Segmentation and Classification of Brain Tumor MR Images through Advance Machine Learning Methods

Lubna Farhi

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

Doctor of Philosophy

(Electrical Engineering)

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,
PAKISTAN
March 2018

Abstract

The main objective of this research work is to develop, test and evaluate an identification support system that is able to provide accurate, fast and reliable diagnosis of brain tumor in MR images. Keeping in consideration that human decision making skills are mainly dependent on experience and prone to error due to fatigue, Artificial Intelligence (AI) can be utilized as an effective aid in the field of medicinal sciences for tumor diagnosis through image recognition. Therefore, this thesis strives to develop such an intelligent system that can be used for the segmentation and classification of infiltrative brain tumors known as Low Grade and High Grade in MR images.

In order to tackle the complex task of brain tumor segmentation in MR images, we present an adaptive algorithm that formulates an energy based stochastic segmentation with a level set methodology. This hybrid technique efficiently matches, segments and determines the anatomic structures within an image by using global and local energies. After evaluating the algorithm on low and high grade images, it was noted that there was an improvement in the resultant similarity between segmented and truth (original) images.

Once effective segmentation was achieved we could then work on the next step of tumor identification; classification. In the second part of the process we proposed two classification frameworks, machine learning and deep learning. In machine learning, we first extracted 22 probabilistic features using gray level co-occurrence matrix methodology that served as input features for the classifiers. Then we showed the improvement in classification (through machine learning) accuracy by providing two methodologies in which the first one involved

classification directly after feature extraction whereas in the second we reduced the extracted features using principal component analysis and then applied those reduced features to several classifiers.

The second framework that we proposed was the brain tumor classification of segmented MR images through optimized CNN-Deep belief learning model. It scales to various image sizes by distributing the hyper-parameters and weights among all locations in an image. The presented model is translation invariant and is compatible with top-down and bottom-up probabilistic inference. This hierarchical classifier was optimized by regularization, that mitigates the effect of overfitting for small datasets, stochastic gradient decent, which works efficiently by utilizing only a small set of samples from a whole training set to infer the gradient and fine tuning of constraints. A comparative analysis, based on accuracy, error/loss and computation time, was carried out between the pre-processed non-segmented and segmented MR images after classification was completed. The results showed that the accuracy of proposed optimized CNN-deep belief learning classifier with segmented MR images was higher while the loss and execution time were reduced.

These methodologies transcend the confines of MR image processing due to their effective modularity allowing them to be suitable for other medical imaging and computer vision tasks.

Chapter 1

Introduction

1.1 Research Motivation

Recent statistics from 2016 show that cancer related to the brain and nervous system is amongst the top ten leading causes of cancer-related deaths. The survival rate of men and women with brain or central nervous system cancer is 34% and 36% respectively (Cancer.Net, 2017). As technological innovation becomes more prevalent, a need for fast, reliable and non-invasive techniques using computer aided diagnostics (CAD) is becoming ever-more necessary. A foundation of CAD frameworks is the field of computerized histopathology, which could catapult the current research in the field of artificial intelligence through the automatic analysis of great amount of input data.

Brain tumors present themselves in the most subtle way such that they exhibit very general clinical symptoms and on top of that, the skull prevents any chance of early detection through physical means rendering the detection of brain tumors an extremely difficult task. The general symptoms of brain tumors can present themselves in different ways, including headaches, dizziness, vomiting, mental impairment, behavioral impairment and lastly fatigue or seizures. These symptoms may collectively indicate the presence of a tumor however they are not a definitive indication of cancer as these symptoms can also be found in other ailments.

Therefore, in order to determine if any tumorous tissue exists, non-invasive techniques such as MRI, CT or PET scans are employed.

Present practice involves a radiologist to manually segment and identify a tumor in every slice of a patient's scan. This is not only very time consuming but it is also prone to error due to fatigue experienced by humans. In addition, with the ever-increasing rate of cancer amongst people, a need for fast, reliable, accurate and autonomous method of tumor identification has risen. Therefore, researchers have turned to digital recognition solutions that can help reduce the time of detection and classification while improving the accuracy and reliability of these results.

To address this challenge, the research tries to create an innovative tool for fast segmentation and classification of brain tumor from MR images. The purpose of this tool is to be used for close monitoring of the growth or shrinkage of the brain tumor and the evolution of its treatment. As a result, image processing, machine learning and deep learning methods form the basis of the framework of this thesis.

1.2 Research Objectives and Scope

This research aims at highlighting the immense potential behind digital segmentation and classification of brain tumor. It centers around providing a digital structure of algorithms that employ novel technologies like computer vision, machine learning and the new developed deep machine learning to segment and identify brain tumors in a MR image.

The main goal behind this research is to improve the accuracy of digital diagnosis through the use of optimized segmentation and classification techniques that are robust enough to be able to detect high and low grade tumors with ease. In order to provide reliable assistance to doctors, this research also aims at improving the reproducibility of tumor identification. Hence to achieve such results, the following research objectives were set:

1. To classify and detect tumors in MR images using existing algorithms and machine

learning techniques.

2. To identify the region of interest using segmentation techniques.

3. To recognize high and low grade tumors in MR images after developing a comparison technique using feature extraction and various machine learning classifiers.

4. To apply new technology like deep machine learning to reduce the number of misdiagnosis and improve the overall accuracy of the classifier.

5. To optimize the classification techniques for brain tumor classification in MR images thereby reducing the computation time without compromising accuracy.

6. To improve the robustness and reliability of tumor identification of the developed model.

The methodologies proposed in this research depend mostly on image processing, computer vision and machine learning techniques. For a detailed description of the proposed techniques refer to the information provided in chapter one onwards.

1.3 Thesis Overview

Before delving into the further intricacies of the research, it is imperative to understand the basic framework of this thesis. The contributions made during this research are of noteworthy importance as they have the potential to be fundamental pillars of advance machine learning algorithms in the future.

Three major tasks associated with computer aided diagnostic include acquisition of data, segmentation of regions of interest, and the classification of these ROI. The thesis makes notable contributions in segmentation (fourth chapter) and machine learning classification (fifth chapter), with a potential extension with advance deep learning to classification (sixth chapter). Fig.[1.1] gives a high level overview of the broad research fields used in artificial intelligent systems.

1.3.1 Basic idea of Research

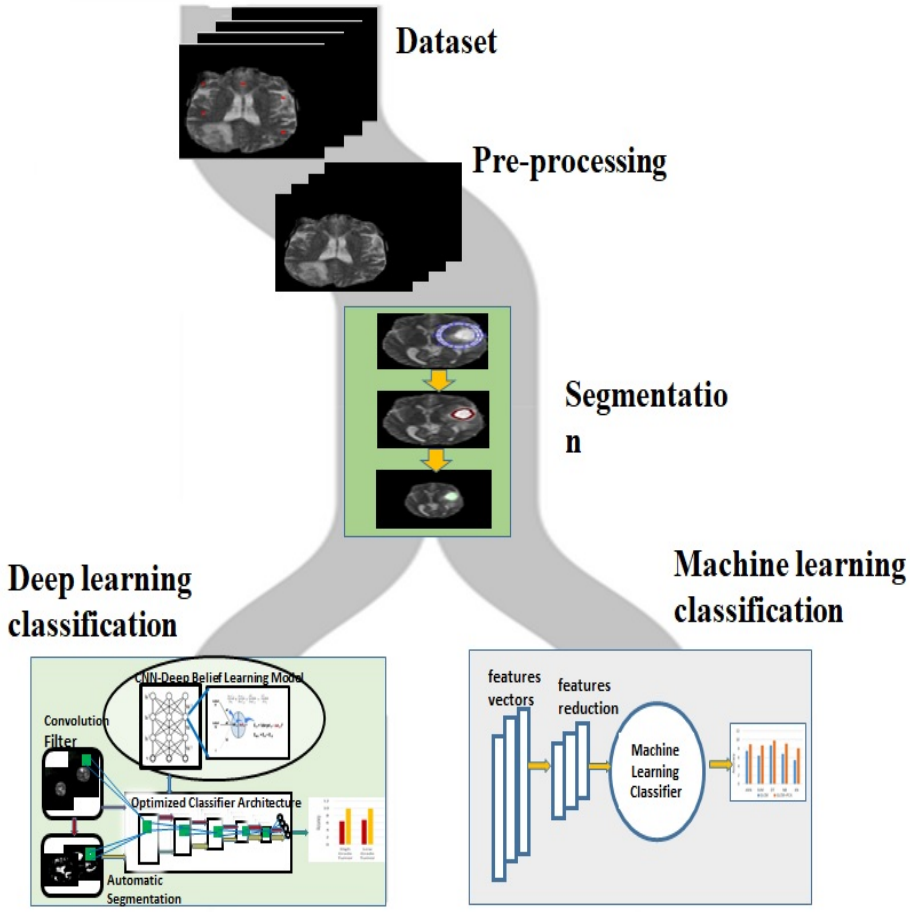


Figure 1.1 Overview of research

1.3.2 Research Contributions

The problem domain we have chosen is the identification of cancerous tumor and its grade. This thesis contributes the necessary theory and implementation for the chain of three novel algorithms. Our first contribution is to use an adaptive segmentation methodology to segment a necessary region of interest. The proposed approach provided a resultant accuracy that came up to 89.5% by dice method, and minimum distance of 0.5(mm) by Hausdorff algorithm. Lastly, it can also be utilized for segmentation even when the binary thresholding

level is greater than 0.2. Second, we contributed "Comparison of Brain Tumor MRI classification methods using Probabilistic features" to develop the benchmark methodology for tumor classification. The proposed technique improved the overall accuracy of any machine learning classifier as it was noted by the increase in classifiers accuracy from 10 to 27%. Third, we presented the approach, "Framework for Brain tumor Classification of MR images using Optimized CNN-Deep Belief Learning Model ". The proposed technique involves a fully automatic deep learning method for MR brain high and low grade tumor images' classification. The result of the application of this methodology is an increase in the accuracy of classification by 19.7% while a reduction in the loss and execution time by 60% and 10% respectively.

1.4 Thesis Organization

In this chapter 2, we initially review some of the brain and its anatomical structures, their functions and importance with respect to everyday activities. We also review how they are affected by tumors in specific regions of the brain. Then in section 2.2, we focus on Brain tumor types and their classification while section 2.3 presents the imaging aspects required to detect brain tumors in MR images. Similarly, section 2.4 provides a detailed description of how a MR image is captured and important concepts such as T1, T2, TR and TE, which are used throughout the chapter. In section 2.5 we explain the characteristics of MRI are illustrated by emphasizing the differences, and explaining their advantages and disadvantages. Finally, section 2.6 provides a summary of this chapter.

Chapter 3 presents a detailed analysis of the previous works in the field of segmentation, machine learning and deep learning. It goes into the depth of classical approaches, highlighting their strengths and weaknesses with respect to medical application for MR images. This chapter outlines the flaws in the current approaches thus providing us with a clear set of objectives that will enable us to create a viable solution. This also helps to contrast the distinctiveness of our contribution with respect to the previous work done.

Chapter 4 provides a holistic explanation of the proposed segmentation technique and this chapter comprises of six sections. Section 4.1 provides a preface to the concept of segmentation and its importance relating to imaging. Section 4.2 highlights the challenges and contribution covered by the proposed model meanwhile Section 4.3 presents a solution through the mathematical model of the adaptive stochastic segmentation via energy convergence. In addition, section 4.4 discusses the implementation of the proposed algorithm using MRI brain images. Section 4.5 has the results and discussion in which the performance of the proposed algorithm is evaluated and the results are shown. Finally, section 4.6 provides summary of presented technique.

Moving on, chapter 5 presents a proposed technique in machine learning classification. This chapter is organized such that section 5.1 provides a detailed introduction to the concept of machine learning and classification. Section 5.2 describes the challenges and contributions tackled by the presented technique while section 5.3 explains the methodology behind it and its implementation flow chart. The results are also presented and discussed in section 5.4 after which the summary is presented in section 5.5.

Furthermore, chapter 6 describes the proposed deep learning methodology for brain tumor MR images. It is organized into six sections where section 6.1 provides the introduction to deep learning. Section 6.2 then discusses the challenged faced and contributions made by the proposed algorithm using MRI brain images. The methodology has also been described in detail in section 6.3 after which section 6.4 highlights the implementation of the model. Section 6.5 has the experimental results in which the performance of the proposed algorithm is evaluated and the results are shown. Lastly, section 6.6 presents the summary regarding the proposed deep learning algorithm.

Finally, Chapter 7 concludes the thesis with summary of main results of this thesis and presents the future work.

Chapter 2

Background

This chapter introduces Brain anatomy and Magnetic resonance imaging (MRI)for the study and monitoring of brain tumors and in particular the High grade and Low grade tumors. This is carried out by focusing on two challenging and important problems in medical imaging: first, segmentation of tumor in MR image and second classification of type of tumor.

2.1 Basic of Brain Anatomy

The human body's nervous system can be separated into two separate parts; the central nervous system (CNS) and the peripheral nervous system (PNS). While the CNS covers the brain and the spinal cord, the PNS comprises of the spinal nerves extending from the spinal cord and the cranial nerves that extend from the brain (Waxman, 1999). This section mainly focuses on the anatomy of the brain and the respective structures of the cell. The brain mostly contains two different types of tissues that are known as gray matter (GM) and white matter (WM), as shown in Fig.[2.1]. The gray matter is responsible for controlling the brain activity as it is made up of neuronal and glial cells that are also known as neuroglia or glia. There are two main parts of the brain that consist of gray matter, the cortex that covers the surface of the brain and the basal nuclei that are embedded deep within the white matter as shown in

Fig.[2.1]. However, the white matter is mostly made of fibers of myelinated axons that provide a connection between the outer cerebral cortex and other regions inside the brain. White matter fibers also form a thick band known as corpus callosum that allows communication between the right and left hemispheres of the brain (Waxman, 1999). Furthermore, there is another aspect related to the brain that is very important to its proper functioning and it is called the cerebrospinal fluid (CSF). The CSF is not only found inside the brain and spinal cord but it also circulates the exterior of the brain and spinal cord where it acts as a protective layer to prevent injury (Woolsey et al., 2003). It consists of glucose, salts, enzymes and white blood cells that provide the necessary nutrients while warding off any cells or viruses that could be harmful to the brain. Lastly, to provide added protection against any physical injury to the brain or the spinal cord, there exists a tissue between the skull and the brain known as meninges.

The brain anatomy revolves mainly around three main parts; the cerebrum, cerebellum and the brain stem Fig.[2.2]. Firstly, the cerebrum being the largest part of the brain, is separated into two halves known as the left and right cerebral hemispheres along the longitudinal fissure. Then, each half is further divided into 4 areas or lobes that are known as the frontal lobe, the parietal lobe, the temporal lobe and the occipital lobe. The frontal lobe, as the name suggests, is in the front of the brain while the parietal lobe is right behind the frontal lobe but towards the top of the brain. The temporal lobe is at the side of the brain below the frontal and parietal lobes, whereas the occipital lobe is at the back of the brain right underneath the parietal lobe (as illustrated in Fig.[2.2]) (Waxman, 1999). Secondly, right below the occipital lobes, at the back of the brain exists the cerebellum which is separated from the cerebrum by a layer known as tentorium (fold of dura). Moreover, the cerebellum is similar to the cerebrum such that they share a resemblance when it comes to the spread of gray and white matter. The cerebellum, just like the cerebrum, has an outer surface cortex of gray matter, which is linked internally with white matter and also small bits of gray matter embedded deep within the white matter

inside the cerebellum. Lastly, the third part of the brain which is called the brainstem is the extension beneath the brain. It is located before the cerebellum and connects to the spinal cord. The brain stem is also made of gray matter, covered in white matter fibrous tracts and comprises three substructures called, the midbrain, pons and medulla oblongata (as shown in Fig.[2.2]). The midbrain can be found underneath the hypothalamus while the pons connect the medulla and midbrain. Finally, the medulla is partially merged with the spinal cord to act as a relay between the brain and the spinal cord.

In order to further understand the intricacies of the brain, it is imperative to know the central structure of the brain. This part includes but is not limited to the thalamus, pineal gland, hypothalamus and the pituitary gland. The thalamus is responsible for relaying any and every bit of relevant sensation/information to the cortex and plays an important role in experiencing sensations, alertness and memory. The pineal gland serves the function of regulating the internal clock and circadian rhythm of the body by releasing special hormones. Additionally, the hypothalamus is one of the most important section of the brain as it is in-charge of controlling the body's autonomous functions like sleep, thirst, hunger, blood pressure, body temperature etc. Lastly the pituitary gland, also known as the master gland, is responsible for controlling other glands in the body. It secretes hormones that are responsible for bone and muscle growth, stress relief and preventing diseases. However, since all of these structures exist in the confined center of the brain, it makes this area very sensitive to damage. If any one of these structures is damaged, the others will be influenced almost immediately. Therefore to protect the brain from any physical injury or other type of damage, the brain itself employed a preventative mechanism known as the ventricular system. The ventricular system is made up of four cavities, also called ventricles, which exist in different areas inside the brain. They contain a special structure that is responsible for producing the cerebrospinal fluid (CSF) which is mainly responsible for protecting the brain. These ventricles, are connected to each other via holes also known as foramen, and tubes. There are two ventricles that exist inside the cerebral

hemisphere and are called the lateral ventricles. The two lateral ventricles are connected to the third ventricle, present in the center of the brain, through a tube known as the foramen of Monro. Then the third ventricle is joined to the fourth via a narrow tube referred to as the aqueduct of Sylvius since the fourth ventricle is relatively smaller and exists towards the bottom side of the brain. Finally, from the fourth ventricle, the CSF spread to the outside of the brain where it encompasses the whole organ and serves as a physical cushion against physical shock (Waxman, 1999; Woolsey et al., 2003).

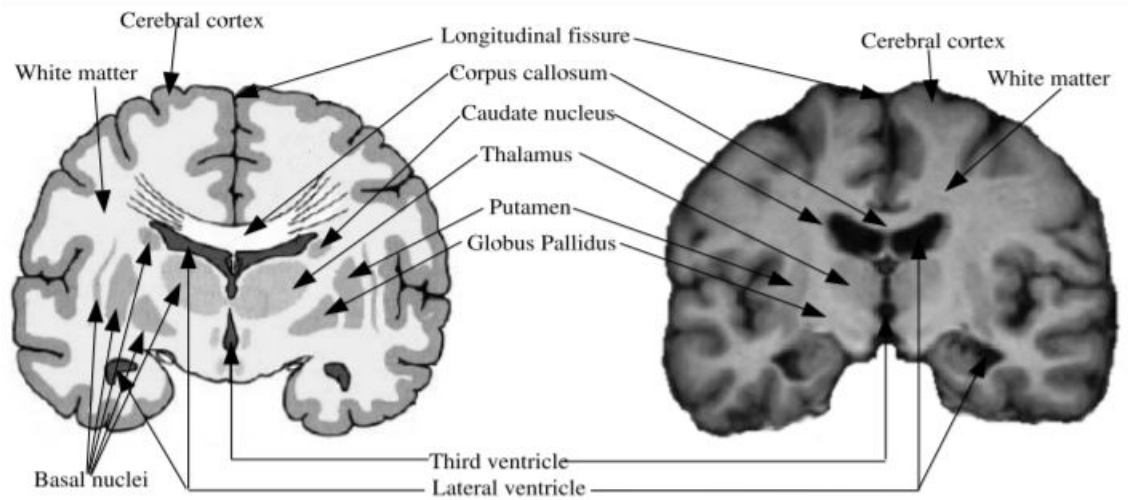


Figure 2.1 Brain anatomy described on a MR image schematic (left) and on a slice (right) (Marieb, 2000)

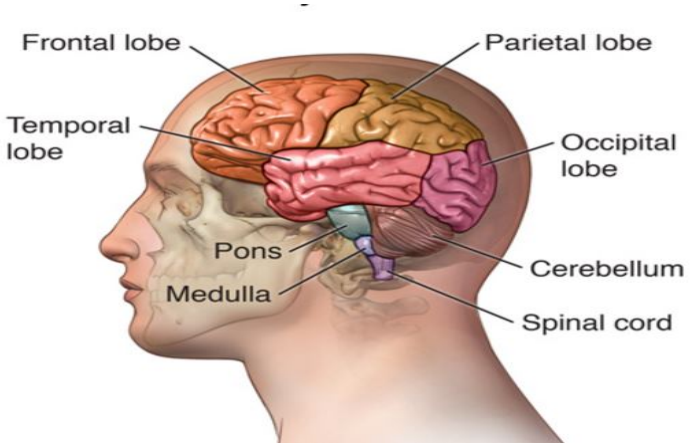


Figure 2.2 Anatomy of the brain (reproduced from (Waxman, 1999))

2.2 Brain tumor Classification

Cancer is a disease referred to the uncontrollable growth of abnormal cells in the human body. These rapidly growing abnormal cells cause an interference with the essential bodily functions performed by the healthy cells which can lead to death. There are two important types of brain cancers; primary brain tumor which forms in the brain and spreads outwards and secondary brain tumor which forms somewhere else in the body and spreads to the brain. Tumors are classified based on how the cells appear microscopically. They are assigned a grade which defines the type of tumor and provides insight as to the cell's growth rate. The first type of tumor is known as grade I in which the tumor is benign and has slow growth while the tumorous cells resemble ordinary brain cells. However in Grade II, the tumor becomes malignant and the cells have a more deformed look as compared to the tumorous cells in Grade I. In Grade III, the tumorous region is not only growing actively but the cells also have a considerable different appearance from the normal cells. Whereas in the last type, Grade IV, the malignant tumor is a rapidly growing region of abnormal-looking tissue that differs from the ordinary brain cells the most. Consequently, the four types of tumors are categorized based on their severity where the low grade tumors are not as severe i.e. grade I and II while the high grade tumors are critically severe i.e. grade III and IV (Farhi and Yusuf, 2017a).

The World Health Organization (WHO) has provided a detailed list of the different types of CNS tumors according to genetics and histology. However, majority of the tumors mentioned in this thesis can be categorized as a class of meningiomas, metastases or gliomas.

2.2.1 Meningiomas

One of the most common types of brain tumor in adults (more common in women than in men) are known as meningiomas. This type of cancer occurs when cells from specific layers of tissue, which cover the brain and spinal cord and are called meninges, start to grow abnormally. Once this tumorous tissue starts to grow, it presses on nearby tissue or the spinal cord which

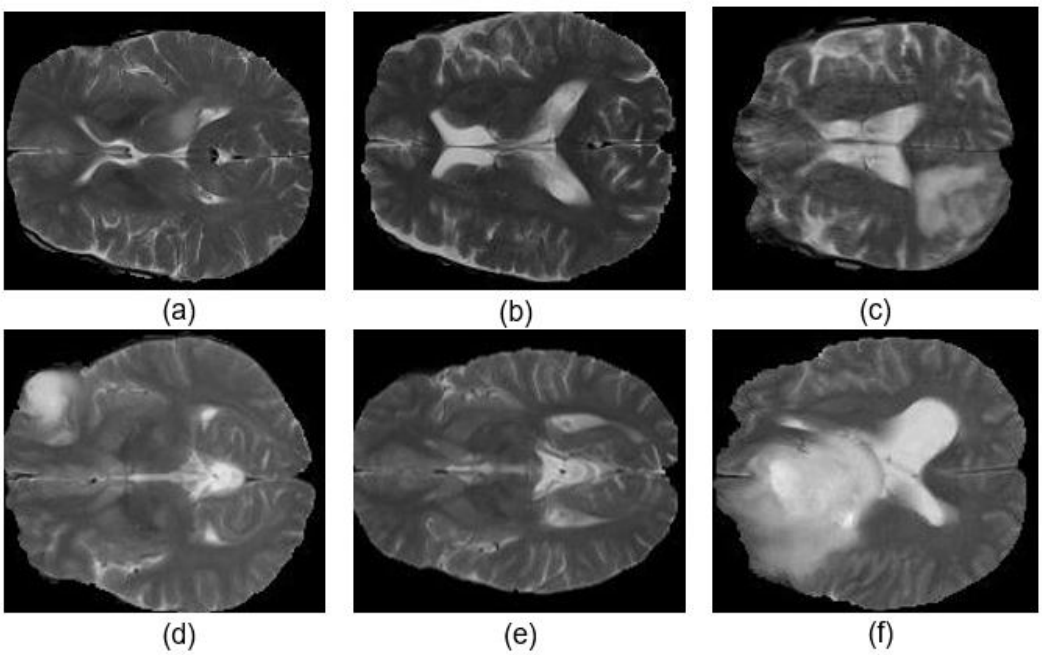


Figure 2.3 Brain tumor images (a)(b)(c) High-grade (Glioblastoma) and (d)(e)(f) Low-grade (Meningioma) MICCAI (2016)

can lead to headaches or limpness in arms or legs. Meningiomas are mostly benign (they are usually grade I) which means they grow very slowly and can be removed through surgery. However sometimes (approximately 20 %), this type of cancer can evolve to become grade II (atypical) or grade III (malignant) in which case it can spread externally inside the body and have disastrous effects (Duffau and Capelle, 2004).

2.2.2 Gliomas

Another type of brain tumor that forms about half and one-fifth of all primary brain and spinal cord tumors respectively is known as glioma. It forms due to the abnormal growth of glial cells in the cerebral hemisphere, brain stem, optic nerves, spinal cord or cerebellum. Due to such wide spread presence of glial cells in the upper part of the body, gliomas are classified according to the origins of their glial cells. Consequently there are numerous categories of different types of gliomas amongst which, some like astrocytoma, ganglioglioma, oligoden-

droglioma and ependymoma are the most common. Here we discuss astrocytoma due to its relevance to our research.

Astrocytoma Astrocytomas are a type of glioma that are formed due to the abnormal growth of a specific type of glial cell, a star-shaped looking connective tissue cell known as astrocyte Fig.[2.4]. Astrocytomas are referred to as one of the most common types of brain tumors since out of all the primary brain tumors they make up approximately 40% of the cases. Moreover, not only are astrocytomas very common, they are also very dangerous as they have been classified malignant by WHO and St-Anne grading system. This system grades a type of tumor based on some specific characteristics like atypia, mitoses, endothelial proliferation, and necrosis (Daumas-Duport, 1992; Daumas-Duport et al., 2000; Lopes and Laws Jr, 2002; Smirniotopoulos, 1999). The presence of these aspects signifies the danger potential of the tumor in relation to rate of growth and invasiveness. Resultantly, grade I tumors are those which do not exhibit any of these characteristics while grade II tumors are those which show at least one of these features (mostly atypia). Similarly, tumors are classified as grade III or grade IV when they appear to have 2 or more features respectively. Hence it can be noted that the grade I and II tumors are the low grade group of astrocytoma while grade III and IV make up the high grade astrocytomas.

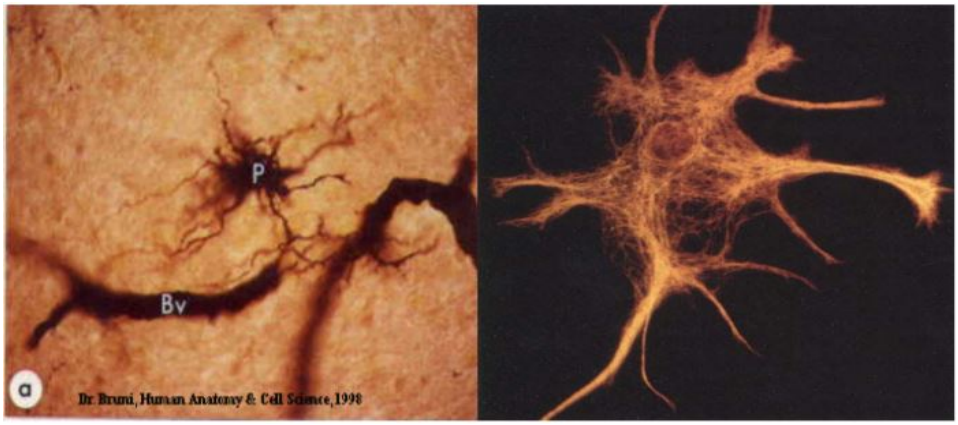


Figure 2.4 Typical Astrocyte Cell (Wen and Teoh, 2003)

Low grade astrocytomas can although be differentiated quite well from normal cells, they can grow to other neighboring tissue despite having slow growth. Moreover, due to their sluggish growth and little vasogenic edema they cause relatively less mass effect as compared to high grade astrocytomas. They are usually found in the cerebral hemisphere, more specifically in the frontal region or the sobcortical white matter, the cerebellum or brain stem of young adults and children. Also, the two most common types of low grade astrocytomas are pilocytic astrocytoma and diffuse astrocytoma (Henson et al., 2005; http://www.emedicine.com, 2005; Wen and Teoh, 2003). And both of these can be detected via CT scan or MRI, although MRI is often the preferred choice due to several reasons. One reason is that low grade gliomas can be differentiated well from normal tissue since they show a highly mitigated signal relative to the neighboring brain tissue in T1 sequences. However, in T2 and FLAIR sequences, this differentiation disappears because a higher signal can imply either the tumor or the surrounding edema thus making the diagnosis less accurate. However, pilocytic astrocytomas are usually found along with a cyst which can be easily detected via T2-weighted sequences. (Daumas-Duport, 1992; Henson et al., 2005; Kantor et al., 2001; Wen and Teoh, 2003).

The most typical forms of high grade astrocytoma that make up about 30% of all primary brain tumors are known as anaplastic astrocytoma and glioblastoma multiform (GBM). Just like other high grade tumors, these also multiply extremely quickly and are also very invasive towards the neighboring cells. Since these tumors do not show a lot of differentiation, they appear almost similar in their features. However, a general feature that has been noted in both that can be used to separate them apart from low grade astrocytomas is that they are usually less circumscribed and contain more edema in the surroundings. Moreover, the fundamental difference between the two high grade astrocytomas, anaplastic astrocytomas and GBMs, is the appearance of necrosis. High grade astrocytomas also exhibit a variable appearance in a radio-image such that the tumor may show itself as an amalgamation of both low and high density regions inside one lesion. Amongst the two high grade astrocytomas, GBM is the most

malignant and apparently also the most common type of glial tumor. GBMs are most often found in adults in the cerebral hemispheres and they show invariant neoplastic astrocytes as compared to the surroundings. Despite being mostly found in adults, GBMs can also crop up in children and be found near the brain stem or spinal cord areas. Finally, GBMs can also evolve from lower-grade astrocytomas (grade II) or anaplastic astrocytomas (grade III) (Patel and Tse, 2004; Wen and Teoh, 2003).

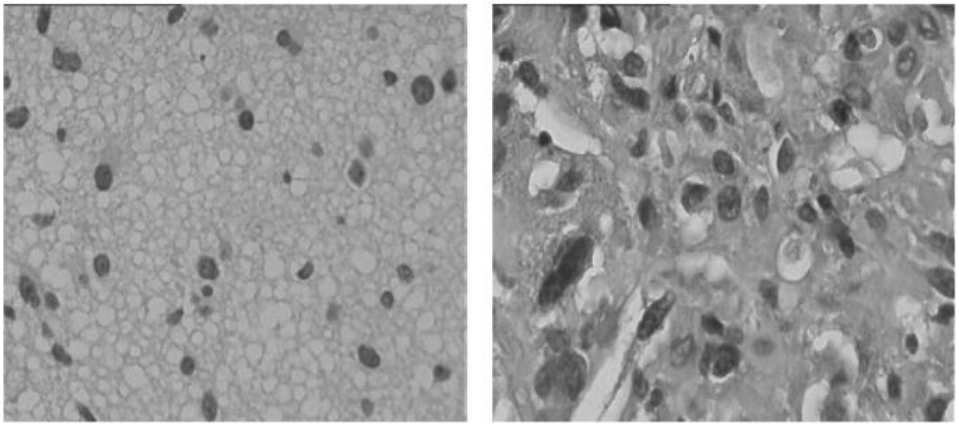


Figure 2.5 Left grade II (Low-grade) and Right grade IV (high-grade) astrocytomas (Patel and Tse, 2004)

2.2.3 Metastatic

Apart from primary brain tumors, a large portion of the general population also suffers from secondary brain tumors also known as Metastatic tumors. They form in another part of the body (its primary site) but then spread elsewhere via blood or adjacent tissue. The primary sites that are most commonly found to develop into metastatic cancer in adults are lungs, breasts, colon and kidney. Since a metastatic tumor is basically a periphery of the primary tumor, it exhibits similar characteristics like being malignant and rapidly growing. However sometimes, during the spreading of the tumor, a metastatic tumor may spread to various sites at once and resultantly be found in different locations around the brain. This type of tumor infects mostly women or middle-aged and elderly men. Once diagnosed with this type of

cancer, a patient usually has around 2 to 16 months to live while he undergoes surgery or chemotherapy to tackle it (Duffau and Capelle, 2004).

2.3 Brain tumors in Magnetic resonance imaging (MRI)

In order to properly treat the brain tumors within patients, the imaging aspect required to detect the tumor plays a vital role. Imaging of the brain is not only important because it can be used to find the tumor, but it is also used in other stages during the diagnosis, treatment and results. Some of these stages are discussed below:

- identification or verification of the presence of an abnormality in the brain,
- determination of the size of the abnormality,
- categorization of the type of abnormality,
- discerning the tumor's stage i.e. how critical it is,
- deciding on whether the current diagnosis needs further clarification, or otherwise proceeding with surgery or other therapies,
- assessment and control of the rejection rate of the body to the current therapy,
- observing the results of the approved therapy.

As a result of such a high level of involvement in the prognosis of a tumor, non-intrusive imaging has become a very important aspect of medical science. Therefore, over the year several different types of imaging techniques have emerged which include Magnetic Resonance Imaging (MRI), Computed Tomography (CT) Scan, Single Photon Emission Computed Tomography (SPECT) Scan, Positron Emission Tomographic (PET) Scan etc. However, amongst all the aforementioned techniques, MRI and CT scans are used the most often since they are readily available and produce better quality images of human organs and tissues relative to other imaging techniques. In addition, CT offers the fastest modality thus making it suitable for imaging those patients who are in need of immediate attention or if they are critically ill. Furthermore, PET and SPECT imaging techniques are more suitable for smaller roles where

their ability to determine tissue physiology and biology can be especially needed. As a result of this ability, PET scanning is also able to identify the tumor grade.

2.4 Physics of MRI

The main reason why MRI is so safe to use, is because it relies mostly on the use of magnetic fields and our body's reaction to it. MRI mostly depends on the presence of large quantities of hydrogen atoms in our bodies and how it reacts to a magnetic field. Since the hydrogen atoms are spinning around their own axis naturally, they have their own magnetic fields. However these magnetic fields cancel each other out due to the random orientation of the atoms in space. Therefore, when these atoms come under the influence of a noticeably large magnetic field B_o, the atoms align themselves and spin with (positive spin) or against (negative spin) the applied magnetic field. Apart from spin on their axes, the atoms also have a precession frequency (Lamor frequence), ω_o, with which they precess around the field and the amplitude of precession is proportional to the magnetic field's magnitude:

$$\omega_o = \gamma B_o \tag{2.1}$$

Once the atoms are spinning in the orientation of the magnetic field, their own magnetic fields' vectors can be defined with two components; a longitudinal component P_z that is parallel to the external field B_o and a transverse component P_{xy} that is perpendicular to B_o. Moreover it is mostly noted that in reality the number of spins in positive alignment is higher than the negative spins thus resulting in a net longitudinal magnetization P_z in the direction of the magnetic field. However the net transverse component of all atoms is mostly zero because the atoms do not precess in phase which causes the perpendicular components to nullify each other.

Then, upon the application of an external radio-frequency pulse B_1, that has the same

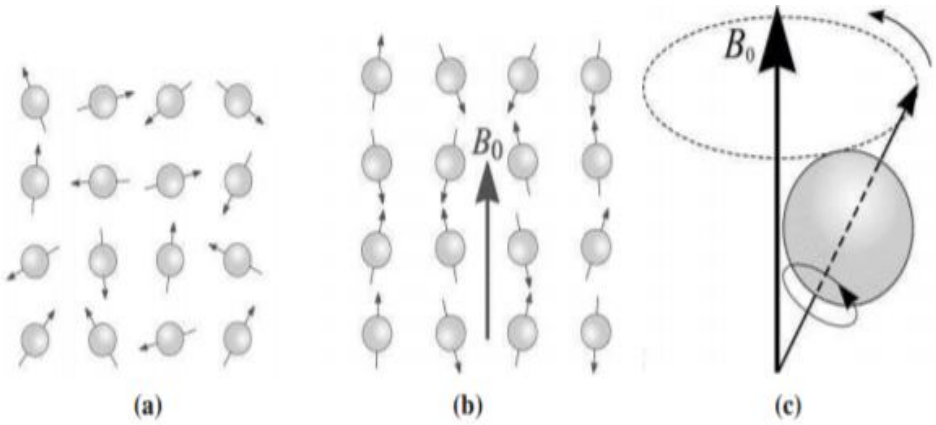


Figure 2.6 Magnetic field impact on the spins' magnetization. (a) Orientation of random spins, (b) Spins alignment with magnetic field, (c) Spin precession around magnetic field (reproduced from (Puddephat, 2002))

magnitude as the natural frequency ω_o, the spinning atoms exchange energy in which they experience resonance resulting in an increase in their amplitude of precession. The effect of this increase in precession causes the net magnetization vector to deviate away from its initial axis (longitudinal axis) and towards the transverse axis. This deviation creates a transverse component which can be detected by a receiver coil. Moreover, to maximize the transverse signal, the pulse B_1 is kept perpendicular to the original magnetic field B_O in order to eliminate the effects of the longitudinal component as much as possible. Once the externally applied RF signal is switched off, the atoms undergo the relaxation phenomenon in which they start to return to their natural state (equilibrium) and re-emit the previously absorbed energy which is detected as the MR signal. In practice there are two different relaxation processes that are detected; the longitudinal relaxation and the transverse relaxation. In the longitudinal relaxation, the atoms return to their natural state by dispersing the energy through the lattice. Since the longitudinal relaxation is related to spin-lattice interaction, the atoms re-align themselves with the original magnetic field B_O. In addition, the longitudinal magnetic vector recovers according to an exponential curve where the specific time constant T1 signifies when 63% of

the original vector has recovered.

$$P_z = P_o(1 - exp^{(-t/T)}) \tag{2.2}$$

Where, the initial magnetization vector or the net magnetization at equilibrium is P_z, which is mostly dependent on the magnitude of the external magnetic field and the proton density. Furthermore, in the transverse relaxation phase, the atoms lose their precession amplitude and they start to get out of phase. This results in a reduction in the transverse component of the magnetic field in an exponential manner where the component is defined via a tissue specific time constant T_2.

$$P_{xy} = P_{xyo}(1 - exp^{(-t/T)}) \tag{2.3}$$

Where, the amplitude of the transverse signal after applying the RF pulse is signified by P_{xyo}. Moreover, if the RF pulse is applied perpendicularly to the external magnetic field B_o then P_{xyo} becomes equal to P_o. It should also be noted that T_1 is always larger than T_2 and that these two constants are very important in defining the image contrast and differentiating the different MRI sequences.

2.4.1 MR image Sequences

In order to obtain an image from the re-transmitted signal from the atoms during relaxation, the MRI technique uses two parameters known as the Echo Time (TE) and the Repetition Time (TR). TE is the time recorded between the emission of one RF pulse and its return like an echo, whereas the time between two successive RF pulses is defined as TR. Since the RF pulse is applied perpendicularly to the magnetic field, each MR signal received represents the relaxation evolution of the transverse component. In addition, the signal has a frequency that matches the resonance frequency and it diminishes according to the exponential rule.

Table 2.1 Comparison of T1-weighted vs. T2-weighted vs. Flair (reproduced from Preston (2006))

Type of tissue	T1 Weighted scan	T2 Weighted scan	Flair
CSF	Dark	Bright	Dark
White Matter	Light	Dark Gray	Dark Gray
Cortex	Gray	Light Gray	Light Gray
Fat(within bone marrow)	Bright	Light	Light
Inflammation (infection, demyelination)	Dark	Bright	Bright

The amplitude of the first signal peak relates to the longitudinal component's magnitude and affects the recovery status. Finally, the received signal will be noticeable if the longitudinal component has been recovered properly.

The second parameter, TR, is used to identify the time duration between two consecutive pulses thus determining the time for longitudinal relaxation. In case a short TR is used, the tissues with long T1 for longitudinal magnetization will not be able to recover in time resulting in a weaker signal meanwhile the tissues with a short T1 will provide a higher magnitude due to their proper recovery (larger magnetization). This dependence of the tissues on their T1 values is very important in developing image contrast thus defining the T1-weighted modality. However, if the TR is kept long enough (same initial magnitude P_o) that all of the tissues are able to recover, then the image contrast will no longer depend on TR. In that case, a long TE can be used to develop an image which is dependent on the transverse relaxation T2. Unlike T1, the signal from transverse recovery will be weaker for tissues with a smaller T2 (quicker transversal decay) therefore providing us with another way of obtaining an image. These images are called T2-weighted images. However, in a case where TR is kept long and TE is kept short the differences between T1 and T2 disappear resulting in a new type of image knowns as the Proton Density weighted image. In this image, the received signal's strength relies on the proton density and consequently the acquired magnetization P_o.

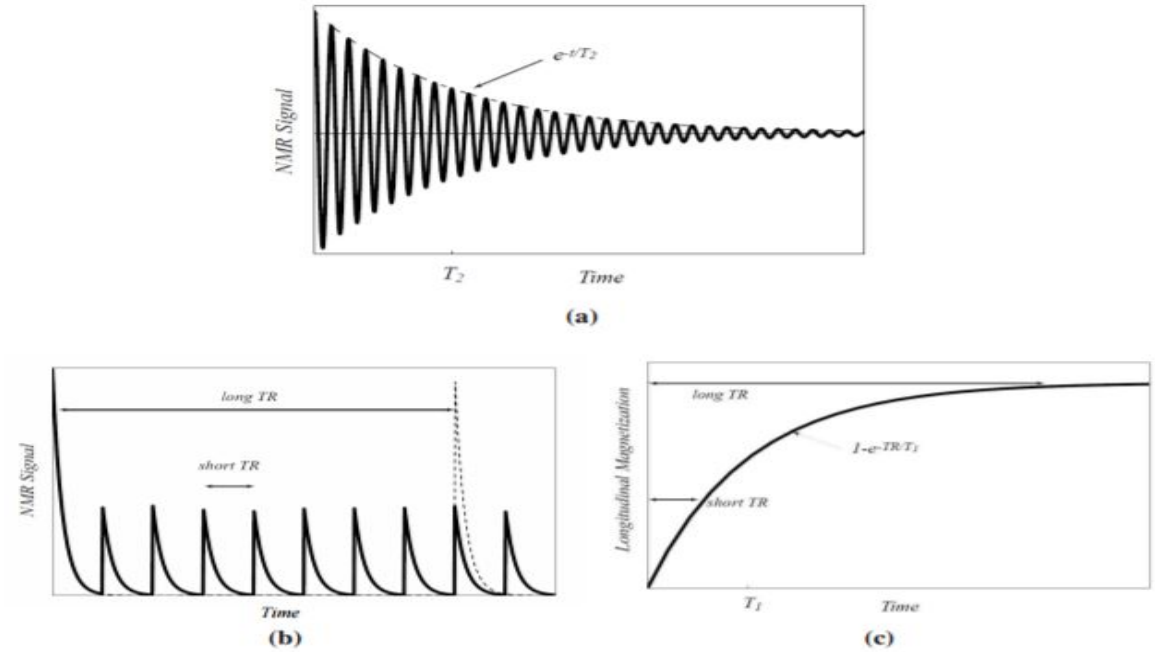


Figure 2.7 Effect of TE and TR on the MR signal (reproduce from (Buxton, 2009)) (a) MR signal, amplitude varying with time, (b) Effect of TR on the initial amplitude of the signal, (c) longitudinal magnetization with respect to TR and T1

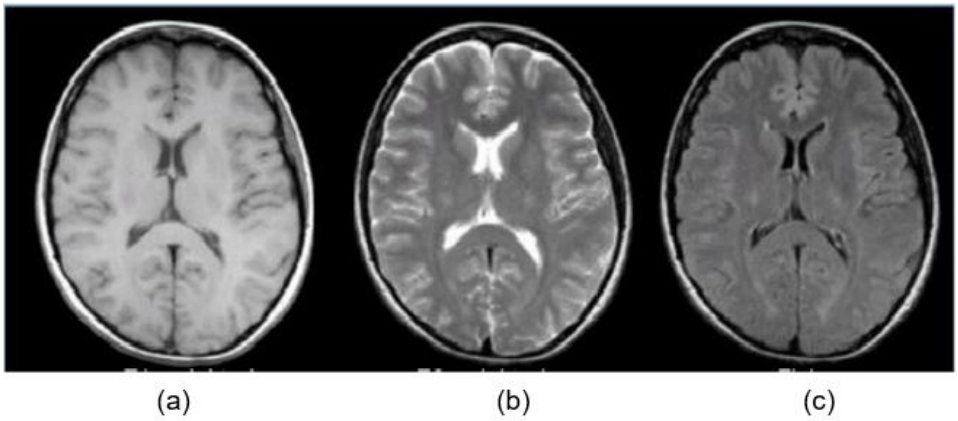


Figure 2.8 Brain images sequences (a) T1 weighted scan, (b) T2 weighted scan, (c) FLAIR (MICCAI, 2016)

2.5 MRI advantages and limitations

Amongst the various techniques mentioned in the previous section, MRI is the most commonly used methodology for obtaining images of the brain starting from the evaluation up till the assessment phase of a patient's diagnosis. MR imaging offers various advantages that allow it to be more suitable for neuroimaging, for example it does not require ionizing radiation for imaging unlike PET, CT and SPECT scans hence it is safer to use. The contrast resolution of a MR image is also higher as compared to the images of other techniques, thus making it easier to detect minute lesions or defects. It is also more suitable for lesion enhancement due to its higher sensitivity relative to CT and other techniques. Moreover, MR imaging has the ability to attain a 3D image which can be divided into three planes, sagittal, axial and coronal. This ability provides a holistic view of the whole target organ which makes it easier to detect a tumor more accurately. The beam-hardening artifact that is produced near the skull in a CT scan is also removed in a MR image which allows for improved detection of defects near the posterior fossa or in the frontal and temporal lobes. Finally, MR imaging also provides a convenient way to obtain functional and anatomical information from just one scan of the tumor effectively reducing the time needed to get that information, energy spent on each scan and also being more cost-effective (Ricci and Dungan, 2001).

Despite having numerous advantages, MR imaging technique has several drawbacks that need to be taken into account. One of the most prominent limitation of MRI is a lack of specificity. Additionally, in MR images several pathological lesions present themselves as hypointense in a T1-weighted image or hyperintense in a T2-weighted image (Kufe et al., 2003). It must be noted that there is no specifically defined trend between the level of enhancement and the histologic tumor grade, however high grade tumors usually show more enhancement on MR images relative to low grade. Although, this phenomenon can be seen for most lesions, there is an exception like the juvenile pilocytic astrocytoma (JPA) which is a slow-growth tumor showing a higher level of enhancement. Conversely, there are other excep-

tions in which some high grade tumor does not show enhancement (Kufe et al., 2003). Thus, diagnosis through features of a lesion in a MR image is not a satisfactory method. In fact, it is normal practice to acquire histologic verification in light of the MR image to come to a conclusive diagnosis. Furthermore, one of the major disadvantages of MR imaging is its inability to clearly differentiate the boundary of a tumor. Sometimes, the MR image shows the borders of abnormal contrast enhancement, but upon further inspection tumor cells are found beyond the proposed border of the tumor thus reducing the reliability factor of MRI technique (Kufe et al., 2003). Lastly, imaging abnormalities that are observed after a treatment for a patient seem to be indistinguishable. For example, in an MR image, effects of radiation therapy have a resemblance with tumorous tissue that might occur due to regrowth. Therefore, it can be noted that the MRI technique cannot provide conclusive evidence about the presence of tumor after radiation therapy.

However, it must be realized that the benefits of MRI outweigh its limitations thus making it the preferred imaging method in neuro-oncology.

2.6 Summary

In this chapter, we first reviewed the anatomy of brain, Brain tumor classification, MR imaging of brain tumor and MRI pros and cons with physics of MRI. The main result of this section is that the enhancement of tumors in T2-weighted images is a key sequence for malignant tumors. The T1-weighted and Flair images are not a proper modality for accurate edema segmentation, while T2-weighted image is more adequate for accurate edema segmentation and classification. Based on these results and tumors characteristics review, we decided to use the T2-weighted images for brain tumors segmentation and classification.

Chapter 3

Literature Survey

Ever since the advent of medicinal practice, doctors and practitioners have emphasized their choice of treatment after extensive examination. Even in the early years of human civilization, physicians and doctors provided a diagnosis of the ailment before actually starting any treatment. As time moved on, innovations in the field of engineering also had an impact on the treatments offered by the doctors. However, development in the detection and diagnosis aspect of medicine was necessary before the treatment methods could be improved. Once this was realized, medical diagnosis was revolutionized through the invention of several equipments such as the electrocardiograph machine, ultrasound machine etc (Health, 2017). These machines not only expedited the process of a patient's recovery but they also improved the diagnosis accuracy while using non-invasive techniques. As this luxury of medicinal technology became a norm, the burden of human dependence on technology increased. Therefore, researchers are constantly looking to state-of-the-art means of medicinal diagnosis and care.

3.1 Introduction

In this chapter we discuss the previous approaches for segmentation and classification of brain tumor. This allows for the study of previous shortcomings and thus motivates the need for

novel solutions.

The goal of present day researchers is to provide personalized medicinal care to every individual patient in the future. The aim of this idea is to reduce the chance of error and collateral damage while improving the overall standard of care holistically. The key to a tailored treatment for each patient is the collection, analysis and management of vast amounts of data in minimum time and therefore to address such an issue, researchers have taken the aid of digital algorithms. These algorithms will possess the ability to process and analyze medical data with immense speed and accuracy thus allowing us to reach our goal of personalized autonomous care.

3.2 Related Work in Segmentation

In order to tackle the complex problem of image segmentation, several methods have been proposed by researchers amongst which, the existing level set function technique has been widely accepted. A summary of the significant work on brain tumor segmentation is presented in Table [3.1]. The level set function can be categorized into two classes: edge and region based. There have been some function models of the two classes presented to solve a variety of problems. One of the approach proposed by Zhu and Yuille (1996) suggested a region based flow, which is based on competition work. This algorithm combines the snake and region growing approaches. They claimed that it is a multi-band segmentation algorithm for gray level images, color images and texture images. There are numerous benefits of region-based methodology such as, vigor against initial curve position and obtuseness to noise of image. However, this method uses the global statistics which are typically noneffective for heterogeneous objects segmentation because segmented objects can not be easily differentiated due to global statistics. Heterogeneous objects are often present in medical images, therefore to accurately perform the segmentation of objects, new energies should be considered that utilize local statistics as well as include the benefits of region-based model. To solve this problem,

Table 3.1 Summary of previous work related to Segmentation of brain tumor

Author(year)	**Title**	**Methodology description**
Ho et al. (2002)	"Level-Set Evolution with Region Competition: Automatic 3-D Segmentation of Brain Tumors"	Developed 3D Level set method is based on the region competition framework with smoothness constraints and changeable topology
Prastawa et al. (2004)	"A brain tumor segmentation framework based on outlier detection"	A methodology based on three steps,first detects the abnormal region based on intensity parameters,finally by using spatial geometric properties detect the ROI.
Lee et al. (2005)	"Segmenting Brain Tumors with Conditional Random Fields and Support Vector Machines"	The Method is based on DRFs (Discriminative random fields) for segmentation and SVM for classification.
Dou et al. (2007)	"A framework of fuzzy information fusion for the segmentation of brain tumor tissues on MR images"	Automated fuzzy segmentation based on fuzzy models and fuzzy fusion techniques.
Khotanlou et al. (2009)	"3D brain tumor segmentation in MRI using fuzzy classification, symmetry analysis and spatially constrained deformable models"	Proposed method for 3D MR images. In this method first tumor detection based on selecting asymmetric areas with respect to brain symmetry plane and apply fuzzy classification.
Wang et al. (2011)	"Automatic MRI brain tumor contour models"	It is a multi - threshold algorithm,by using dilation and erosion operators to obtain an initial contour from the tumor domain image.

Chan and Vese introduced active contours without edges (Chan and Vese, 2001). This methodology was based on curve evolution, Mumford-Shah function for segmentation and level set techniques. This model does not depend on the gradient of image, therefore the initial curve can be anywhere in the image, but the final contour will always build around the region of interest. Similarly, Yezzi and Willsky in 2002, concentrated on a global statistics of image

segmentation by using the equations of curve. The suggested algorithm separated the values of certain image statistics over identified various region types (Yezzi et al., 2000). This methodology was based on statistics of deterministic principle of maximally separating those values that are within a set of curves and geometric constraints of active contour of background subject. It can be noted that these models consider the image region information as a whole while neglecting the influence of local information. As a result, these can detect objects with weak boundaries and would not segment images with intensity inhomogeneities.

The most recent approach was presented by Taheri et al. (2010), merging the level-sets and threshold-based. They proposed a level set segmentation based on global threshold scheme. The author claimed that the results were better when compared with the region competition method. Various models have used region-based active contours that were generally used for image regions of constant intensity. Region-based approaches have numerous advantages over edge-based methods. However, region-based techniques employ global statistics and thus are not suitable for segmenting heterogeneous objects. Another adaptable approach for segmentation are the deformable models. It can be categorized into parametric and geometric models. Parametric deformable models permit interaction directly with the model during deformation, therefore it can be used for fast real-time implementation. However, splitting or merging during deformation while implementing model topology can hinder the use of parametric models. Geometric models represent curves as a scalar function centered on the philosophy of energies and the level set method can handle topological changes easily (Zhang and Freedman, 2003). Recently in 2017, Ilunga et al. proposed a localized active contour model with background intensity compensation (LACM-BIC) applied on automatic MR brain tumor segmentation (Ilunga-Mbuyamba et al., 2017). The author used Hierarchical centroid shape descriptor (HCSD) by applying the k-dimensional tree algorithm for detecting the region of interest which is enclosed in a bounding box and balances the mean intensity distance between an image foreground and background. In the results, they compared the proposed

algorithm (LACM-BIC) with local binary fitting (LBF) model, Gaussian distribution fitting (LGDF), Chan-Vase (C-V) and localized MS (LMS) models and also compared the evaluation parameters between the models mentioned above. However, it can be noted, that the different energies and forces that are imperative for convergence, stabilization and optimization of the local boundary were neglected and for accurate analysis of algorithm, comparison of ground truth images and updated methodologies with results to support the proposed methodology.

Although deformable models have been popular in medical image analysis, they have some noteworthy drawbacks. There are major issues related to segmentation of MR images i,e. noise during the acquisition process, intensity homogeneity,object-region which have heterogeneous measurements can not be segmented with global-energies, since the energies fizzle and instability of the contour around the region of interest.

3.3 Literature Review of Classification with Classical Machine Learning

Over the recent years, emphasis has been put on identifying different brain tumors by combining MR/CT imaging aspects into pattern recognition techniques via machine learning methodologies. A number of approaches have been used to classify the brain tumor types in MR images. Table [3.2] shows the prior studies on brain tumor classification. From the literature review, it is evident that as the data increases dimensionally, analysis and classification of data becomes harder. Furthermore, it can also lead to the data becoming increasingly sparse in its occupied space, which can result in complications for both supervised and unsupervised learning. This limitation along with its predicaments is referred to as the curse of dimensionality by Powell (2007). In the case of supervised learning, the training data may be too small (few data objects) to allow the establishment of a consistent model for assigning a class to all possible objects. Where as in the case of unsupervised learning models or clustering method-

Table 3.2 Comparison of past work.

Author(year)	Tumor Type	Total Images	Feature Extraction	Classifier	Accur.(%)
Zhang et al. (2011a)	Normal,Abnormal	66	wavelet	NN	98
EL-Sayed A.El-Dahshan et al. (2009)	Normal,Abnormal	70	wavelet	ANN	98
				KNN	93
et al (2015)	Grade I-III	30	GLCM	PNN	97
Gonal and Kohir (2015)	Normal,Abnormal	25	wavelet	k-mean	98
(Zacharaki et al., 2009)	High,Low grade	102	texture	SVM	87
Abdullah et al. (2011)	Normal,Abnormal	32	wavelet	SVM	85
John et al. (2012a)	Normal,Abnormal	20	GLCM	PNN	98
Al-Badarneh et al. (2012)	Normal,Abnormal	275	Wavelet	ANN,	98
				KNN	92
Vidyarthi and Mittal (2014)	Normal,Abnormal	80	DWT	SVM,	91
				ANN,	93
				KNN	95
Al-Badarneh et al. (2012)	Class I-IV	200	GLCM	sBAM	95

ologies, increased dimensionality may cause the data to lose its credibility since significant definitions such as distance between points or density become more uniform. Resultantly, it can be noted that high dimensional data (features set) leads to inferior accuracy and clusters of poor quality. Not only is the quality affected but they have high computational cost and memory usage. Apart from this, it is also mentioned by that an improved comprehensible model and simple use of different visual techniques can be achieved by a reduction of the attribute space. Numerous extensive surveys of various feature selection and feature reduction techniques can be found in the literature. For example, Gonal and Kohir (2015) provided a solution to classify normal and abnormal brain MR images using wavelet component and k-Means classifier. GLCM was used to extract the texture feature vector and the Euclidean distance of 24 (normal and abnormal) MR images. They found that normal brain images had a smaller Euclidean distance as compared to abnormal brain images. As a result, they claimed that the proposed system had 100% accuracy. However, a neural network with probabilistic

approach was used in (EL-Sayed A.El-Dahshan et al., 2009; et al, 2015; John et al., 2012a), using the same texture based features without dimensionality reduction. After reviewing these papers, it was noted that when the classifier was used for a small dataset of 20-60 images the accuracy achieved was 100%. This occurs because in machine learning models if we have relatively few data points (small dataset), bias-variance becomes a problem and generates over fitting, and therefore to avoid over fitting, the size of the dataset should increase.

Many researchers gives hybrid approach for accurate classification. For instance, in Zhang et al. (2011a) took a different approach and proposed "A hybrid method for MR brain image classification". This paper is based on a neural network (NN) and scale conjugate classification approach for normal and abnormal tumor on MR brain images. They applied wavelet transform method for feature extraction, and PCA for feature reduction. NN classifier with back propagation and scaled conjugate (SCG) function is used for classification of 66 MR images (18 normal and 48 abnormal images). Resultantly, they claimed that the SCG method hybrid with BPNN showed 100% accuracy with less execution time. In the same way, Al-Badarneh et al. (2012) also published a hybrid approach by using the ANN and KNN classifiers. They used DWT (Discrete Wavelet Transform) for feature extraction, feature selection via PCA and classification through ANN and KNN. For their research they used 275 MR brain images, the size of 256 x 256 pixels amongst which 94 were normal and 181 abnormal. By using the GLCM method, they extracted 278 textures and used them as input data for ANN and KNN classifiers separately. In their results, they concluded that KNN achieved the highest accuracy of 100%, while ANN achieved an accuracy of 98.92%.

As it can be noted , the authors either used the Discrete Wavelet transformation (DWT) (Zhang et al., 2011a) or GLCM method for feature extraction (Al-Badarneh et al., 2012; Machhale et al., 2015) , however the resultant dimensionality of feature vector was further reduced by using PCA. Therefore it can be conclusively stated that feature extraction with reduction method improves the accuracy and performance of classifier even if the dataset is small or

large. Additionally, Abdullah et al. (2011) compared the SVM classifier results with principal components analysis (PCA) and without PCA for normal and abnormal brain MR images . First, they gathered 32 patients' (22 abnormal and 10 normal) T2 FLAIR weighted images of 256 x 256 size. Then, they applied wavelet transform to remove the noise. Mean and standard deviation were used as a feature vector. Later on, the feature set was reduced using PCA for the transformation of feature vector into new, lower dimension feature space. The resultant reduced feature vector was used as input data of SVM classifier. Finally, the results indicated that SVM classifier without PCA had 65% accuracy, whereas with PCA, it was 85% accurate. She conclusively proved in this research that a classifier could not work efficiently for huge feature datasets without using dimensionally reduced feature dataset of MR images.

3.4 Related work in Classification with Deep Learning

In recent years, various machine learning domain models have been presented for classification of brain tumors from MR images, which involve classical pattern recognition methods, neural networks, support vector machine, clustering classifiers, and probabilistic and statistical techniques. Some notable models that utilize these techniques include Fisher linear Discriminant analysis (Sun et al., 2012), multilateral perceptron (Gholami et al., 2013), artificial neural network and support vector machine. Similarly, Sun et al. (2012) showed that the Eigen gene mined by independent component analysis was an effective feature that could be used for tumor classification. The author successfully applied Eigen gene with SVM classifier to increase the diversity of weaker classifiers. In addition,Farhi and Yusuf (2017b) represented five classifiers performances for high and low grade tumor classification with respect to MR images. The author first extracted the probabilistic features and then proceeded to reduce the dimensionality of features with the execution of Principal component analysis (PCA). Finally, in the classification stage five classifiers performances, with the emphasis, on supervised machine learning were compared. In the presence of all the pre-existing techniques, Chua and Yang

(1988) developed a new technique in signal processing called the "Cellular Neural Network (CNN) model" that could be implemented as an integrated circuit. The CNN function is a 2D filter that performs parallel processing of the input image space while delivering a continuous output unlike any other conventional 2D digital filter. This allows the filter to process a larger image size in real time hence making CNN-based imaging an effective option to process complex images. Hence, the accuracy of classification is dependent on the effectiveness of the preliminary process.

Various brain tumor automatic segmentation and classification methods used hand designed features (Menze et al., 2015). These strategies execute an established machine learning technique according to which feature vector are extracted first and then apply to classifier without effecting the nature of those extracted features. An other approach for designing task adapted feature representations is to learn a architecture of increasingly complex features directly from in-domain data.

Bengio et al. (2013) showed that deep neural network have been known to excel at learning such feature hierarchies. As technology progressed, machine learning algorithms evolved into more effective identification and classification tools. One such technology that stemmed from this advancement was Deep learning. It is a growing phenomenon that has been identified as one of the top emerging technologies of MIT (ning).

Deep learning is an upgraded form of artificial neural networks since it comprises of more layers that allow higher levels of abstraction and improved data predictions (LeCun et al., 2015). Among the Deep learning technologies, convolution neural networks (CNNs) have specifically proven themselves to be useful tools that are capable of handling a wide variety of computer vision tasks (Weng et al., 2017). The success of deep belief convolution network lies in its ability to autonomously learn mid-level and high-level abstractions from the analyzed image. This makes it a very effective tool that can be used for recognizing tumors and localizations in natural images. A common CNN structure comprises of multiple layers of

convolution filters that are mixed with a multitude of data reduction or pooling layers. This allows the CNN to typically provide one or more probabilities or class labels as its output. In addition, the convolution filters are developed using a training data set that educates the filters by allowing them to pre-process the image automatically without the need of tediously hand-crafting features that are application specific. Resultantly, CNNs have been employed in numerous identification problems, amongst which some are shown in Table [3.3].

Table 3.3 Summary of previous work related to Classification by Deep learning

Author(year)	Title	Methodology description
Xu et al. (2015)	"Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation."	Proposed a CNN- ImageNet based solution for classification and segmentation by using small dataset. The training set comprises of 23 GBM and 22 LGG images
Barata et al. (2013)	"Two Systems for the detection of melanomas in dermoscopic images using texture and color Features."	The author compared two methodologies to classify melanomas in dermoscopic images and claimed a global method to classify skin lesions.
Krizhevsky et al. (2012a)	"ImageNet Classification with Deep Convolutional Neural Networks."	A methodology based on deep convolutional neural network, it is equipped for accomplishing better outcomes on a very difficult dataset utilizing purely supervised learning.
Chato and Latifi (2017)	"Machine Learning and Deep Learning techniques to predict Overall Survival of Brain Tumor Patients using MRI Images."	This paper used BRATS2017 dataset, which included 163 images. Extracted features were volumetric, statistical and intensity texture, histograms and deep features; classification techniques were support vector machine, k-nearest neighbors, linear discriminant, tree, ensemble and logistic regression.

3.5 Summary

After an extensive survey of the research done in the three aforementioned categories, it is evident that currently there exists no such comprehensive model capable of addressing all aspects of the problem at once. In order to have a pragmatic solution that is viable for clinical use, it is essential that the proposed solution should be accurate, fast, scalable and cost-effective. Therefore, in this thesis we present Advance classification with Deep learning framework in the last chapter with the aim of providing a milestone, if not a complete solution, in the field of computer aided diagnostics.

Chapter 4

Brain Tumor Segmentation

In this chapter, the challenging and important problems in the segmentation for brain tumor MR images have been discussed. The segmentation procedure aims at creating a boundary around a particular object of interest to determine its location in an image. This task is extremely difficult once the boundaries become irregular or the appearance of the object of interest become more heterogeneous with the surroundings.

This chapter proposed an adaptive algorithm that formulates an energy based stochastic segmentation with a level set methodology. The hybrid method uses global and local energies, which are efficient in matching, segmenting and tracing anatomic structures by exploiting constraints computed from the data of the image. The algorithm performs autonomous stochastic segmentation of tumor in Magnetic Resonance Imaging (MRI) by combining region based level sets globally and three established energies (uniform, separation and histogram) in a local framework. The local region is defined by the segmentation boundary which, in the case of level set method, consists of global statistics and local energies of every individual point and the local region is then updated by minimizing (or maximizing) the energies.

4.1 Introduction

The brain has a very complex structure and by nature it is tightly bounded within the skull that makes it more complex to diagnose its diseases. The abnormal growth in brain cells creates a cluster known as brain tumor. Diagnosis of a brain tumor is a very intricate process. There are two general classifications of tumors; they could either be benign or malignant (John et al., 2012b). In the case of a benign tumor, the tumorous mass lacks the ability to attack adjacent healthy cells which means that it is unable to metastasize therefore it is termed as non-cancerous. Nevertheless, in some cases it can transform into a malignant tumor which is basically a rapid growth of cancerous cells. If this type of tumor is left untreated it can lead to death (Charfi et al., 2014a).

MRI aid to identify whether the growth is a primary brain tumor or if it is a secondary cancer that has spread to the brain from another part of the body. Sometimes partial or full tumor removal is the only way of accurate medical diagnosis. The prime task in all this process is to acquire precise locality of the tumor in MR images. Image segmentation here can aid in the identification and isolation of such effected regions. Taking medical image processing into consideration, the hefty query set and the intricacy of the anatomic shapes of interest, it can be noted that the process of image segmentation is challenging where it structures and then recreates an accurate depiction of these structures. Digital recognition and computer vision techniques heavily rely on image segmentation in the area of medical image research. Segmentation is the process implemented on an image to delineate certain regions of interest for further processing. Manual tumor segmentation performed by neurosurgeons or radiologists can not only suffer from the subjective nature of human judgment and errors caused by over-whelming amount of work but it is also a very time intensive task. Due to this, automatic segmentation methods can prove very beneficial as they can be developed to be more precise and mitigate all of the sources of errors. However this task is not easy, especially in the case of brain tumors, due to the similarity in the appearance of cancerous

tissue with healthy tissue and also the irregularities with respect to the locations of the tumor in the brain. Moreover, the unpredictable size, shape and irregular boundaries add to the complexity of detection of tumors in images through digital means (Chang et al., 2008). As a result, methods like thresholding or region growing are not able to handle this complex task Gibbs et al. (1996). Keeping in consideration all the aspects mentioned above, reliable and accurate segmentation of brain tumors in MR images remains a challenge despite the best efforts of researchers and improvement in computational informatics.

4.2 Challenges and Contribution

Although prior models have been popular in medical image analysis, they have some noteworthy drawbacks. There are four major issues related to segmentation of MR images. The first issue is related to noise during the acquisition process that can change the pixel intensity such that region of interest (ROI) recognition becomes indeterminate . Secondly, intensity homogeneity, i.e. a distinct type of tissue can have the same intensity level progressively along the image. This adds complexity in precise detection of the desired boundary. Thirdly, object-region which have heterogeneous measurements can not be segmented with global-energies, since the energies fizzle. The fourth issue is the instability of the contour around the region of interest that corresponds with the speed of each point of contour. The advantages and disadvantages of commonly used segmentation methods for brain MR images are summarized in Table [4.2] (El-Dahshan et al., 2014). In this chapter an active region-based contour model for localization and global set curve evolution adaptive stochastic approach is presented. This helps create an entire region based energy minimization. This algorithm is based on two optimized approaches, level set and region-energy. The algorithm utilizes a single level set as an initial curve to simplify the progression process and improve each point along the curve by local region energies. Four main contributions have been proposed in this research. Firstly, the framework can be used to localize any energy based region. Secondly, a localized active

Table 4.1 Advantages and Disadvantages of the previous approaches related to brain tumor Segmentation

Approach	**Advantages**	**Disadvantages**
Level set	Level set approach is a powerful method that can be used for the medical image segmentation, since it can deal with any of holes, concavities, convolution, splitting, or merging.	In any case, this approach needs initial seed for starting curves and can only provide better results if these curves are put close symmetrically with respect to object boundary.
Deformable model	This model readily extends or contracts after some time inside an image and conform to particular image features.	If these models are connected to noisy images which have ill-define boundaries then the resultant shapes may have conflicting topology when compared to the original object.
Threshold method	These methods are especially helpful for image linearization which is an exceptionally fundamental task for segmentation.	This approach does not work appropriately for all MR images due to a high variation in the foreground and background image intensity.
Region based	Region growing methods can correctly separate the regions that have same properties. It also performs well in the presence of noise.	It requires a seed point that is manually chosen and eliminates all pixels associated with the preliminary seed, based on some predefined condition.

contour is produced to inter-relate with one another to create n-array partitions. Thirdly, the control over stability of the contour around the region of interest using energies and forces of each point. Finally the impact of local radius on a tumor segmentation is analyzed. Therefore, the presented approach showed a capability of reaching maximum accuracy in minimum computation time for the brain tumor segmentation in MRI.

4.3 Methodology of Adaptive stochastic segmentation via energy-convergence

The proposed methodology, is a hybrid model of level set and region based energies. Level set is used for initialization of mask and region based energies converge the mask around the desired region.

4.3.1 Level set

Introduced by Osher and Sethian (1988), level set method is a technique used to describe shapes and track affecting interfaces. It produces the track of 3D surface by changing the movement of a planar curve. According to this method, level set boundary is the zero-crossing of the level set function φ which can change front boundary $\gamma(t)$ (Zhang and Freedman, 2003). Partial differential equations (PDE) are used to determine the implicit level set function φ:

$$\frac{\partial \varphi}{\partial t} = S.|\nabla \varphi| \tag{4.1}$$

The speed function S is the scalar velocity which is based on the curvature and external parameters like the image gradient which is related to the image data. The gradient operator is denoted by ∇. The algorithm expresses the speed function as $S = S(C)$, in which the local

curvature C is given by:

$$C = div\left(\frac{\nabla\varphi}{|\nabla\varphi|}\right) = \nabla.\frac{\nabla\varphi}{|\nabla\varphi|}$$

$$= \frac{\varphi_{ii}\varphi_j^2 - 2\varphi_i\varphi_j\varphi_{ij} + \varphi_{jj}\varphi_i^2}{(\varphi_i^2 + \varphi_j^2)^{\frac{3}{2}}} \tag{4.2}$$

The zero level set function $(\varphi = 0)$, at time t defines the developed surface. Therefore, the iterative deformation of $\gamma(t)$ is in the normal direction with S. While, at each iteration, $\gamma(t)$ position is calculated by the following equation:

$$\gamma(i,j,t) = \frac{(i,j)}{\varphi(i,j,t)} = 0 \tag{4.3}$$

The initial function φ is formulated by using the Euclidean distance d among a signed image point of initial γ_0 which is:

$$\varphi_o(i,j) = \pm d(i,j) \tag{4.4}$$

If the point i lies inside the boundary then the distance d is assigned a negative sign, whereas if the point i lies outside the boundary, it is assigned a positive sign. Then, the region of interest manifests on positions where $\varphi = 0$ (Osher and Fedkiw, 2003).

4.3.2 Region-based Framework

As an alternative of global statistics, every point's local energy along the curve will also contribute with foreground and background for establishing the local regions. Furthermore, to improve these local energies, each point will play an individual role by moving so as to minimize (or maximize) the computed energy in its respective local region. In order to calculate these local energies, the evolving curve is used to segregate the local neighborhoods into local interior and local exterior. Hence, energy optimization is achieved by assigning an appropriate mode to every discrete local region. Therefore, C can be describe as a closed curve,

represented as the zero level set of function φ

$$C = \{i | \varphi(i, j, t) = 0\} \tag{4.5}$$

Heaviside function has the following approximation (Tsai et al., 2003) for C interior

$$H\varphi(i) = \left\{ \begin{array}{ll} 1 & \varphi(i) < -\in \\ 0 & \varphi(i) > \in \\ \frac{1}{2}\{1 + \frac{\varphi}{\in} + \frac{1}{\pi}\sin(\frac{\pi\varphi(i)}{\in}) & otherwise\} \end{array} \right\} \tag{4.6}$$

and for the exterior C is $(1 - H\varphi(i))$

In order to detect the area just around the mask take the derivative of $H\varphi(i)$. The derivative of the Heaviside function is known as Dirac delta function $\delta\varphi(i)$:

$$\delta\varphi(i) = \left\{ \begin{array}{ll} 1 & \varphi(i) < -\in \\ 0 & \varphi(i) > \in \\ \frac{1}{2}\{1 + \cos(\frac{\pi\varphi(i)}{\in}) & otherwise\} \end{array} \right\} \tag{4.7}$$

Now introduce second independent variables i and j, that represent every single position of the curve. By establishing a characteristic function $M(i, j)$, as a mask of local region function in terms of a radius parameter r

$$M(i, j) = \left\{ \begin{array}{ll} 1 & \| i - j \| < r \\ 0 & otherwise \end{array} \right\} \tag{4.8}$$

When the point j (centered at i) is within a curve of radius $r, M(i, j)$, will be 1 otherwise it will be 0. This interaction produces the interior and exterior regions. Now, the energy function is

illustrated in terms of force function F, which is given as follows:

$$E(\varphi) = \int_{\alpha_i} \delta\varphi(i) \int_{\alpha_j} M(i,j).F(I(j).\varphi(j))djdi \tag{4.9}$$

The force function $F(I(j))$ defines the internal energy measure that represents each point along the contour for a given model. For finding $E(\varphi)$, consider only those points which are near the curve region. Considering eq.(4.9), it is obvious that a wide range of an object is covered by multiplying the Dirac function $\delta\varphi(i)$ in the outer integral over i. This term ensures that the mask will not change the topology by sudden development of latest contours, even though it still allows contours to split and merge. For each point i selected by $\delta\varphi(i)$, and using characteristic function $M(i,j)$ to guarantee that force drives only on local image information about i. Thus the total contribution of the first term in eq.(4.9) is the sum of $F(I(j).\varphi(j))$ values for every $M(i,j)$ neighborhood along the zero level set according to a parameter r (radius). Finally, for curve smoothness, add a parameter λ. which compensates for the removal of arc length of the curve. The final statistical energy is as follows:

$$E(\varphi) = \int_{\alpha_i} \delta\varphi(i) \int_{\alpha_j} M(i,j).F(I(j).\varphi(j))djdi + \lambda \int_{\alpha_i} \delta\varphi(i) \parallel \nabla\varphi(i) \parallel di \tag{4.10}$$

Variation of energy is given as:

$$\frac{\partial\varphi}{\partial t}(i) = \delta\varphi(i) \int_{\alpha_j} M(i,j).\nabla\varphi(j)F(I(j).\varphi(j))djdi + \lambda\delta\varphi(i)div\left(\frac{\nabla\varphi(i)}{|\nabla\varphi(i)|}\right) \tag{4.11}$$

Variation can be computed only to ensure that all regional energies can be put into defined region.

4.3.3 Internal energies

Specifically, there are three internal energies that contribute to localization: the uniform modeling energy, means separation energy, and histogram separation energy.

In this algorithm, segmentation frame is computed, in terms of internal energies. The methodology describe how localization can be improved by these energies while maintaining the stability of the curve. Now briefly present how every individual global energy gives an insightful representation of its response and how it contributes in a framework. The global mean intensities of the interior and exterior regions are utilized by common methodologies which are represented here as as x and y, respectively (Prastawa et al., 2004).

$$x = \frac{\int_{\alpha_j} H\varphi(j).I(j)dj}{\int_{\alpha_j} H\varphi(j)dj} \tag{4.12}$$

$$y = \frac{\int_{\alpha_j} (1 - H\varphi(j)).I(j)dj}{\int_{\alpha_j} (1 - H\varphi(j))dj} \tag{4.13}$$

The basis of the internal energy function is the local mean intensities that separate the regions. Now localized versions of the interior mean intensity x_i, and exterior mean intensity y_i, in terms of characteristic function $M(i, j)$ at a point i is given as:

$$x_i = \frac{\int_{\alpha_j} M(i,j).H\varphi(j).I(j)dj}{\int_{\alpha_j} M(i,j).H\varphi(j)dj} \tag{4.14}$$

$$y_i = \frac{\int_{\alpha_j} M(i,j).(1 - H\varphi(j)).I(j)dj}{\int_{\alpha_j} M(i,j).(1 - H\varphi(j))dj} \tag{4.15}$$

To formulate local energies at every individual point on the curve, these localized statistics are used, which reduce the instability of every point of the contour.

Uniform Modeling Energy (UM)

Uniform modeling energy is the Chan constant intensity model (Chan and Vese, 2001),

$$E_{UM} = \int_{\alpha_j} H\varphi(j)(I(j)-x)^2 + (1-H\varphi(j))(I(j)-y)^2 dj \tag{4.16}$$

The foreground (x) and background (y) are constant mean intensities. Therefore the related internal energy force function F, is formed by interchanging the global x and y with their respective local means as x_i and y_i.

$$F_{UM} = \int_{\alpha_j} H\varphi(j)(I(j)-x_i)^2 + (1-H\varphi(j))(I(j)-y_i)^2 dj \tag{4.17}$$

It is entirely local energy force. Moreover, to find the implicit level set function φ, take the derivative of eq.(4.17) with respect to $\varphi(j)$.

$$\nabla_{\varphi(j)} F_{UM} = \delta\varphi(j)((I(j)-x_i)^2 - (I(j)-y_i)^2) \tag{4.18}$$

By substituting above in eq.(4.11), then localized uniform energy model is:

$$\begin{aligned}\frac{\partial\varphi}{\partial t}(i) = \delta\varphi(i)\int_{\alpha_j} M(i,j)\delta\varphi(j).((I(j)-x_i)^2 - (I()j-y_i)^2)dj \\ +\lambda\delta\varphi(i)div\left(\frac{\nabla\varphi(i)}{|\nabla\varphi(i)|}\right)\end{aligned} \tag{4.19}$$

The uniform energy determines the minimum energy of the exterior and interior regions when they are calculated by global x and y. Next, after the relocation of every point on the curve surface, x_i and y_i is estimated for local interior and exterior of each of those points and the uniform models compute minimum energy locally.

Mean Separation Energy (MS)

Mean separation energy is given as:

$$E_{MS} = \int_{\alpha_j} (x_i - y_i)^2 \tag{4.20}$$

This energy relies on the fact that both of the background and foreground regions should have extremely distinct mean intensities. The interior and exterior means need to have a large variance in order to update the energy produced so as to move the curve. The force F has been formed by confining the universal energy with the local mean as follows:

$$F_{MS} = (x_i - y_i)^2 \tag{4.21}$$

By inserting this in eq.(4.11), then get localized region flow

$$\begin{aligned} \frac{\partial \varphi}{\partial t}(i) = \delta\varphi(i) \int_{\alpha_j} M(i,j)\delta\varphi(j).\left(\frac{(I(j)-x_i)^2}{A_x} - \frac{(I(j)-y_i)^2}{A_y}\right) dj \\ +\lambda\,\delta\varphi(i) div\left(\frac{\nabla\varphi(i)}{|\nabla\varphi(i)|}\right) \end{aligned} \tag{4.22}$$

Here, A_x and A_y are local interior and exterior regions areas:

$$A_x = \int_{\alpha_j} M(i,j).H\varphi(j)dj \tag{4.23}$$

$$A_y = \int_{\alpha_j} M(i,j).(1 - H\varphi(j))dj \tag{4.24}$$

This energy finds very fine edges of the image, when the interior and exterior regions are not constant at every i along the curve. Therefore, it is desired that the local background and foreground means should be different.

Histogram Separation Energy (HS)

Lastly, the proposed algorithm consider a composite energy that is computed from the histograms of the image background and foreground. Its integration into the model is as simple as that of the prior energies. Now consider that $P_x(n)$ and $P_y(n)$ are two intensity histograms using n intensity bins that are calculated from the global external and internal regions of a segmented image. Energy of segmented image, based on minimizing HS statistics reported by Michailovich et al. is given as Malladi et al. (1995).

$$E_{HS} = \int_n \sqrt{P_x(n) - P_y(n)} dn \tag{4.25}$$

This will be called HS (histogram separation) energy. It works by separating intensity histograms of external and internal area of the mask, and hence permits outside and inside regions to be varied provided that their intensity profiles are different.

Similarly $P_{(x,i)}(n)$ and $P_{(y,i)}(n)$ signify the intensity histograms in the local image regions $M(i,j)$. Force F is computed by $P_x(n)$ and $P_y(n)$ as:

$$F_{HS} = \int_n \sqrt{P_{x,i}(n) - P_{y,i}(n)} dn \tag{4.26}$$

For localized version, substitute this in eq.(4.11) and obtain the following:

$$\begin{aligned} \frac{\partial \varphi}{\partial t}(i) = \delta\varphi(i) \int_{\alpha_j} \frac{M(i,j).\delta\varphi(j)}{2} \times \Bigg[F_{HS} \left(\frac{1}{A_x} - \frac{1}{A_y} \right) + \int_n K(n - I(j)) \\ \times \left(\frac{1}{A_x} \sqrt{\frac{P_{x,i}(n)}{P_{y,i}(n)}} - \frac{1}{A_y} \sqrt{\frac{P_{y,i}(n)}{P_{x,i}(n)}} \right) \Bigg] dj + \lambda \delta\varphi(i) div \left(\frac{\nabla\varphi(i)}{|\nabla\varphi(i)|} \right) \end{aligned} \tag{4.27}$$

where K is Gaussian kernel.

4.4 Implementation

In this algorithm, energies are introduced as a signed distance function φ which allows easy implementation flow in the level set framework. The efficiency, in this methodology, is improved by computing values of zero level set function φ and local region based energies for every point along the surfacing mask. The proposed hybrid method builds the complexity of algorithm and reduces the computation time. In this algorithm, every pixel is dependent on the local region based energies in the narrow band with local exterior and interior statistics. During simulation, local statistical models are stored in the memory for every initialized pixel. Then the statistical representations of every pixel in the mask $M(i, j)$ and its neighborhood are updated globally. Furthermore every individual pixel keeps track of the amount of pixels outside and inside the mask along with the sum of pixel intensities in those regions. Therefore, in order to upgrade this model, values needed to be exchanged between the interior and exterior regions. A global region based method updates the statistics of every m number of pixels in the global region while the related local region based energy flow updates $m.z$ times, where z signifies the number of pixels inside the mask $M(i, j)$. The process described above is formalized into an algorithm as follows:

Algorithm 1: ASSVEC

Result: Tumor area segmentation in selected image I_{xig}

Data: 3D volume of MR brain imagesI_x , where $X = [x_1, x_2, x_3, \ldots, x_n]$

Initialization;

For $i = k : K$ (slices) **do**

Acquire 2D slice image I_{xi} from 3D dataset;

Convert I_{xi} to gray scaleI_{xig};

Input radius(r) = a;

Input weight(w) = b;

If $(r \ \ and \ \ w \ \leq \ 0)$ **then**

Input again $r \ \ and \ \ w$;

Else

Initialize level set function φ;

Compute area of curvature based on image gradient;

Determine speed function S ;

Calculate iterative deformation $\gamma(t)$;

Compute signed distance from mask;

Update the local mean intensities by using $eq.(4.14) and (4.15)$;

$i = i + 1$;

Initial contour built by Level set function;

For p = 1:T (iterations) **do**

For q = m:M (mask) **do**

Compute uniform energy and force of exterior and interior regions by using $eq.(4.19)$;

Calculate mean separation energy and force to find very fine edge of boundary by using $eq.(4.22)$;

Compute histogram energy and force to separate the intensities bins of mask by using $eq.(4.27)$;

End

Update the boundary of contour till convergence;

End

Stable mask converge around tumor region of interest;

End

Compute the Dice index and Hausdorff distance between segmented and truth image

4.5 Results and Discussion

Obtaining real brain image data, for research, is a precarious task, because of privacy issues, Hence real data was acquired from BRATS (MICCAI, 2015).

The pathological T2 weighted images are useful for locating the lesioned region in the brain due to higher resolution. For experimentation purposes, the 3D volume dataset was first converted into 2D dataset which comprises of x,y and z slice. Then, these images were converted into pure gray scale intensity images. However since slice z is holistic and has complete information, it is used in the experiment Fig.[4.1].

The initialization of each pixel in the restricted region with internal and external statistics starts the process of adaptive segmentation. An ellipse that relates to a level set initializes the algorithm. The mask changes the topology while the elliptical boundary captures the ROI. As the level set is only updated in the area specified by $\varphi(i)$, it is not likely for new contour to appear in this region Fig.[4.2].

In the developed hybrid adaptive segmentation method, the mask is stable even during initialization. It converges in 100-150 iterations and it is so robust that it does not leak during the remaining solution. The proof of such stability is evident from the preliminary tests with different initializations that indicate that the mask can cover the entire tumor even if placed over a small part of tumor Fig.[4.3].

The corresponding results of the three patients are shown in Fig.[4.4]. Visually, they are satisfying for all the three patients in spite of different tumor shapes. Fig.[4.5] shows the comparison between ground truth segmented image of tumor and simulated segmented region by the adaptive hybrid method.

The algorithm calculates the similarity by dice method that is 89.5% and distance 0.5(mm) by Hausdorff algorithm, depending on the contrast, perfusion levels of image and initialization

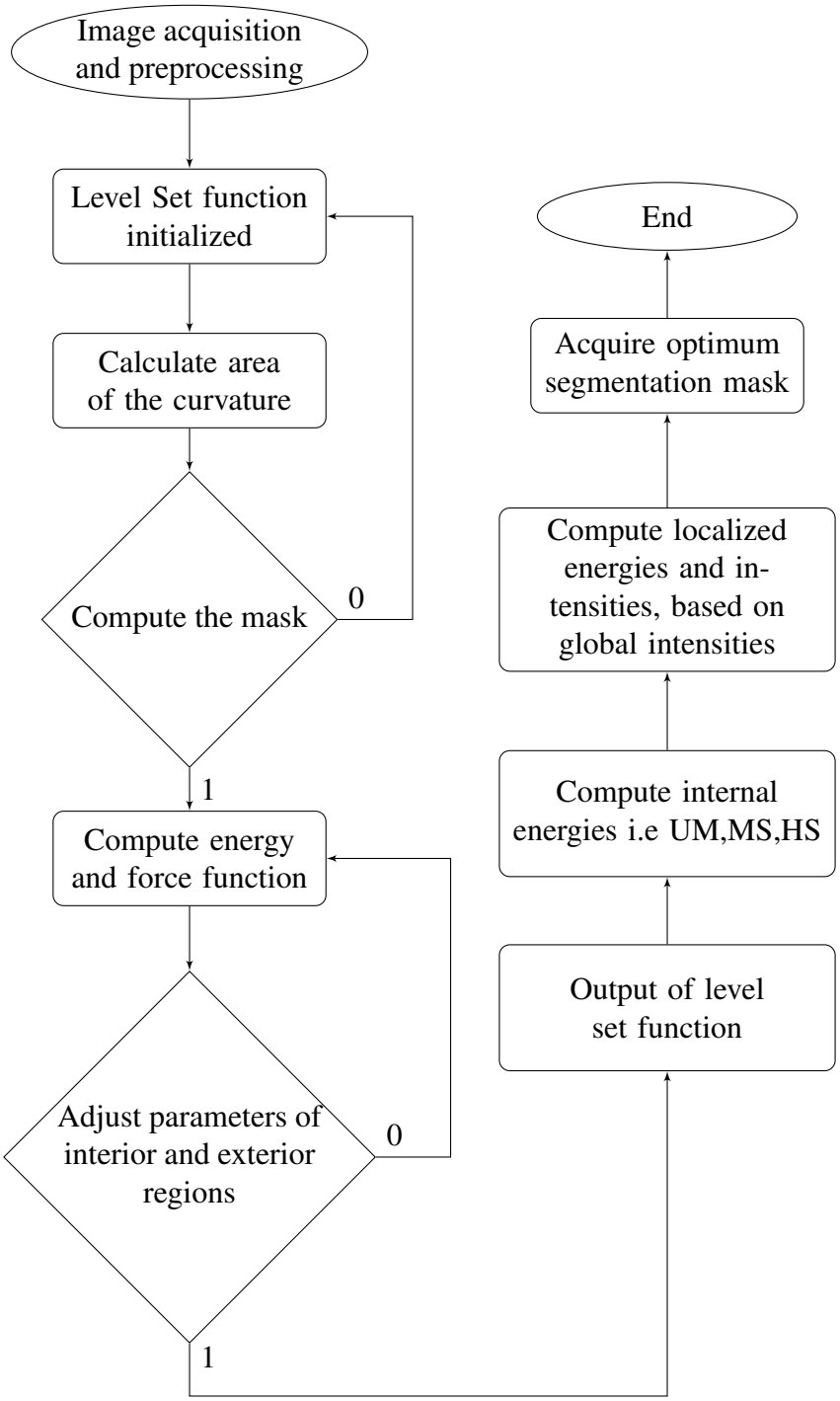
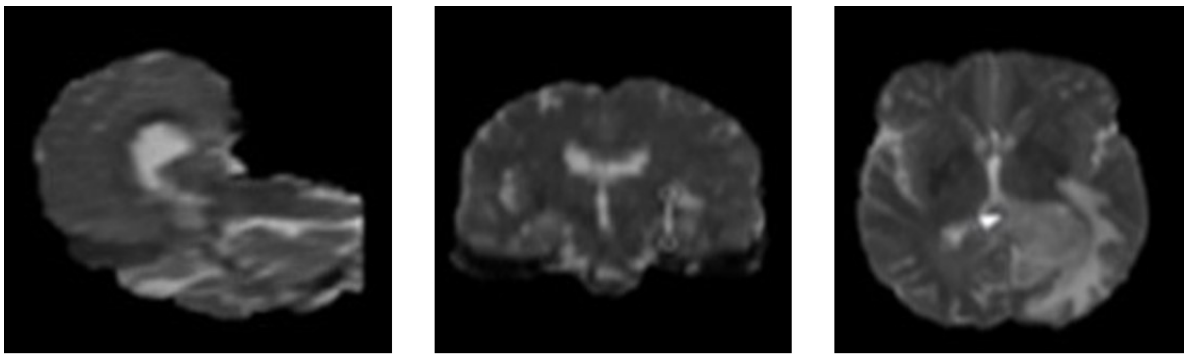


Image acquisition and preprocessing
Level Set function initialized
Calculate area of the curvature
Compute the mask
0
1
Compute energy and force function
Adjust parameters of interior and exterior regions
0
1
Output of level set function
Compute internal energies i.e UM,MS,HS
Compute localized energies and intensities, based on global intensities
Acquire optimum segmentation mask
End

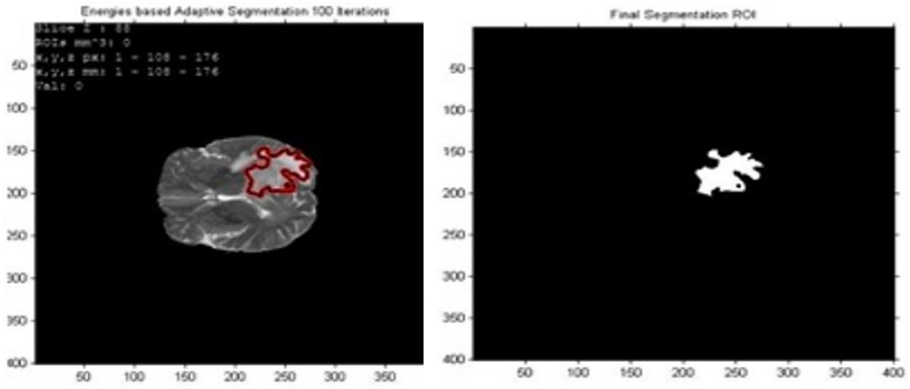


(a) (b) (c)

Figure 4.1 T2- Weighted images (a) x-slice, (b) y-slice, (c) z-slice (MICCAI, 2015)

(a) (b) (c)

Figure 4.2 Level Set Method (4a) Initialization and segmented region (4b) and (4c) show the ellipse mask capturing the global region of interest (ROI)



(a) (b)

Figure 4.3 Hybrid adaptive Segmentation Method (4a) shows behavior of a corresponding Region based energies over the level set mask and sharp changes in intensities within foreground and background, (4b) Final segmented region

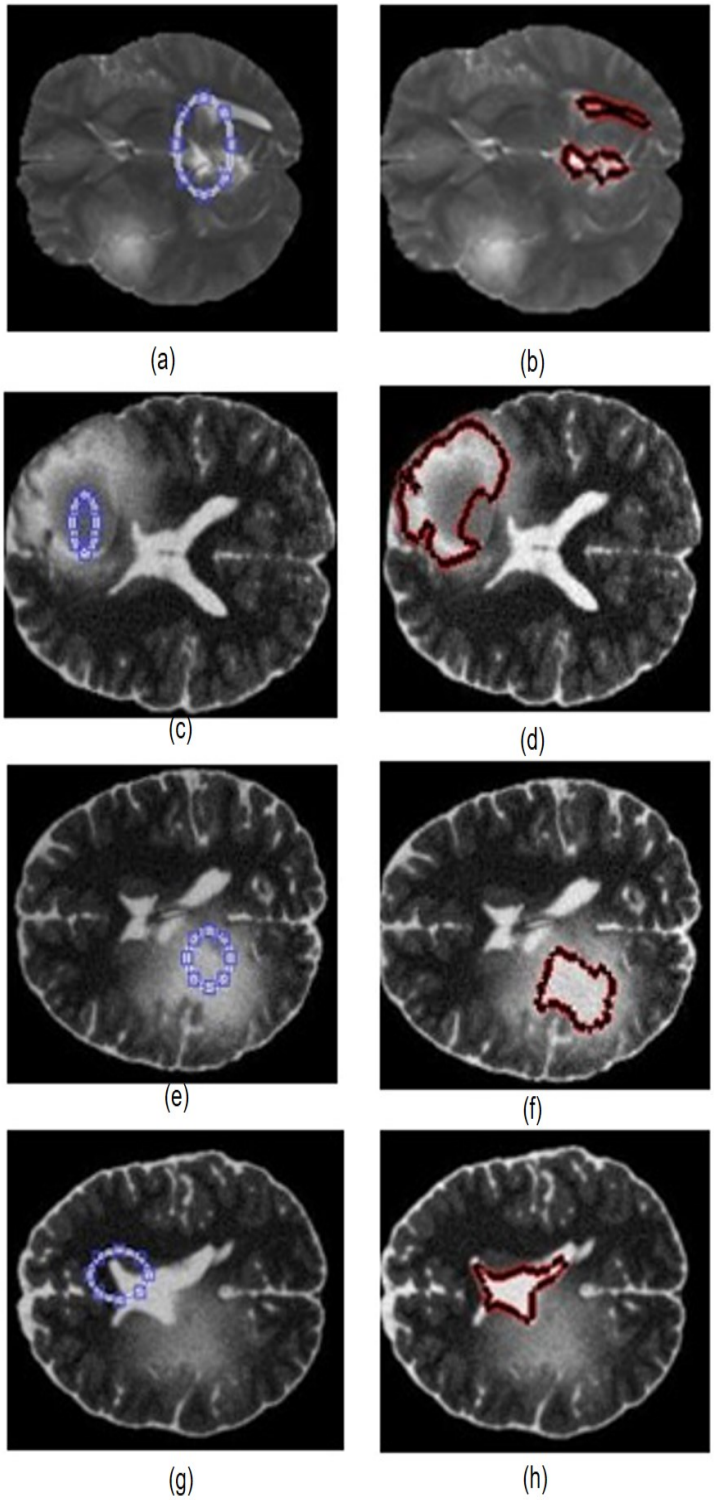
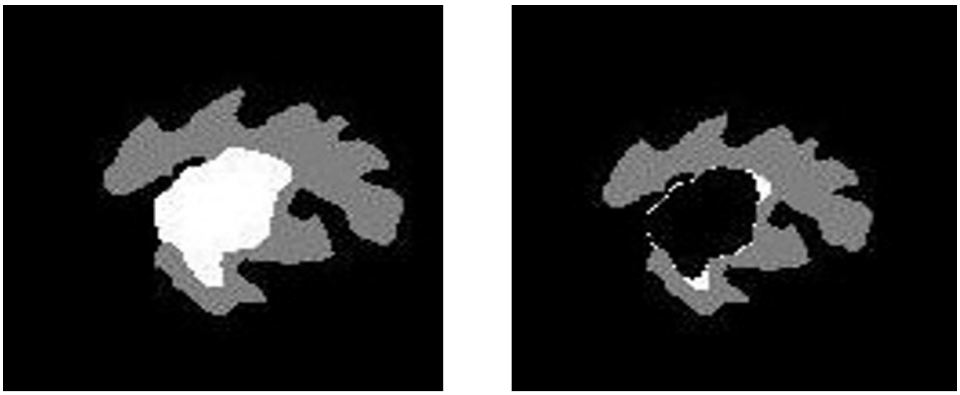


Figure 4.4 Hybrid adaptive Stochastic Segmentation Method (a,c,e,g) shows behavior of a corresponding Region based energies over the level set mask and sharp changes in intensities within foreground and background, (b,d,f,h) indicate the final segmented region



(a) (b)

Figure 4.5 Comparison (a) Ground truth segmented image of tumor, (b) Mapped image, black foreground is the segmented region by adaptive method and white background segment is truth region

seed of segment algorithm. If image has M binary thresholding levels and M $>$ 0.2 then this adaptive algorithm is likely to compute an M-binary level segmentation more quickly than commonly-used segmentation algorithms.

The suggested algorithm was successfully run on fifteen MR images. The output of each segmentation was a binary image on the same voxel grid as the original MR Image. The Dice similarity and Hausdorff distance methods were used to show the accuracy of the segmented image.

The Dice similarity coefficient (DSC) was used as a statistical validation metric to evaluate the performance of segmented image and the spatial overlap accuracy of automated probabilistic fractional segmentation of MR images. The Dice coefficient lies in the range [0, 1] and has value 0 if there is no overlap between the two images and 1 if both images are identical.

Hausdorff distance computes the shape similarity between segmented image and ground truth image. The function computed the average distance from a point on truth image to the closest point on the segmented image for forward and reverse distances and the output distance was the minimum value from both distances, the lower the distance value, the batter match. That method gives promising results, even in the presence of noise or occlusion. The results

for eight tumor datasets are shown in Table [4.2].

Table 4.2 Comparison with truth image

Dataset	**Dice similarity index(%)**	**Hausdorff distance(mm)**
HG0003	72.8	0.767
HG0004	63.8	0.581
HG0005	45.6	0.990
HG0006	77.8	1.16
HG0008	80.3	1.07
HG0022	89.51	1.005
HG0025	75.9	1.25

In this approach, smoothness of contour is optimized by the internal forces while the external forces help guide the contour towards the contour of ROI. The proposed algorithm solved the issue of stability and convergence related to the smoothness of the contour for a segmented object.

The stability condition in level set contour is

$$F_{max}\Delta t \leq min(h_i, h_j) \tag{4.28}$$

Here, F_{max} is the maximum absolute speed of all points on the grid and h_i and h_j are grid spacing in i and j direction. Therefore, for numerical stability the contour can cross only one grid at each time step Δt. This adaptive algorithm, improves the stability by controlling the speed function through internal and external energies and force functions. The speed function guides the dynamics of the mask to slow down and finally reach the steady state.

The other problem in formation of contour is convergence, if the iteration does not terminate at the optimal instant, the curve boundary gives an undesirable output. The energies solve this problem with the help of the radius size of the curve. Fig.[4.6] shows the speed of convergence at different radii.

The results demonstrate that at the largest radius, segmentation curve converges quickly to an incorrect energy value, whereas at an intermediate radius, the segmentation converges

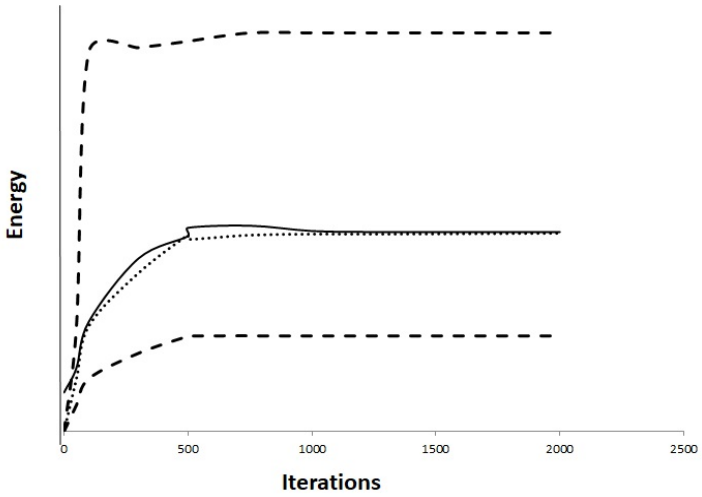


Figure 4.6 Energies for segmentation at different radii

smoothly to a correct energy value. Thus proving that the convergence actually depends on the radius, if the radius is too big or too small then energies convergence would fail and the result would be incorrect. The convergence performance comparison of existing algorithms with proposed algorithm is shown in Table [4.3]. In the reference algorithms, the iteration count was chosen based on the image size and the length of the initial contour.

Table 4.3 Comparison of Performance

Methods	**Accuracy %**
Region growth method with manual selection of initial points (Kavzoglu et al., 2016)	82.48
Improving Brain Magnetic Resonance Image(MRI) Segmentation via a Novel Algorithm based on Genetic and Regional Growth (Javadpour and Mohammadi, 2016)	80.2
Proposed Method	**89.1**

4.6 Summary

In this chapter, a novel adaptive and intuitive hybrid level set with energies based active contour algorithm has been implemented. The results prove that local energies can interact simultaneously on region of interest for making the segment and they also show significant improvement in accuracy of the segmentation. The proposed method has flexible topology and improved iteration speed and segmentation accuracy. The segmented results are close to the desired object region and real target edges as shown in Fig.[4.4]. This could be helpful for surgeons to detect the boundaries of the tumor in MR images. Finally, this segmentation approach proves that when global region based energies are insufficient, then the combination of localized energies provides the optimum solution for accurate segmentation.

Chapter 5

Classification by Classical Machine Learning

In this chapter, we review machine learning and its classical classification algorithms that are useful for the detection, identification and interpretation of brain tumors and their surrounding structures in magnetic resonance (MR) images.

The aim of this chapter and the next one is to contribute to this domain, by proposing an original method, which is general enough to address high and low types of tumor. We proposed two methodologies that were used to establish the significance of feature reduction in classification accuracy. In the first methodology, the extracted feature set from GLCM was applied to the classifiers for comparison of performance. In the second methodology, principal component analysis (PCA) was used to reduce the extracted features and afterwards the uncorrelated reduced vector was applied to the same classifiers.

5.1 Introduction

A number of researchers are working from decades on detection and proper classification of brain tumors so, that many lives can be saved by providing in time and right treatment (El-

Dahshan et al., 2014). There are more than hundred different types of brain tumors that mainly include primary (arising from supportive tissue of the brain) and secondary tumors (generate outside the brain) as discussed in 2.2. The standard technique for classifying brain tumor is based on visual examination of MRI followed by surgical biopsy or resection. The visual examination means that the radiologist classify tumor by observing the texture uniformity or non-uniformity of tumor region. An appropriate mathematical model based on these visual features of an MR image can be used for designing a computer assisted diagnosis (CAD) system for brain tumor classification.

For medical image analysis, especially in MR images, texture is used to determine distinct tissues (normal or abnormal), and classify tissue (Machhale et al., 2015). Texture analysis established that it is helpful to identify the objective (tumor) type by using textural feature vector. Textural features are determined from the statistical distribution of intensities at specific location relative to the neighboring position in the image. According to the number of pixels (intensity points) in each combination, statistics are classified into 1st, 2nd or higher-order statistics. The 2nd order statistics are used in gray level co-occurrence matrix (GLCM). The GLCM is basically a 2D histogram that considers the spatial relationship between various gray levels pixels (Haralick et al., 1973a). Texture is used to classify the tumor in brain MR images. It offers information about the surface pattern of image structure. Additionally, the feature extracted from GLCM is textural features that are based on gray-tone dependencies and are used for image classification (Soh and Tsatsoulis, 1999).

5.1.1 Machine Learning

When the databases being used involve complicated information that is constantly evolving, searching for any desired information manually becomes extremely difficult and tedious. In order to solve this problem, machine learning is used since it is also a part of artificial intelligence. An intelligent system means that the used system has the ability to evolve and learn

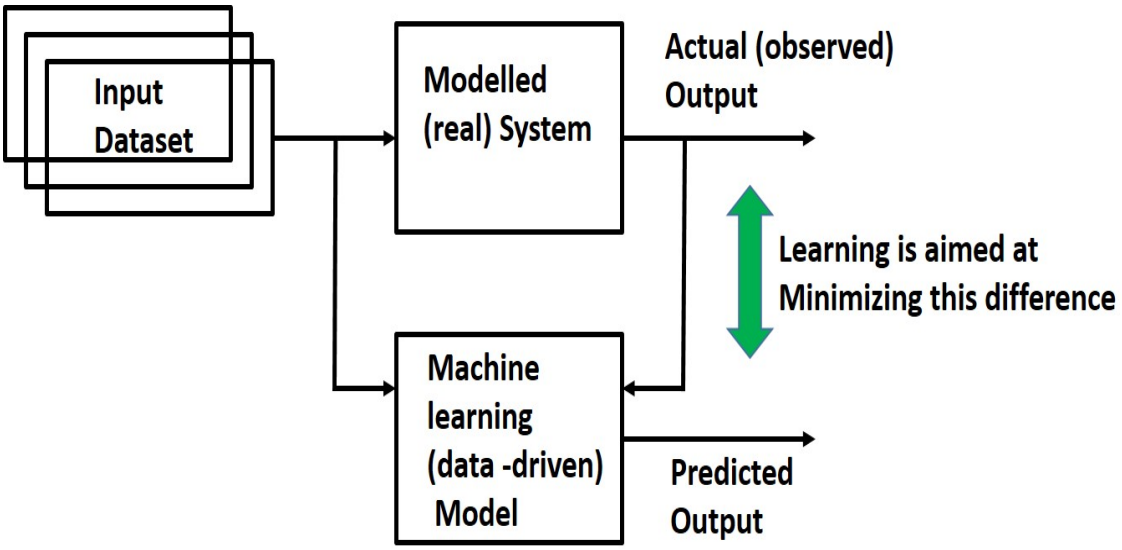


Figure 5.1 Machine Learning

according to the changing information hence eliminating the need for time-consuming manual tasks such as searching for specific information or finding a solution among a plethora of possible situations.

Machine learning implies that a digital model has the ability to learn from a set of training data and previous inputs using computational statistics. After learning, the model must able to extrapolate and make predictions regarding future input data or extract information from the data to not only learn but also provide valuable information. Since the basic principle behind machine learning is making inferences on new data using the knowledge and experience gained from previous data, it heavily relies on building mathematical models using the theory of statistics. However, in order for machine learning to be an effective tool for handling data and solving problems it needs to address two major points required by the modern computer sciences; first, it must have the ability to process and store huge amounts of data and second while processing the input data, the algorithm in use must be efficient and fast. It must be noted that in some cases, the speed at which the algorithm determines the solution with as little effort as possible, is as important as its accuracy of prediction.

The algorithms used for machine learning are organized based on the outcome desired from it. Common algorithm types include as shown in Table [5.1].

Table 5.1 Machine learning algorithm

Algorithm	Description
Supervised learning	The algorithm matches the inputs to the desired outputs by learning a particular function. The standard method behind supervised learning is that it looks at various samples of input-output relations to learn a function that is able to map a vector into one of the several desired classes.
Unsupervised learning	This model has a dataset of inputs without labels.
Semi-supervised learning	This model produces a suitable function by utilizing a set of both labeled and unlabeled input data.
Reinforcement learning	This algorithm learns through the process of understanding how every action leaves an impact on the environment as a whole. An action is performed which effects the surroundings and then the surroundings provide a feedback that enables the model to learn.
Transduction	This algorithm behaves partly like supervised learning such that it tries to guess the outputs based on the inputs and vice versa without actually developing a function that maps the input-output relation.
Learning to learn	It uses previous experience to develop its inference abilities.

As mentioned previously machine learning heavily relies on statistics, especially computational learning theory. It is necessary to understand that the learning involved by the machine learning algorithms is not meant to mimic that of the learning ability of human consciousness. Instead, it employs mathematics in order to find patterns and repetitions in the data to statistically make future inferences thus emulating the ability to learn. Therefore, even though the algorithms developed for machine learning will not imitate a human's ability to learn, they can offer valuable information regarding the process of learning in complex environments and

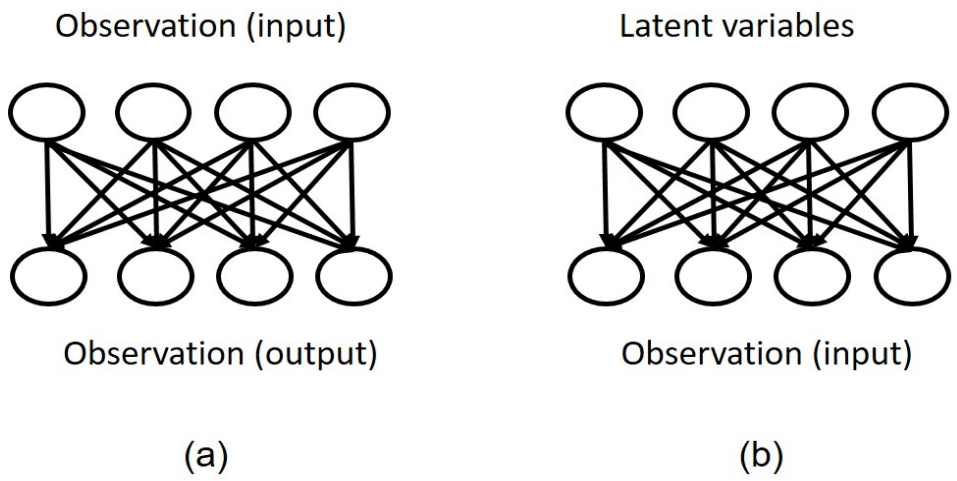


Figure 5.2 Causal structure of Supervised and Unsupervised learning (a) supervised learning, (b) unsupervised learning

with large amounts of data.

Supervised Learning: It is one method of machine learning, in which the model determines the function that produces a given output, a set of observations, after taking in some specific inputs, another set of observations. Basically, the inputs that are provided at the beginning of the casual chain and outputs that are obtained at the end are known while the function that creates those outputs is inferred. Therefore, it can be concluded that supervised learning algorithms can make predictions based on examples and then be tested on by a set of testing data that only has inputs but no outputs. The Supervised process is shown in Fig.[5.3].

Un-supervised Learning: Unlike supervised learning, unsupervised process involves latent variables that dictate the observations that are at the end of the casual chain. The advantage unsupervised learning has over supervised learning is that in the latter process all inputs (or its variations) must be known in order to generate a desirable function. In case all inputs are available, supervised learning is preferred over unsupervised learning however on the other hand if the inputs are not known then the model cannot infer the outputs properly. Lastly, if modeled inputs are used for machine learning then the difference between unsupervised and supervised learning disappears since those inputs can act as latent variables.

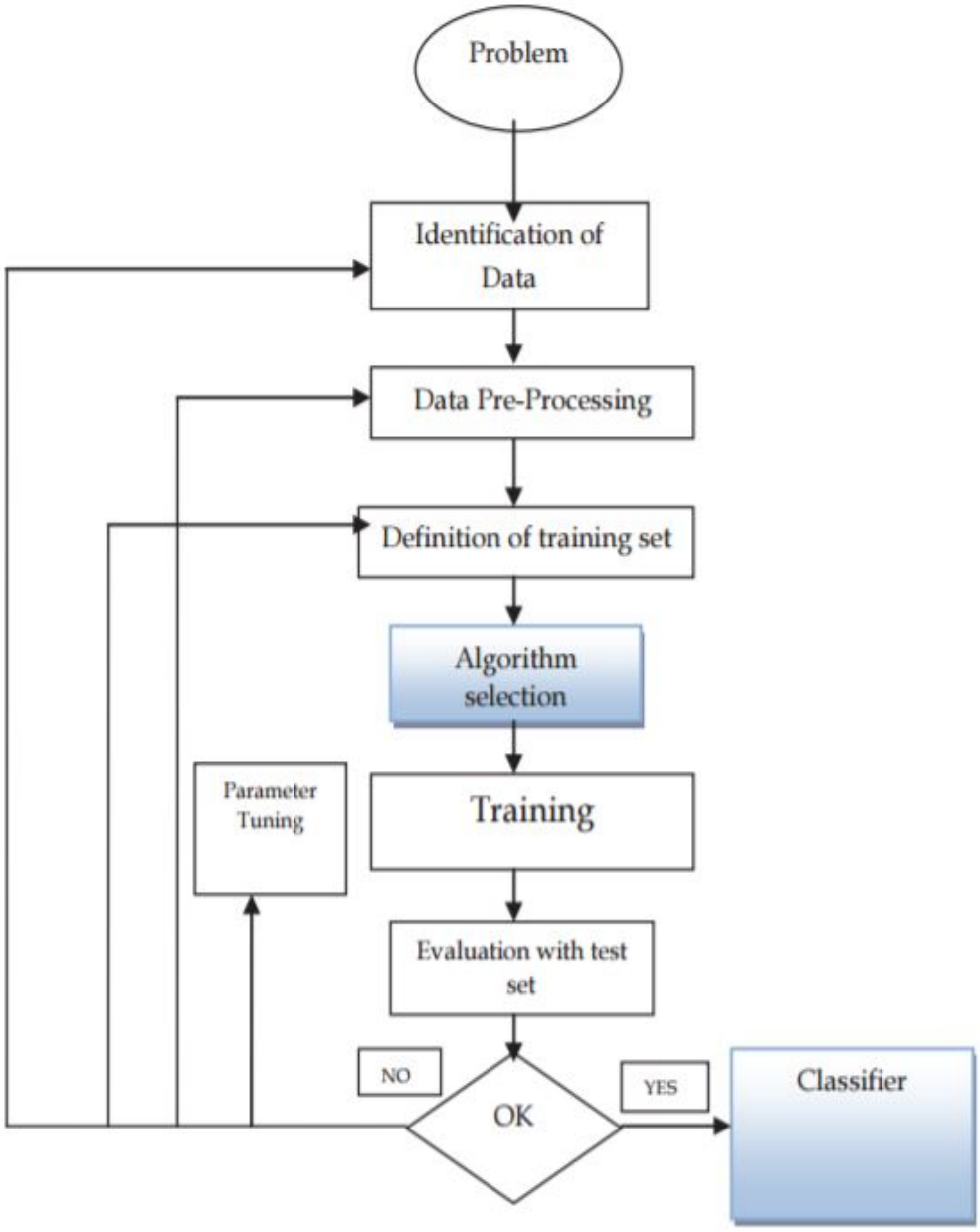


Figure 5.3 Supervised Process

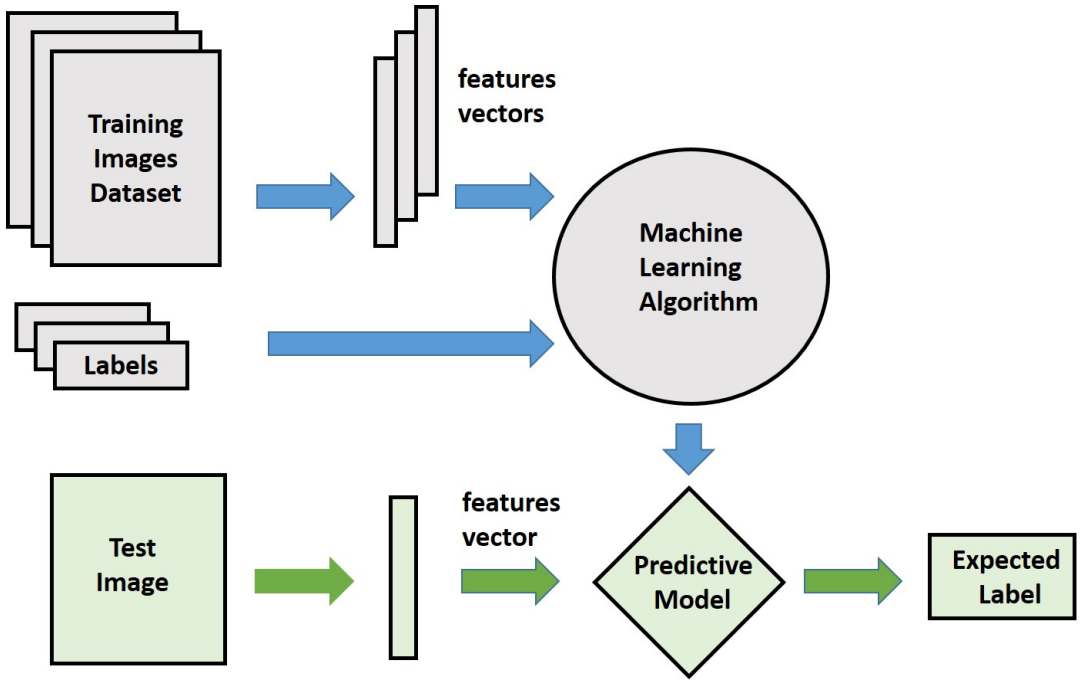


Figure 5.4 Classification with Machine learning

5.1.2 Classification

The process of classification is usually defined as identifying an output after receiving and analyzing a recognizable input (Voznika and Viana, 2007). This process consists of two main steps that are used to determine the outcome accurately. In the first step, the algorithm learns to infer by developing a function after receiving a certain set of attributes known as inputs and their corresponding outputs that are also known as goal or prediction attributes. Using these attributes and their prediction attributes, the model is able to learn the inputs and find a relationship that describes how the two attributes are related. Then, in the second step the learning model is given another set of data that is non-classified (unknown). This set contains the input attributes but does not provide the corresponding prediction attributes and the job of the learning model is to analyze these inputs and provide a prediction of the outputs (Fig.[5.4]). Lastly, the prediction attributes produced by the learning model is evaluated through prediction accuracy that compares these attributes with the ones that were provided with the data set (but not given to the model) in order to prove the effectiveness of the algorithm.

Algorithm Types

The algorithms that are utilized in supervised learning for classification are mentioned below:

- Quadratic Classifiers
- Neural networks
- Linear Classifiers
  - Perceptron
  - Bayes Classifier
  - Logical Regression
  - Support Vector Machine
- Bayesian Networks
- Decision Tree
- K-Means Clustering

5.2 Challenges and Contribution

The features extracted by feature extraction methods have two issues , first is correlation and other is more feature space. PCA method is a transformation that produces uncorrelated effective features and reduces feature vector dimensionally thus minimizing the feature space. The PCA is not only for dimension reduction , but it also decreases the computational time and complexity. Authors surveyed the different brian MRI classification methods. However they do not compare the results based on similar features and dataset to give an unbiased statistical comparison of the algorithm performance.

In this chapter, the comparison study shows whether the classifiers performance is better with GLCM feature extraction only or with GLCM feature extraction along with feature reduction method(PCA). We apply the five classifier models for brain tumor classification and compare their performance. Each classifier is evaluated using only GLCM features and PCA reduced GLCM features.

5.3 Methodology of Comparison of Brain Tumor MRI classification methods using Probabilistic features

5.3.1 Feature Extraction

In order to accurately discriminate between the types of human brain tumors, 1st order and 2nd order probabilistic statistical features have been proven powerful and used in the present thesis to provide the correct information to classification algorithms.

Probabilistic GLCM features (Clausi, 2002) were extracted from dataset. The amount of gray levels in the image are equal to the number of columns and rows in GLCM. Furthermore, the relative frequency represents an element $P(i, j|\triangle x, \triangle y)$ of the matrix, where two pixels are present in a neighborhood. The intensity of each pixel is represented by i and j and the distance between the pixels is indicated by $\triangle x, \triangle y$. In this chapter, we have used GLCM to extract 22 features from each image. For an image with GLCM P, we have assumed that $P(i, j)$ is the (i, j) value in the P matrix. The image is quantized to have N_g number of gray levels where μ_x, μ_y and σ_x, σ_y are the mean and standard deviations for the GLCM (Machhale et al., 2015).

1st order statistical probabilistic features

Information pertaining to the number of times each gray level (histogram) appears in an image is provided by the first order statistical probabilistic features. Some of those features are listed

below:

1. Mean Value

$$\mu_x = \sum_i \sum_j P(i,j) \tag{5.1}$$

$$\mu_y = \sum_i \sum_j P(i,j) \tag{5.2}$$

2. Standard Deviation

Standard deviation formulate the variation of gray levels in image from the mean value.

$$\sigma_x = \sum_i \sum_j (i-\mu_x)^2 P(i,j) \tag{5.3}$$

$$\sigma_y = \sum_i \sum_j (i-\mu_y)^2 P(i,j) \tag{5.4}$$

3. Skewness

Skewness represent the distribution asymmetry around the mean gray-tone value.

$$S_k = \frac{\sum_i \sum_j (i-\mu_{(x,y)})^3 P(i,j)}{\sigma^3_{(x,y)}} \tag{5.5}$$

4. Kurtosis

Kurtosis describe the gray-tone distribution sharpness as compared to the normal distribution

$$K_u = \frac{\sum_i \sum_j (i-\mu_{(x,y)})^4 P(i,j)}{\sigma^4_{(x,y)}} \tag{5.6}$$

2nd order statistical probabilistic features

Even though the first order features may provide the frequency of occurrence of gray level within an image, they do not offer any information relating to the location of those gray levels

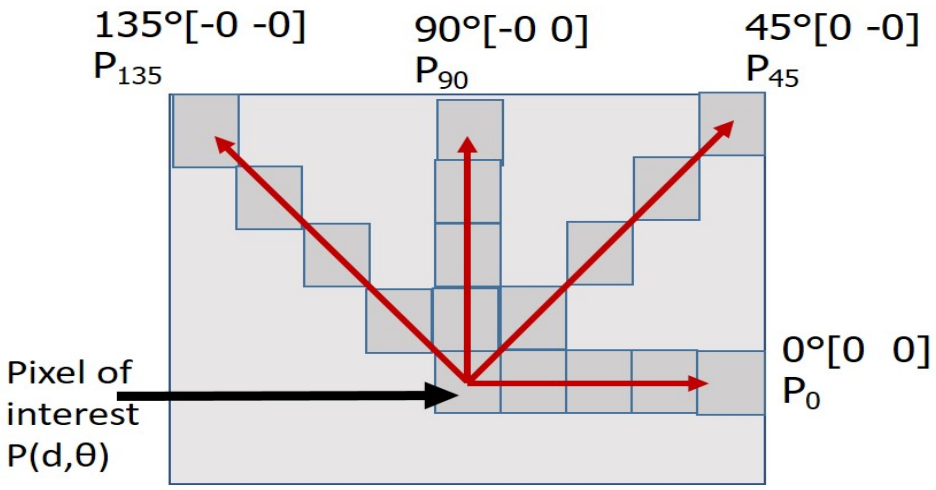


Figure 5.5 Geometrical relationship of GLCM measurements made for four distances d and angles $0^o, 45^o, 90^o and 135^o$ are under the assumption of symmetry

in that image. Instead, matrices like the co-occurrence and run-length matrices are used to obtain this kind of data (Galloway, 1975; Haralick et al., 1973b).

Co-occurrence matrix Features

The co-occurrence matrix provides information regarding the number of pixels that have similar gray tones at a distance d in an image and this distance is called inter-pixel distance Lorensen and Cline (1998). The co-occurrence matrix is used by dividing it into four sub-matrices that can be calculated in four dimensions $\theta(0^o, 45^o, 90^o, 135^o)$ Fig.[5.5] to determine the orientation and distribution of pixels. For example, if Fig.[5.6 (i)] is considered then it comprises of four gray levels $g = 0, 1, 2, 3$, and if the pixel pair $(2, 3)$ is chosen then the co-occurrence matrix element value assigned to it will be equal to the number of times two continuous gray tones of value 2 and 3 appear in a forward and reverse scan direction θ with $d = 1$ (Fig.[5.6 (ii)]). If this element is considered further, then the (2,3) pair appears one time in the horizontal direction with $\theta(0^o)$ and one time in the diagonal $\theta(45^o)$ direction. However, it occurs two times in both the vertical $\theta(90^o)$ and the anti-diagonal $\theta(135^o)$ directions. Therefore, the resultant sub-matrices can be found in Fig.[5.6] (iii).

$$\begin{matrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 2 & 2 & 3 & 3 \end{matrix}$$

(i)

	Gray Tones			
	0	1	2	3
0	(0,0)	(0,1)	(0,2)	(0,3)
1	(1,0)	(1,1)	(1,2)	(1,3)
2	(2,0)	(2,1)	(2,2)	(2,3)
3	(3,0)	(3,1)	(3,2)	(3,3)

Gray Tones

(ii)

Co-occurrence (0^O)	**Co-occurrence** (45^O)	**Co-occurrence** (90^O)	**Co-occurrence** (135^O)
$\begin{matrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{matrix}$	$\begin{matrix} 4 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 \\ 0 & 2 & 4 & 1 \\ 0 & 0 & 1 & 0 \end{matrix}$	$\begin{matrix} 6 & 0 & 2 & 0 \\ 0 & 4 & 2 & 0 \\ 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 0 \end{matrix}$	$\begin{matrix} 2 & 1 & 3 & 0 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{matrix}$

(iii)

Figure 5.6 Co-occurrence matrix calculation example

The classic definition of 2-dimensional co-occurrence matrix of distance $d$and angle $\theta$is given by :

$$P(d,\theta)_{x,y} = \frac{1}{2}[\varphi(d,\theta)_{x,y} + \varphi^T(d,\theta)_{x,y}] \tag{5.7}$$

where $P(d,\theta)_{x,y}$ is a squared matrix of dimension $N_g \times N_g$(number of gray levels), symmetrical around its diagonal axis. $\varphi^T(d,\theta)_{x,y}$is the transpose matrix of $\varphi(d,\theta)_{x,y}$ which gives the sub-matrix in the reverse direction.

In the present thesis for the purpose of quantifying textural volume properties of brain tumors were also applied . The features that can be extracted from the co-occurrence matrix are (Haralick et al., 1973b) :

1. Angular Second Moment (ASM):

Angular second moment encodes image smoothness and takes minimum values for smooth textures.

$$ASM = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}(P(i,j))^2 \tag{5.8}$$

2. Contrast (CON):

Contrast increases for high contrast of image. The factor n^2 is enhances big differences

$$CON = \sum_{n=0}^{N_g-1} n^2\{\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}(P(i,j))^2\}, \quad |i-j| = n \tag{5.9}$$

3. Inverse Different Moment (IDM):

IDM increases for low contrast of image due to the dependence on $(i-j)^2$.

$$IDM = \frac{\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}P(i,j)}{1+(i-j)^2} \tag{5.10}$$

4. Entropy (ENT):

Entropy is a measure of randomness and takes low values for smooth image.

$$ENT = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (P(i,j))log(P(i,j)) \tag{5.11}$$

5. Correlation (COR):

Correlation describes the gray tones dependencies in image.

$$COR = \frac{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (P(i,j))(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{5.12}$$

6. Sum of Squares (SSQ):

$$SSQ = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (P(i,j)((1-\mu)^2 \tag{5.13}$$

7. Sum Average (SAVG):

$$SAVG = \sum_{i=2}^{2N_g} P_{(x+y)}(i) \tag{5.14}$$

where $P_{(x+y)}$ is

$$P_{(x+y)}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (P(i,j), (i+j) = k, \quad k = 2,3,4 \cdots 2N_g \tag{5.15}$$

8. Sum Entropy (SENT):

$$SENT = -\sum_{i=2}^{2N_g} P_{(x+y)}(i)log(P_{(x+y)}(i) \tag{5.16}$$

9. Sum Variance (SVAR):

$$SVAR = -\sum_{i=2}^{2N_g} P_{(x+y)}(i)(1-SEnt)^2 \tag{5.17}$$

10. Difference Variance (DVAR):

$$DVAR = \sum_{i=2}^{2N_g} P_{(x-y)}(i)(1 - Savg)^2 \tag{5.18}$$

11. Difference Entropy (DENT):

$$DENT = -\sum_{i=0}^{N_g-1} P_{(x-y)}(i) log(P_{(x-y)}(i) \tag{5.19}$$

where $P_{(x-y)}$ is

$$P_{(x-y)}(k) = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (P(i,j), (i-j) = k, \quad k = 2,3,4 \cdots N_g - 1 \tag{5.20}$$

The other extracted features (Haralick et al., 1973a) in proposed methodologies are: Energy, Homogeneity, Cluster prominence, Cluster shade, Auto-correlation, Dissimilarity, Maximum probability, Information measures of correlation 1, Information measures of correlation 2, Inverse difference and Inverse difference normalized.

5.3.2 Feature Reduction

Feature reduction was used to verify the improvement in overall accuracy of classification as shown in (Abdullah et al., 2011; Al-Badarneh et al., 2012; John et al., 2012a).

Principal component analysis is a popularly used method for reducing dimensionality. It is able to achieve this by determining a k subspace with a lower dimension than the original dimension n of the input data. Most of the data points exist mainly in this linear space thus retaining the variability of the data. In addition, this method obtained its name because of the principal components that form a new coordinate system based on the k orthogonal vectors in the linear subspace. Since these principal components are linear transformations of the original data points and are also orthogonal, they always have to be less than n. However, in

effect the principal components required to span the space of n original axes are only $k < n$.

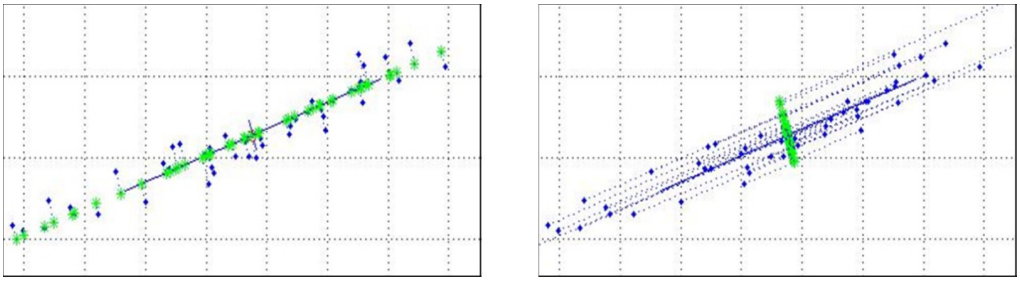


Figure 5.7 Illustration of PCA (a) 2-Data projection onto the first eigenvector, (b) 2-Data projection onto the second eigenvector

Hotelling (1933) describes the definition of PCA to be those k orthonormal, principal axes, for a set of particular input data vectors $x_i, i\varepsilon 1...p,$, that have the maximum variance under projection.

Let us calculate the maximum variance by choosing the 1st principal component, represented by U_1, to have the maximum variance. In this case, all centered observations are accumulated in the columns of a matrix M that has the dimensions $n \times p$. The columns in this matrix correspond to observations in n-dimensions while having p observations. Finally, let the coefficients $w = w_1...w_n$ define a linear combination of M that is the first principal component. The matrix form is:

$$U_1 = w^T M \tag{5.21}$$

$$var(U_1) = var(w^T M) = w^T P w \tag{5.22}$$

where P is the $n \times n$ sample covariance matrix of M.

It is evident that increasing the magnitude of w increases the $var(U_1)$ so we chose w to $max.w^T Pw$ when $w^T w = 1$.

A Lagrange multiplier ζ_1 was introduced to solve the optimization problem:

$$L(w,\zeta) = w^T P w - \zeta_1 (w^T w - 1) \tag{5.23}$$

Differentiating with respect to w gives n equations,

$$Pw = \zeta_1 w \tag{5.24}$$

Multiplying both sides by w^T :

$$w^T P w = \zeta_1 w^T w = \zeta_1 \tag{5.25}$$

if ζ_1 is the largest eigenvalue of P then $var(U_1)$ is maximum.

Therefore, ζ_1 and w are an eigenvalue and an eigenvector of P. Differentiating Eq.(5.23) with respect to the Lagrange multiplier ζ_1 gives us back the constraint $w^T w = 1$.

This illustrates that the normalized eigenvector with the biggest associated eigenvalue of the sample covariance matrix M gives the 1^{st} principal component. Similarly, we can find the first k principal components by the k prevalent eigenvectors of covariance matrix P.

The other characteristic of PCA, given by Pearson (1901), is that the squared reconstruction error is minimized by projecting on to the principal subspace, where $\sum_{i=1}^{p} \|x_i - \hat{x}_i\|^2$. In other words, the principal components provide a sequence of best linear approximations to data, for all ranks $k \leqslant n$. The rank-k linear approximation model as follows :

$$f(y) = \bar{x} + U_k y \tag{5.26}$$

It is the parametric representation of a hyperplane of rank-k.

Let $\bar{x} = 0$,otherwise $\tilde{x} = x_i - \bar{x}$. Suppose the rank-$k$ linear model would be

$$f(y) = U_k y \tag{5.27}$$

where U_k is a $(n \cdot k)$ matrix with k orthogonal unit vectors and y is a vector of parameters.

For minimizing the reconstruction error, use this model to the data by least squares.

$$\min_{U_k, y_i} \sum_{i}^{p} \|x_i - U_k y_i\|^2 \tag{5.28}$$

By partial differential for y_i optimization we get:

$$\frac{d}{dy_i} = 0 \Rightarrow y_i = U_K^T x_i \tag{5.29}$$

We require to determine the orthogonal matrix U_k:

$$\min_{U_k} \sum_{i}^{p} \|x_i - U_k U_k^T x_i\|^2 \tag{5.30}$$

if H_k is a projection matrix $n \times n$, it projects every data point x_i onto its rank-k reconstruction. Hence,

$$H_k = U_k U^T k \tag{5.31}$$

$H_k x_i$ is the orthogonal projection of x_i onto the subspace spanned by the columns of U_k. A unique solution U can be obtained by calculating the singular value decomposition of X, U_k consists of the first k columns of U (Friedman et al., 2001). The solution for U can be expressed as singular value decomposition (SVD) of X.

$$X = U \sum V^T \tag{5.32}$$

since the columns of U in the SVD contain the eigenvectors of XX^T. The PCA algorithm is:

Algorithm 1: PCA algorithm

Result: Reconstruct reduced feature vector data

Data: Feature extraction co-occurrence matrix

Calculate $XX^T = \sum_i^p x_i x_i^T$; Suppose U eigenvectors of XX^T corresponding to highest k eigenvalues. [Recover basis]

$Y = U^T X$; [Training data encoding]

where Y is a $k \times t$ matrix of encodings of the training original data.

$\hat{X} = UY = UU^T X$; [Training data reconstruction]

$Y = U^T x$; [Test example encode]

where y is a k-dimensional encoding of x.

$\hat{x} = Uy = UU^T x$; [Test example reconstruction]

5.3.3 Classical Classifiers

Image classification is known as the process of extracting relevant attributes from an image and mapping each set of attributes to a predefined class. The resulting classification can be used to create a thematic map. Images with numerical values close to one another are considered to be members of the same class.

There are two main types of classification techniques; Unsupervised classification, where image classes are not predefined by the user, instead the classification algorithm finds the relation between pixels and creates clusters to define them into groups. Supervised classifiers

use the information it has learned during training to identify image class (Gadkari, 2004). The classifier uses the spectral signature of the classes to identify each class. These signatures are created during the training phase. In this chapter, we compare the classification accuracy by using following algorithms:

ANN (Artificial neural networks)

ANN is a machine learning model which simulates the biological nervous system. A network is a set of nodes and connections. In ANN the nodes are called 'neurons'. The path of data flow is determined by the connections between neurons and their interactions define the global behavior of the network (Charfi et al., 2014b). ANN has two modes of operations, namely the training and the testing phase. Network, in the first phase learns input data patterns and the outputs associated with them. In the testing phase, the trained network is used to predict the output for an unknown dataset. There are several methods which are used to train a neural network. The most efficient method is the Back Propagation algorithm. The learning method used in Back Propagation is supervised learning which means that in the training stage the inputs and outputs of the network are provided and the error between expected and actual result is computed. Back Propagation algorithm essentially reduces the error between the two results by adjusting weights (Gershenson, 2003).

The activation function of the artificial neuron is the addition of all inputs x_i multiplied by the relevant weights w_{ij} of the neurons.

$$A_j(x,w) = \sum_{i=0}^{n} x_i w_{ij} \tag{5.33}$$

Let us define y as a k-dimensional vector of real numbers such that $y \in R^k$ and $h_{\theta(x)}$ also k-dimensional vector, so $h_{\theta(x_i)}$ refers to the ith value in that vector. $h_{\theta(x)} \in R^k h_{\theta(x_i)} = ith$ output. If m is the number of training data and λ is the regularization parameter, then the cost function of L layered artificial neural network, having k distinct classes and sl neurons in each

layer, such that $sl = k$, is defined by the following equation whose output is a k dimensional vector

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m}\sum_{k=1}^{k} y_{k_{(i)}} \log(h_{\theta(x_{(i)}k)}) + (1 - y_{k_{(i)}}) \log(h_{\theta(x_{(i)}k)})] + \frac{\lambda}{2m}\sum_{i=1}^{L-1}\sum_{i=1}^{sl}\sum_{j=1}^{sl+1}(\theta_{ji(i)})^2 \tag{5.34}$$

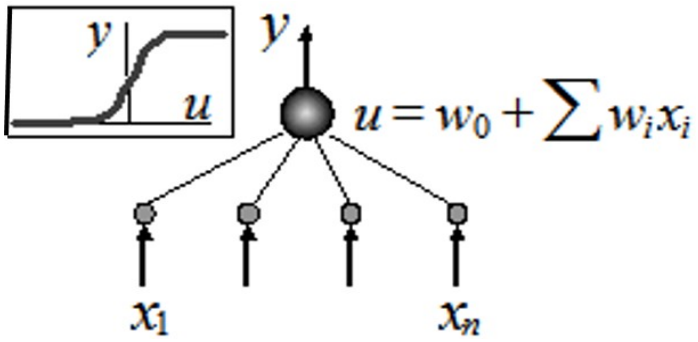


Figure 5.8 Single Artificial Neuron

Decision tree

Decision tree arrives at a class membership by repeatedly dividing datasets into smaller and more uniform datasets. The subsets are broken down into further subsets via different variables and attributes (Kamavisdar et al., 2013). The data is partitioned based on maximum reduction in deviance over all splits of all nodes to choose the next split. If a node s is divided into nodes u and y, and then the reduction in deviance is calculated by:

$$D = D_s - D_u - D_y \tag{5.35}$$

where, $D_i = -2\sum_k n_{ik} log(p_{ik})$, n_{ik} is the number of cases of class k in node i and p_{ik} is the probability distribution of class k in node i. Decision tree requires very little computational load and does not require extensive network training. It is best suited for non-parametric data.

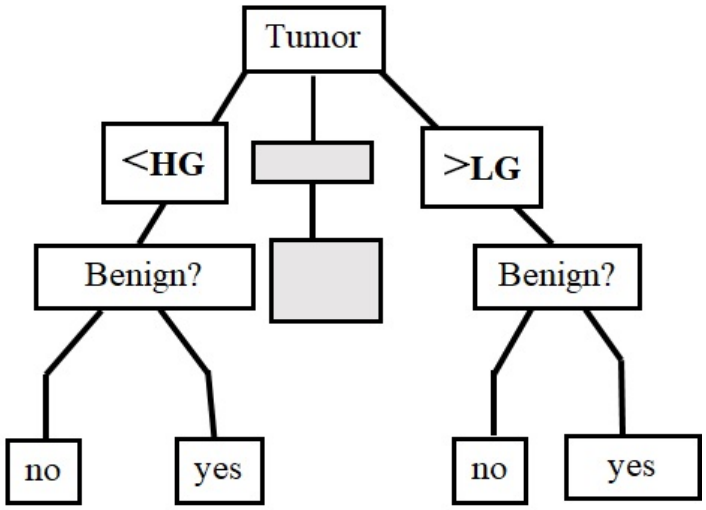


Figure 5.9 Decision tree

KNN (K-nearest neighbor)

Another method that is used for classification and regression is KNN which is a non parametric classifier. This classifier provides output in the form of class membership and the object is related to the class with maximum resemblance (Zhang and Zhou, 2005).

In the training stage of the KNN classifier the input dataset is divided into k classes, with each class containing only the inputs belonging to its class. This is done by finding the spearman distance between all inputs and distributing them into k classes. Each input consists of p features. Different distance measures can be adopted as the reference measure for classification. The spearman distance between L_i and any ordered list δ is defined as:

$$S(\delta, L_i) = \sum_{t \in L_i, \bigcup \in \delta} |r^{\delta(t)} - r^{L_i(t)}| \tag{5.36}$$

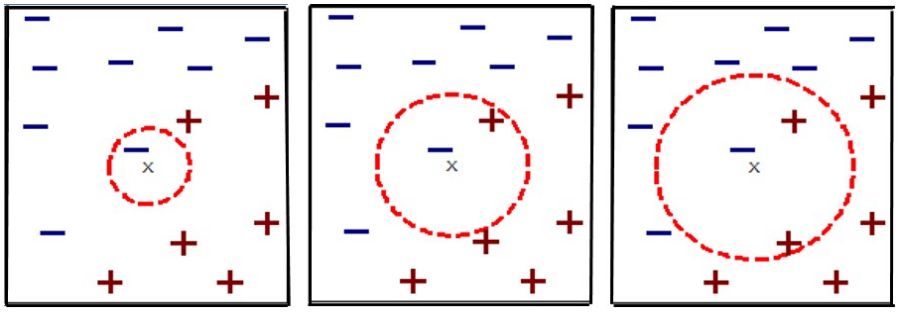


Figure 5.10 K-nearest neighbor (a) One-nearest neighbor, (b) Two- nearest neighbor, (c) Three-nearest neighbor

In the testing phase, the sparman distance is found between input and all points in the KNN classifier. Input is then matched to the class with the smallest distance Chang et al. (2003).

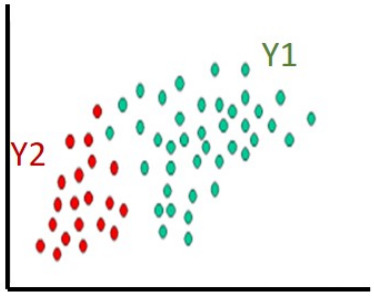
Naïve Bayes

It is the member of family of probabilistic classifiers. It uses Bayes' theorem to calculate the probability of a sample to belong to an exacting class Kamavisdar et al. (2013). Let suppose that all attributes of the samples in a class are not dependent of each other given the context of the class. If x is an input sample with n features such that:

$$x(n) = (x_1, x_2, \cdots x_n) \tag{5.37}$$

According to Bayes' theorem, for a class $P(x|c)$ is the posterior probability of C conditioned on x, $P(C)$ is the prior probability of C and $P(x)$ is the prior probability of x. The probability of x belonging to class C is given by:

$$P(C|X) = \frac{P(x|c)P(c)}{P(x)} \tag{5.38}$$



$$\frac{P(y_1 \mid \mathbf{x})}{P(y_2 \mid \mathbf{x})} > 1$$

Figure 5.11 Two categories, y1 and y2 using n features x1, x2, . . . , xn. If above condition true, then the object holds to the category y1

Support vector machine (SVM)

SVM is a non-parametric classifier used for classification and regression. It builds a hyperplane or a set of hyperplanes in an infinite or high dimension space in order to work. Generally SVM is applied to classify data into binary classes, however it can be modified to classify data into more than two classes using 1-vs-1 or 1-vs-all technique (Burges, 1998; Chu et al., 2005). SVM is capable of using linear, polynomial or sigmoid kernel functions for decision function which makes it very versatile.

Fig. [5.12] provides a pictorial representation of how SVM separates classes in a feature space using hyperplanes. The mathematical representation of this methodology is as follows (Vapnik, 2013):

$$y = sign(w.x + b) \tag{5.39}$$

Furthermore, (Vapnik and Chervonenkis, 1974) provided a method of restricting the k-dimension using a particular quantity known as a margin. This margin can be described as the

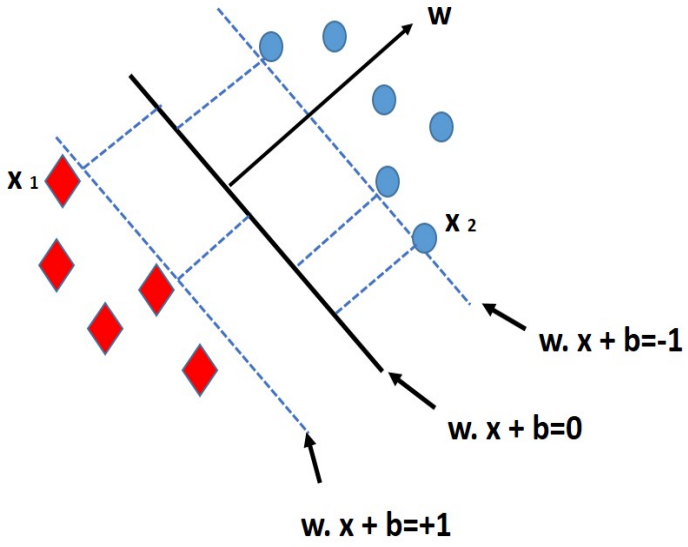


Figure 5.12 Different classes separated by a hyperplane. The margin (dashed lines) is the minimum distance between the class and hyperplane

smallest distance between the patterns and the hyperplane.

The dashed lines in Fig. [5.12] represent the margin which can be shown to be a function of w by tweaking the variables w and b so that the points near the hyperplane satisfy $|w.x+b| = 1$.

If x_1 and x_2 are considered as two patterns that represent two distinct classes such that $|w.x_1 + b| = +1$, and $|w.x_2 + b| = -1$, then the perpendicular distance between those two points can be used to calculate the margin using:

$$\frac{w}{\|w\|}(x_1 - x_2) = \frac{2}{\|w\|} \tag{5.40}$$

Then the following expression can be used to show the inequality that links the margin to the k-dimension of the separating hyperplanes' class:

$$h^2 \leq \upsilon^2 R^2 + 1 \qquad and \|w\| < \upsilon \tag{5.41}$$

Where, the radius of the smallest sphere is represented by R. Moreover, the inverse proportionality in eq.(5.40), between $\|w\|$ and the margin, signifies that if there is a large margin then

a small k-dimension will be obtained and vice versa. A small error can then be expected if and only if both the k-dimension and the training error are kept small, as described by the bound. The separating hyperplanes can also be constructed with minimal errors in training patterns and maximum margins when linear learning machines are used with small k-dimensions and training errors.

5.3.4 Performance evaluation parameters

Table 5.2 Confusion matrix

		Predicted Class	
		+	-
Actual Class	+	TP	FN
	-	FP	TN

Accuracy, sensitivity and specificity were used for measuring the robustness of classifiers (Charfi et al., 2014b):

$$Accuracy = \frac{(TP+TN)}{TP+FP+TN+FN}$$

$$Sensitivity = \frac{(TP)}{TP+FP}$$

$$Specificity = \frac{(TN)}{TN+FN}$$

Confusion matrix is the best tool to describe the performance of classifier models. Where, TP = true positive, TN = true negative, FP = false positive, and FN = false negative as shown in Table [5.3]. TP can be said to occur when classification results are positive in the presence of clinical deviation while TN is when the result is negative in the absence of deviation. Furthermore, FP proves that the result is positive in the absence of clinical deviation, where FN is when the result is negative in the presence of clinical deviation.

5.3.5 Flow Chart

The following flowchart outlines the process of tumor classification of high grade and low grade images through five classifiers by using only GLCM and GLCM + PCA.

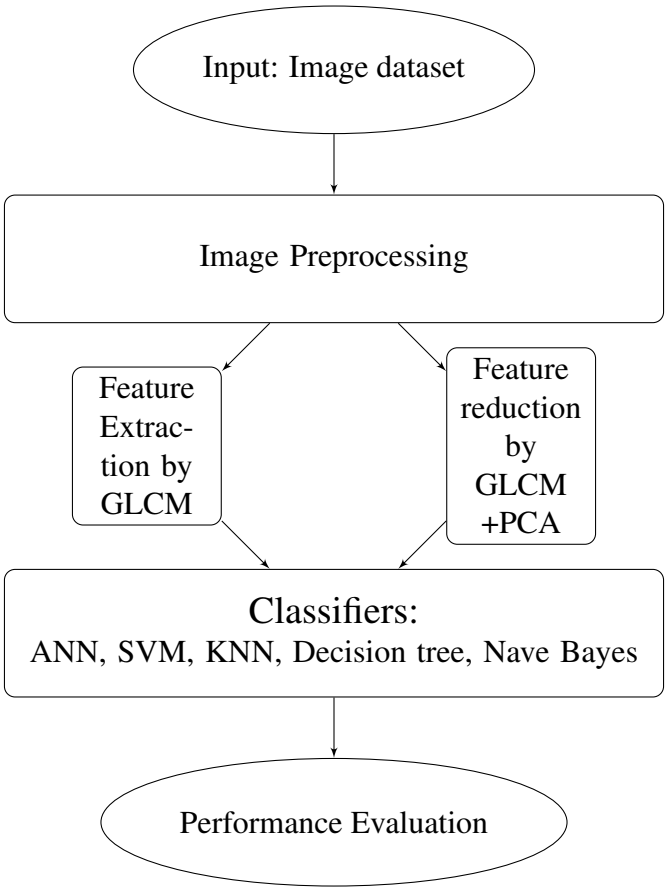


Figure 5.13 Flow diagram of the methodology

5.4 Results and Discussion

Obtaining real brain image's data, for research is a precarious task, because of privacy issues. In this chapter, real data is acquired from BRATS (MICCAI, 2016) .

BRATS dataset has a total of 104 brain tumor images, that are comprised of four grades. These grades are classified by doctors based on their appearance under a microscope. The four

grades are classified as: Grade I, II, III and IV. When the tumor is benign, such that most of the cells look like normal brain cells and have slow growth, the tumor is identified as being Grade I. Moreover, when the tumor is malignant and the cells hardly resemble Grade I conditions, the tumor is known as Grade II. However, a tumor is considered as a Grade III tumor when the malignant tumor is made of cells that are unlike healthy cells and the tumorous cells are growing rapidly (anaplastic). Whereas, a Grade IV tumor is when the malignant tissue is completely made up of tumorous cells that grow uncontrollably. Therefore, Grade I and Grade II tumors are known as low-grade tumors while Grade III and Grade IV are regarded as high-grade tumors.

In these 104 images, 61 images were high-grade while 43 were low-grade. However, in order to avoid biasness of classifiers we selected 43 high-grade and 43 low-grade images randomly.

The pathological 'T2' weighted images are useful for locating the lesioned region in the brain due to higher resolution. For experimentation purposes, the 3D volume dataset was first converted into 2D dataset which comprises of x,y and z slices. Then, these images were converted into pure gray scale intensity images. However since slice 'z' is holistic and has complete information, it is used in the experiment Fig.[4.1].

It is common practice amongst the researchers to extract 4-10 features from GLCM, but in this study we extracted 22 features of each slice by GLCM. Thus their input vector is 86 x 44. After the reduction by applying PCA, our input vector is reduced into 86 x 43.

Table [5.4] shows the confusion matrix which describes the performance of the five classifier models with GLCM and GLCM with PCA.The Table [5.5] compares the accuracy,sensitivity and specificity of each classifier with GLCM and with reduced GLCM vector by PCA respectively.

Computation time is another critical factor needed to assess the classifier. Consequently, 86 images were passed into the classifiers, corresponding computation time were recorded,

Table 5.3 Confusion matrix Comparison

Method	**Confusion Matrix (GLCM)**		**Confusion Matrix (GLCM+PCA)**	
Artificial Neural Network	30	9	41	6
	13	34	2	37
Decision Tree	37	6	43	0
	5	38	2	41
KNN (spearman)	43	0	43	0
	39	4	16	27
Naïve Bayes	28	15	36	7
	12	31	1	42
SVM (RBF)	10	33	20	23
	9	34	7	36
SVM (Linear)	20	23	39	4
	8	35	8	35

Table 5.4 Performance Comparison

Method	**GLCM**			**GLCM+PCA**		
	Accur. (%)	**Sensi. (%)**	**Speci. (%)**	**Accur. (%)**	**Sensi. (%)**	**Speci. (%)**
ANN	74.4	69.7	79.1	90.7	95.3	86.0
DT	87.2	88.1	86.4	97.6	95.6	100.0
KN(spearman)	54.6	52.4	1.0	81.4	72.8	100.0
NB	68.6	70	67.3	90.6	97.2	85.7
SVM (RBF)	51.1	52.6	50.7	65.1	74.0	61.0
SVM (Linear)	63.9	71.4	60.3	86.0	82.9	89.7

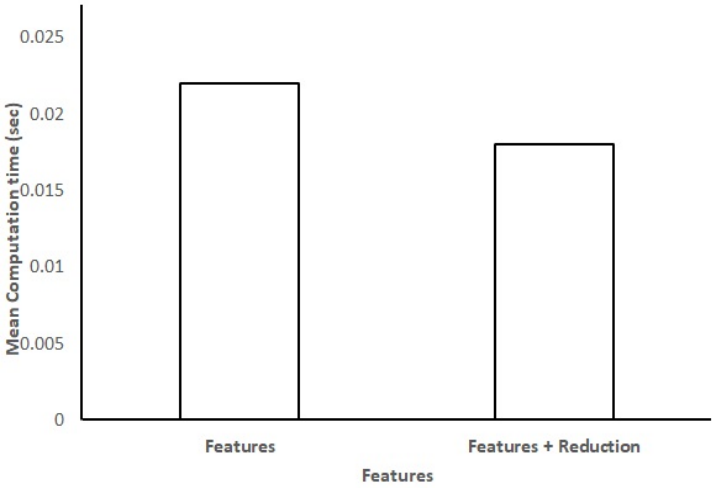


Figure 5.14 Computation time

the mean value was computed and the consumed time was observed with GLCM features and with reduced features by PCA as shown in Fig.[5.14]. GLCM with PCA according to its low computational complexity, low computational time and high accuracy leads to better results with each classifier rather than only using GLCM.

5.5 Summary

In this chapter, compared the performance of machine learning classifiers. However the novelty of this research lies in the fact, that feature extraction along with feature reduction methodology was successfully applied on five classifiers and each classifier provided better accuracy and classification results. It was observed that GLCM features with PCA produced more accurate results as compared to GLCM feature set alone.

We conclude that the quality of features has a significant influence in improvement of results. A hybrid approach of feature selection which uses multiple features may further improve the accuracy of classification. For future work we suggest employment of advance machine learning methods like deep neural network models with large dataset for further improvement.

Chapter 6

Classification by Deep Learning

Machine learning classification methods were presented in Chapter 5 comparing classification with or without feature reduction technique. In this chapter, an Optimized CNN-Deep Belief Learning Model for Brain tumor Classification on MR images has been proposed. This approach consists of, namely pre-processing, segmentation, CNN-Deep belief learning classifier and optimization constraints of that model. Firstly, pre-processing is carried out in which linear filters like the Gaussian filter de-noise the image thus making the MR image suitable for segmentation. Once pre-processing has been completed, the image is then segmented using automatic fuzzy C-Means (FCM) clustering method. The segmented image is then sent to a flexible, high capacity CNN-deep belief learning classifier model that is employed for classification of High and Low grade brain tumor. This model is key to our solution as it has the ability to scale to various image sizes by distributing the hyper-parameters and weights among all locations in an image. It is also translation invariant and is compatible with top-down and bottom-up probabilistic inference. Then, the classifier is optimized by regularization, stochastic gradient decent and fine tunning constraints. Lastly, a comparative analysis is carried out between the pre-processed non-segmented MR images and segmented MR images after classification. In that analysis, the presented technique is evaluated based on accuracy, error/loss and prediction of individual images.

6.1 Introduction

Cancer is a disease referred to the uncontrollable growth of abnormal cells in the human body. These rapidly growing anomalous cells cause an interference with the essential bodily functions performed by the healthy cells which can lead to death. There are two important types of brain cancers; primary brain tumor which forms in the brain and spreads outwards, and secondary brain tumor which forms somewhere else in the body and spreads to the brain. Tumors are classified based on their histopathology and are assigned a grade which defines the type of tumor and provides insight regarding the cell's growth rate as discussed in 2.2. In order to detect and classify these tumors, stages in digital detection are feature extraction and feature selection. Once the features have been extracted and chosen, a feature set is formed (John et al., 2012a). Nonetheless, in order to have the ideal feature set, it must contain distinct and effective features which are formed after feature reduction to eliminate the sparsity. Furthermore, it was established through texture analysis that the textural feature vector can be very helpful in identifying the tumor type and that it can be determined using the statistical distribution of intensities at specific locations compared to their neighboring positions. The statistics are categorized on the basis of the amount of pixels (intensity points) in each combination, into 1st, 2nd or higher-order statistics (Farhi and Yusuf, 2017b).

Before the features can be extracted, the tumor has to be clearly captured by an image and in order to achieve this, Magnetic Resonance Imaging (MRI) technology is utilized. MRI based classification has emerged as a difficult task due to the complexity and divergence of a tumor (Zhang et al., 2011b). MRI captures the interior images of a body through the use of electromagnetic waves (like radio waves) and magnetic fields as discussed in 2.4. However, T2-weighted scan provides the best resolution and clarity and is therefore used in neuroimaging processing.

6.2 Challenges and Contribution

Most of the researchers used standard models i.e AlexNet (Krizhevsky et al., 2012b), the VGG net (Simonyan and Zisserman, 2014), and GoogLeNet (Szegedy et al., 2015) for non-medical imaging tasks. However, these architectures are not suitable for MR images because they are grayscale images that are often of low image resolution. Neighboring pixels in a grayscale image have similar intensity which leads to redundant computation while the low resolution of the image reduces the ability of the network to learn global features in small patch sizes (Kayalibay et al., 2017).

Recently, digital imaging studies have been focused at learning as many features as possible in order to increase the accuracy of detection through 3D-CNN architecture (Dou et al., 2016; Kleesiek et al., 2016). They claim that high-level features can be extracted via 3D convolutional neural network (CNN) architecture most effectively. However, the challenges faced during the training of a CNN-deep model from scratch are existent in both 2D-CNN and 3D-CNN. The first, and most prominent, challenge faced by CNNs is the scarcity of training data that is labeled. CNNs need large amounts of data that is clearly annotated for training, however it is very difficult to come by this kind of data especially in the medical field due to the high cost of annotation from radiologists and high variability of the same medical disease. Deep CNN usually require multiple adjustments in the architecture or parameters of the network in order to maintain comparable learning speed between all layers, because of the presence of overfitting and convergence issues.

In this chapter, the proposed methodology is based on CNN-Deep belief learning model with constraints optimization, which classifies the gliomas (low grade) and glioblastomas (high grade) tumors in MR images. It has a unique approach which allows the units of higher layer to cover maximum locations of the input probabilistically. In this framework, convolution deep belief network is used whose hyper-parameters are distributed among whole area of the image. The presented model scales to realistic image sizes because inference can be done

smoothly using convolution. The main power of CNN lies in its deep architecture, which allows the extraction of a set of distinct features at multiple levels of abstraction. The first, second and third layers of CNN deep belief model learn edge detectors, all locations of objects and the objects themselves respectively. In order to tackle these three aspects, the proposed approach consists of four modules: Pre-processing, FCM segmentation, CNN-deep belief network classification and optimization of constraints. These modules provide an optimized solution to classify the segmented MR images by using "regularization" and “fine-tuning optimization constraints”. Hence, it provides a more accurate solution by solving overfitting and convergence problems, and also minimizes the computation time.

The experimental analysis was focused between non-segmented and segmented MR images using the proposed CNN-deep belief learning classifier, and the performance was compared between two methodologies.

Our contributions in this research are as follows:

1. We proposed a fully automatic segmentation and classification approach of Brain MR images through Fuzzy C-means (FCM) clustering algorithm and CNN Deep Belief classifier for small datasets.

2. Presented a comparison of the results of segmented and non-segmented images showing that the former provide improved features for better accuracy in tumor classification.

6.3 Methodology of CNN-Deep Belief Learning Model

6.3.1 Pre-processing

MR images cannot be directly given as the input as they are contaminated by the effects of different noises, like the Rician noise. Therefore, to improve the image quality they are first pre-processed where the noises are removed and different pixel intensities are normalized. It is necessary to have a good quality image before segmentation is done to obtain accurate

and precise results. In presented approach, input image is passed through low pass Gaussian filter which prevents high frequency artifacts thus improving the image quality through noise elimination, contract enhancement, intensity equalization and outliers removal. Additionally, the Gaussian distribution in 1D and 2D cases are shown in Eq. (6.1) and (6.2):

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \tag{6.1}$$

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{6.2}$$

where σ is the standard deviation of the distribution. In 2-D cases, an isotropic Gaussian is circularly symmetric.

6.3.2 FCM Segmentation

Fuzzy C-Mean (FCM) clustering method is broadly used to separate data into two or more clusters. It uses a process known as clustering to gather all the data points that have similar feature vectors in one cluster while grouping all the other data points that have dissimilar feature vectors in different clusters. FCM algorithm is the evolved form of K-mean algorithm such that it provides soft classification where the value of an element belonging to a cluster can vary between 0 and 1 unlike K-mean, which gives hard classification where it strictly confines an element to a cluster by giving it a value of 1 or 0 otherwise.

Brain anatomy is characterized on the basis of signal intensity received from different tissues in a MRI. FCM has a vital role in determining the difference between signal intensities of pixels in different clusters, providing a change estimation of intensities after evaluating parenchymal regions and finding connections between the morpho variations and the variations of strict parameters.

The fuzzy C-Mean segmentation method optimizes the objective function using an itera-

tive process as shown below:

$$\Phi_m = \sum_{i=1}^{q}\sum_{j=1}^{C} \mu_{ij}^{m}\|x_i - c_j\|^2, \qquad 1 \leq m < \infty \tag{6.3}$$

where $m > 1$ and is the membership degree of the value x_j in the j-th cluster, x_i is the i^{th} measured data, c_j is the n-dimension of the center and norm $\|*\|$ representing the resemblance between measured data and the center of the cluster. The membership μ_{ij} and the cluster centers c_j are updated iteratively using Eqs.(6.4) and (6.5):

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|x_j - c_j\|}{\|x_j - c_k\|}\right)^{\frac{1}{m-1}}} \tag{6.4}$$

$$c_j = \frac{\sum_{j=1}^{q}\mu_{ij}^{m}\cdot x_i}{\sum_{j=1}^{q}\mu_{ij}} \tag{6.5}$$

The iteration will stop when $max_{ij}\{|\mu_{ij}^{k+1} - \mu_{ij}^{k}|\} < \varepsilon$ where ε is a stopping criterion between 0 and 1, and k are the iteration steps. The segmentation described above is formalized into an algorithm as follows:

Algorithm 1: FCM Segmentation

Result:Automatic tumor segmentation in selected image I_{xig}

Data: 3D volume of MR brain imagesI_x , where $X = [x_1, x_2, x_3, \ldots, x_ig]$

Initialization;

For $i = 1 : q$ **do**

For $j = 1 : C$ **do**

1. $P = [\mu_{ij}]$ (matrix),$P^{(0)}$

2. Calculate center vector $C^{(k)} = [c_{ij}]$ with $P^{(k)}$

at k-step [Eq.(6.5)];

3. Update with $P_{(k)}, P^{(K+1)}$, [Eq.(6.4)];

4. **if** $\|P^{(k+1)} - P^{(K)}\| < \varepsilon$ then stop; **else**, return to step 2;

5. Update with $P_{(k)}, P^{(K+1)}$, [Eq.(6.4)];

6. Calculate $\Phi_m = \mu_{ij}^m \|x_i - c_j\|^2$

End

Find the objective function Φ_m of fuzzy clustering segmentation;

End

The fuzzy C-Mean (FCM) algorithm mitigates the intra-cluster variance, since the result is dependent mostly on the weights chosen initially. Therefore, the convergence properties of partial membership in clusters in FCM is better as compared to K-mean algorithm.

6.3.3 CNN-deep belief learning model

Convolutional neural networks (CNNs) are very important architectures as they are able to extract extremely complex features by converting the output feature maps of one layer to input channels of the subsequent convolutional layer. A feature map, in the case of neural networks, is a layer of hidden neurons or units in which each coordinate represents an individual neuron. The receptive field of that neuron corresponds to the kernel$'s$ size that also indicates that weights of the connections between the layer$'s$ neurons and the neurons in the previous layer.

It is also usually determined that the learned kernels are similar to edge detectors, where every kernel has adjusted to a different spatial scale, orientation and frequency according to the statistics of the training data.

Furthermore, Deep learning algorithms use a different approach to accomplish their task which involves a network of parameters that are organized by layers. These layers are known as hidden layers with the exception of the input and output layers. (Schmidhuber, 2015).

Algorithm

Deep networks are trained via gradient descent to reduce a predefined cost function which is usually illustrated as the negative log likelihood function of the output layer. Amongst a plethora of deep architecture, there exists a variant known as deep belief network (DBN). In this architecture, each layer is initialized as a Restricted Boltzmann Machine (RBM) thus mitigating the energy function of the inputs of RBM.

However, both DBNs and RBMs ignore the 2-D structure of images so that the weights that are needed to extract a given feature must be learned individually for each pixel. This limitation causes the scaling of these models to full MR images very strenuous due to increased computational complexity. To overcome this hindrance we utilized the CNN-deep belief learning model (CNNDBL) (Lee et al., 2009) for Brain MR images. In this model, the hyper-parameters are distributed among all locations in an image thus allowing inferences to be done effectively using convolution. It is because of this aspect that this model can scale well to full images. The CNN-deep belief learning model employs the aid of convolution deep restricted Boltzmann machine (CDRBM), which is similar to RBM in most manners, but the weights in CNNDBL $'s$ layers (visible and hidden) are shared amongst all pixels in an image. This model has L number of layers with input layer X and hidden layers H. The input layer is made of a $N_X \times N_X$ binary units array which causes the hidden layers to be of $N_H \times N_H$ binary unit's array. The hidden layers are comprised of K groups were each group contains

the aforementioned arrays, thus resulting in $N_H^2 K$ hidden units. Each of the K groups is associated with $N_W \times N_W$ filters, where $N_W = N_X - N_H + 1$ and W is the weight matrix between layers. Additionally, the hidden units in a group share the filter weights. The probabilistic statistics for RBM with binary hidden units and visible units are h and V respectively. Each possible joint configuration of the visible and hidden units has an energy and the energy of a joint configuration of the visible and hidden units determines its probability:

$$P(V,h) = \frac{1}{Z} exp(-E(V,h)) \tag{6.6}$$

where Z is the partition function. Therefore, the energy function is defined as follows:

$$E(V,h) = -\sum_{k=1}^{K} \sum_{j=1}^{N_H} \sum_{r,s=1}^{N_W} h_{ij}^k W_{rs}^k V_{i+r-1,j+s-1} - \sum_{k=1}^{K} b_k \sum_{i,j=1}^{N_H} h_{ij}^k - d \sum_{ij=1}^{N_v} v_{ij} \tag{6.7}$$

where d_{ij} and b_k are visible and hidden unit biases respectively.

From the energy function, Gibbs sampling using the following conditional distributions over hidden and visible and layers:

$$P(h_{ij}^k = 1|v) = \sigma \sum_k ((W^k * v)_{ij} + b_k) \tag{6.8}$$

$$P(v_{ij} = 1|h) = \sigma \sum_k ((W^k * h^k)_{ij} + d) \tag{6.9}$$

where logistic sigmoid function is σ.

6.3.4 Architecture

The basic architecture lies with the CNN-deep belief learning model which is a continuation of the multi-layer perceptron (MLP) architecture analogous to DBNs. A typical convolutional neural network consists of several feature extraction layers, each of which contains numerous

convolutional nonlinear activation layers, and pooling or sub-sampling layers.

Convolution Layer

The core layer of CNN are the convolutions. Convolution layer is used to compute feature maps in a deep learning model. In this network every unit has its own respective field, a grid of units in the previous layer which it receives input from. The receptive fields of units typically overlap. In convolution layer, the convolution filters are implemented on small parts of the image where they start detecting the simple image features. The filters work similar to the low-level vision processing ability of the human brain such that it first detects simple features like straight edges or circles for organs or round objects and then move on to more complex and higher order features that include texture, and global and local shape. Resultantly, the output given by CNN is usually a number of probabilities or class labels. Moreover, a set of training data is used to train the convolution filters after which they are sent through an activation function in order to obtain an output feature map.

$$Z_j^k = \sum_{k=1}^{K} \sum_{j=1}^{N_H} \sum_{r,s=1}^{N_W} h_{ij}^k W_{rs}^k V_{i+r-1,j+s-1} + \sum_{k=1}^{K} b_k \sum_{i,j=1}^{N_H} h_{ij}^k \tag{6.10}$$

$$Y_j^k = f\left(\sum_{i \in M_j} \left(Z_j^k \right) \right) \tag{6.11}$$

where M_j represents a numeral of input feature maps and Y_j^k is the value of the j^{th} output layer unit in response to the k^{th} input pattern.

Rectified Linear Unit(ReLU)

To form a neuron's output function f of its input x is with $f(x)$. Gibbs sampling forms the basis of learning algorithm as a logistic sigmoid activation function.

$$f(x) = \frac{1}{1 + exp^{-}x} \tag{6.12}$$

The activation function maps $[-\infty, +\infty] \rightarrow [0,1]$.The maximum nonlinearity take place at $f(\pm 1) = \pm 1$ and will hence keep away saturation during training if the wanted training targets are normalized (LeCun et al., 2015). During training time with gradient decent, these saturating nonlinearities are much slower than the non-saturating nonlinearities.

Probabilistic Pooling

Higher-level features detectors require information from progressively larger input regions. Existing translational invariant networks such as convolution, comprise of two types of layers called, detection and pooling. Representation of detection layer can be reduced by using a constant scale. In particular, in pooling layer, each unit calculates the maximum activation in a small area of detection layer. Minimizing the representation with pooling allows higher layer illustration to be invariant to small translations of the input and reduces the computation cost.

In this model, pooling layer p and detection layer H, both have K groups of units and each group of pooling layer has $N_p \times N_p$ binary units and each pooling layer p^k reduces the detection layer H^k by a factor of r along every dimension. The detection layer H^k is partitioned into blocks of size $r \times r$ and each block is connected to exactly one binary unit p^k_α in the pooling layer i.e. $N_p = N_H/r$. Therefore, the energy function of probabilistic pooling is:

$$E(V,h) = -\sum_k \sum_{i,j} \left(h^k_{i,j} \left(\tilde{W}^k * v \right)_{i,j} + b_k h^k_{i,j} \right) - d \sum_{i,j} v_{i,j} \tag{6.13}$$

If $\sum_{(i,j)\in B_\alpha} h_{i,j}^k \leq 1, \quad \forall k, \alpha$ and $B_\alpha \cong (i,j) : h_{i,j}$ belongs to the block α.

6.3.5 Optimization Constraints for training CNNDBL model

Multi-layers parameters(MLPs), which are generally called hyper-parameters, can not be optimized through gradient descent technique only. Hence to optimize them, the following constraints are used:

Zero-One Loss

The main purpose of training a classifier is to reduce the various errors, also known as zero-one loss. In order to mitigate the error for a model as much as possible, the log-likelihood of classifier in a training set needs to be maximized. Since the likelihood of a correct class is not equal to the number of right predictions, the cost function becomes the negative log-likelihood:

$$NLL(\theta, D) = -\sum_{i=o}^{|D|} logP\Big(Y = y^{(i)} | f(x^{(i)}), \theta\Big) \tag{6.14}$$

where $\theta = [W, b]$ and D is the training set.

The NLL is the replacement for zero-one loss, and gradient of the function over the training data as a supervised learning signal for deep learning of classifier. Moreover in CNN-deep belief learning model, the cost function is :

$$f(Z_j^k) = \frac{1}{1 + exp^{(Z_j^k)}} \tag{6.15}$$

Stochastic Gradient Descent (SGD)

In convolution deep belief network (Lee et al., 2009), the author used an ordinary gradient decent method that estimates the gradient of an entire training data set resulting in a higher variance and computation time. In this architecture, stochastic gradient decent is used to update the weights by mini-batch of the dataset. This technique reduces the variance in the estimate

of the gradient. The use of smaller batches of the datasets also minimizes the computation time.

Regularization

Additionally, in order to prevent overfitting, the training data is optimized through regularization. In this process, the absolute value of the kernel weights are bounded and, L1 and L2 regularization is applied on all layers to prevent overfitting. This is achieved by fusion a regularization term to the negative log loss function. Furthermore, L1 and L2 regularization is also called "weight decay" thus making the regularized loss is:

$$E(\theta, D) = NLL(\theta, D) + \lambda_1 \|\theta\|_1 + \lambda_2 \|\theta\|_2 \tag{6.16}$$

where $\|\theta\|_n = \left(\sum_{j=0}^{|\theta|} |\theta_j|^n \right)^{\frac{1}{n}}$ which is the L_n norm of θ.
Regularization requires that the norm of the incoming weight vector at each hidden unit be upper bounded by a fixed constant λ. In other words, if θ represented the vector of weights incident on any hidden unit, the neural network was optimized under the constraint $\lambda_1 \|\theta\|_1$ and $\lambda_2 \|\theta\|_2$. In addition, the constant λ is a tunable hyper parameter for L_1 and L_2 regularization terms respectively, and it is determined using a validation set. While, L_1 encourages sparsity, L_2 encourages small values and they both work together to improve the performance of SGD of deep neural networks.

Learning rate and Epoch

The Learning rate is controlled by:

$$\rho = \frac{\mu_o}{1 + \gamma t} \tag{6.17}$$

where μ_o is the initial rate and γ is a "decrease constant", after which the learning rate ρ decreases and t is the epoch/stage. Finally, choosing the learning rate for each parameter of

the network is adaptively based on the error of the classifier.

Batch size

The batch size is the amount of training samples considered when the optimization solver is updated once. Before choosing a certain batch size, several factors are kept in consideration i.e. the computational cost and the uncertainty of update from a small batch to a large batch.

A smaller batch size produces more noise as compared to a larger batch size. However in the presence of a large number of minima in the error function, the model may get stuck in the first minima it encounters. Hence an ideal batch size needs to be chosen that will optimize the model by introducing more noise in the model estimate of the gradient. This noise can then be utilized to push the model out of some shallow valleys in the error function.

6.3.6 Performance evaluation parameters

Accuracy was used for measuring the robustness of classifiers :

$$Accuracy = \frac{(TP+TN)}{TP+FP+TN+FN}$$

Error/Loss is interrelated with the selection of the output layer activation function. For standard regression problems, the output-layer activation function is the identity $f(x)$ as in Eq.(6.12), which is also called linear activation function, due to the fact that the output is then a simple linear combination of the inputs. The linear activation function does not constrain the output to a defined range and is therefore suitable for regression problems. Assuming that the training data set is given as $X = [x_1, x_2, x_3, \ldots, x_{ig}]$, with the corresponding output vectors as $Y = [y_1, y_2, y_3, \ldots, y_{jg}]$, then back-propagation error/loss is formalized into an algorithm as follows:

Algorithm 2: Error Backpropagation

Result: Evaluate the "error term $\delta_j^{(k)}$

Data: 3D volume of MR brain imagesI_x , where $X = [x_1, x_2, x_3, \ldots, x_{ig}]$

Initialization;

For $l = 1 : k$ **do**

For $j = 1 : h$ **do**

1. Calculate "error term" by using "back-propagation formula"

$\delta_j^{(l)} = f'(x_j^l) \sum_l W_{rs}^{(l+1)} \delta_k^{(l+1)}$

where f' is the inverse of activation function.

2. Calculate the gradient of the error function with respect to the

weight vector w_{rs}

$\frac{\partial E(W,b)}{\partial w_{rs}^l} = \delta_j^{(l+1)} f(x)_i^l$; where $f(x)$ is logistic sigmoid activation function

3. Repeat step (2) by using "back-propagation formula"

End

End

The output of network is defined as $h_{W,b}(x)$, the error function can be defined as:

$$SSE(\theta) = \frac{1}{2} \sum_{i=1}^{n} \|h_{W,b}(x_i) - y_i\|^2 \tag{6.18}$$

It is known as sum of squares error function. Where in binary classification, error function can be chosen as a "cross entropy error function":

$$CE(\theta) = -\sum_{i=1}^{n} \{y_j lnh(x_i) + (1 - y_i) ln(1 - h(x_i))\} \tag{6.19}$$

6.3.7 Flow Chart

The flowchart Fig.[6.1] outlines the process of tumor classification of high grade and low grade images through deep belief learning classifier by using optimization constraints.

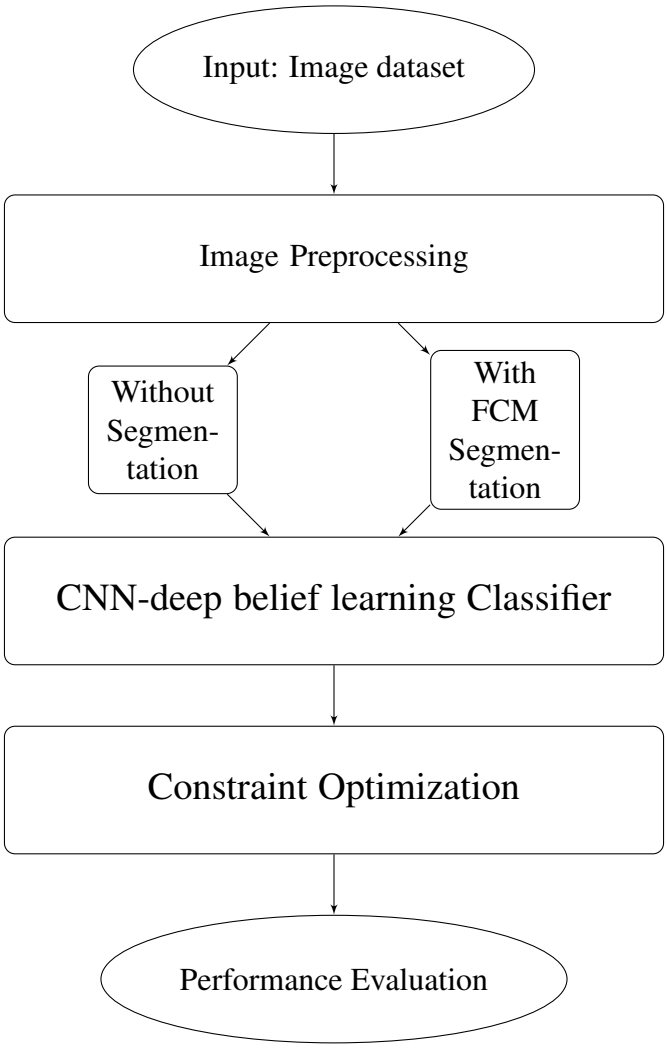


Figure 6.1 Flow diagram of the methodology

6.4 Implementation

Before implementing the proposed model, a MRI dataset was acquired from BRATS (multi-model brain tumor segmentation) challenges (MICCAI, 2016). It consisted of four MRI

modalities, T1, T1-C, T2 and FLAIR, amongst which the pathological T2 weighted images were used due to their higher resolution. However since the dataset was a 3D volume dataset, it had to be converted into 2D form. This resulted in 3 slices of 2D images, specifically x, y and z slice. Since, the z slice provided a holistic representation with complete information, it was preferred as shown in Fig.[4.1].

Then the implementation of our methodology was based on NVIDIA deep learning GPU training system. This technique trains and uses multi-layered artificial neural networks to help accomplish a relevant task without any involvement from humans. Furthermore, DNNs that are used for image classification normally use an amalgamation of CNN and fully connected layers. Where the fully connected layers respond to overlapping visual field regions through the use of tiled artificial neurons.

In this study, we used a dataset that comprised of 280 original MR images of size 256 x 256. Then amongst those images, 70% were used for training, 15% for validation and the rest for testing.

After the dataset was created and set up, the images were pre-processed, employed a low pass digital filter known as the Gaussian filter which removed the highest and lowest intensities from the images effectively. Once pre-processing was complete, we used two methodologies to compare the accuracy and efficiency of deep belief network learning models. In the first methodology, we used the original MR images for the training and testing of the deep belief learning classifier. Whereas in the second methodology, the images were segmented first via adaptive FCM automatic segmentation method and then the segmented images were applied to the same classifier. However, to achieve the best accuracy, the classifier was first optimized by tuning parameters which include learning rate, batch size, epochs and kernel size for convolution. The resultant classification from the two methodologies was then evaluated based on several performance-related aspects.

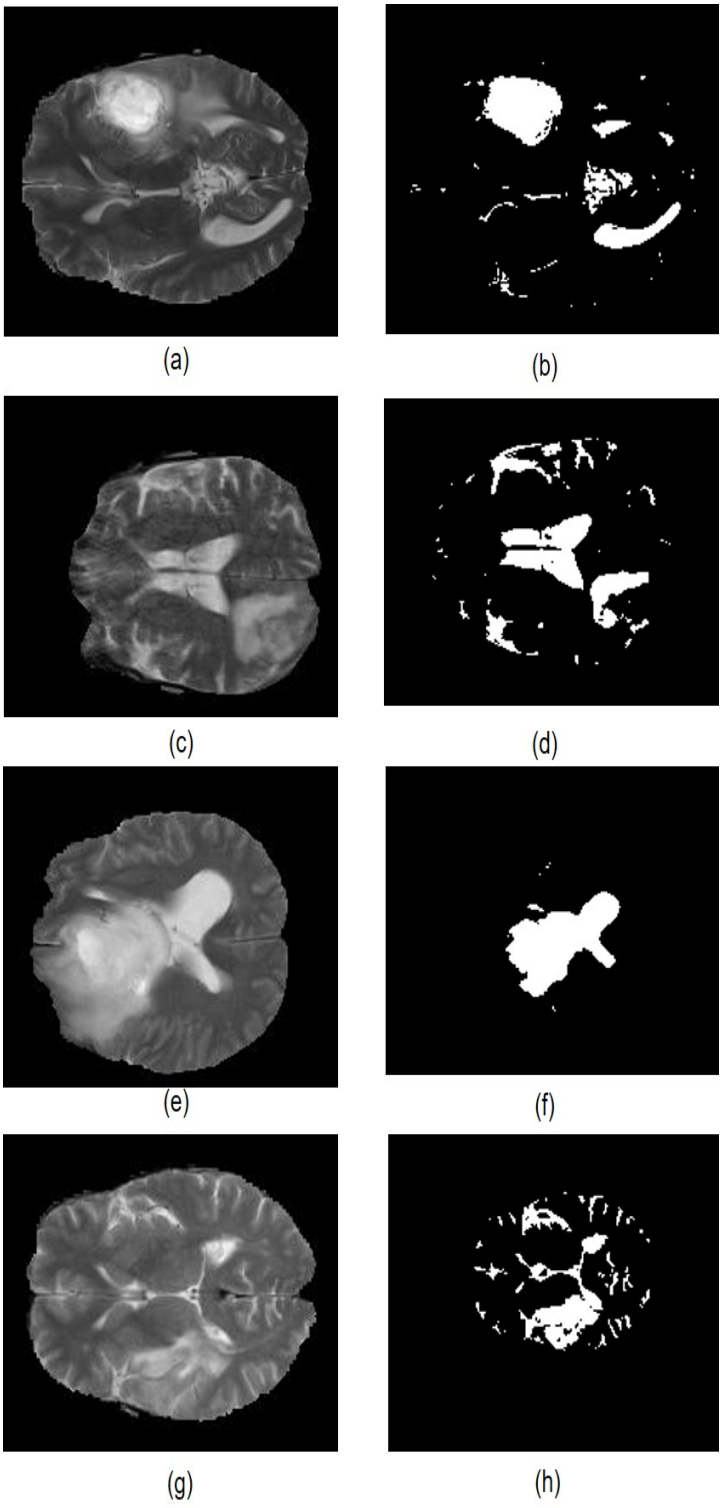


Figure 6.2 Original and FCM Segmented MR images (a,c) High grade tumor MR images (b,d) High grade segmented images (e,g) Low grade tumor MR images (f,h) Low grade segmented images

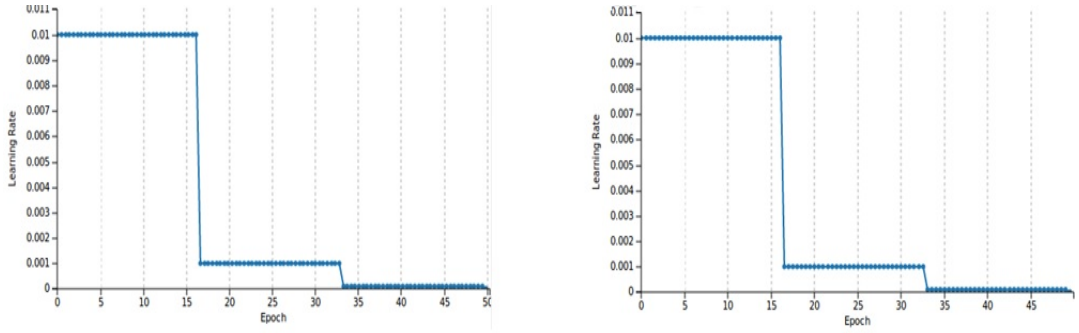


Figure 6.3 Segmented and Non-segmented classifier learning rate (a) Learning rate and Epoch graph of non-segmented images, (b) Learning rate and Epoch graph of segmented images

6.5 Results and Discussion

The experimental results obtained after going through the proposed methodologies have been presented. First, the outcome of the segmentation procedure, achieved through FCM algorithm, has been compared with the non-segmented images in Fig.[6.2]. The architecture employed for the classifier (for both segmented and non-segmented images) had five convolutional layers with kernel size 9, 7, 5, 3 and 3 for each layer respectively.

Then, we used the stochastic gradient descent (SGD) to train the classifier model by setting the momentum as 0.9 and weight decay value as 0.0005. The importance of weight decay with regards to the models ability to learn since it reduces the training error encountered in the model. A zero-mean Gaussian distribution with standard deviation 0.01 was also used to initialize the weights after which the same learning rate was used for all layers by the model. By following a heuristic process in which the learning rate was divided by 10 as soon as no improvement was noticed in the validation error rate along with the current learning rate. As it is indicated in Fig.[6.3], the learning rate was terminated after being reduced three times when it was initialized at 0.01.

Moreover, during the experiment, the batch size was chosen to be 10 for 50 epochs. After optimizing all the constraints, the classifier was implemented on both the segmented and non-segmented images. The Fig.[6.4] gives a comparison between epochs, accuracy and loss

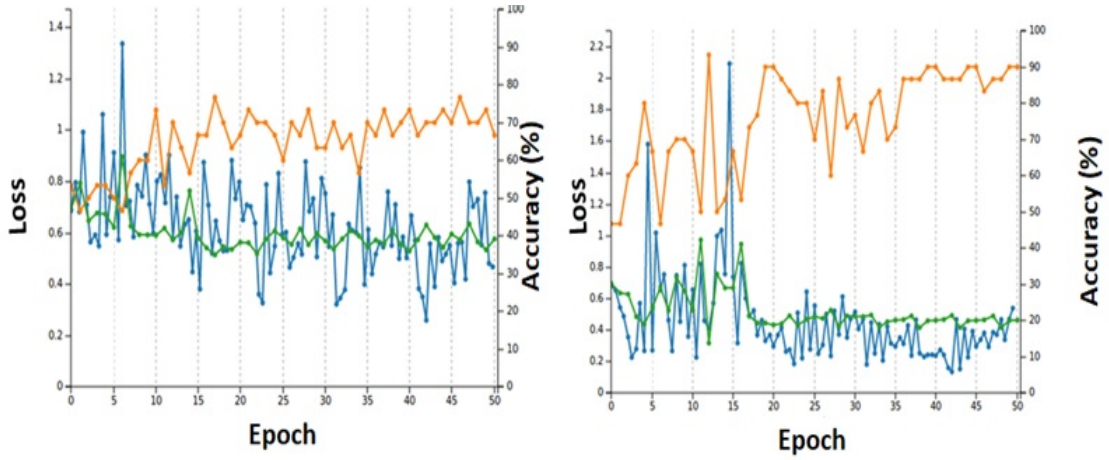


Figure 6.4 Segmented and Non-segmented classifier accuracy (a) Accuracy and loss graph of non-segmented images, (b) Accuracy and loss graph of segmented images

(validation and testing loss) for segmented and non-segmented images. As it can be seen from the general trend and the figures provided in Table [6.1] that the maximum and average accuracy of the segmented MR images is higher than non-segmented images while the loss is also lower. Moreover, the execution time for segmented images is also lower than original images for the same epochs thus signifying a lower computational load for the former type of images.

Table [6.2] provides a comparative analysis of models that are currently in use in the field of deep convolution network. The results show that for grayscale images, CNNDBL model outperforms most of the recent approaches for this dataset.

After the models were trained, 20 high grade and low grade images were tested to evaluate their performance individually. A comparison between six individual images for the segmented and non-segmented methodologies was shown in Table [6.3]. The results indicated in the table signify the models better performance while processing segmented images due to the higher percentage of prediction as compared to the original, non-segmented MR images for both high and low grade.

The proposed model utilized undirected connection amongst all layers which differs from the model proposed by Hinton et al. (Hinton and Salakhutdinov, 2006). That model involved only the top two layers using undirected connections whereas the other layers worked by uti-

Table 6.1 Comparison of Performance

Model	**Accuracy (%)**		**Error/Loss (%)**		**Execution time (Sec.)**
	Maximum	**Avarage**	**Maximum**	**Avarage**	
Original MR images	73.3	76.7	76	53	370
Segmented MR images	93	90	31	46	340

Table 6.2 Comparison of CNNDBL model with the state-of-art models

Model	**Non-segmented MR images**		**Segmented MR images**	
	Accuracy (%)	**Error (%)**	**Accuracy (%)**	**Error (%)**
CNNDBL Model (Proposed)	**76.7**	**53.0**	**90.0**	**46.0**
DCNN Model (AlexNet)Krizhevsky et al. (2012a)	56.6	59.1	62.2	60.1
CNN Model (LeeNet)LeCun et al. (2015)	50.1	90.0	49.2	90.0

Table 6.3 Test results of individual images

Dataset	**Non-Segmented MR image Prediction**		**Segmented MR images Prediction**	
	HG (%)	**LG (%)**	**HG (%)**	**LG (%)**
HG0001	64.58	35.42	99.6	00.40
HG0002	63.17	36.83	99.5	00.41
HG0004	66.64	33.36	99.6	00.40
LG0001	29.55	70.45	0.42	99.58
LG0002	39.68	60.32	0.41	99.59
LG0004	32.22	67.78	0.38	99.62

lizing the top-down directed connection. The author also used a single bottom-up pass to estimate the posterior distribution. However in this method the higher layers cannot work together with the lower layers to estimate the posterior in the presence of occlusions or uncertainties. On the other hand, the proposed approach can smoothly combine bottom-up and top-bottom

information between the layers because of the use of undirected edges. Furthermore, this model has been optimized by probabilistic max-pooling and convolution. The former helps by shrinking the higher layers to make the model scalable while the speed is improved by weight-sharing through convolution.

The hyper-parameters of both models (segmented and non-segmented) such as i.e. kernel and maximum pooling size of each layer and number of layers, were kept the same in both models. However they were tuned and optimized using grid search and cross validation on a validation set and chosen based on their performance on that set; the hyper parameters upon which the model performed best were chosen. For maximum pooling, we always used a stride of two in order to retain per pixel accuracy during full image prediction.

The combination of these CNNDBL model algorithm and optimization of constraints approach ensures a degree of shift, scale and distortion invariance while improving the durability against variations in the location of distinctive features in the input data. Hence, this framework was the most suitable option and capable of tackling the complexities of medical radiology images.

In the first proposed technique, the training of non-segmented MR images takes about 7.5 seconds per epoch for the CNN deep belief learning model, on a NVIDIA GeForce 940MX graphics card, meanwhile it takes about 6.8 seconds per epoch for the same model to train itself on segmented MR images. During the training and testing time, we ran our models on the GPU in order to exploit its computational speed, reducing the execution time further. We noted that the segmented images used in the classifier model were processed 8-10 times faster than non-segmented classified model. Predictions for one image with CNNDBL model take on average 4 minutes.

In addition, as mentioned in the previous section (Experimental results) that the classifier model learns the features more accurately in the case of segmented images. Conversely, it is unable to distinguish the features of the tumor in non-segmented images because brain tumors

can be of any size and shape, they can appear at different locations, can have varying image intensities and may also deform the nearby structures and appear together with edema, thus changing the intensity properties. Therefore, it is very difficult for the model to detect the tumor without any help. FCM has the ability to detect abnormal regions, where the intensity characteristics don't match the expected result, and then determine the presence of tumor or edema in that particular region. This allows the segmented image to have more recognizable features with appropriate intensity parameters for the tumor and edema regions resulting in a more accurate classification.

It is a common medical practice for radiologists to identify the tumors in medical images by observing and detecting the tumor region in different 2D image slices of a patient's scan Huang et al. (2012). Most medical images do not require complex processing tasks as compared to natural images thus eliminating the need for complicated CNN architectures such as that of 3D data. 3D-CNN has also some open issues which have not been resolved like the number of images in every case is different so an automatic model, especially using machine learning, without bias is difficult to design. It is an infeasible task to develop an extremely complex network of 3D architecture and volumetric data that faces the same issues as 2D-CNN (Dou et al., 2016). Therefore, it is better to use an optimized classifier model for 2D-CNN which is not only more relevant to radiologists, due to its practicality, but it also maintains detection accuracy and reliability.

6.6 Summary

In this chapter, framework for brain tumor classification by CNN deep belief learning model classifier (CNNDBL) has been presented. High and low grade brain tumor classification results were compared by using non-segmented and FCM based segmented MR images. The proposed approach comprises of four steps for accurate identification and characterisation of the tumor: Pre-processing, segmentation, constraints optimization and classification. Once

the neuroimaging classification was implemented, the results were evaluated on the basis of average and maximum accuracy and loss, and the execution time. During the performance analysis, it was determined that the segmented images were classified about 19.7% more accurately than the non-segmented images while the loss (validation and training loss) in the segmented images decreased by approximately 60%. In addition to correct classification, the segmented images also required 10% less computation time as compared to the original images. Lastly, individual high and low grade images were tested for true prediction and it was determined that the segmented MR images (in both HG and LG cases) were 40 % more accurately predicted than the non-segmented images.

Chapter 7

Conclusion and Future Work

There were two main goals of this thesis, first was the development of methods for tumor segmentation through level-set and, external and internal energies while the second goal was classification via machine learning. In the second part of the thesis, where classification was presented, two approaches were proposed: one that exploited specifically machine algorithm with feature reduction technique and the second exploited deep belief learning model with specific architecture for brain tumor in MR images. Moreover, the application of the contributions presented in this paper are not limited to MR images but also have the potential to be used for other radiology gray scale images in the field of medicine.

7.1 Conclusion

This thesis first presents the anatomy of brain, brain tumors in Magnetic Resonance Imaging (MRI) and the physics behind MRI. In our research, we are using T2-weighted images for brain tumors segmentation and classification because it provided the best visualization of the tumor amongst the three modalities leading to more accurate segmentation and classification. The characteristics of brain tumors were also discussed, especially in MR images where we investigated those aspects which are clearly distinguishable in MR images. The WHO has

classified the tumors according to their severity which resulted in two grades of tumors; high and low grade. We used these two categories of tumors for machine learning and deep learning classifiers.

We then moved on to identifying the state-of-the-art methodologies for the digital diagnosis of cancer in brain MR images. This process involved searching numerous research works that attempted to provide a solution through segmentation and classification. We discovered that in every aspect of the process of diagnosis; whether it is classification, machine learning or deep learning, the solutions always had some deficiencies with regards to MR images. The complexity associated with the homogeneity, image acquisition noise and randomness of the tumor were a major hindrance for previous segmentation approaches. Moreover, classical machine learning techniques were limited due to their heavy reliance on computing power, biasness in the presence of a small dataset and a long computation time. The literature survey further presents a relatively modern approach to classification called deep learning. It highlights the benefits of deep learning over machine learning and emphasizes on its ability to reduce computation costs while maintaining accuracy. However, deep machine learning also has its restrictions specifically the inability to scale to realistic image sizes used in MR imaging. This chapter therefore helps differentiate our contributions by highlighting the improvements required in previous works.

Perfect segmentation is the main problem, especially for brain tumors. We investigated this issue in the presented research and provided information relating to the detection and segmentation of tumors in MR images, specifically brain tumors, for medical diagnosis. The main target that was intended in this research, was to create such an algorithm that is able to identify a region of interest within a MR image by detecting the anatomical anomalies and abnormal tissues in the image. We came up with an active region-based contour model for localization and global set curve evolution adaptive stochastic approach which helps create a complete region based energy minimization. In order for this model to be effective, it performed four

basic steps that allowed it to segment the region of interest. First of all, an energy based region was localized using this framework, then a localized active contour was created that kept stable over the region of interest through energies and forces at each point. Finally, the tumor segmentation created was analyzed based on the local radius. This methodology was able to achieve maximum accuracy in which the segmented region matched the desired region. It also improved the iteration speed, hence the computation time decreased, and allowed for a flexible topology. It was also noted that a combination of global and local energies provides the most optimized solution for segmentation in case the global region based energies are insufficient by themselves.

Various classification methods were applied for the purpose of novel, and effective tumor discrimination. For more accurate and less computation cost, we proposed two classification approaches.

Our first proposed classification technique of brain tumors in MR images is based on machine learning classifier and probabilistic features. This approach provides a comparison between two fundamental proposed methodologies that employ various machine learning classifiers such as SVMs, ANNs, Decision tree, Naive Bayes and k-mean. In the first methodology, we performed classification using the extracted features obtained from GLCM. However, in the second methodology, we first performed feature reduction using Principal Component Analysis (PCA) on the extracted features and then we sent those reduced features to the classifier. It was observed from the experimental results, that the second technique showed a marked improvement in the classification of high and low grade tumors. The classification accuracy of GLCM extracted features and then reduced with PCA was more than 97%. The results prove that the use of PCA can make any classifier more accurate and robust while increasing its potential applications.

Our second approach is based on deep machine learning classification which is the new innovation after classification through machine learning. This research particularly focused on

the proposed methodology which was based on CNN-Deep belief learning model with constraints optimization on segmented and non-segmented MR images. The specialty of CNN-deep belief learning model is that the higher layer units have the ability to cover larger areas of the input image in a probabilistic way. Not only this, but the proposed model can also be used on realistic image sizes due to its ability of making inference using convolution. In order to create the most efficient version of this model, four modules were proposed, which include Pre-processing, FCM segmentation, CNN-deep belief network classification and optimization of constraints, in that order. These modules can be optimized through regularization and fine tuning of constraints to produce an optimized solution, of classification of segmented MR images, that is not only more accurate but also takes less computation time. After testing CNNDBL, the model's results were evaluated based on accuracy, loss and execution time. It was observed that the accuracy of the segmented images was higher than the accuracy of non-segmented images by about 19.7% whereas the loss decreased by 60%. Moreover, the computation time for segmented images was also approximately 10% than the non-segmented images. The last performance parameter that was evaluated was the prediction accuracy of individual high and low grade images. The result of this test was that there was an increase in prediction accuracy of 40 % for segmented MR images hence proving that segmenting the images before running through the classifier results in a better performance. Finally, this thesis concludes that the presented work can improve a physician's ability to detect and analyse pathologies leading to a more reliable diagnosis. This can make the diagnostic process easier and faster, and can overcome the problem of variation in the reported diagnosis.

7.2 Future Work

In the future, the aim of this project would be to improve the accuracy and variability of this model by incorporating other modalities (like CT, PETS etc) and constructing multi-modal classification models. Moreover, this project could also be veered in a different direction

by including qualitative features, along with quantitative features already in use, for a more distinct classification and detection of tumors in images. It is also necessary to develop the current model through the use of the most innovative and state-of-the-art techniques that could make this process more effective without compromising the accuracy. Additionally, it must be kept in mind that the main purpose behind this study is to provide such a system to doctors and clinicians that can expedite their diagnosis period. Therefore it is necessary to work on improving the execution time of these types of models; the less computation time required, the better the model. Lastly, keeping in consideration the development of other aspects in technology, specifically 3D images and holograms, it must be noted that this model could be developed further to work with 3-Dimensional images in order to provide a more holistic classification of the tumor after using a more detailed image. Finally, the algorithms and models developed must undergo extensive testing and performance evaluation before they can be deemed hospital-worthy to ensure proper functionality.