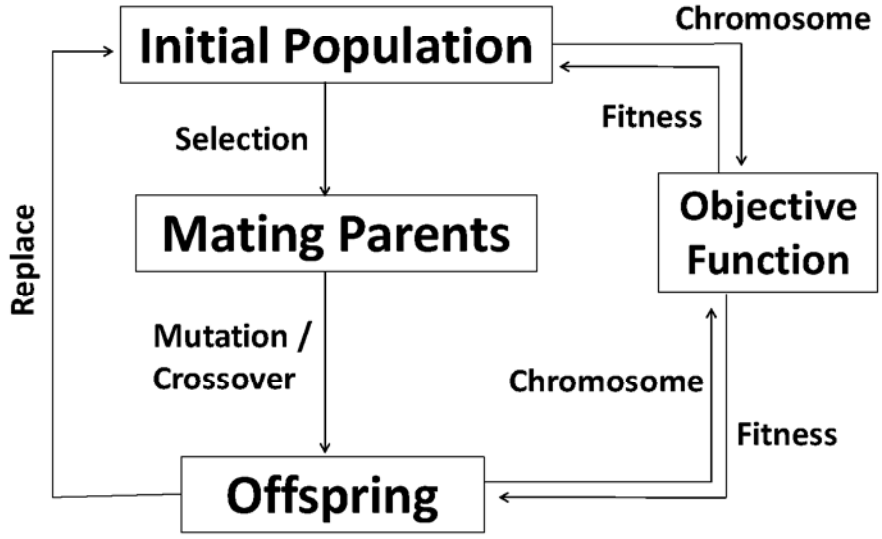

Fig.4.1. A Genetic algorithm model

Genetic Algorithm is used to select optimal inputs and also to find optimal weights for the neural network.

4.2.3 Neural Network Design

A neural network (NN) is an interconnected group of artificial neurons that uses a mathematical or computational model for information processing based on a connectionist approach to computation. In most cases a NN is an adaptive system that changes its structure based on external or internal information that flows through the network. In a NN

model simple nodes, which can be called "neurons" or "Processing Elements" (PE) are connected together to form a network of nodes.

In recent years there has been increasing interest in the use of Neural Networks (NN) in the diagnosis of various types of cancers. Many work were proposed to elaborate medical images interpretation systems whose majority is based on data- processing tools present an inspiration of biological and physiological metaphors like the Neural Networks (NNs), Fuzzy Logic(FL), Expert Systems and the Genetic Algorithms (GAs). Although these metaphors perform the task of interpretation successfully, they present much weakness and disadvantages. In a hybrid method, two or more of these methods are used together to mitigate their disadvantages.

Most NN algorithms are designed to alter the strength (weights) of the connections in the network to produce a desired signal flow. NNs have been applied successfully in many medical images interpretation problems where there input layer presented the relevant descriptions of the various healthy or suspected areas of the image (which present anomalies), and its output layer gives a decision on the nature of those areas (benign or malign tumors) [Garcia and Moreno 2004]. A typical model of NN used in the model is shown in the Figure 4.2.

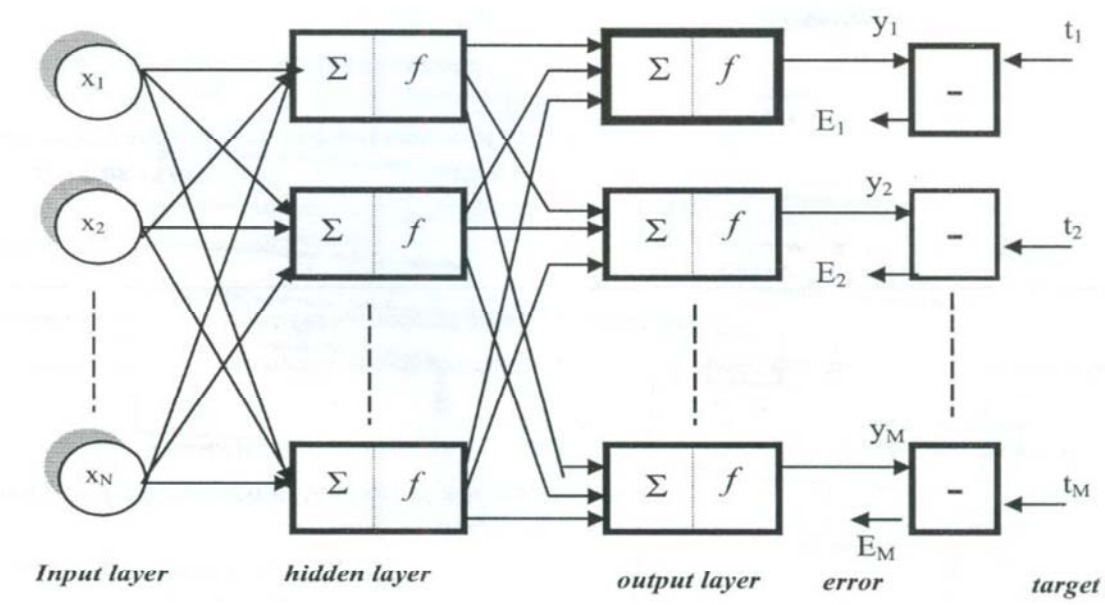

Fig.4.2. Model of a typical Neural Network

The three layers NN here is capable of producing a decision function with enough hidden units. Sigmoid function is used as the transfer function. The NN uses the best chromosome obtained from the GA as the initial weight vector. It consists of following steps:

1. Create weight vector from the most suitable chromosome obtained from GA.
2. Compute output of the feed-forward NN.
3. Compute Error between desired output and actual output.
4. Update weight using back-propagation.

4.2.4. Image segmentation

Some neural-based approaches perform segmentation directly on the pixel data, obtained either from a convolution window (occasionally from more bands as present in, e.g., remote sensing and MR images), or the information is provided to a neural classifier in the form of local features.

Image segmentation based on pixel data

Many ANN approaches have been proposed that segment images directly from pixel. Several different types of ANNs have been trained to perform pixel-based segmentation. Some of these ANNs include feed-forward ANNs, SOMs, Hopfield networks, probabilistic ANNs radial basis function networks , CNNs [Vilarino 1998] etc. Red- dick *et al.* developed a pixel-based two-stage approach where a SOM is trained to segment multispectral MR images. The segments are subsequently classi1ed into white matter, gray matter, etc., by a feed-forward ANN.

Image segmentation based on features

Several feature-based approaches apply ANNs for segmentation of images. Different types of ANNs have been trained to perform feature-based image segmentation: feed-forward, recursive SOMs, variants of Radial Basis Function (RBF) networks, Hop1eld ANNs, Principal Component Networks and a dynamic ANN. Hierarchical network architectures have been developed for optical character recognition and for segmentation of range images.

4.2.5 Proposed Model:

The proposed framework is shown in the Figure 4.3. It facilitates to search similar images in a large scale database of MRI with reasonable computational complexity. Searching the entire database for similar images will increase the time required to retrieve similar images. Based on the character of background tissue MR images can be classified as N= Normal or T= Tumor. Hence, the proposed model is divided in to two stages. In the first stage, the MR images present in the database is classified as N or T using SOM. In the second stage, the query image is obtained and the Region of Interest (ROI) is selected by the radiologist. As and when the ROI is selected the system acquire the length l and breadth b and then the class of the query image is identified using the same SOM network. Once, the class of query image is identified searches for similar suspicious region in the corresponding database with the same dimension of the ROI lxb will be followed.

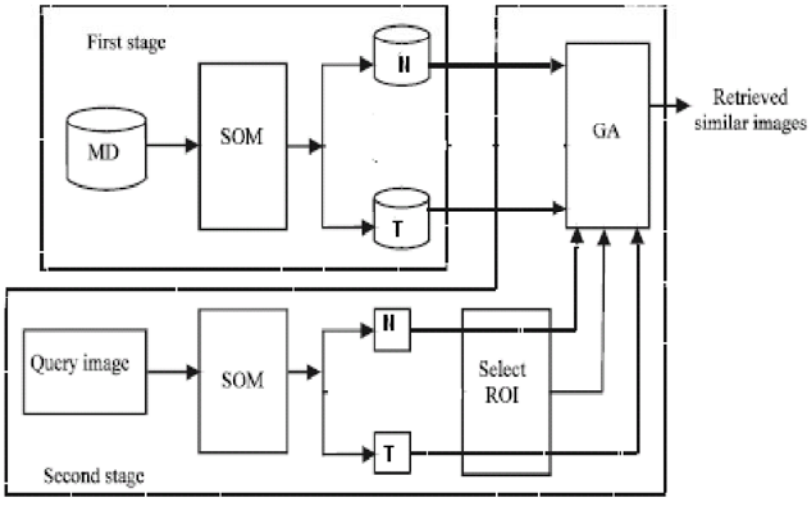

Fig.4.3. Functional Diagram of the proposed model

Classification of MRIs:

The overall image classification model is shown in Figure 4.4. To classify the MRI we used the SOM Neural Network. SOM algorithm [Kohonen 1995] is a neural network algorithm based on unsupervised learning. Basically it performs a vector quantization on the histogram of the images in the database and simultaneously organizes the quantized vectors on a regular low-dimensional grid. Histogram of the image is chosen as input to the SOM since it is very simple to calculate.

While finding, the histogram of a MRI images, the continuous black background as well as the over exposed white regions will add considerable amount of error in the histogram output. To overcome this, the pixels which make the MRI shape alone are considered for calculating the histogram. To achieve this, as a preprocessing the pixels corresponding to the black background as well as the over exposed white regions were removed by using a suitable lower and upper threshold and the histogram corresponding to the remaining pixels were calculated.

Application of GA:

In the second stage, after the classification of the query image and its ROI identified, every image from the specific class is taken and a random search is performed using GA over the entire image to locate the suspicious region matching with the ROI. The initial population for GA consists of chromosomes which represents x and y, the random position of ROI on the MRIs. Binary chromosomes have been used for this purpose. Furthermore, the GA makes use of Roulette wheel selection and single point crossover A

fitness function proportional to correlation has been identified to evaluate the chromosomes as given in equation 4.1.

Following algorithm was used to obtain the optimal weight and optimal learning rate:

Step 1: Randomly generate an initial population: $W^0.=(w_1^0, w_2^0 \ldots..w_n^0)$

Step 2 : Compute the fitness $f(w_i^t)$ of each chromosome in W^t

Step 3: Create new chromosome W'^t of mating current chromosomes, applying mutation and recombination.

Step 4: Delete numbers of the population to make room for new chromosomes

Step 5: Compute the fitness of w_i^t, and insert these into population.

Step 6: Increment number of generation.

The above is repeated till the desired fitness value is achieved or for a fixed number of iterations. The following algorithm was used for image pattern classification:

First Stage

Step 1: Initialize 'fitness value' to zero and 'child chromosome' as null.

Step 2: Set the objective function (system error for the neural network).

Step 3: Compute fitness value of the child chromosome.

Step 4: Compare the child chromosome with parent chromosome.

Step 5: Select parent using selection

Step 6: Apply Crossover and Mutation and Check fitness value of the child chromosome.

Step 7: Replace the old population by the new (best chromosome).

Second Stage

Step 1: Set the best chromosome as the initial weight vector.

Step 2: Compute the error between desired output and actual output.

Step 3: Update the weights or learning rate using neural network algorithm.

Since, only the x and y are present in the chromosome the remaining coordinates can be calculated as (x+l, y),(x, y+b),(x+l, y+b) where l and b are length and breadth of the ROI which is fixed. Hence, ROI is always a rectangle with size lxb.

$$Fitness = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2 (\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \quad (4.1)$$

where, A_{mn} is a 2D matrix representing the ROI selected by the radiologist, $\bar{A}$ is the mean of A_{mn} . B_{mn} is a 2D matrix representing a portion of the image in the database bounded by (x, y),(x+l, y),(x, y+b),(x+l, y+b) and $\bar{B}$ is the mean of B_{mn}. The initial population consists of different x and y which correspond to different regions on a single MRI in the database.

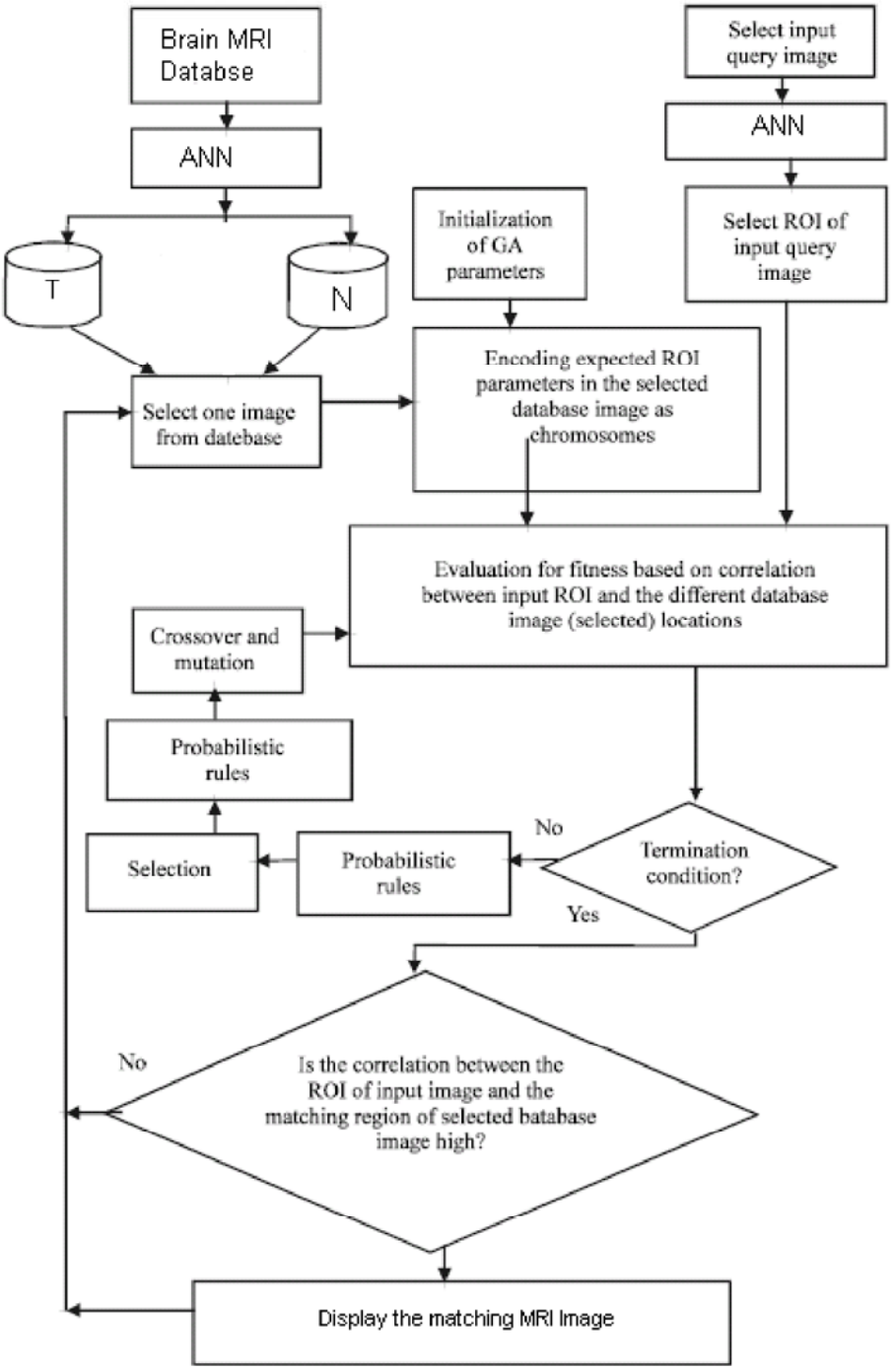

Fig.4.4. The overall image classification model

4.3 Gaussian RBF kernel based Support Vector Machine

This section presents a model based on classification by SVM which can be used to identify the tumor pattern from MRI.

4.3.1. Proposed model.

Classification is the most widely used strategy used for recognizing the objects based on its features. Figure 4.5 shows the classification model proposed to identify the tumors present in brain. Two major parts of the model are feature extraction and classification. The textural features obtained from the MR images are given as input to the classification part of the model. The classifier used is the Support Vector Machine (SVM) with Gaussian RBF kernel.

4.3.2 Overview of SVM:

SVMs are the most well known learning systems based on kernel methods. First introduced by Vapnik[1995], it is as an alternative to neural networks, and that has been successfully employed to solve clustering problems, specially in biological applications. SVM (Support Vector Machine) is a useful technique for data classification. It performs classification by constructing an *N*-dimensional hyperplane that optimally separates the data into two categories.

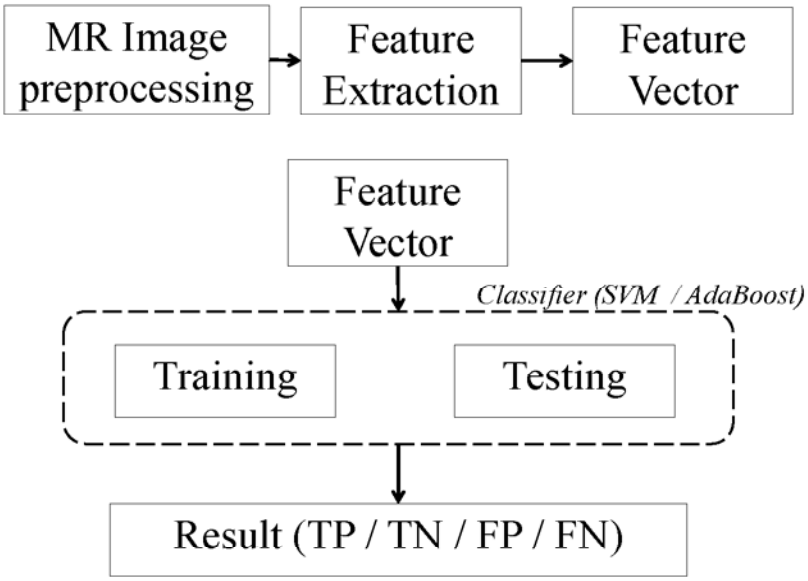

Fig.4.5. Proposed model for tumor detection from MR images.

A classification task usually involves training and testing data which consist of some data instances. Each instance in the training set contains one "target value" (class labels) and several "attributes" (features). The goal of SVM is to produce a model which predicts target value of data instances in the testing set which are given only the attributes. Given a training set of instance-label pairs $(x, y) = \{(x_1, y_1), (x_2, y_2) \ldots\ldots (x_n, y_n)\}$ where $x_n \in R^y$ and $y_n \in \{-1, +1\}$, SVM requires the solution of the following optimization problem [Boser *et al.* 1992]:

$$\min_{w,b,\xi} \quad \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$

Subject to $y_i(w^T \varnothing(x_i) + b) \geq 1 - \xi_i$, $\xi_i \geq 0$. (4.2)

Here training vectors x_i are mapped into a higher (maybe infinite) dimensional space by the function Φ. Then SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. $C > 0$ is the penalty parameter of the error term. The effectiveness of SVM lies in the selection of kernel and soft margin parameters. For kernels, different pairs of (C, γ) values are tried and the one with the best cross-validation accuracy is picked. Following are the four basic types of kernel functions :

Linear: $$K(x_i, x_j) = x_i^T x_j \quad (4.3)$$

Polynomial $$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0 \quad (4.4)$$

Radial Basis Function (RBF):

$$K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2), \gamma > 0 \quad (4.5)$$

Sigmoid: $$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \quad (4.6)$$

Here γ, r and d are kernel properties.

The kernel used in the proposed model is termed as Gaussian RBF and given by

$$K(x_i, x_j) = \exp\left(-\frac{\| x_i - x_j \|^2}{2\sigma^2}\right) \quad (4.7)$$

Where σ = width of kernel in SVM .Our interest is to vary the width σ to eliminate the inconsistency of the coexisting over-fitting and under-fitting in SVM.

4.4 AdaBoost:

AdaBoost, short for Adaptive Boosting, is a machine learning algorithm, formulated by Yoav Freund and Robert Schapire [1997]. It is a meta-algorithm, and can be used in conjunction with many other learning algorithms to improve their performance.

AdaBoost is adaptive in the sense that subsequent classifiers built are tweaked in favor of those instances misclassified by previous classifiers.It consists of generating an ensemble of *weak* classifiers (which need to perform only slightly better than random guessing) that are combined according to an arbitrarily strong learning algorithm [Friedman]. It has been applied with great success to several benchmark machine-learning problems using rather simple learning algorithms such as decision trees or linear regression.

The AdaBoost algorithm takes as input a labeled training set, $(x, y) = \{(x_1, y_1), (x_2, y_2).....(x_n, y_n)\}$ where $x_n \in R^{y}$ and $y_n \in \{-1, +1\}$ and calls a weak or base learning algorithm iteratively. At each iteration , a certain confidence weight $D_t(x_i)$ is given (and updated) to each training sample . The weights of incorrectly classified samples are increased so that the weak learner is forced to focus on the hard patterns in the training set. The task of the base learner reduces to find a hypothesis $h_t : x \to y$ for the distribution D_t. At each iteration, the goodness of a weak hypothesis is measured by its error

$$\epsilon_t = P[h_t(x_i) \neq (y_i)] = \sum Di \qquad (4.8)$$

Once the weak hypothesis has been calculated, AdaBoost chooses a parameter $\alpha_t = (1/2)\ln((1-\epsilon_t)/\epsilon_t))$.The distribution is next updated in order to increase the weight of samples misclassified and to decrease the weight of correctly classified patterns.

4.5 Implementation

For the GA-SOM model, experiments were performed for different values of String Size, Mutation level, and Crossover rate of the GA. The various best parameters came out to be as follows:

String Size= 40

Mutation level = 0.3

Crossover rate = 0.65.

The threshold level to identify the fittest elements= 0.8.

Selection method used: Roulette wheel selection scheme

Crossover algorithm used: 2-point crossover

Mutation operator: binary traditional mutation scheme with random choice of bits to be muted.

Offspring replacement scheme used: Selective Breeding

The proposed system was successfully implemented and evaluated using MATLAB© on a normal 2G.Hz Pentium IV computer with 512 MB RAM. The proposed algorithm performed well even for moderate number of images.

4.5.1. Feature extraction: For all the MRI images the pixels were extracted and their spatial coordinates and intensities were used for constructing the Gray Level Co-Occurrence Matries (GCM). Haralick method [Haralick *et al.* 1979] was used for the extraction of features from MRI. The texture measures computed are

i) Entropy: A measure of nonuniformity in the image based on the probability of co-occurrence values.

ii) Energy: A measure of homogeneity.

iii) Difference moment: A measure of contrast.

iv) Inverse Difference Moment: A measure of local homogeneity.

v) Correlation: A measure of linear dependency of brightness.

4.5.2. Performance Measure:

The classification of the images may give four types of results namely, True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN). The contingency table for classification is given in Table 4.3.

Table 4.3. Contingency table

Real Group	Classification Result	
	Normal	Abnormal
Normal	TN	FP
Abnormal	FN	TP

The most commonly used performance measures in classification are Precision and Recall. Precision can be seen as a measure of exactness or fidelity, whereas Recall is a measure of completeness. In a statistical classification task, the Precision for a class is the number of true positives (i.e. the number of items correctly labeled as belonging to the positive class) divided by the total number of elements labeled as belonging to the positive class (i.e. the sum of true positives and false positives, which are items incorrectly labeled as belonging to the class). Recall in this context is defined as the number of true positives divided by the total number of elements that actually belong to the positive class (i.e. the sum of true positives and false negatives).

In a classification task, a Precision score of 1.0 for a class C means that every item labeled as belonging to class C does indeed belong to class C (but says nothing about the number of items from class C that were not labeled correctly) whereas a Recall of 1.0 means that every item from class C was labeled as belonging to class C (but says nothing about how many other items were incorrectly also labeled as belonging to class C) [Makhoul *et al.* 1999]. The formula for precision and recall are given by

Precision= TP / (TP+FP) * 100

Recall= TP / (TP + FN) * 100

Also,

Accuracy= (TP + TN) / (TP + TN + FP +FN) * 100

and Error= 1- Precision

4.6 Results

The MR images were used for the four models mentioned earlier. The classification results have been summarised in tables 4.3(a) and 4.3(b).

Table4.4 (a) Classification Results for Set 1

Classifier	Precision (%)	Recall(%)	Accuracy(%)
GA-SOM	84.6	87.1	86.2
SVM (Gaussian RBF kernel)	93.33	95.28	92.71
AdaBoost	90.25	91.66	89.31

Table4.4 (b) Classification Results for Set 2

Classifier	Precision %	Recall (%)	Accuracy(%)
GA-SOM	85.4	86.2	88.2
SVM (Gaussian RBF kernel)	91.62	93.56	91.75
Ada Boost	89.05	90.52	88.26

4.7 Conclusion

As described in section 4.5, two different image sets were taken as biased and unbiased ones and experiments were performed. The performance of all the three proposed classifiers were evaluated and compared. The comparative performance results tabulated in section 4.6 indicate that SVM approach is better than other classifiers. Therefore, it can be concluded Gaussian RBF kernel based SVM is a promising technique for image classification in a medical imaging application. This automated pattern detection system can, therefore, be further used for classification of images with different pathological condition, types and disease status.

Chapter 5

Particle Swarm Optimization

5.1. Introduction

In this chapter the fundamental concepts of PSO is presented. Next section deals with the introduction to PSO and its basic algorithm. PSO suffers from a serious limitation of premature convergence. Section 5.3 addresses this and other limitations of using PSO in image clustering. In the next section 5.4 some proposed improvements in PSO have been presented. Finally, based on some experimental results, a selection is made as which version of PSO to be used in the present work and conclusion of this study is presented.

5.2 Particle Swarm Optimization

Particle swarm optimization (PSO) is a swarm intelligence optimization technique that was inspired by the behaviour of flocks of birds [Kennedy and Eberhart,1995]. It is a kind of intelligence that is based on social-psychological principles and provides insights into social behavior, as well as contributing to engineering applications. The recent years have seen tremendous growth in the use of PSO in various areas of computing ranging from networking, classification, scheduling, and training of artificial neural network to feature extraction and image processing.

The advantage of PSO over many of the other optimization algorithms is its relative simplicity and ease of use. PSO shares many similarities with evolutionary computation

techniques such as Genetic Algorithms (GA) [Goldberg 1993]. The system is initialized with a population of random solutions and searches for optima by updating generations. However, unlike GA, PSO has no evolution operators such as crossover and mutation. The potential solutions, called particles, fly through the problem space by following the current optimum particles.

In the basic PSO a problem is given, and some way to evaluate a proposed solution to it exists in the form of a fitness function. A communication structure or social network is also defined, assigning neighbors for each individual to interact with. Then a population of individuals defined as random guesses at the problem solutions is initialized. These individuals are called the candidate solutions. They are also known as the particles, hence the name particle swarm. An iterative process to improve these candidate solutions is set in motion. The swarm of individuals (called *particles*) flies through the search space.

The position of a particle is influenced by the best position visited by itself (i.e. its own experience) and the position of the best particle in its neighborhood (i.e. the experience of neighboring particles). When the neighborhood of a particle is the entire swarm, the best position in the neighborhood is referred to as the global best particle, and the resulting algorithm is referred to as a ***gbest*** PSO. When smaller neighborhoods are used, the algorithm is generally referred to as a ***lbest*** PSO [Shi and Eberhart, 1998a]. The performance of each particle (i.e. how close the particle is from the global optimum) is measured using a fitness function that varies depending on the optimization problem. Each particle in the swarm is represented by the following characteristics:

$\boldsymbol{x}_i$: The *current position* of the particle;

$\boldsymbol{v}_i$: The *current velocity* of the particle;

$\boldsymbol{y}_i$: The *personal best position* of the particle.

$\hat{\boldsymbol{y}}_i$ :: The *neighborhood best position* of the particle.

The personal best position of particle *i* is the best position (i.e. the one resulting in the best fitness value) visited by particle *i* so far. Let *f* denote the objective function. Then the personal best of a particle at time step *t* is updated as

$$\boldsymbol{y}_i(t+1) = \begin{cases} \boldsymbol{y}_i(t) & \text{if } f(\boldsymbol{x}_i(t+1)) \geq f(\boldsymbol{y}_i(t)) \\ \boldsymbol{x}_i(t+1) & \text{if } f(\boldsymbol{x}_i(t+1)) < f(\boldsymbol{y}_i(t)) \end{cases} \tag{5.1}$$

The *gbest* model: In this model, the best particle is obtained from the entire swarm by selecting the best personal best position. The position of the global best particle is given by,

$$\hat{\boldsymbol{y}}(t) \in \{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_s\} = \min\{f(\boldsymbol{y}_0(t)), f(\boldsymbol{y}_1(t)), \ldots, f(\boldsymbol{y}_s(t))\} \tag{5.2}$$

where *s* denotes the size of the swarm. The velocity of particle *i* is updated using the following equation:

$$v_{i,j}(t+1) = wv_{i,j}(t) + c_1 r_{1,j}(t)(y_{i,j}(t) - x_{i,j}(t)) + c_2 r_{2,j}(t)(\hat{y}_j(t) - x_{i,j}(t)) \tag{5.3}$$

Where w is the inertia weight, c_1 and c_2 are the acceleration constants, and $r_{1,j}$ and $r_{2,j}$ are factors lying between (0,1).

In the equation 5.3 there are additional three terms defined as follows:

The *inertia weight* term (w)*:* [Shi and Eberhart 1998b]. This term serves as a memory of previous velocities. The inertia weight controls the impact of the previous velocity: a large inertia weight favors exploration, while a small inertia weight favors exploitation.

The *cognitive component* $(\mathbf{y_i}(t) - \mathbf{x_i})$ **:** This term represents the particle's own experience as to where the best solution is.

The *social component*,$(\mathbf{y(t)} - \boldsymbol{x_i})$, which represents the belief of the entire swarm as to where the best solution is.

The position of particle i, xi, is then updated using the following equation:

$$x_i(t+1) = x_i(t) + v_i(t+1) \qquad (5.4)$$

Algorithm 5.1 presents the main steps of basic PSO algorithm. The particles in the swarm are updated according to equations 5.3 and 5.4. This updation takes place for a specified number of iterations or when the velocity updates are close to zero. There have been many versions of PSO proposed time to time on the basis of accuracy, speed or overall performance. Most important of them are Binary PSO [Kennedy and Eberhart 1997], Gauranteed Convergence PSO (GCPSO), Clamped PSO, Hybrid PSO,

Coevolutionary PSO, Repulsive PSO, Multi-objective PSO, Adaptive PSO (APSO), Discretized PSO etc.

Initialize population

Do

For particle i=1 to Swarm size S

if $f(x_i) < f(p_i)$, then $p_i = x_i$

$p_g = \{ p_i \mid \min f(p_j) \quad j=1,2,\ldots N \}$

For d =1 to Dimension D

Update v_{id} using equation 5.3

If $v_i > v_{max}$ then $v_i = v_{max}$

else if $v_i < -v_{max}$ then $v_i = -v_{max}$

Update x_{id} using equation 5.4

Next d

Next i

Until termination criteria is met.

Algorithm 5.1 The basic PSO

5.3 Limitations of PSO

PSO and other stochastic search algorithms have a major drawback of premature convergence. Although PSO finds good solutions much faster than other evolutionary algorithms, it usually can not improve the quality of the solutions as the number of iterations is increased [Angeline 1998]. As the swarm iterates, the fitness of the global best solution improves (decreases for minimization problem). It could happen that all particles being influenced by the global best eventually approach the global best, and from there on the fitness never improves despite however many runs the PSO is iterated thereafter. The particles also move about in the search space in close proximity to the global best and not exploring the rest of search space. This phenomenon is called 'convergence'. PSO usually suffers from premature convergence when strongly multi-modal problems are being optimized.

The rationale behind this problem is that, for the *gbest* PSO, particles converge to a single point, which is on the line between the global best and the personal best positions. This point is not guaranteed to be even a local optimum. Another reason for this problem is the fast rate of information flow between particles, resulting in the creation of similar particles (with a loss in diversity) which increases the possibility of being trapped in local optima [Riget and Vesterstrom 2002]. If the inertial coefficient of the velocity is small, all particles could slow down until they approach zero velocity at the global best. The selection of coefficients in the velocity update equations affects the convergence and the

ability of the swarm to find the optimum. One way to come out of the situation is to reinitialize the particles positions at intervals or when convergence is detected.

Numerous techniques for preventing premature convergence have been proposed. Some research approaches investigated the application of constriction coefficients and inertia weights. Many variations on the social network topology, parameter-free, fully adaptive swarms, and some highly simplified models have been created. Most important of them are discussed in the nest section.

5.4 Improvements in Convergence Behavior of PSO

5.4.1 Inertia weight model: The inertia weight term, c_0, which was first introduced by Shi and Eberhart [1998a], serves as a memory of previous velocities. The inertia weight controls the impact of the previous velocity: a large inertia weight favours exploration, while a small inertia weight favours exploitation [Shi and Eberhart, 1998b]. The modified velocity update equation is given by equations 5.3 and 5.4.

5.4.2 Constriction factor model: A constriction factor can be used to choose values for w, c1 and c2 to ensure that the PSO converges. The modified velocity update equation is defined as follows:

$$v_{i,j}(t+1) = \chi(v_{i,j}(t) + c_1 r_{1,j}(t)(y_{i,j}(t) - x_{i,j}(t)) + c_2 r_{2,j}(t)(\hat{y}_j(t) - x_{i,j}(t))) \quad (5.5)$$

Here, χ is the constriction factor defined as follows:

$$\chi = \frac{2}{\left|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}\right|} \quad (5.6)$$

and ,

$$\varphi = c_1 + c_2, \quad \varphi > 4$$

Use of the constriction factor and velocity clamping together generally improves both the performance and the convergence rate of the PSO [Eberhart and Shi 2000].

5.4.3 Guaranteed Convergence PSO (GCPSO)

The basic PSO converges prematurely because the velocity update equation depends only on the term wv_i (t). GCPSO [Van den Bergh 2002] avoids this by using a different velocity update equation,5.7.

$$v_{\tau,j}(t+1) = -x_{\tau,j}(t) + \hat{y}_j(t) + wv_{\tau,j}(t) + \rho(t)(1-2r_{2,j}(t)) \qquad (5.7)$$

The resulting equation for position update is given by the following the equation 5.8.

$$x_{\tau,j}(t+1) = \hat{y}_j(t) + w\boldsymbol{v}_{\tau,j}(t) + \rho(t)(1-2r_{2,j}(t)) \qquad (5.8)$$

The term ρ(t) defines the area in which a better solution is searched. The value of ρ is given is initialized to 1.0, with ρ (t+1) defined on the basis of number of successes and failures.

$$\rho(t+1) = \begin{cases} 2\,\rho(t) & \text{if } \#\text{ sucesesses} > s_c \\ 0.5\,\rho(t) & \text{if } \#\text{ failures} > f_c \\ \rho(t) & \text{otherwise} \end{cases} \qquad (5.9)$$

Van den Bergh suggests repeating the algorithm until ρ becomes sufficiently small, or until stopping criteria are met. Stopping the algorithm once ρ reaches a lower bound is

not advised, as it does not necessarily indicate that all particles have converged – other particles may still be exploring different parts of the search space. It is found that GCPSO has guaranteed local convergence whereas the original PSO does not.

5.4.4. Attractive and Repulsive PSO (ARPSO)

There are two phases between which ARPSO [Riget and Vesterstrøm 2002] alternates. In the attraction phase, PSO is used for fast information flow, as such particles attract each other and thus the diversity reduces. In this phase 95% of fitness improvements can be achieved. This observation shows the importance of low diversity in fine tuning the solution. In the repulsion phase, particles are pushed away from the best solution found so far thereby increasing diversity. ARPSO was found to give better results than PSO and GA in most of the test cases.

5.4.5. Multi-start PSO (MPSO)

Proposed by Van den Bergh [2002] MPSO tries to make GCPSO a global search algorithm. It works as follows:

1. Randomly initialize all the particles in the swarm.

2. Apply the GCPSO until convergence to a local optimum. Save the position of this local optimum.

3. Repeat Steps 1 and 2 until some stopping criteria are satisfied.

In Step 2, the GCPSO can be replaced by the original PSO. Several versions of MPSO were proposed by Van den Bergh [2002] based on the way used to determine the convergence of GCPSO.

5.4.6. Techniques using Mutation

Some techniques have been proposed which use a hybrid of PSO with Gaussian mutation [Higashi and Iba 2003] or hybrid lbest- and gbest- PSO with a non-uniform mutation Operator [Esquivel and Coello 2003]. These hybrid techniques have given better results than PSO and GPSO in all the experiments conducted. One more similar technique is Dissipative PSO (DPSO) [Xie *et al.* 2002]]that adds random mutation to PSO in order to prevent premature convergence. DPSO introduces negative entropy via the addition of randomness to the particles. The results showed that DPSO performed better than PSO when applied to the benchmarks problems.

5.4.7. Differential Evolution PSO (DEPSO)

DEPSO [Zhang and Xie 2003] uses a differential evolution (DE) operator [Storn and Price 1997] to provide the mutations. A trait point is calculated as follows:

If ($r_1(t) < p_c$ OR $j = k_d$) then

$$\ddot{y}_{i,j}(t) = \hat{y}_j(t) + \frac{(y_{1,j}(t) - y_{2,j}(t)) + (y_{3,j}(t) - y_{4,j}(t))}{2} \qquad (5.10)$$

where $r_1(t)$, k_d, $y_1(t)$, $y_2(t)$, $y_3(t)$ and $y_4(t)$ are randomly chosen from the set of personal best positions. To avoid disorganization of the swarm, $y_i(t)$ is mutated instead of $x_i(t)$. DEPSO works by alternating between the original PSO and the DE operator such that equation 5.3 and 5.4 are used in the odd iterations and equation 5.9 is used is in the even iterations. The performance of DEPSO has been found to be better than other popular versions of PSO in case of benchmark functions.

5.4.8. Fitness-Distance Ratio PSO (FDR-PSO)

A new term was added to the velocity update equation by Veeramachaneni *et al.* [2003]. The new term allows each particle to move towards a particle in its neighborhood that has a better personal best position. The modified velocity update equation is given by:

$$v_{i,j}(t+1) = wv_{i,j}(t) + \psi_1(y_{i,j}(t) - x_{i,j}(t)) + \psi_2(\hat{y}_j(t) - x_{i,j}(t)) + \psi_3(y_{\eta,j}(t) - x_{i,j}(t)) \quad (5.11)$$

where, ψ1 , ψ2 and ψ3 parameters specified by users. Here each y_n is chosen in a way such that the following is maximized.

$$\frac{(f(x_i(t)) - f(y_n(t)))}{|y_{n,j}(t) - x_{i,j}(t)|} \quad (5.12)$$

The performance of FDR-PSO is found to be better than PSO, ARPSO, DPSO, SOC PSO etc. In addition to above techniques, a scoring based method was developed by Chandra *et al.* [2009] to control the convergence of PSO. The following algorithm was used for this purpose:

5.5 Experimental results

Out of the PSO versions discussed in section 5.4 three promising algorithms on the basis of previous results [van den Berg and Engelbrecht 2005] were selected and experimented with some benchmark functions as detailed in table 5.1. These algorithms are PSO, Attractive and Repulsive PSO (ARPSO), Guaranteed Convergence PSO

(GCPSO) and Fitness-Distance Ratio based PSO (FDR-PSO). The results are shown in table 5.2.

Table 5.1 : Some benchmark functions used for testing PSO

Function	**Definition**	**Type**	**Range**	**Initial Population**
Sphere	$f_s(x) = \sum_{d=1}^{D} x_d^2$	Unimodal	[-50,50]	[25,40]
Quartic	$f_q(x) = \sum_{d=1}^{D} dx_d^4$	Unimodal	[-20,20]	[10,16]
Rosenbrock	$f_{Ro}(x) = \sum_{d=1}^{D-1} 100(x_{d+1} - x_d)^2 + (x_d - 1)^2$	Unimodal	[-100,100]	[50,80]
Griewank	$f_g(x) = \frac{1}{4000}\sum_{d=1}^{D} x_d^2 - \prod_{d=1}^{D} \cos(\frac{x_d}{\sqrt{d}}) + 1$	Multi-modal	[-600,600]	[300,500]
Rastrigin	$f_{Ra}(x) = \sum_{d=1}^{D} x_d^2 + 10(1 - \cos(2\pi x_d))$	Multi-modal	[-5.12,5.12]	[1,4.5]

5.5.1 Parameter Setting: The various parameters of PSO are to be defined by user. These parameters include population size, inertia weight, acceleration constants etc. and are vary from one problem to another.

Population Size: The PSO researchers have suggested the population size to be 2n to 5n where n is the number of decision variables. But, most researches have taken it to be constant. In this study the population size was varied from 20 to 50 and best results were obtained at population size of 35.

Inertia Weight: A linearly decreasing inertia weight of the order (0.9-0.4) is found to give good result.

Acceleration Constant: c1=c2= 1.8

Initialization of Swarms: The particle swarms were initialized by the Gaussian distribution given by, $f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}}$ (5.13)

Diversity: Diversity is measured by the following equation:

$$Div(S(t)) = \frac{1}{n_s} \sum_{i=1}^{n_s} \sqrt{\sum (x_{ij}(t) - \overline{x_j(t)})^2} \quad (5.14)$$

From table 5.2, it is evident that GCPSO and FDR-PSO give better results than others in most of the benchmark functions. These two versions of PSO are hybrid with Scoring based method (SBM) and tested again on the same functions. The results obtained are shown in table 5.3. From table 5.3 it appears that FDR – PSO, when clubbed with scoring based method, gives better results in most of the benchmark functions, as long as the convergence behavior is concerned.

Table 5.2. Results of PSO, ARPSO, GCPSO and FDR-PSO on benchmark functions (First row shows mean, second row shows diversity while the third one indicates deviation). The best mean values are shown in **bold**.

Function	PSO	ARPSO	GCPSO	FDR-PSO
Sphere	1.1956e-46	**0.9566e-46**	0.9566e-46	**1.0956e-46**
	1.9868e-18	1.9868e-18	1.9868e-18	2.59868e-18
	4.8961e-8	4.8961e-8	4.8961e-8	4.8741e-8
Quartic	12.3426e-12	11.4628e-11	**28.3426e-14**	14.4427e-12
	11.6743e-16	14.7645e-16	21.6743e-15	15.5482e-16
	16.9812e-6	17.67412e-6	36.9812e-14	19.8773e-7
Rosenbrock	21.9265	4.4057	**3.9951**	9.9928
	4.2378	2.9044	1.8737	2.5279
	8.6274e+03	4.1212	3.9593	3.1689
Griewank	0.031646	0.0110	0.0116	**0.0047**
	0.000710	0.0008	0.0015	1.6313e-08
	0.025322	0.0191	0.0197	0.01266
Rastrigin	19.2213	26.5612	14.5419	**11.6548**
	0.01002	0.4279	0.0002	0.5437
	17.2321	13.3956	11.9854	8.3498

Table 5.3 Results of GCPSO and FDR-PSO on benchmark functions by applying scoring based method.

Function	**GCPSO + SBM**	**FDR-PSO + SBM**
Sphere	0.8956e-47	**0.0756e-48**
	2.0021e-19	1.2564-18
	3.4561e-10	3.7684e-8
Quartic	**13.4526e-15**	16.2157e-15
	19.3743e-15	15.5482e-16
	31.9812e-14	19.8773e-7
Rosenbrock	1.9954	**1.8799**
	1.0834	2.4537
	2.6955	2.4589
Griewank	0.0111	**0.0022**
	0.0011	0.9314e-08
	0.0112	0.0112
Rastrigin	10.2741	**8.6516**
	0.0001	0.5125
	9.8837	6.7354

5.6 Conclusions

This chapter presented a brief introduction to PSO. Then the some modified versions of PSO were detailed to tackle the major drawback of PSO, i.e. premature convergence. The four algorithms namely PSO, GCPSO, ARPSO and FDR-PSO described in the chapter use different Gaussian initialization scheme for generating the swarm population. Two best versions were chosen from the preliminary comparative study of these versions. These versions were then hybrid with SBM to get the best version to be used in this work. It was found that FDR-PSO, when combined with SBM gives considerably better results.

Chapter 6

PSO Based approaches for Detection of Brain Tumors

6.1 Introduction

In the previous chapter we have discussed the ongoing research in the field of Swarm Intelligence and its application to various real life problems. There we also compared the relative performance of few important versions of Particle Swarm Optimization. It was found that FDR-PSO, when combined with SBM gives considerably better results.

In the present chapter three models are proposed based on Particle Swarm Optimization (PSO). The first model described in Section 6.2, is an image clustering algorithm based on PSO. This algorithm finds the centroids of user specified numbers of clusters, where each cluster groups together similar image primitives. Results of this PSO based clustering is compared with the standard conventional classifier *K-Means*. In the second model in Section 6.3, a hybrid of k Means and PSO is used for MR image segmentation. While the third model is based on the hybrid of kNN and PSO. In section 6.4 the results and performance analysis of the proposed models are presented. A model for grading of brain tumors is presented in section 6.5.

6.2 PSO based Image Clustering Algorithm

The basis of the classification of MRI images in this model [Chandra 2009, A PSO] is that different feature types manifest different pixel values based on spectral reflectance and emittance properties. This type of classification, which is based on pixel-by-pixel spectral information, is referred to as *spectral pattern recognition* [Lavine *et al.* 2002]. Table 6.1 shows the list of variables and their meanings in the proposed algorithm.

Table 6.1 Variables used in PSO based Image Clustering algorithm

Variable	Meaning
p_i	Particle , $p_i = (m_{i1}, m_{i2}, \ldots m_{in} \ldots, m_{iN})$: Represents *N* cluster centroids.
m_{ik}	k^{th} cluster centroid of i^{th} particle
X_i	Pattern matrix of i^{th} particle
f( )	Fitness function $f(p_i) = w_1 d_{max}(X_i, p_i) + w_2(X_{max} - d_{min}(p_i))$
d_{max}, d_{min}	Maximum and minimum Euclidean distances
X_i	a matrix representing the assignment of patterns to the clusters of particle i.
N_c	Number of clusters
w_1, w_2	User defined constants(weight factors)
S_i	Silhautte validity index

We propose the following image clustering algorithm which is based on PSO:

While ((S_i -1 $\geq \delta$)

1. ***For each particle i***
   a. For each pattern X_p

      Compute the Euclidean distance $d(X_p, m_{i,k})$ for all clusters $C_{i,j}$..

      Assign y to C_{ij}. Such that d-value is minimized.

   b. Calculate fitness of the particle, $f(p_i)$
2. ***Compute the personal best and global best solution y'(t)***
3. ***Update the cluster centroids.***

Algorithm 6.1: PSO based Image Clustering Algorithm

6.2.1 Performance measure: According to the definition of the fitness function, a small value of *f* is desirable as it means well-separated clusters. The goodness of a clustering algorithm can as well be described by Error of Quantization given by,

$$Q_e = \frac{\sum_{k=1}^{K}[\sum_{\forall x_p \in C_k} d(x_p, m_k)]/n_k}{K} \qquad (6.1)$$

where, C_k is the k^{th} cluster and n_k is the number of pixels in C_k.

The fitness function in table 6.1 has as objective to simultaneously minimize the intra-distance between patterns and their cluster centroids, that is d_{max} and to maximize the inter-distance between any pair of clusters, that is d_{min}. So, a smaller value of d_{max} and a higher value of d_{min} is desirable for good clustering. Another quality measure used in this

algorithm is the cluster validity which measures goodness of a clustering algorithm. Cluster validation is very important issue in clustering analysis because the result of clustering needs to be validated in most applications. In most clustering algorithms, the number of clusters is set as user parameter. There are a lot of approaches to find the best number of clusters. In this work silhouette validity index [Rousseeuw 1987] has been used which is given by,

$$S_i = (b_i - a_i) / \max(a_i, b_i) \tag{6.2}$$

Where a_i is the average dissimilarity of i-particles to all other particles in the same cluster; b_i is the minimum of average dissimilarity of i-particles to all particles in other cluster (in the closest cluster). A silhouette value close to 1 implies a good clustering, and it means that the pixels are assigned to an appropriate cluster. On the other hand, if silhouette value is about zero, it means that that sample could be assigned to another closest cluster as well, and the sample lies equally far away from both clusters. If silhouette value is close to −1, it means that sample is "misclassified" and is merely somewhere in between the clusters. The overall average silhouette width for the entire plot is simply the average of the S_i for all objects in the whole dataset. δ is the permissible tolerance defined by the user. δ value was set to 0.04 in this work. Q_e is calculated only after the segmentation has been done, while fitness function and silhouette index are computed and used in the algorithm itself.

6.3 An Image Segmentation Model Based on Hybrid of K-Means and and PSO

The *K-Means algorithm* [McQueen 1987] is one of the simplest unsupervised learning algorithms, which is used for data classification and analysis. It is based on the minimization of a performance index. The algorithm first chooses K random cluster centers, and then assigns each sample to a cluster based on the minimum distance to the cluster centers. Then, it updates the cluster centers with the new average of the values in each cluster. In image processing, the data set would be a set of pixel vectors. Hence, each pixel of the image will be classified into a cluster. Steps in the K-Means algorithm are listed below:

Step 1: Choose K initial cluster centers either randomly or taking from samples.

Step 2: Calculate the distance for each pixel to each of the cluster centers and assign the pixel to the cluster which has the minimum distance to its center.

Step 3: Update the new cluster center with the average of pixel values in each cluster.

Step 4: Repeat steps 2 and 3 until the clustering converges.

The K-Means algorithm tends to find the local optimal rather than the global optimal solutions. When the initial cluster centers are chosen relatively far apart, the result becomes more acceptable. If the main clusters are close in the feature space, the K-Means algorithm fails to recognize them with the unsupervised mode. To improve the performance of the K-Means algorithm, optimization techniques are usually employed. In

addition, the Fuzzy CMeans algorithm (FCM) [Bezdek *et al.* 1984] was proposed to separate data clusters with fuzzy means and fuzzy boundaries. The FCM is less dependent on the initial state of clustering.

In this work two efficient versions of PSO namely, *GCPSO* and *FDR-PSO* are proposed for optimization purpose.

6.3.1 GCPSO - K-Means algorithm:

The proposed GCPSO - K-Means algorithm is formulated as follows:

Step 1: Initialize the number of clusters to K and number of particles to m.

Step 2: Initialize m sets of K random cluster centers to be used by m particles.

Step 3: Assign each pixel to a cluster with the minimum Euclidean distance.

Step 4: Calculate new cluster center; if the new cluster centers converge to the old ones, go to the next step. Otherwise, go to Step 3.

Step 5: Save the best solution found performed by each particle. Call it pbest or personal best solution.

Step 6: Save the best solution among the m personal best solutions found. Call it gbest or global best solution.

Step 7: Update cluster centers of each particle according to the cluster center values of the pbest and gbest solution using equations 5.7 and 5.8.

Step 8: If the termination criterion is satisfied, go to the next step. Otherwise, go to Step 3.

Step 9: Output the optimal solution

6.3.2 FDR-PSO - K-Means algorithm:

The proposed PSO-C-*K-Means algorithm* is presented as follows:

All steps are exactly same as those described in 5.3.1., except for Step 7.

Step 7: Update cluster centers of each particle according to the cluster center values of the pbest and gbest solution using (5.11) and (5.4).

6.4 An Image Classification Model Based on Hybrid of kNN and PSO

In this section an algorithm based on the hybrid of kNN and PSO is presented for the classification of MRIs. The k-nearest neighbors algorithm (kNN) is a method for classifying objects based on closest training examples in the feature space. kNN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. The k-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of its nearest neighbor.

The neighbors are taken from a set of objects for which the correct classification is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. The k-nearest neighbor algorithm is sensitive to the local structure of the data. Nearest neighbor rules compute the decision boundary in an implicit manner. It is also possible to compute the decision boundary itself explicitly, and to do so in an efficient manner so that the computational complexity is a function of the boundary complexity [Bremner *et al.* 2005].

The best choice of k depends upon the data; generally, larger values of k reduce the effect of noise on the classification, but make boundaries between classes less distinct. A good k can be selected by various heuristic techniques, for example, cross-validation. The special case where the class is predicted to be the class of the closest training sample (i.e. when k = 1) is called the nearest neighbor algorithm. The accuracy of the kNN algorithm can be severely degraded by the presence of noisy or irrelevant features, or if the feature scales are not consistent with their importance. Much research effort has been put into selecting or scaling features to improve classification. In the proposed model, PSO has been used for optimizing the feature scaling.

The problem of using highly imbalanced dataset for pattern recognition is that the classification model built on the training data tends to be biased on preferring the majority class while ignoring the samples from the minority class. Data sampling method tries to remedy the skewed class distribution by either increasing the sample size of minority class or decreasing the sample size of majority class. However, algorithms that modify the

sample distribution with greedy measures can introduce undesired bias. In this study we re-apply the techniques in feature selection to data sampling using a PSO based hybrid system. The schematic flow of the proposed hybrid system is illustrated in Figure 6.1.

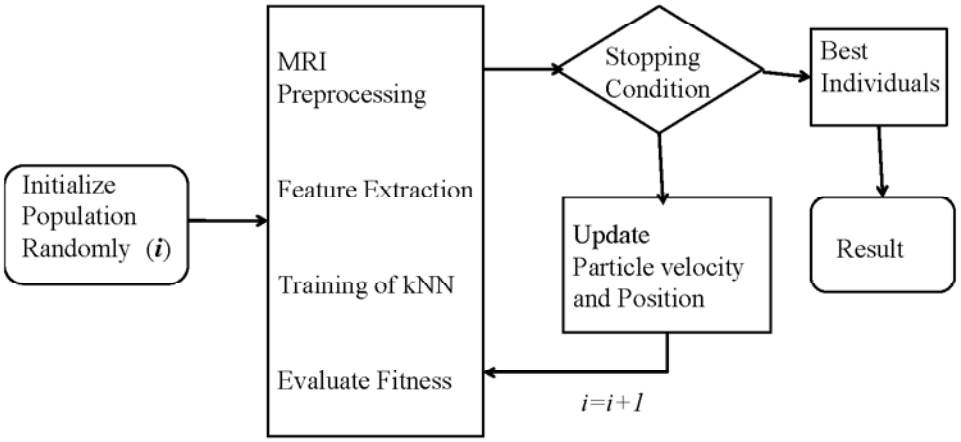

Fig.6.1. Schematic flow of the proposed hybrid system based on PSO and kNN.

6.5 Results and performance analysis

All the images detailed in the image data set table 4.1 were used for the three PSO based methods. Some of the segmentation results are shown in figure 6.2. Table 6.2 summarizes the average values of Q_e, d_{max} and d_{min} for the two versions of PSO, K-Means

and PSO-K-Means Hybrid models which been used for the clustering algorithms as described above.

Table 6.3 shows the values of Q_e , d_{max} and d_{min} obtained for GCPSO, K-Means and PSO – K-Means Hybrid.

Table 6.2 Performance of PSO, K-Means and PSO-K-Means Hybrid

Algorithm	Q_e	d_{max}	d_{min}
GCPSO	9.3265	12.2356	20.2356
FDRPSO	8.8862	11.2354	19.8857
K-Means	5.6689	20.2356	12.5742
GCPSO-K-Means Hybrid	5.5246	17.5832	14.2568

From table 6.2 it is evident that as far as Q_e is concerned the *K-Means* based approaches give better results. Whereas, the comparative values d_{max} and d_{min} show that PSO based approaches are better. The comparative results of precision, recall and accuracy for all the three models have been summarized in table 6.3. It appears from table 6.3 that the overall performance of PSO- kNN Hybrid model is very poor compared to PSO and Hybrid PSO kMeans; therefore PSO- kNN Hybrid model is not used for experiments hereafter.

Table 6.3 Classification Results for the three proposed algorithms

Classifier	Precision	Recall	Accuracy
GCPSO	92.76	96.24	94.42
Hybrid PSO kMeans	93.33	95.28	96.71
PSO kNN Hybrid	91.12	90.24	91.15

6.5.1 Effect of PSO parameters

In order to achieve best results from PSO it must be fine tuned. For this purpose the PSO based clustering algorithms were executed with varying values of swarm size, velocity, inertia weight and acceleration constants. The results obtained for Q_e, d_{max} and d_{min} against varying values of swarm size, s, have been plotted in figures 6.3(a), (b) and (c) respectively. The classification accuracies of GCPSO and PSO-*K-Means* Hybrid methods against different values of iterations and swarm sizes have been plotted in figures 6.4(a) and 6.4(b) respectively. The plots of figures 6.3 and 6.4 reflect that the performance obtained is the best when swarm size lies between35 and 45. Furthermore, it is also evident from the plot that in order to get better results, total number of iterations should be around 500-700. Table 6.4 summarises the various values of PSO parameters used.

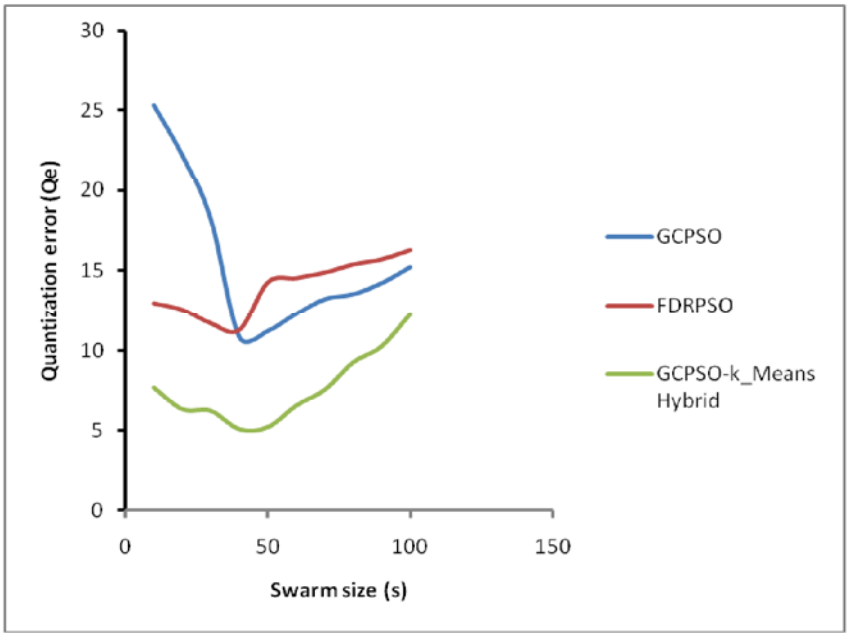

Fig.6.2(a). Effect of swarm size on quantization error

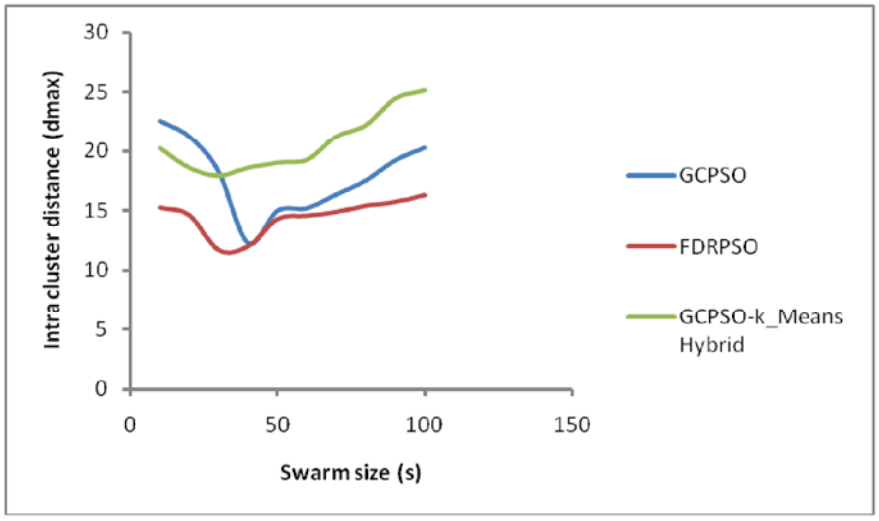

Fig. 6.2(b). Effect of swarm size on intra cluster distances

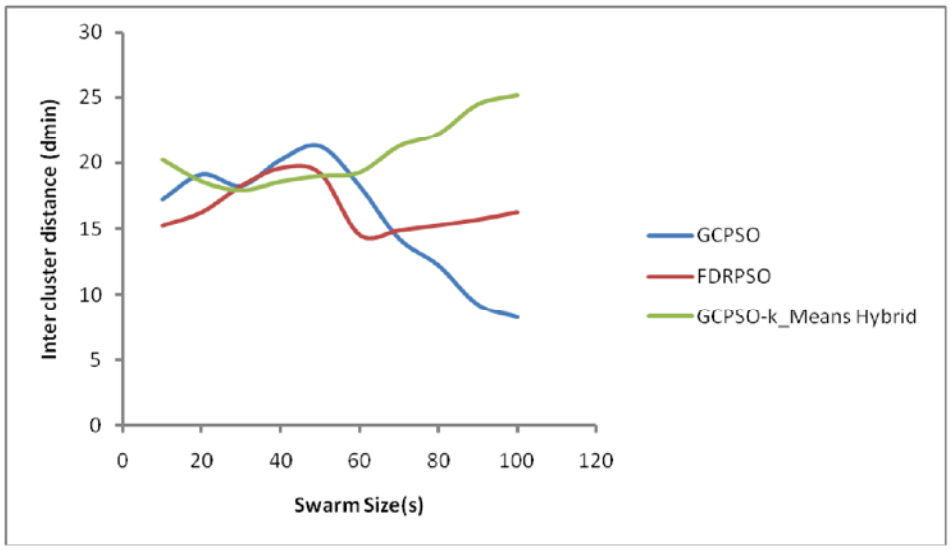

Fig. 6.2(c). Effect of swarm size on inter cluster distances.

Table 6.4 PSO parameters used

Parameter	**Value**
Size of Particle Population	40
Iteration	500
Update Rule	Sigmoid Function
Cognitive Constant	1.6
Social Acceleration Constant	1.5
Inertia Weight	0.7
Velocity Range	0.2-1.1
Fitness Weight	0.34

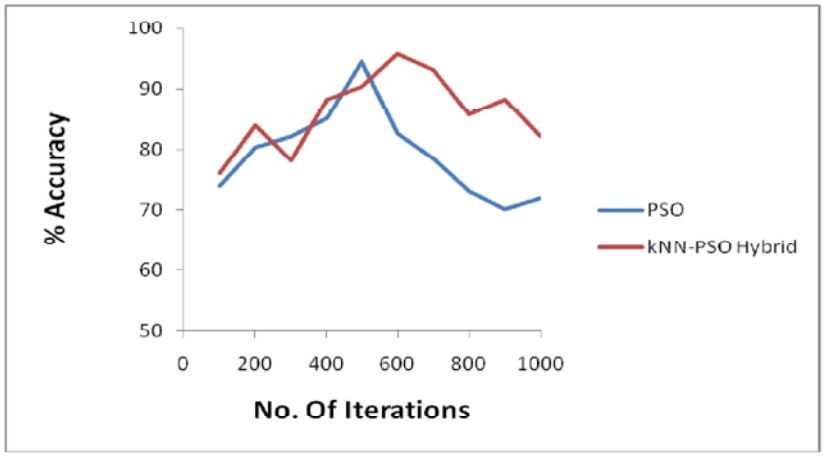

Fig.6.3(a). Percentage classification accuracy of PSO and Hybrid PSO against different iterations.

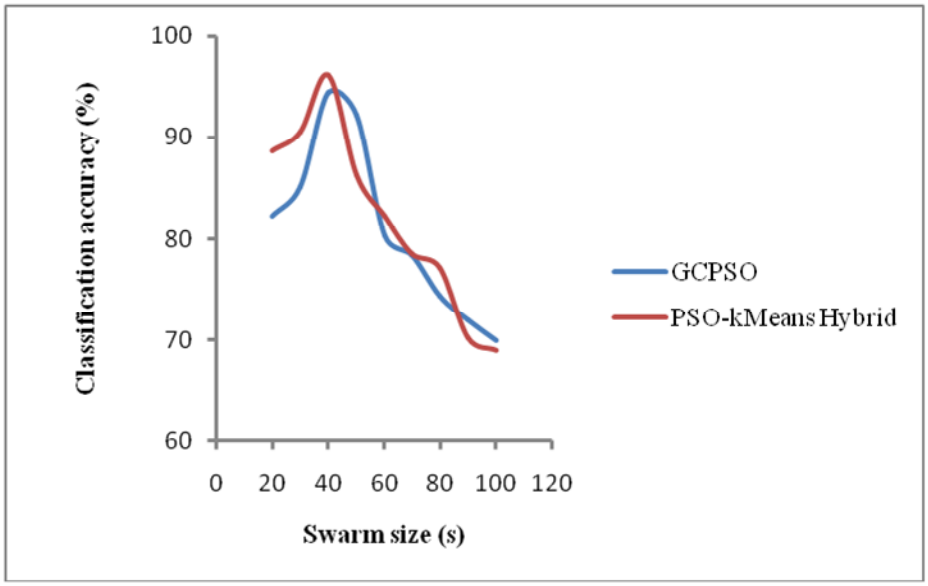

Fig.6.3(b). Percentage classification accuracy of PSO and Hybrid PSO against different values of Swarm size.

6. 6. Grading of Tumors

The main problem in designing an automated tool for grading of tumors is that the traditional MR imaging is often not adequate in telling the tumor grade. Therefore, medical practitioners rely on further investigations such as biopsy for determining the grades of tumors. In this work some features of MRI have been found that may be suggestive of some grade of brain tumor. Then a model is developed for automated grading of tumors.

For the grading of tumors we follow the following steps:

1. ROI definition
2. Feature extraction

3. Feature selection

4. Classification based on hybrid of kNN-PSO coupled with Cross validation by Leave One Out [Bo *et al.* 2006].

Firstly the heterogeneous regions of brain tumors are explored by combining imaging features from several sequences. Then morphological and textural characteristics, such as rotation invariant texture features based on Gabor filtering are extracted and the significance of each feature is assessed in classification. Multi class classification is applied for differentiating between the most common brain tumors: metastasis, meningioma (usually grade I) and glioma (grade II, II and IV).

6.6.1 Feature Extraction:

For feature extraction first of all, appropriate ROIs were manually selected. A number of features were selected which included tumor shape characteristics, image intensity characteristics within several regions of interest and Haralick texture features.

1) Shape and statistical characteristics of tumor:

2) Image intensity characteristic.

3) Haralick textural features. [Miyamoto 2005].

6.6.2 Grading Implementation and Results

For grading purpose, 156 images were considered for which the grading had been confirmed by needle biopsy method. The brain masses were graded on the basis of WHO standard as grade I, grade II, grade III and grade IV. MR images of different tumor types are shown in Fig. 8. The results of grading are tabulated below in table 8.

Table 6.5 Result of Grading of Brain Tumors

Tumor Type	No. of Samples	TP	TN	FP	FN	Precision	Recall	Accuracy
GL1	44	41	5	5	7	89.1	85.4	79.3
GL2	25	22	8	6	8	78.5	73.3	68.1
GL3	47	39	11	7	11	84.7	78.0	73.5
GL4	40	36	8	6	9	85.7	80.0	74.5

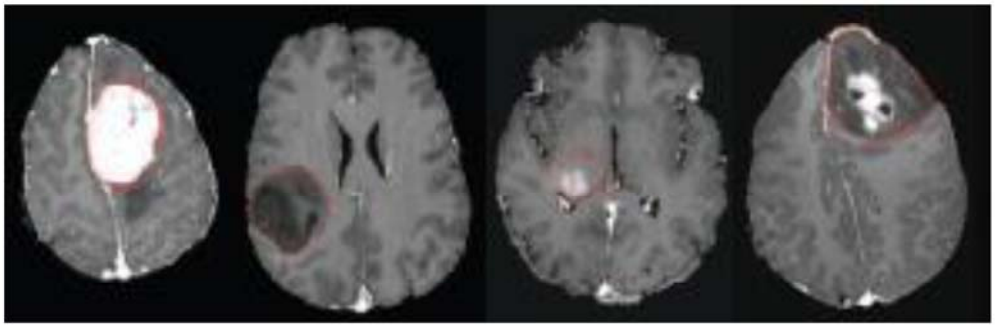

Fig.6.4. Different types of Brain tumors : from L to R: Grade1, 2, 3 &4

6.7 Conclusions

The application of different versions of PSO algorithms to the MR image segmentation and classification was explored. An algorithm based on PSO was presented for image clustering. A method using hybrid of K-Means with SBM-PSO was presented for MR image segmentation. All the proposed methods were implemented and tested with MR images obtained from various sources. It was found that PSO can be used successfully for image segmentation as well as for optimizing the clustering algorithm K-Means. Finally,

one more method using hybrid of PSO and kNN was presented, which was used for classification of MR images as normal or abnormal, on the basis of absence or presence of brain tumor. The PSO-K-Means hybrid model was further used in the grading of MR Images as per WHO scheme of brain tumors. The experimental results showed that the proposed algorithms have better segmentation results and, in general, larger inter-cluster distances and smaller intra-cluster distances. Based on the results shown in the experiments, it appears that the K-Means algorithm can be improved by the PSO technique. The combination of PSO and the K-Means algorithms performed better and had better stability than the K-Means algorithm or PSO based method alone. The "not so good" results of grading concluded that the present MRI technique is not sufficient enough for the grading purpose.

List of Publications

1. **Satish Chandra**, D.S. Chauhan "Pattern Recognition using a Hybrid method based on Neural Networks and Genetic Algorithms", *In Proceedings of International Conference of Advanced Computing ,(ICACT)* ,Pp.502-509, 2008.

2. **Chandra, S**.; Bhat, R.; Chauhan, D.S. "A Score Based Method for Controlling the Convergence Behavior of Particle Swarm Optimization," *In proceedings of 11th International Conference on Computer Modelling and Simulation, 2009. UKSIM '09.* , vol., no., pp.19-24, 25-27,March 2009, doi: 10.1109/UKSIM.2009.96 [IEEE Xplore]

3. **Chandra, S**.; Bhat, R.; Singh, H. "A PSO based method for detection of brain tumors from MRI," *In proceedings of World Congress on Nature & Biologically Inspired Computing, NaBIC 2009.* vol.V,no1 pp.666-671,doi:10.1109/NABIC.2009.5393455. [IEEE Xplore]

4. **Satish Chandra** , Rajesh Bhat , Harinder Singh , D.S.Chauhan , "Detection of Brain Tumors from MRI using Gaussian RBF kernel based Support Vector Machine ", IJACT: International Journal of Advanced Computing Technology, Vol. I, No. 1, pp. 46-51, 2009.

5. **Satish Chandra**, Rajesh Bhat, D.S. Chauhan "Detection of Brain Tumors from MRI using Polynomial kernel based Support Vector Machine", Knowledge Engineering Society Journal, England, UK. (*communicated.)*

6. **Satish Chandra** , Rajesh Bhat , Harinder Singh "Particle Swarm Optimization Based Hybrid methods for Detection of Brain Tumors from MRI", Journal of Digital Content Technology and Application (*In press*).

References

Angeline P. Evolutionary Optimization versus Particle Swarm Optimization: Philosophy and Performance Difference. In *Proceedings of the Seventh Annual Conference on Evolutionary Programming*, pp. 601-610, 1998.

Avants Brian B., Tustison Nick and Gang Song, Advanced Normalization Tools (ANTs) Release 1.5, March 30, 2010, Penn Image Computing And Science Laboratory, University of Pennsylvania. (http://www.picsl.upenn.edu/ANTS)

Barbara Zitova, Jan Flusser: Image registration methods: a survey. Image Vision Comput. 21(11): 977-1000, 2003.

Bargmann V., Michel L., and Telegdi V. L. , Precession of the Polarization of Particles Moving in a Homogeneous Electromagnetic Field, Phys. Rev. Lett. 2, 435, 1959.

Bishop C. Neural Networks for Pattern Recognition. Clarendon Press, Oxford, 1995.

Bo L., Wang L., Jiao L., Feature scaling for kernel fisher discriminant analysis using leave-one-out cross validation, Neural Computation 18: pp961-978. 2006.

Bremner D, Demaine E, Erickson J, Iacono J, Langerman S, Morin P, Toussaint G. Output-sensitive algorithms for computing nearest-neighbor decision boundaries. *Discrete and Computational Geometry 33* (4): 593-604, 2005.

Busch, C. Wavelet based texture segmentation of multi-modal tomographic images. Computer and Graphics, 21(3):347–358. 1997.

Capelle, A., Alata, O., Fernandez, C., Lefevre, S., and Ferrie, J. Unsupervised segmentation for automatic detection of brain tumors in MRI. In *proceedings of IEEE International Conference on Image Processing*, pages 613–616. 2000.

Chaa S., Update on Brain Tumor Imaging: From Anatomy to Physiology, *American Journal of Neuroradiology* 27:475-487, March 2006.

Chandra**, S**.; Bhat, R.; Chauhan, D.S. A Score Based Method for Controlling the Convergence Behavior of Particle Swarm Optimization, *In proceedings of 11th International Conference on Computer Modelling and Simulation, 2009. UKSIM '09.* , vol., no., pp.19-24, 25-27,March 2009, doi: 10.1109/UKSIM.2009.96 [IEEE Xplore]

Chawla N, Japkowicz N and Kotcz A: Editorial: Special Issue on Learning from Imbalanced Data Sets. SIGKDD Explorations 6:1–6, 2004.

Cheng H., Jaing X., Sun Y. and Wang J. Color Image Segmentation: Advances & Prospects. *Pattern Recognition*, vol.34, pp. 2259-2281, 2001.

Clark, M., Hall, L., Goldgof, D., Velthuizen, R., Murtagh, F., and Silbiger, M., Automatic tumor segmentation using knowledge- based techniques. *IEEE Transactionson Medica Imaging*, 17:238–251,1998

Dempster, A.P., Laird, N.M., Rubin, D.B. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B 39 (1): 1–38. 1977.

Dhawan, A.P. A review on biomedical image processing and future trends, *Computer Methods and Programs in Biomedicine,* vol. 31(3-4), pp. 141-183, 1990.

Dickson, S. and Thomas, B.Using neural networks to automatically detect brain tumours in MR images. *International Journal of Neural Systems,* 4(1):91–99. 1997.

Duvernoy H. M., Vannson J. L. The Human Brain: Surface, Three-Dimensional Sectional Anatomy with MRI, and Blood Supply, Springer, 1999.

Eberhart R. and Shi Y. Comparing Inertia Weights and Constriction Factors in Particle Swarm Optimization. In *Proceedings of the Congress on Evolutionary Computing*, San Diego, USA, pp. 84-89, 2000.

Egmont-Petersen M., de Ridder D., Handels H. Image processing with neural networks - a review, *Pattern Recognition,* Vol. 35, No. 10, pp. 2279-2301, 2002.

Esquivel S. and Coello C. "On the use of Particle Swarm Optimization with Multimodal Functions".*Proceedings of IEEE Congress on Evolutionary Computation,* 2003, pp 1130-1136.

Evans, A. and Collins, D. A 305-member mri-based stereotactic atlas for cbf activation studies. In 40th Annual Meeting of the Soceity for Nuclear Medicine. 1993.

Fletcher-Heath, L., Hall, L., Goldgof, D., and Murtagh, F. R. Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. *Artificial Intelligence in Medicine*, 21:43–63.2001.

Freund Yoav and Schapire Robert E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, no. 55. 1997.

Fuh C., Cho S. and Essig K. Hierarchical Color Image Region Segmentation for Content-Based Image Retreival Systems. IEEE Transactions on Image Processing, vol. 9, no. 1, pp. 156- 162, 2000.

Garcia, C. and Moreno, J. (2004). Kernel based method for segmentation and modeling of magnetic resonance images. *Lecture Notes in Computer Science*, 3315:636–645,2004.

Gering, D. Diagonalized nearest neighbor pattern matching for brain tumor segmentation. R.E. Ellis, T.M. Peters (eds), Medical Image Computing and Computer Assisted Intervention. 2003.

Gibbs, P., Buckley, D., Blackb, S., and Horsman, A. Tumor volume determination from MR images by morphological segmentation, *Physics in Medicine and Biology*, 41:2437–2446, 1996.

Goldberg. D. Genetic Algorithms in search, optimization and machine learning. Addison-Wesley, 1993.

Gonzalez Rafael C. and Woods Richard E. Digital Image Processing, 3rd ed, Prentice Hall,1992.

Gosche, K., Velthuizen, R., Murtagh, F., Arrington, J., Gross, W., Mortimer, J., and Clarke, L. Automated quantification of brain magnetic resonance image hyperintensisites using hybrid clustering and knowledge-based methods. *International Journal of Imaging Systems and Technology*, 10(3):287–293.1999.

Greenberg Harry S., Chandler William F. and Sandler Howard M. Brain Tumors, Oxford University Press, 1999.

Hajnal Joseph V. Hill Derek lg and Hawkes David J.(Ed) Medical Image Registration. CRC Press, Boca Raton, 2001.

Haralick Robert M., Shanmugam K and Dinstein I Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics,* 1973.

Higashi N. and Iba H. Particle Swarm Optimization with Gaussian Mutation. In *Proceedings of the IEEE Swarm Intelligence Symposium* Indianapolis Indiana, USA. pp. 72-79.2003.

Ho, S., Bullitt, E., and Gerig, G. Level set evolution with region competition: automatic 3D segmentation of brain tumors, In *proceedings of 16th International Conference on Pattern Recognition*, pages 532–535, 2002.

Jain R., Kasturi R. and Schunck B.. Machine Vision. McGraw-Hill, Inc., New York, USA, 1995.

Jain A., Murty M. and Flynn P. Data Clustering: A Review. *ACM Computing Surveys*, vol. 31,no. 3, pp. 264-323,1999.

Kaus, M., Warfield, S., Nabavi, A., Black, P., Jolesz, F., and Kikinis, R. Automated segmentation of MR images of brain tumors. Radiology, 218:586–591, 2001.

Kennedy J. and Eberhart R. Particle Swarm Optimization. In *Proceedings of IEEE International Conference on Neural Networks*, Perth, Australia, vol. 4, pp. 1942-1948, 1995.

Kohonen T. Self Organizing Maps, Springer, 3rd edition, 2001

Kwok James S. H. and Constantinides A.G. A Fast Recursive Shortest Spanning Tree for Image Segmentation and Edge Detection, IEEE Transactions on Image Processing, 6, 2, 328-332.1997.

Lai Shang-Hong and Ming Fang, "A new variational shape-from-orientation approach to correcting intensity inhomogeneities in magnetic resonance images", *Medical Image Analysis*, Volume 3, Issue 4, Pages 409-424 , December 1999

Lavine B. K. *et al.* Genetic algorithms for spectral pattern recognition, Vibrational Spectroscopy, Volume 28, Issue 1, 28 February 2002, Pages 83-95.

Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):897–908. 1999a.

Leemput, K., Maes, F., Vandermeulen, D., and Suentens, P. Automated model based bias field correction of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):885–896.1999b.

Lucchese L. and Mitra S. Color Image Segmentation: A State-of-the-Art Survey. In *Proceedings of the Indian National Science Academy* (INSA-A), New Delhi, India, vol. 67, no. 2, pp. 207-221, 2001.

Martin John H. Neuroanatomy: Text and Atlas, McGraw-Hill, United States, 2003.

Mazzara, G., Velthuizen, R., Pearlman, J., Greenberg, H., and Wagner, H. Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. *International Journal of Radiation Oncology,* 59(1):300–312, 2004.

McQueen J, Some Methods of Classification and Analysis of Multivariate Observations, *In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistic and Probability,* University of California Press, Berkley, CA, 1967.

Miyamoto Eizan , Merryman Thomas , Fast Calculation of Haralick Features, Technical Report, Carnegie Mellon University, www.ece.cmu.edu/pueschel /teaching /18-799B- CMU- spring05/ material/eizan-tad.pdf.2005.

Montillo Albert, Udupaa Jayaram, Leon Axelb and Dimitris Metaxasc, Interaction between noise suppression and inhomogeneity correction in MRI, *Medical Imaging, Proceedings of SPIE* Vol. 5032, 2003.

Murgasova Maria, Dyet Leigh, David Edwards,Rutherford Mary ,Hajnal Jo and Daniel Rueckert. Segmentation of brain MRI in young children. Academic Radiology, vol. 14,no. 11, pp. 1350–1366, 2007.

Norvig Peter and Russell, Artificial Intelligence: A Modern Approach, Prentice Hall, 2002.

Nyul LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. *IEEE Trans Med Imaging*. 19(2):143-50. Feb 2000.

Ozkan, M., Dawant, B., and Maciunas, R. Neural-network-based segmentation of multi-modal medical images: a comparative and prospective study. IEEE Transactions on Medical Imaging, 12(3):534–544.1993.

Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

Riget J. and Vesterstrom J. A Diversity-Guided Particle Swarm Optimizer – The ARPSO. EVALife Technical Report no. 2002-2, 2002.

Rousseeuw P.J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics.* 20. 53-65, 1987.

Rueckert D., Sonoda L. I., Hayes C., Hill D. L. G. , Leach M. O. and Hawkes D. J. Non-rigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712-721, 1999.

Russ J. C., "The Image Processing Handbook", Raleigh, NC: CRC Press, 1995.

Sammouda, R., Niki, N., and Nishitani, H. A comparison of hopfield neural network and boltzmann machine in segmenting mr images of the brain. *IEEE Transactions on Nuclear Science,* 43(6):3361–3369.2004.

Schad, L., Bluml, S., and Zuna, I. MR tissue characterization of intracranial tumors by means of texture analysis. *Magnetic Resonance Imaging*, 11(6):889–896.1993.

Schmidt Mark, A Method for Standardizing MR Intensities between Slices and VolumesTechnical Report, Department of Computing Science, University of Alberta, Edmonton, Alberta, T6G 2E8, CANADA, http://webdocs.cs.ualberta.ca/TechReports/2005/TR05-14/TR05-14.pdf

Shapiro Linda G. and Stockman George C. : Computer Vision, pp 279-325,Prentice-Hall, New Jersey,2001.

Shattuck D. W. , Sandor-Leahy S. R. ,Schaper K. A. ,Rottenberg D. A. and Leahy R. M. , Magnetic resonance image tissue classification using a partial volume model, *Neuro Image*, vol. 13, no. 5, pp. 856–876, 2001.

Shi Y and Eberhart R. Parameter Selection in Particle Swarm Optimization. *Evolutionary Programming VII: Proceedings of EP 98*, pp. 591-600, 1998a.

Shi Y and Eberhart R. A Modified Particle Swarm Optimizer. In *Proceedings of the IEEE International Conference on Evolutionary Computation*, Piscataway, New Jersey, pp. 69-73, 1998b.

Sled J. G., Zijdenbos A. P., and Evans A. C., ``A non-parametric method for automatic correction of intensity non-uniformity in MRI data, *IEEE Transactions on Medical Imaging, vol. 17, pp. 87-97, February 1998.*

Soltanian-Zadeh, H., Peck., D., Windham, J., and Mikkelsen, T. Brain tumor segmentation and characterization by pattern analysis of multispectral NMR images. *NMR Biomed,* 11(4 - 5):201–208.1998.

Storn R. and Price K., "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces", *Global Optimization*, vol. 11, 1997,pp. 341-359

Studholme C., Hill D.L.G. ,Hawkes D.J., A Normalised Entropy Measure of 3D Medical Image Alignment, In *Proceedings of Medical Imaging* 1998,San Diego, SPIE Press. pp. 132-143.

Tou J.and Gonzalez R. Pattern Recognition Principles. Addison-Wesley, Massachusetts, USA, 1974.

Trivedi M. and Bezdek J. Low-level Segmentation of Aerial Images with Fuzzy Clustering. IEEE Transactions on Systems, Man and Cybernetics, vol. 16, no. 4, pp. 589-598, 1986.

Van den Bergh F. An Analysis of Particle Swarm Optimizers, PhD Thesis.Department of Computer Science, University of Pretoria, South Africa, 2002.

Vapnik, V. N., The nature of statistical learning theory, Springer-Verlag, New York, 1995.

Veeramachaneni K, Peram T, Mohan C. and Osadciw L.. Optimization Using Particle Swarm with Near Neighbor Interactions. Lecture Notes Computer Science, vol. 2723, Springer Verlag, 2003.

Vilarino D.L. et al., Discrete-time CNN for image segmentation by active contours, Pattern Recognition Lett. 19 (8) pp. 721–734.1998.

Vinitski, S., Gonzalez, C., Mohamed, F., Iwanaga, T., Knobler, R., Khalili, K., and Mack, J.Improved intracranial lesion characterization by tissue segmentation based on a 3D feature map. *Magnetic Resonance in Medicine*, 37:457469. 1997.

Wells, W., Kikinis, R., Grimson, W., and Jolesz, F. Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging*. 1996.

Xie X, W. Zhang and Z. Yang. "A Dissipative Particle Swarm Optimization", *IEEE Congress on Evolutionary Computation*, Honolulu, Hawaii, USA, 2002.

Yoon, O.K., Kwak, D.-M., Kim, D.-W., and Park, K.-H. MR Brain Image segmentation using fuzzy clustering. In Fuzzy Systems Conference Proceddings, 1999.

Zhang, J., Ma, K., Er, M., and Chong, V. Tumor segmentation from magnetic resonance imaging by learning via one-class support vector machine. International Workshop on Advanced Image Technology, pages 207–211. 2004.

Zhang W and Xie X. "DEPSO: Hybrid Particle Swarm with Differential Evolution Operator", *IEEE International Conference on Systems,* Man and Cybernetics, Washington DC, USA, 2003, pp. 3816-3821.

Zhu, Y. and Yan, H. Computerized tumor boundary detection using a hopfield neural network, *IEEE Transactions on Medical Imaging*, 16:55–67,1997.