Classification of Malignant Brain Tumors using MRI by Machine Learning

A THESIS SUBMITTED

in

in the Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

ANKIT VIDYARTHI

2012RCP9514

Under the supervision of

Dr. Namita Mittal

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Malaviya National Institute of Technology Jaipur

INDIA - 302017

March, 2017

Declaration

I, **ANKIT VIDYARTHI**, declare that the work done in this thesis titled, ***Classification of the Malignant Brain Tumors using MRI by Machine Learning*** is my own contribution. The written submission embodies my ideas in my own words.

I confirm that:

1. This work is done wholly while in candidature for a Ph.D. degree at MNIT Jaipur, Rajasthan, INDIA.
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at MNIT or any other institution, this has been clearly stated.
3. Where I have consulted the published work of others, this is always clearly attributed.
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this Dissertation is entirely my own work.
5. I have acknowledged all main sources of help.

Signed:

Date: March 6, 2017

Certificate

This is to certify that the thesis entitled **"Classification of Malignant Brain Tumors using MRI by Machine Learning "** is being submitted by **Mr. Ankit Vidyarthi** (Registration No. 2012RCP9514) to the Department of Computer Science & Engineering, Malaviya National Institute of Technology Jaipur, Rajsthan, INDIA, for the award of the degree of Doctor of Philosophy in Computer Science & Engineering. It is an original research work carried out by him under my supervision and guidance.

(Dr. Namita Mittal)
Assistant Professor
Department of Computer Science & Engineering
Malaviya National Institute of Technology Jaipur

Acknowledgement

First of all I would like to express my heartily gratitude to **Dr. Namita Mittal** for her enormous contribution in the realization of this thesis. She stood by me not only as a supervisor but also as a mentor. She also taught me Data Mining, Machine Learning, programming, and the art of writing scientific articles for various standard digital repositories.

I am very much grateful to **Malaviya National Institute of Technology Jaipur** for providing me the opportunity to carry out my Ph.D work in the Department of Computer Science and Engineering. I am thankful to **Prof. M.C. Govil** and **Prof. M.S. Gaur**, Heads, Department of Computer Science & Engineering, all faculty members and technical staff, for the fruitful conversations we had all these years.

I am also thankful to all the members of my *Department Research Evaluation Committee* (DREC) **Dr. Girdhari Singh, Dr. Dinesh Goplani, and Dr. Mushtaq Ahmed** for giving their valuable time and suggestions in every semester during progress presentation.

I would like to acknowledge the guidance of **Dr. Sunil Jakhar**, M.D., Department of Radiology, Sawai Man Singh Medical College Jaipur for providing us the MR dataset of Malignant Brain Tumors.

My special gratitude also goes to all the persons from whom, I have received critical reviews, Prof. R.K. Agarwal (Professor, JNU New Delhi), Prof. V.C.V Rao (Associate Director, CDAC Pune), Prof. Senthil Kumaran (Associated Professor, AIIMS New Delhi), Dr. Kuntal Ghosh (Associate professor, ISI Kolkata), Dr. Pradipta Maji (Associate professor, ISI Kolkata), Dr. Subrahmanyam Murala (Assistant Professor, IIT Ropar), during this period.

I also thank to the research scholars, especially Dr. Basant Agarwal, Mr. Lokesh Kumar Sharma, Mr. Chandra Prakash, Mr. Vijay Sharma, Mrs. Mithlesh Arya, and Ms. Akanksha Gaur, from the Department of CSE, Malaviya National Institute of Technology Jaipur, for helping me in my thesis editing. I would also like to thanks all my departmental colleagues for helping me through out the journey of my research progress.

Last but not least, I would like to thank my family for believing in me and supporting me with all possible means.

Ankit Vidyarthi

Dedication

To
My Loving Parents, My Wife,
&
My Daughter

Abstract

Brain Tumor is one among the most noxious diseases in the arena of medical science. The process of brain tumor identification requires a moderately sophisticated assessment of the various *Magnetic Resonance Imaging* (MRI), *Computer Tomography* (CT) Imaging and *Positron Emission Tomography* (PET) Imaging. The assessment is performed by experienced radiologists using heuristic learning and machine intelligence. Despite certain subjective features associated with tumors, an expert radiologist accomplishes the assignment with the higher degree of accuracy. However, to enhance the accuracy diagnosis for better treatment of the tumor patient's, medical science is continuously seeking the machine learning algorithms for tumor cells identification and categorization. These machine learning algorithms help in the segmentation of the medical images, extracting features from the segmented part, selection of the relevant and informative features among the extracted feature set and classification of the segmented regions based on *World Health Organization* (WHO) grading.

The main problems with the state-of-art machine learning algorithms are as follows.

1. Most of the existing machine learning algorithms for segmentation suffers from the selection of the predefined variables like threshold value, the number of required clusters and initial starting point. These algorithms depend on the pre-initialization of input data and thus the performance of these algorithms has been proved to be limited.

2. Machine learning based classification works on the set of the extracted features from the segmented regions. These features are used to describe the image in the digital system either by using spatial properties or spectral properties. To select the appropriate feature extraction method is an issue in machine learning.

3. Machine learning classifiers perform well in the presence of an efficient feature selection technique. Feature selection is used to eliminate the noisy and irrelevant features and to reduce the size of the high dimensional feature vectors. To investigate an effective feature selection technique for spatial and spectral features is a challenge.

4. In the classification problems, the selection of the number of nearest neighbor for decision making is still an open issue. This nearest neighbor selection is

based on the choice of the user and it varies from dataset to dataset. The selection of an optimal value for nearest neighbors is another problem in machine learning based classification problems.

The objective of this thesis is to design, implement and improve the performances of machine learning algorithms that identify and categorize various types of malignant brain tumors in MRI. This thesis focuses on finding a solution to two main issues i.e. (i)Automatic segmentation of the tumors cells from MRI, and (ii) Finding the proper combination of algorithms for feature extraction, feature selection and classifier, that improves the classification accuracy. Besides it, the dynamic selection of the optimal number of nearest neighbor for classification is also the target of this thesis.

To fulfill the objective, firstly in the thesis, the hierarchical segmentation technique is proposed. The hierarchical approach uses the concept of tree formation to extract the *Region Of Interest* (ROI) from MR image. These segmented ROI's are used in the extraction of features in machine learning system. Secondly, for discrimination among the types of malignant tumor dataset, this thesis proposes two different algorithms *Counting Label Occurrence Matrix* (CLOM) and *Texture Occurrence Matrix* (TOM) for extracting spatial and spectral textural features. Further, the relevant and the informative features are selected from the overall extracted features using the proposed *Cumulative Variance Method* (CVM) and hybrid dimension reduction approach. The statistical validation of the proposed CLOM and TOM algorithms is given by using T-test method.

Finally, at the end of the thesis, a new voting based mathematical rule is proposed that automates the selection of the number of nearest neighbors in classifiers like *K-Nearest Neighbor* (KNN). The proposed rule provides the optimal number of nearest neighbors using space reduction approach. The proposed approach don't use any selection of the predefined variable like threshold. The experimentation results of the proposed rule are also compared with the existing state-of-art classifiers and found that by using proposed methodology the advancement in the classification accuracy is gained. At the last, the proposed rule is statistically validated by using the McNemar's test.

All the experiments are performed using the clinical dataset consists of six hundred sixty (660) malignant brain tumor MRI. All the images are of T1-weighted post-contrast axial modality taken from 3.0 T GE MR Scanner. The dataset includes

five classes of malignant brain tumor i.e. *Central Neuro Cytoma, Glioblastoma Multiforme, Gliomas, Intra Ventricular Malignant Mass and Metastasis* collected from Department of Radiology, SMS Medical College Jaipur, Rajasthan, INDIA .

List of Tables

List of Figures

Acronyms & Notations Used

Acronyms

MRI - Magnetic Resonance Imaging
CT - Computer Tomography
PET - Positron Emission Tomography
MR Magnetic Resonance
CBTRUS - Central Brain Tumor Registry of the United States
US - United States
CNS - Central Nervous System
ROI - Region of Interest
GLCM - Gray Label Co-occurrence Matrix
DWT - Discrete Wavelet Transformation
CLOM - Counting Label Occurrence Matrix
TOM - Texture Occurrence Matrix
KNN - K-Nearest Neighbor
SVM-RBK - Support Vector Machine with Radial Basis Kernel
CVM - Cumulative Variance Method
AVNM - A Voting based Novel Mathematical Algorithm
GE - Global Electronics
PNET - Primitive Neuro-Ectodermal Tumors
NMR - Nuclear Magnetic Resonance
PRESS - Point Resolved Spectroscopy
STEAM - Stimulated Echo Acquisition Mode
WHO - World Health Organization
VOI - Volumes of Interest
FCM - Fuzzy C-Means
DCE - Discrete Curve Evolution
PD - Proton Density
PSO - Particle Swarm Optimization
PDF - Probability Density Function
DCT - Discrete Cosine Transformation
PCA - Principal Component Analysis
LLE - Locally Linear Embedding
ICA - Independent Component Analysis

mRMR - Minimum Redundancy Maximum Relevance
LDA - Linear Discriminant Analysis
ANN - Artificial Neural Network
BPNN - Back Propagation Neural Network
SOM - Self-Organizing Map
PS - Probabilistic Sum
FDR - Fisher Discriminant Ratio
LOO - Leave-One-Out

Contents

Chapter 1

Introduction

Brain is one of the most sensitive part of human body. It comprises a collection of numerous soft tissues and cells. The development of the brain depends on the repetitive formation of new cells and destruction of old cells. A brain is called *normal brain* or a healthy *brain* if the formation and destruction of the cells are in controlled manner. An uncontrollable formation of the cells in any part of human body results in the formation of solid mass of cells, called *tumor* in medical science. Depending on the characteristics of tumor, they are characterized into *benign* and *malignant* tumors. Based on the growth rate of tumor during specific time interval, if the size of tumor remains constant or grows very slow then the tumor is called *benign tumor*. On contrary, if the size of tumor grows rapidly then the tumor is called *cancerous* or *malignant* tumor. *World Health Organization* (WHO), based on the ground statistics of medical science study and research, classifies these tumors in four grades, based on the individual characteristics that each tumor holds.

Brain tumors identification and their classification is the challenging task in medical science. To extract the proper regions from the imaging modality having abnormality is the key concern for the radiologists. Earlier in past, radiologist used the naked visualization methodology for the analysis of the brain tumor imaging modality e.g. X-ray, *Computer Tomography* (CT), *Magnetic Resonance Imaging* (MRI), and *Positron Emission Tomography* (PET). With the advancement in the technology, as the size and dimensionality of the imaging voxels increases, the visual identification becomes quite complicated. With the use of latest research techniques in machine learning, the analysis of these high dimensional imaging modalities becomes easier. Radiologists use the machine learning algorithms to make the decisions for visualization of images with different orientations, scales, and segmentation of tumors regions.

The main challenge for the medical science is to find the best appropriate machine learning algorithm for radiological application like tumor segmentation. The advancement in the research of machine learning formed the various algorithms which exist to solve a particular problem of real life. To select an algorithm that solves the problem of tumor segmentation from MRI with high accuracy is the concern of medical science. Further, medical practitioners are continuously seeking the computer assisted machine learning algorithms that classify the brain tumor types. These algorithms may provides the idea to radiologist about the type of brain tumor present in MRI without intervention of the clinical tests at initial stage.

Motivated by the prescribed research scope of machine learning in radiology and various statistics related to the formation of tumors (in respective ages of humans and their survival rate), this thesis aims in developing new machine learning algorithm to solve the problem. The thesis aims in segmentation of the regions from brain MRI having tumor cells. Further, the thesis intends to provide the best possible combination of the feature extraction, feature selection, and classifier to predict the higher classification accuracy for distinguishing among the malignant brain tumors.

1.1 Motivation

A brain tumor is most deadly disease in medical science, which is categorized into benign and malignant brain tumors. In the report of *Central Brain Tumor Registry of the United States* (CBTRUS), brain tumors are the (i) Prominent source of cancer-related deaths in children under age 20. (ii) The second foremost cause of cancer-related deaths in males ages 20-39 and (iii) Fifth leading cause of cancer allied deaths in females ages 20-39. Brain tumor occurrence rates in the *United States* (US) provided by CBTRUS utilize the US standard population and are reported per 100,000 population from 2008-2012. Certain facts and statistics provided by CBTRUS related for brain tumor are:

1. The incidence rate of all primary malignant and non-malignant brain and *Central Nervous System* (CNS) tumors is 21.97 cases per 100,000 for a total count of 356,858 incident tumors; (7.23 per 100,000 for malignant tumors for a total count of 117,023 incident tumors and 14.75 per 100,000 for non-malignant tumors for a total count of 239,835 incident tumors).

2. An estimated 77,670 new cases of primary malignant and non-malignant brain and CNS tumors are expected to be diagnosed in the United States in 2016. This includes an estimated 24,790 primary malignant and 52,880 non-malignant that are expected to be diagnosed in the US in 2016.

3. The incidence rate of childhood primary malignant and non-malignant brain and CNS tumors in the US is 5.37 cases per 100,000 for a total count of 16,366 incident tumors.

4. The incidence rate of childhood and adolescent primary malignant and non-malignant brain and CNS tumors in the US is 5.57 per 100,000 for a total count of 23,113 incident tumors.

5. The average annual mortality rate in the US between 2008 and 2012 is 4.31 per 100,000 with 71,831 deaths attributed to primary malignant brain and CNS tumors. An estimated 16,616 deaths will be attributed to primary malignant brain and CNS tumors in the US in 2016.

6. From birth, a person in the US has a 0.62% chance of ever being diagnosed with a primary malignant brain/CNS tumor and a 0.46% chance of dying from the primary malignant brain/CNS tumor.

7. For males in the US, the risk of developing a primary malignant brain/CNS tumor is 0.69%, and the risk of dying from a primary malignant brain/CNS tumor is 0.51% while for females in the US, the risk of developing a primary malignant brain/CNS tumor is 0.55%, and the risk of dying from a primary malignant brain/CNS tumor is 0.41%.

8. The fiveyear relative survival rate in the US following diagnosis of a primary malignant brain and CNS tumor is 34.4% (31.7% for males and 34.4% for females) (1995-2012 data).

9. Five-year relative survival rates following diagnosis of a primary malignant brain and CNS tumor by age of diagnosis (1995-2012 data):

    a. Age 0-19 years: 73.6%

    b. Age 20-44 years: 59.0%

    c. Age 45-54 years: 32.1%

    d. Age 55-64 years: 17.9%

    e. Age 65-74 years: 10.8%

f. Age 75 or older: 6.1%

In India, the incidence (newly diagnosed cases of cancer in a year) of brain tumors in India is about 2 patients per 1,00,000 population, while the morality rate (deaths due to brain cancer) is a little less than 2 patients per 1,00,000 population. In the year 2006 at TATA Memorial Hospital in Mumbai, India 372 people are diagnosed with brain and CNS Tumors, out of which 250 (67%) are males and 122 (33%) are females. The death rate is at higher side among which most of the cases are of malignant brain tumors.

The primary reason, which is diagnosed for higher side of mortality rate, is malignant tumor, which is unpredictable at the first site. The proper identification of malignant brain tumor is the need of radiologist for better treatment of patients suffering from this deadly disease.

1.2 Research Objectives

This thesis is carried out in the exploration of new algorithms to improve the classification accuracy among various grades of malignant brain tumor. The broad research objectives of the thesis are:

1. To find the *Region Of Interest* (ROI) having abnormal tissue cells in MRI.
2. To find the best possible combination of the feature extraction, feature selection, and classifier to get the precise class of malignant brain tumors.
3. To find an optimized solution to improve the accuracy of the selected classifier.

1.3 Research Contributions

By focusing on the research objectives of the thesis, certain contributions are presented to achieves the required goal of the thesis. These contributions are summarized as follows.

1. Two algorithms are proposed to segment the tumor region from the T1-weighted post contrast axial brain MRI. One of the proposed algorithm is the hybrid algorithm, that segments the MR image into two clusters, and second one is the

hierarchical algorithm that segments the image into multiple clusters. Given algorithms are quantitatively and qualitatively better as compared with state-of-art algorithms. [**Appendix I: Publication C2, C4, C5]**

2. Two new feature extraction algorithms are proposed named *Counting Label Occurrence Matrix* (CLOM) and *Texture Occurrence Matrix* (TOM) to classify the segmented tumor regions. CLOM and TOM are used to extract the spatial and spectral domain textural features. The experimentation results proved that the new algorithms outperform when compared with state-of-art algorithms like GLCM and Run length matrix. [**Appendix I: Publication J1, C1, C3]**

3. To enhance the classification accuracy, a new feature selection algorithm is proposed named *Cumulative Variance Method* (CVM) that is based on the statistical computations of variance between features. Additionally, a hybrid algorithm based on univariate and multivariate feature selection algorithm is also presented. The proposed algorithm enhances the performance of the machine learning in terms of tumor classification accuracy. [**Appendix I: Publication J3]**

4. A new mathematical algorithm is proposed, namely *A Voting based Novel Mathematical* (AVNM) algorithm that finds the number of nearest neighbors in KNN. The selection of the nearest neighbors is based on the iterative reduction of the space having data samples. The proposed algorithm avoids the dependency of the initial selection of the number of nearest neighbors. [**Appendix I: Publication J2**]

All the experiments are performed using the dataset consists of six hundred sixty (660) malignant brain tumor MRI's. All the images are of T1-weighted post-contrast axial modality taken from 3.0T GE MR Scanner. The dataset includes five classes of malignant brain tumor i.e. Central Neuro Cytoma, Glioblastoma Multiforme, Gliomas, Intra Ventricular Malignant Mass and Metastasis. The dataset is collected from Department of Radiology, *Sawai Man Singh* (SMS) Medical College Jaipur, Rajasthan, INDIA.

1.4 Organization of the thesis

In chapter 1 the motivation of the thesis and the key challenges of the brain tumor diagnosis is presented. Further, in the chapter, the broad objective of the thesis is

given. Later in the chapter, the main contributions in the thesis is summarized.

Chapter 2 presents a literature, related to the introduction of the brain tumors and its diagnosis in MRI. Further, the important part of machine learning in radiology is explained, with more accent given to segmentation and classification algorithms, as the thesis centers around the computer-assisted diagnosis of malignant brain tumors. More precisely, at the end of the chapter a literature review of the brain tumors in MRI is presented.

In chapter 3, proposed machine learning methods for tumor segmentation is given, that aim towards extracting ROI by examining T1-weighted post contrast axial brain MR images. Additionally, the chapter presents the qualitative and quantitative analysis of the given algorithm for segmentation of tumor region in MR images.

In chapter 4, classification of the malignant brain tumor MRI using textural features is presented. The classification is performed with KNN and Support Vector Machine using Radial basis Kernel (SVM-RBK) classifiers with hybrid collection of spatial and spectral textural features. The chapter demonstrates the two new proposed feature extraction algorithms, *Counting Label Occurrence Matrix* (CLOM) and *Texture Occurrence Matrix* (TOM) for extracting features from input MRI. Further, to enhance the classification accuracy, the selection of relevant features are determined using proposed *Cumulative Variance Method* (CVM) algorithm and Hybrid algorithm consisting multivariate and Univariate dimension reduction algorithms.

Chapter 5 demonstrates the classification of the malignant brain tumors with variant of nearest neighbor algorithm, named *A Voting based Novel mathematical* (AVNM) algorithm. AVNM is based on the reduction of the size of sample space. In addition, the chapter present the details of the given rule that automates the selection of the number of nearest neighbor in classification models. Finally, at the end of chapter, the comparative analysis of AVNM with state-of-art algorithms and its validation is presented.

At the end, chapter 6 encapsulates important findings with conclusions and future perspectives of the thesis. The list of publications as a result of research work accomplished for the thesis is listed in Appendix 1.

1.4 Organization of the thesis

The overall thesis organization diagrammatically is shown in the following Figure 1.1.

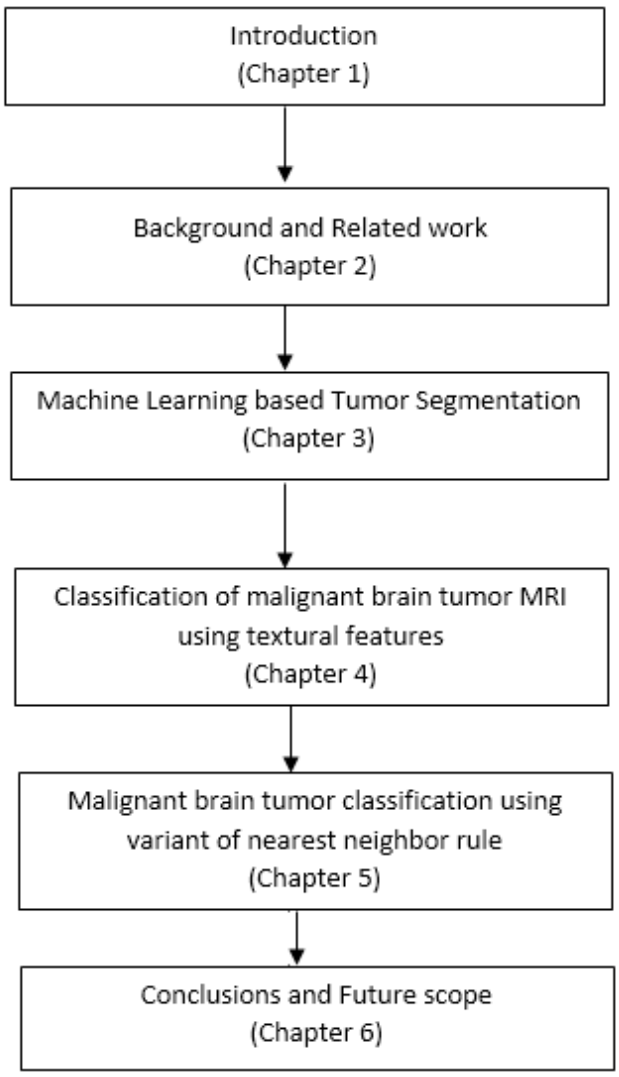


Figure 1.1: Thesis Organization Chart

Chapter 2

Background and Related work

This chapter presents the basic background knowledge of the general concepts about the human brain and brain tumors. Later in the chapter, the related work is presented for the segmentation and classification of brain tumors using machine learning approach. The related work in presented in the order of the phases used in machine learning, i.e., *segmentation, feature extraction, feature selection and classification*. Rest of the chapter is organized as follows: In section 2.1, an introduction of the basics about human brain and brain tumors (*symptoms, causes, types*) is presented. In section 2.2, a interconnection of the machine learning and radiology is presented followed by the literature review of the machine learning in radiology in section 2.3. Finally in section 2.4, an overall summary of the chapter is presented.

2.1 Introduction to brain tumor

Brain is one of the complex organ of the human body. It consists of variety of divisional sections which were termed and recognized by different name in medical science. This section demonstrates the basics of human brain before getting into the introduction of brain tumors.

2.1.1 Basics about human brain

A brain is the most sensitive part of human body which controls the movement of body and sense organs. The human brain is alienated into six lobes [1] i.e. Frontal lobe, Temporal lobe, Parietal lobe, Occipital lobe, Cerebellum and Brain stem as shown in Figure 2.1 [1].

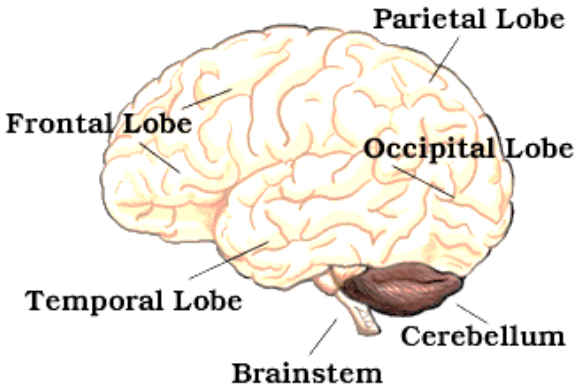


Figure 2.1: Brain Lobes description of human brain

A. *Frontal Lobe* is the largest lobe of the brain, located behind forehead i.e. inside front of the skull and near rough bony ridges. This lobe participates in planning, organizing, problem-solving, memory, impulse control, decision making, controlling behavior and emotions.

B. *Parietal Lobe* is located behind the frontal lobe. It integrates sensory information from various parts of the body and contains the primary sensory cortex, which controls sensation.

C. *Occipital Lobe* is located at the lower back of the head. It helps to receive and process visual information and contains the area that helps in perceiving shapes and color.

D. *Cerebellum Lobe* is the smallest lobe among all lobes, located at the back of the brain. It controls balance, movement, and coordination. It also allows us to stand upright, keep our balance, and move around.

E. *Brain stem Lobe* is present at the base of the brain. It includes the midbrain, the Pons, the medulla. It regulates basic involuntary function necessary for survival such as breathing, heart rate, blood pressure, swallowing.

F. *Temporal lobe* is located on the side of the brain under the parietal lobes and behind the frontal lobes at about the levels of the ear. This lobe is responsible for recognizing and processing sound, understand and producing speech, various aspects of memory etc. [1]

2.1.2 Basics about brain tumor

The main fundamental building block of human body constitutes the collection of nerves and cells. Inside the human body, the cells grow and destroy with time. But

if the cells have a strange growth inside the human body, this results in the formation of a solid mass of lesion called *tumor* [2]. The formed lesion inside the brain called *brain tumor*, is categorized into *benign tumor, non-cancerous tumor, and malignant tumor, cancerous tumor*.

Benign tumors are non-cancerous tumors as it resides inside the human body and mostly do not invade neighbor cells. Although benign brain tumors can form in any of the lobes of the brain and press healthy cells which potentially affects the functioning of sense organs, even then the person can survive for a long duration. While *malignant tumors* invade nearby cells as it grows very rapidly. Due to speedy growth of the tumor cells, the human sense organs get affected and patient cant survive for a long duration. Within a short span of time malignant tumor grows in size and circumstance becomes uncontrollable.

By generating the position of tumors, these brain tumors are categorized into *primary brain tumors* and *secondary brain tumors*. *Primary brain tumors* are the ones which evolve inside the brain itself, covering brain membranes (meninges cells), cranial nerves and pituitary gland. In general, *primary brain tumors* found less common than that of secondary brain tumors, in which cancer begins elsewhere and spreads to the brain. Several examples of primary brain tumor includes, on basis of associated cell name, are:

1. **Gliomas -** These type of tumors formulate in the brain or spinal cord with glial cells and include tumors like astrocytomas, ependymoma, glioblastomas, oligoastrocytomas and oligodendrogliomas.

2. **Meningiomas -** Such tumor arises from the membranes that surround the brain and spinal cord (meninges).

3. **Acoustic neuromas (schwannomas)-** These type of tumors develop on the nerves that control balance and hearing leading from your inner ear to brain.

4. **Pituitary adenomas -** These tumors develop in the pituitary gland at the base of the brain. These tumors can affect the pituitary hormones whose effects is seen throughout the human body.

5. **Medulloblastomas -** A cancerous brain tumor which is mostly found in children. A medulloblastoma starts in the lower back part of the brain and tends to spread through the spinal fluid. These tumors are less common in adults, but they are chances of having such tumor in adult as well.

6. **PNETs -** *Primitive Neuroectodermal Tumors* (PNETs), cancerous tumors that start in embryonic (fetal) cells in the brain. They can occur anywhere in the brain.

Secondary brain tumors, also known as *metastatic tumors*, originate somewhere else in the body and move towards the brain. These tumors are more common than the primary brain tumors and most often occur in people who have a history of cancer. Several cancers which are found responsible for generating brain tumors are breast cancer, colon cancer, kidney cancer, lung cancer, and melanoma.

2.1.3 Causes of brain tumor

To find the primary cause of commencing brain tumor is an area of research for medical society. The literature of past decade suggests that even after contributing lot of efforts in research of finding tumor cause, no such evidence is presented which evaluated the cause of brain tumor. Some of the parameters which are identified for tumor formation are chemical agent based generation of the tumor, ionizing radiation based tumor generation [3] and genetic inheritance based tumor formation.

The main chemical agent that led to the formation of brain tumor is exposure of the human body to vinyl chloride. Other than this no other known chemical or environment agent identified which cause the base formation of the brain tumor. Some literature suggests that there is a concern related to electromagnetic fields that cause the formation of tumors through glial cells [4]. The sources of such glial tumors are considered as the use of mobile phone but no such evidence exists which support this theory. But people who have been exposed to ionizing radiation have an augmented risk of brain tumor. Examples of ionizing radiation include radiation therapy which is used to treat cancer in past and radiation exposure caused by atomic bombs.

Another cause of brain tumors is related to known genetic conditions. People who have experienced one of these rare syndromes have an increased risk of getting a brain tumor. Since the cells building in human body starts before the birth, using DNA of parents, thus the risk of tumor cells get inherited. These syndromes cause a number of different medical problems [5].

If a person has a parent, brother or sister diagnosed with a brain tumor, then the risk of getting brain tumor to a person become higher than other people in the

general inhabitants.

2.1.4 Symptoms of brain tumor

A person suffering from the brain tumor may experience with several symptoms which illustrate the presence of any kind of brain tumor. But, sometimes, a person having brain tumor dont show such symptoms based on his/her medical conditions. These symptoms can be categorized into general symptoms and specific symptoms. General symptoms of the brain tumor are based on the pressure created by the tumor part of the brain and spinal cord while specific brain tumor symptoms are caused by the formation of the tumor at some specific position inside brain due to which some specific part of the brain stops working [6].

The general symptoms of the brain tumor include a headache that gradually become more frequent and more severe at the time of early morning. Seizure, or convulsions, are another symptom of brain tumor that includes sudden involuntary movements of a persons muscles. Due to seizure, persons sensory movement get affected and includes a change in sensations, vision complications like blurred vision, double vision or loss of peripheral vision, smell, difficulties with body balance and hearing without losing consciousness [7]. In addition of these, sometimes nausea and vomiting are also seen with brain tumor patients.

2.1.5 Types of brain tumor

An abnormal mass of lesion inside a brain called *tumor* is classified into two categories, *primary brain tumors* that originate inside the brain and *secondary brain tumors*, or metastatic, that voyage from their originate location and transferred into the brain.

Primary brain tumors are formed either by neurons, called *brain cells* or by supportive cells called *neuroepithelial cells*. Based on characteristics brain tumors are characterized by *benign tumors* which are non-cancerous in nature and *malignant tumors* which are cancerous [8].

The main characteristics associated with benign tumors is the association of slow growing cells within the specified boundaries. Under the microscopic examination of benign tumors, cells look a lot like normal cells which make the diagnosis of tumor a complicated task. It is noted that benign tumors found approximately 40% of all existing primary brain tumors. On the other hand, malignant tumors are

associated with the uncontrolled growth of the cells. Malignant tumors causes severe complications because of their violent and aggressive nature, results in forming pressure on cells and provoke life-threatening condition.

In this thesis, the primary and secondary category of malignant brain tumors is employed. The primary tumors that are taken into consideration for the experimentation purpose are *Gliomas, Glioblastoma multiforme, Central Neuro Cytoma, Intra Ventricular Malignant Mass* while the secondary tumor that is taken for experimentation is *Metastatic*.

2.1.6 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is considered as the most common application of *Nuclear Magnetic Resonance* (NMR) [9]. Despite other existing imaging modalities, like *Computer Tomography* (CT) and *Positron Emission Tomography* (PET), which uses harmful radiations i.e. X-rays to capture internal structure of human body, MRI gains the advantage above CT and PET. *Magnetic Resonance* (MR) images are developed entirely non-invasively without triggering any risk of harmful radiations to the patients. Additionally, it provides the good contrast to soft tissues of the human body which makes it useful for studying central nervous system and cancers.

The basic principle of MRI for in vivo is based on the water present in human body. The resonance signal obtained by the water molecule or proton (1H) are recorded by the scanner. Besides this, gradients correspond to the magnetic field need to be applied to external magnetic field B_0. If the MRI generation process would apply without gradient then the scanner acquire the signal having a response of all spins with no spatial distribution information.

2.1.6.1 Brain Tumor Diagnosis using Magnetic Resonance Images

Diagnosis of the brain tumor is considered as the challenge in medical science because of higher mortality rate. The prime organization of Medical community, *World Health Organization* (WHO), discriminates among different types of brain tumors, their subtypes and grades by the analysis of the malignancy [10]. On the basis of clinical tested and verified reports WHO classified 120 types of brain tumors till date. In addition, the unordered characteristics of tumor like heterogeneity

at the level of tumor type and grade pretense difficulties in the diagnosis and treatment of brain tumors.

In the detection of the brain tumor and its treatment brain imaging play an important role [9]. In pre-diagnosis of the brain tumor various imaging modalities are used like *Computer Tomography* (CT) Imaging, *Magnetic Resonance* (MR) Imaging, *Functional MR Imaging* (FMRI) and *Positron Emission Tomography* (PET). Among the various imaging modalities, NMR is the most widely used technique for diagnosis of brain tumors. MRI provides 3D voxel of internal structures of the brain. At each pixel location, MRI computes following attributes of the soft tissues being scanned i.e. *Proton Density* (PD), spin to lattice relaxation time (T1), spin to spin relaxation time (T2) and spin to lattice with post-contrast (T1C). These MRI techniques are consistently analyzed by the radiologist to assess the location of the tumor inside the brain or other portion of the human body but they are not able to find the growth of tumor cells and to characterize tumor type or grade on WHO malignancy scale.

Some of the other factors on which brain scan depends are pulse sequence, resolution slice thickness, the distance between inter-slices and signal noise ratio. The gained intensity for the pixel on an MR imaging scan depends mainly on the content of particular pixel to its neighbor tissues and presence of fat and cerebrospinal fluid. A T1-weighted MRI brain scans generate images with gray regions having bone and water while the regions having fat are shown by white regions. On the other hand, T1-weighted post-contrast MR scans show the tumor with white regions due to the contrast agent which mix with the tumor affected blood cells and provides enhancement to blood cells. In T2 weighted images, the generated scan are reverse of that of T1 weighted scans, where gray regions are changed to white and vice versa. It occurs when in T2 weighted scans water and fluid have given high signal intensities while bone and fat has low intensities. This increases the feasibility to detect tumor and edema together in generated T2 scans [11, 12].

Brain imaging scans allow radiologists to track tumor location and its growth. These scans are noticed by most innovative imaging tools that allow the radiologist to inspect the patient brain voxels to detect abnormality regions. The growth of the tumor in a particular scan is quite difficult to predict in a single scan. Patients need to follow up through a sequence of scans taken over a specified duration of weeks or months. These sequential scans help the radiologists to measure the growth and spread of the tumor.

To analyze the brain tumor the MR scanner provides the orthogonal structure of the head which is divided into three section i.e. axial, coronal and sagittal. Each of the section has a different orientation of the scanner plane which is shown in Figure 2.2 [13].

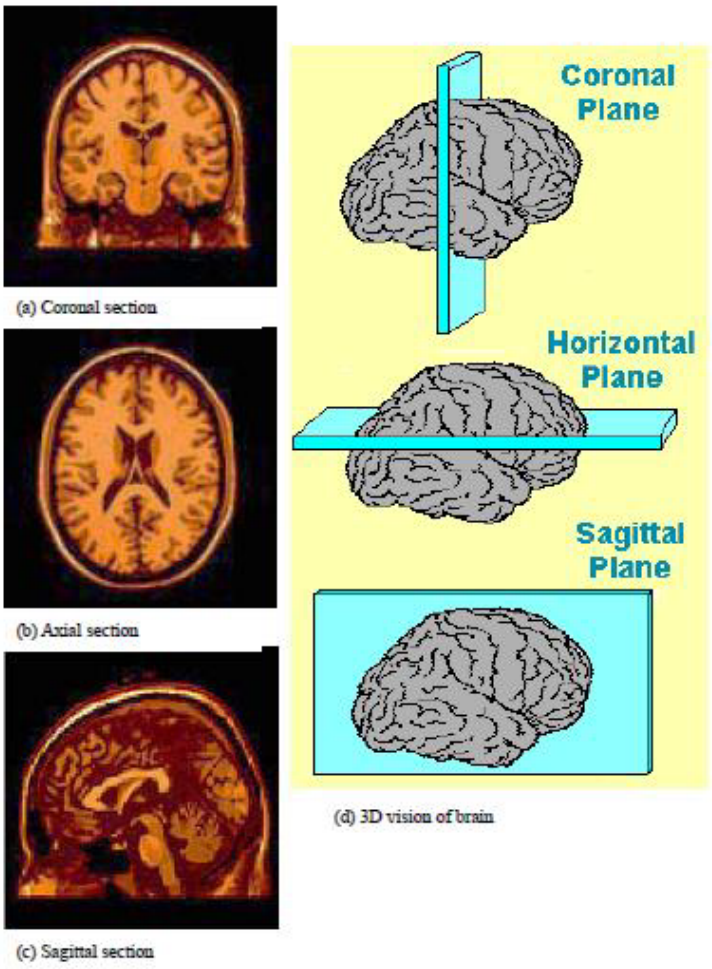


Figure 2.2: Orthogonal planes used for MR image acquisition

2.2 Machine learning and Radiology

In the past, *machine learning* algorithms have been used in the variety of domain applications for performing tedious tasks. These algorithms are not limited to speech and speaker recognition, artificial intelligence, weather forecasting, biometric security analysis, planning, prediction, and disease diagnosis. The *machine learning* is classified into two broad domains i.e. *supervised machine learning and unsupervised machine learning*.

In *supervised machine learning* problems, the initial step is to get the model learning using existing data. Inside the digital machine, each data sample is stored in the form of intensity and pixel information. To represent the data object machine often uses some feature values associated with the data object to get it learn and recognize. A machine then uses the learned model to predict the class for an unidentified instance by classification. The result of the classification is mainly associated with the assessment value or a probabilistic function. The overall classification accuracy is evaluated by the gained confusion matrix parameters i.e. True Positive, False Positive, True Negative and False Negative [14].

In the *unsupervised machine learning problems*, where the class of data instance is unidentified, machine samples the data instance into the number of groups called *clusters*. Each *cluster* represents the group of items having almost similar characteristics or property. The parameter associated with cluster generation is metric, and most appropriate used metric is based on distance computation. The formed clusters from the data instance have the property of minimum intra-cluster distance (maximum similarity) between cluster elements and maximum inter-cluster distance (maximum dissimilarity) between any two clusters [15].

In Medical Science perspective and especially in radiology, machine learning contributes a lot. Machine learning isolates intricate patterns automatically that are based on the learning parameters. It helps the radiologists to make intellectual conclusions on radiological data such as CT, MRI, and PET images. In medical applications, the performance of machine learning-based analysis, prediction and diagnosis systems has shown to be equivalent to that of an experienced radiologist. Some of the challenging application areas of machine learning in radiology are segmentation of medical images, medical image registration, computer-aided diagnosis and detection of specific disease, a study on the functionality of brain, and research study on diagnosis of the neurological disease like Parkinsons disease [16].

Radiology uses various radiations and imaging modalities to detect and diagnose diseases. It uses the advanced research concepts of nuclear physics, electronic engineering, and computer science. Based on the research advances of these fields, imaging technology improved a lot i.e. from 1Tesla MR imaging from past to 12T MR imaging and beyond [17]. This advancement in technology rapidly increases the imaging data, not only in visual clarity and image size but also in the generation of the 3D voxel image slices. In the daily routine of the radiologist, various medical images from different modalities are interpreted by experienced radiologists for

diagnosis of the disease. These imaging slices are interpreted by radiologist usually in a short span of time but with the increase in the imaging slices, this task becomes quite a time consuming that can be resolved with the help of machine learning.

Machine learning provides an efficient approach based on the algorithms to automate the breakdown for imaging modalities. Using such, it somewhere reduces the burden and provides liability to the radiologist in their practice. In the brain tumor analysis, the radiologist endlessly looking for such tools and methodologies that provide a high rate of accuracies in the accurate diagnosis of brain tumor, categorization of normal and tumor affected cells and deliver the most effective treatment to patients. To obtain the desired result, several research studies have employed in machine learning to discriminate between types of human brain tumors. The whole studies in machine learning consist following stages: i) *Segmentation or Region Of Interest* (ROI) extraction, ii) *Feature extraction*, iii) *Feature selection* and (iv) *Classification*.

2.2.1 Segmentation or Region Of Interest extraction

Segmentation is the process in which data instance is divided into several parts such that each separated part initiates the common property. In medical image analysis based on segmentation, the images consist of several structures which are classified as normal and abnormal structures. The key of the segmentation in the medical image is to properly extract the region which contains abnormality section, called *Region Of Interest* (ROI). Sometimes the extracted ROI are also known as informative regions and left over part are considered as background. In medical images by using segmentation various organs of human body, bones and muscles are extracted separately from imaging modalities. Besides this segmentation is also used to find the fractures in bones of the human body and tumor inside imaging modality which differentiates the normal cells and abnormal cells. These segmented regions are also called clusters. In general, the cluster is defined as a group of items having similar characteristics.

In brain tumor detection or machine learning based tumor diagnosis, ROI act as an important and necessary step for processing. When the segmentation is performed, among the segmented regions, some of the most informative region is selected as ROI. These ROIs are considered as meaningful sections and further analysis is required to predict the disease. The extracted ROI consists of abnormal cells

information, which is processed in terms of features extraction. Features are used by the machine to learn the pattern and predict the final classification result.

2.2.2 Feature Extraction

Features are the most important part in machine learning based analysis. Each of the data instance in machine learning is associated with several number of features which are used to represent, recognize, search and categorize the data instance. In practice, it is noticed that there is no limit on the number of features associated with data. Features are termed as a computational assessment using which one data instance is differentiated from others. There are various approaches that help to find the specific features of the data instance. These approaches are probability-based approach, statistics-based approach, visual parameter approach, object description and physical characteristic based approach.

In machine learning, there are two domains in which any of the data instances is stored, i.e., spatial domain and spectral domain. Spatial domain relates to the pixel intensity value, using which varieties of features are computed. These features use the stored intensity value to obtain the features based on the above discussed approaches. Each of the feature extraction approach uses the algorithmic background to custom the original pixel or group of the pixel values to manipulate, process and obtain some new values or characteristics that represent the object inside machine learning.

On the other hand, temporal and spectral domain feature extraction approach uses the time-frequency or time-scaling sampling to store and retrieve data instances. In the spectral domain, the data sample is represented in the scale of time and frequency which is processed using various transformation functions. The transformation function like *Fourier transformation, Cosine transformation, Wavelet transformation and Curvelet transformations*, transform the input data signal into frequency scale with respect to time so that the data will be analyzed in small chunks which provide desirable information.

In machine learning based brain tumor analysis, features are computed from the segmented ROIs from MR images. The features are extracted based on spatial, temporal and spectral features extraction algorithms. The collections of the features from the variety of domains are called *feature set pool* which are used in learning,

recognition, and characterization of the brain tumors.

2.2.3 Feature Selection

Feature selection is an important part of machine learning which is applied after feature extraction. The basic difference between the feature extraction and selection is defined as:
Feature extraction: methods are found transformative. It indicates that after applying the transformation function ϕ to the data object, the data is projected to a new space indicated by the following equation:

$$\phi : \begin{Bmatrix} x_1 \\ x_2 \\ ... \\ ... \\ x_n \end{Bmatrix} \rightarrow \begin{Bmatrix} x_{i_1} \\ x_{i_2} \\ ... \\ ... \\ x_{i_n} \end{Bmatrix} \tag{2.1}$$

where ϕ is the transformation function, set $\{x_1, x_2, ..., x_n\}$ is the set of data objects and set $\{x_{i_1}, x_{i_2}, ..., x_{i_n}\}$ is the transformed set representing extracted features.

Feature selection: method selects the set of features from the original set based on some criteria function ζ, that filter out unwanted or redundant feature and left with relevant set of features. It transform the high dimensional data to lower dimensional and is given by following equation:

$$\zeta : \begin{Bmatrix} x_{i_1} \\ x_{i_2} \\ ... \\ ... \\ x_{i_n} \end{Bmatrix} \rightarrow \begin{Bmatrix} x_{i_a} \\ x_{i_b} \\ ... \\ ... \\ x_{i_k} \end{Bmatrix} \tag{2.2}$$

where ζ is the filter function, set $\{x_{i_1}, x_{i_2}, ..., x_{i_n}\}$ is the transformed set representing original extracted features and set $\{x_{i_a}, x_{i_b}, ..., x_{i_k}\}$ is the reduced set of original feature set.

The objective of feature selection in this thesis is to filter out the most appropriate features among the extracted features in the feature set pool. The reason of filtering is to find and truncate the features which represent either the similar type of information or an information that dont provide sufficient information about data

instance. Based on such, the features are grouped as relevant or irrelevant features and redundant features.

In former section, the various approaches for feature extraction are discussed through which feature set pool is generated. Among the extracted feature set, certain features are termed as redundant if they provide the similar information as given by any of the previous extracted feature in feature set pool. Such features only increase the dimensionality of the space in machine learning and thus these features must be eliminated before proceeding classification. On the other hand, certain features are termed as irrelevant if the features contributions in the feature set do not play any role. For example, the classification performance of the machine learning remains similar after including and excluding these features, then such features are termed as irrelevant to the classification system.

To select the relevant and non-redundant features, machine learning provides three categorical approaches for feature selection: i) *Filter approach*, ii) *Wrapper approach*, and iii) *Embedded approach*. All the feature selection approaches have the slight variation in the working perspective like *filter algorithms* are free from the use of classifier in decision making. This is also categorized into *univariate filter approach*- single feature is used at a time to make the decision e.g. *chi-square test, t-test, information gain, gain ratio* and *multivariate filter approach*- multiple features are used together to make decision e.g. *correlation based feature selection, Markov blanket filter, random forest*. The second approach is *wrapper approach* where a feature selection algorithm uses the classifier output to make the decision. *Wrapper approach* are categorized into *Randomized wrapper approach*- that uses the random feature subsets e.g. *Simulated Annealing, Genetic algorithm, Random hill climb algorithm* and *Deterministic wrapper approach* that uses the greedy search for finding feature dependencies e.g.*sequential forward selection, sequential backward elimination, plus q take-away r algorithm*.

The third approach i.e. *embedded approach* is a special method where features are selected with the help of inbuilt feature selection functionality of the classifier. Here the classifier itself predicts the best features to gain classification accuracy e.g. *decision tree, weighted Support Vector Machine (SVM), nave bayes algorithm*.

2.2.4 Classification

The last state of the machine learning is *classification* or *recognition*. Both the methodologies uses the above-defined phases sequentially but have the difference in final output type. Classification uses the pre-stored knowledge base for learning and predicts one class label, among the k classes that exist in the database. While *recognition* uses the pre-stored knowledge base for learning and predicts one sample among samples that exists in the database during testing. The difference between the two is former is used for an identification of totally unknown sample while later is used for matching the pre-stored data samples.

In machine learning, the goal of classification is to predict one among the existing class labels for an unidentified sample. This is achieved when the sample to be classified (Test sample) is matched with the existing class samples in the pre-stored database. To match the similarity between the samples, distance function is used as a metric for computation and decision making. The various distance functions which are used as a metric for computation are *euclidean distance, city block distance, minkowski distance and mahalanobis distance*. To build the classification system, accuracy is the key which plays an important role. To test the accuracy of the system, the dataset is divided into training set and testing set. The training set is used to learn the machine learning and recognize the data patterns while the testing set is used to test the machine learning.

An important aspect in the division of the dataset is to overcome the issues of *over-fitting* and *under-fitting*. In machine learning, there is a need to fit a model to a given training data to make reliable predictions on unseen data. In *overfitting* a model describes a random error or noise present in the data instead of underlying relationship which exist in the data. This occurs when the model involves to many parameters in comparison to the number of training samples.

While *underfitting* occurs when a model is not able to capture the underlying trend of the data. It happens when the model does not fit the data correctly. Generally it occurs when we consider a simple model to fit the data. Thus to make the model effective, the partition of the dataset is in such a form that makes the learning and testing effective for the model. Several methods have been discussed in past to partition the dataset, but among all, the most effective method found is 80-20 rule. It helps to partition the whole dataset into two groups, such that one group consists of 80% of the data samples, which are used to learn the model (training set) and

another group contain 20% of the samples, which are used to test the model (testing set). By the computed results for testing set, the accuracy of the model is predicted.

2.2.5 Contribution of machine learning in radiology

Machine learning has the variety of application in the radiology. These applications differ from each other and may have different forms regarding the problems to be get solved. They vary in the form of input and output data, functional constraints, prior knowledge, and stored hidden variables. Machine learning algorithms exploited to solve many such problems which seem complicated. Initially, the researchers of radiology found themselves in the confusing state to decide whether to employ machine learning algorithms or not and if yes, then which one is an appropriate algorithm for radiological problems.

Machine learning studies the roadmap of algorithms which make digital systems to recognize composite patterns and make decisions by experimental data. The most substantial influence of machine learning in radiology is that it provides a way to simplify human knowledge obtained from training dataset to predict the test data which is still unidentified. For example, calcification of the human breast in mammography imaging system. Another application of radiology i.e. medical image segmentation uses the original CT, MRI, PET and ultrasound images as an input and segment the volume or region of interest in the input image. The segmented region can include human organs, bones, blood vessels etc. To segment the region, the main requirement is the prior knowledge about the region which is to be segmented e.g. shape and texture. Beside it, another approach would be the extraction of the relevant features, and learn the model that use the feature and find the solution for object description to segment the image.

In radiology, another machine learning technique that is most widely used is *classification*. Many clinical applications of radiology expressed in the form of classification problems. For example, smart classification of the lesions inside the imaging modalities of various parts of human body like lungs, breast, brain, skin, tongue and all such parts of body. Classification of the lesions in various parts of the human body becomes the fundamental task in computerized systems using a variety of machine learning classifiers like *linear classifiers, non-linear classifiers, ensemble classifier and Artificial Neural Networks (ANN)*.

Radiological text report processing is another application of machine learning in radiology. The daily radiology practice reports fill huge text databases. By incorporating machine learning based modern information processing techniques the search and retrieval of the text report from huge database help the radiologists in proper diagnosis of the disease. Even though machine learning has proven its importance in many radiological problems and applications, there still exists certain hurdles to the interpretation of machine learning techniques in clinical practices in radiology. The various challenges that have been faced in the past is:

1. The effect of the dataset size - in the literature, many studies of machine learning in radiology are proposed using small data sets. When machine learning methods integrate to solve radiology problems from small data sets to large data sets then such algorithms are not generalize well. This result in the poor performance on large datasets which earlier is performed well on smaller datasets [18].

2. The second issue relates about the complexities that arises from the perspective of machine learning and radiology. The complexity arises when certain machine learning algorithms or techniques are found quite complicated to incorporate in real time radiological applications. Such algorithms constraint severe assumptions on the problems to be get solved and thus found it hard to map directly in radiological applications [19]. Further, from radiological perspective, there exist certain applications of radiology that are found quite complicated and no known machine learning algorithm exist that could solve such problems. For example, interventional radiological application, where complete automatic procedure requires prior knowledge of human anatomy, real time trailing of needles through blood vessels and treatment of cancers are still too complex for existing machine learning environment. Also, there are certain real time variables that may change with time e.g. scanning rules or protocols. Thus machine learning algorithms or systems which are trained on old data sets may not adjust quite well with latest progressing situations [20].

3. The third concern relates to the human psychology. Even though machine learning proved its necessity and advantages in many applications areas, still several people think that in clinical diagnosis human interpretation and diagnosis are much better. They are uncertain about the accuracy of computerized systems and thus thought that human interpretation outperforms such digitized systems. This psychological myth makes it hard to get the full benefit of

the machine learning in medical science. So current medical scenario mainly treats machine learning or computerized systems as an aid to decision making but not as problem-solving.

2.3 Literature review of machine learning in radiology

As per the discussion in the previous section, the use of the machine learning in radiology for diagnosis purpose is not groundless. In this section, the literature is divided into the two aspects of machine learning. The first section incorporates the work done for the segmentation of the tumors inside brain MR images. The second section demonstrates the work done for classification of the tumors. Classification method internally uses three phases. The first internal phase is relate to the extraction of the features from the brain MR images followed by the feature selection methodologies in second internal phase. The last phase relates to the classifiers of machine learning that is used for classifying the unknown cases of brain tumors in radiology.

2.3.1 Tumor segmentation

Medical images have many modalities like X-ray, CT, MRI and PET which consists of many normal and abnormal sections. Segmentation is one such process using which such regions or sections are identified in medical images. The concept of segmentation is not trivial in medical science because of its complexity and variability of *Volume Of Interest* (VOI). The concept of segmentation in medical images indulge the grouping of the regions that have comparable appearances. But certain parameters that complicate the process includes variation in normal anatomic, incomplete & indefinite boundaries, inadequate contrast, and noise.

Segmentation of the brain tumors in medical images are found either by using *biclustering algorithms* or by using *multicluster algorithms*. *Biclustering algorithms* use the threshold parameter to divide the image into two segments like in the case of OTSU thresholding [21, 22] and morphological operation [23]. OTSU algorithm, uses the threshold value as the baseline by which the intensity values are filtered and the image is segmented into two halves. Each half is the resulting

segment of the algorithm. On the other hand, the morphological operation uses the region growing and region splitting to segment the image based on pixel intensity values. These methods are identified to be highly sensitive when there is an even minor change in color and illumination reported.

Another approach which is widely used for image segmentation for multiclustering is the *pixel oriented clustering* based approach. Although analysis of the clusters in machine learning is not mainly planned for medical images segmentation, but due to the highly overlapping objectives of segmentation and clustering algorithms, many of the algorithms can be applied to solve medical domain problems. k-means clustering algorithm is one such algorithm which is used in machine learning for last 50 years [24]. After proving successful contribution of the k-means algorithm for data clustering, the algorithm is integrated to detect brain tumor in MR images using color distribution [25]. The k-means clustering based segmentation is used to overcome the limitation of the bi-clustering and used to generate k clusters for an input image.

In [26], a new hybrid algorithm is proposed for medical imaging where tumors inside the human brain are segmented using the k-means with the threshold function to gain the clustering accuracy. However, in machine learning, k-means algorithm is considered as a non-deterministic algorithm and is found extremely sensitive to the initial choice of cluster centers. The pre-initialized number of required clusters with the selection of the centroid among the data samples is desirable before the start of the algorithms. To increase the accuracy of the clustering algorithm, a concept of fuzzy is introduced with k-means i.e., *Fuzzy C-Means* (FCM), to find the membership of those data samples which is found on the boundary of clusters. The idea of FCM clustering was developed by J.C. Dunn in 1973 [27] and later it is improved in 1981 by J.C. Bezdek [28]. The concept of FCM gained the advantage above k-means in terms of its cluster accuracy.

The concept of fuzzy is introduced in medical science for the segmentation of the brain tumor in MR imaging [29]. The fuzzy approach helped to gain the better result with respect to existing clustering approaches. The performance of the fuzzy approach is analyzed in [30, 31] for automatic segmentation of the tumor region in MR images. In radiological application fuzzy approach also play an important role in finding the area of the abnormality section segmented from the abnormal MR images [32]. On applying the FCM methodology in tumor segmentation, the initial step is to determine the group of tissue classes. In such, each of the pixels is

assigned a membership value by the fuzzy function which is found in constrained to be in the range of 0 to 1. This membership value reflects the similarity between the pixels to the tissue class centroid. The membership value 0 reflects no similarity, 1 reflects the complete similarity and rest intermediate values reflect the degree of similarity of the pixel to the cluster. The initialization of the cluster centers for FCM clustering is still an issue, it is reported that if the estimation is correct then the algorithm converges faster with improved clustering results. Over such issue, in [33] a splitting technique based on the *Discrete Curve Evolution* (DCE) is proposed to find the most appropriate and accurate cluster centers for T_1, T_2 *and Proton Density* (PD) brain MR images segmentation.

Several research based on the integration of the FCM with other machine learning and statistical approaches are reported in literature of the segmentation of brain tumors in MR images. Techniques like *SVM* [34], *gaussian spatial information* [35], and optimization based on *markov random field* with *ant colony optimization* [36], *genetic algorithm and Particle swarm optimization* (PSO) [37] are extended with FCM to gain better clustering results in brain tumor MR images. The limitations of the FCM related to noise due to non-consideration of the spatial information have been spotted in [36]. Several kinds of research are also associated with FCM to segment the medical image on the multi-sequence data. The primary fuzzy clustering proof on multi-sequence brain MR images is given in [38]. The author demonstrates visually that even for multi-sequence data the intensities distribution of the abnormal and normal tissues overlap a lot. This demonstration directed other researchers to unite various other additional knowledge into the extracted clusters to get exact information about the tumor affected tissues. Thus a knowledge base methodology with the multispectral histogram analysis with FCM is proposed by Clark *et al.* in [39]. With the help of extended knowledge with FCM, a malignant tumor named *Glioblastoma Multiforme* is extracted from the T_1, T_2, and PD multichannel brain MR images. Another knowledge base fuzzy approach is proposed in [40] for the segmentation of the non-enhancing brain tumors with 3D connected components for building the shape of the tumor in T_1, T_2, and PD MR images.

Many researchers of machine learning and radiology have measured that the standard FCM for segmentation of the tumors in MR images is not much efficient as it fails in dealing with the firm correlation of nearby pixels. This lead to the formation of the noise and also several other imaging artifacts. To overcome this problem several solutions are proposed in [41, 42]. Most of the proposed solutions include the consideration of the local spatial information. The concept is based

on the fact that besides gray level pixel information the information present by the neighboring pixels also subsidizes to the assignment of the pixel to the corresponding cluster [43]. The neighboring information about the pixels is analyzed using the degree of attraction of the pixel to the cluster. For such, there are certain methods which are used to gain the degree of attraction like genetic algorithm and optimization functions based on PSO and ACO [44]. Another issue that is reported with FCM is consumption of the time taken by the algorithm because of its iterative nature i.e. the algorithm stops only when it converges to an exact or nearby solution. Keeping the issue in mind to reduce the time taken by the algorithm, an enhanced FCM algorithm is proposed for segmentation of the images [45]. The enhanced approach is the update of the improved FCM algorithm which is proposed by [46] with an average pre-filter for segmentation of the tumors in MR images.

Besides the segmentation of the brain tumors in medical images through clustering, segmentation can be classified into four broader categories [47]:

(a) Manual, semi-automated, and fully automated brain tumor segmentation

(b) Supervised and unsupervised brain tumor segmentation

(c) Model-based brain tumor segmentation

(d) Hierarchical and graph-based brain tumor segmentation

Manual segmentation of the brain tumor includes the manually plotting of the boundary regions around the tumor and other ROIs. In such cases, the trained radiologist used the existing information which is present in the image with certain other additional information like anatomy about the image to draw the boundaries around tumor regions [48]. In semi-automatic segmentation method, the combined effort of human and machine learning algorithm is reported to segment the tumor from MR image. The algorithm helps to find and segment the desired region of interest. While the human operation is required for method initialization, accuracy analysis and to manually correct the algorithmic based segmentation result. Some of the main components of semi-automatic segmentation are reported in [49, 50].

In fully automatic segmentation method, there is no interaction of the human operation while the whole task of segmentation is performed by the machine learning algorithms. Such algorithm incorporates the human intelligence and the anatomy of the brain structures and tumors to segment the abnormality section from the input MR image. However no perfect fully automatic segmentation algorithm have

developed which is incorporated in the medical science. The most successfully implemented algorithms for radiological applications are based on *Discriminative random field* with the SVM [51], a framework for segmentation using fuzzy and optimized region growing approach [52–54],*Self-Organization Map* (SOM) [55], and knowledge-based systems incorporating ANNs [56].

To segment the 3D volumetric MR image, model-based segmentation approach is proposed in [57]. Model-based segmentation mechanisms incorporate initial familiarity about entities present in MR image of specific regions like shape, location, orientation and statistical information extract from training sample dataset. The challenging task in model-based segmentation is to find and extract the boundary regions that belongs to same structures. Incorporating all such parameters, two models of segmentation is proposed, i.e., parametric deformable model and geometric deformable model or level set model. The related work on such models is presented in Table 2.1.

The graphical concept also provides a sophisticated way to extract the information that is found useful in the segmentation of the images. Graph cuts, is one such method based on graphs and exploit flows on the graph [67]. Based on the above said concept, a graph partitioning based segmentation method is proposed named normalized cut based image segmentation [68]. A normalized cut partition consist small remote points and gives more stable partitions as compared with existing ordinary cut. The concept of graph cut has many applications in medical science for segmentation of the medical images. In 3D MR images, the segmentation of the organs using graph cuts is proposed in [69] and segmentation of the lesions using MRI in [70]. There are many segmentation algorithms exist in machine learning, even then no proper algorithm exist that completely provides the optimal solution for segmentation. This issue is discussed in the positional paper [71].

2.3.2 Tumor Classification

In machine learning, for the classification perspective, the system requires a finite set of features which represent the description of an object which was to be classified. The set of extracted features were undergone the feature selection process that filters the relevant set of features which were used for object description. This section provides the description of the *feature extraction algorithms, feature selection algorithms* followed by the state-of-art *classifiers* used in machine learning for

Table 2.1: Summary of related works in Model-based segmentation

Model	Author and Year	Description
Parametric Deformable Model	Chan *et al.* [58]	Search the boundary of the tumor in MR image using contour and snake function.
	Luo *et al.* [59, 60]	Two mechanisms are proposed with improved *snake function, ballon model, and gradient vector flow model.*
Geometric Deformable Models	Caselles *et al.* [61], Malladi *et al.* [62]	Proposed contour with image gradient force based stopping criteria.
	Osher [63]	The main component of the *geometric model* is the implicit representation of the interface.
	kichenassamy *et al.* [64], Yezzi *et al.* [65]	Proposed the diverse approach for segmentation of the regularly shaped medical images.
	Siddiqi *et al.* [66]	Proposed the modified approach of mechanism presented in [64, 65] by adding parameter based on weighted area function based gradient flow.

classification.

2.3.2.1 Feature Extraction

In machine learning, features are termed as the most important and informative analytical aspect which is used to translate the information present inside images. In radiological applications, features are extracted from the segmented volumes which

consist of the valuable information. These informations are encoded with the help of the features. By using machine learning, two categories of features have been incorporated to discriminate among the types of brain tumors with utmost accuracy, which proved their significance in tumor classification, i.e., i) Features from the spatial domain and ii) Features from spectral domain.

Features from spatial domain

The spatial domain in image analysis is considered as the normal image space in which the image is represented in the form of intensity values or pixels. These intensity values are used to extract the features which represent the comprehensive data information into the reduced form. The main advantage of using spatial features is reported in [72], which suggest that spatial features are always meaningful and easy to understand. These features can be extracted from any shape without losing object information. The various approaches which are present in the literature for extracting spatial domain features are by extracting *fractal dimension information* [73, 74], *local binary patterns* [75], *histogram analysis based features* [76], *texture features based on co-occurrence matrix and run-length matrix*.

Among the various feature extraction techniques, *texture-based features* gain the more advantage than other mechanisms in terms of classification accuracies. In brain tumor analysis, texture features are computed with the help of co-occurrence matrix [77]. *Gray Label Co-occurrence Matrix* (GLCM) proved as one of the most informative methods for texture information description. Haralick in [77], suggested several first order and second order two-dimensional features for image representation. There are fourteen statistical features extracted that are used to describe the texture properties from an image. The extracted features are found susceptible to noise and angular rotation in an image. It is proved that angular directions 45^0, 90^0, and 135^0 in medical imaging dataset isnt dominant directions for specific parts in an image [78]. Also, leading directions of different parts are usually different.

In the classification of the brain tumors in MR images, use of texture features is seen in several literatures. Use of GLCM features on the brain tumor dataset having normal and abnormal brain images are proposed in [79–82]. The approach is tested on the dataset of 42 to 100 MR images. An approximately fourteen to thirty-two features are computed based on different orientations of the MR image. Additionally incorporating GLCM features with other texture extraction algorithms like *histogram based features* [83], *shape and intensity features* [84, 85], *run-length*

features [86] is proposed on the different brain tumor datasets having two classes of images i.e. normal and abnormal brain images. The hybrid algorithm extracts the texture features from different regions with different approaches. This results in the addition of varieties of features in features set. These features are later used in classification for discriminant between normal and abnormal brain images.

Moments is another spatial texture feature approach which is proposed in [87], [88], for extracting first, second and third order *moments. Moments* are computed heuristically that describes the shape of an object or a distribution of the *Probability Density Function* (PDF). Among the extracted *moments* we have: (a) first order moment that define the mean or center of mass for an object while in PDF, it is the expected value of a probability distribution in any event. (b) The second order *moment* is the description of the variance that indicates the spread of the data distribution. (c) The third *moment* is the skewness that finds how the data distribution is biased after ignoring scaling factors.

Features from spectral domain

The representation of the image in the space of time and frequency is termed as the spectral domain. Spectral domain acquires an importance in machine learning when the data sample (images) is represented in the plane of time & frequency and then used in classification problems. In medical images classification problems, spectral domain based features support the machine learning architecture to gain classification accuracies even for low contrast medical images. To acquire the spectral features, two approaches are mostly used in literature, i.e., *wavelet based feature extraction* and *gabor filter based feature extraction*.

Wavelet transformation has attained great popularity in the medical image classification system. It is noticed that wavelet transform is not considered as the unique transformation for feature extraction in the spectral domain. There is a wide variety of wavelets each of which define a different transform *e.g. fast wavelet transform* and *curvelet transform*. There are many ways through which wavelets can be explained and defined depending on kind of acceptance required [89]. In radiological application, brain tumor analysis, *Discrete Wavelet* (DW) frequencies based feature extraction is proposed by Farias *et al.* [90] using the dataset of brain MR images having three classes of separability. The dataset consists of normal brain images, primary tumor images, and secondary tumor images. Using wavelet frequencies spectral features are computed that is used in classification later. Wavelet transfor-

mation based extraction of the features in spectral domain is used in [91–94] for discriminating among the normal and abnormal brain MR images. The extracted features consist the time-frequency space information of MR image that describes the structure in a smooth manner.

In literature, an informative approach in spectral domain for feature extraction using *Discrete Cosine Transformation* (DCT) is also used in [95,96]. The extracted features using DCT are used in medical image recognition with application in brain tumor images. The performance of the extracted features using DCT is compared with the DWT based features in [97] where it is shown that the performance with DWT is more that with DCT based features. The main reason for such variation in the accuracies is noticed in the image representation in space. The DCT based image features are found in frequency space only while the DWT based features consists the information of both time and frequency space.

By the advantage of using DWT based features, some hybrid algorithms are used in past in medical image classification system. Those hybrid algorithms used the DWT features with GLCM based features [98], GARCH features [99] and wavelet entropy for spider plots [100] for extracting textural information in the time-frequency domain. Such hybrid algorithms gained the increment in classification accuracies when tested with ensemble classifiers for discriminating normal and abnormal brain MR images.

2.3.2.2 Feature selection

In radiological applications, as the size of the data sample and dimensions increases with the express development in imaging instruments, the need for machine learning is seen for past several years. For example, with the advancement in imaging technique from 1T MR image to 7T MR image (3T MR image in case of literature), the imaging resolution has increased significantly. The higher resolution means increased in the number of voxels in an image that results in the increase in the number of features if all voxels are used. The increased number of features from MR image result in increased number of dimensions that is used to solve many problems in machine learning.

As the number of dimensions increases, the task of machine learning becomes complicated. *Feature extraction and feature selection* are the solution of such problem in which the relevant and informative information is extracted from the feature

set. *Principal Component Analysis* (PCA) is one such technique that is designed for data whose sub-manifold is embedded almost linearly in feature plane [101]. While for nonlinear sub-manifold data, several methods are discussed like *Locally Linear Embedding* (LLE) [102], *Laplacian Eigenmaps* [103], ISOMAPS [104] and *Diffusion Maps* [105]. PCA is the most widely used and known dimensionality reduction technique by which the number of variables are reduced. With the use of PCA, patterns in data are analyzed and expressed the data in such a way that highlight the similarities and differences of the data. The other main advantage of PCA is that once such patterns in the data are identified, then the data can be expressed by reducing the number of dimensions, without much loss of information.

In radiological applications, where the data dimensionality is high enough, PCA plays the significant role. It analyze the number of variables using which the high dimensionality space is transformed to lower dimensionality space. Using the new transformed values of low dimensional space, the data is analyzed with the help of machine learning algorithms. In brain tumors categorization various literature suggests the use of PCA, where high dimensional feature space is transformed to low dimensional feature space. Literature showed the implication of PCA in both spatial and spectral domain feature set pool.

Likewise PCA, another dimensionality reduction technique is *Independent Component Analysis* (ICA) that linearly transforms the original feature sample space to a new linear feature space [106]. Besides the similarity for data transformation, the major difference between the two is that in PCA the data transformation matrix is to preserve the components with maximum variance while for ICA the data transformation matrix is to minimize the statistical dependence between its components. Thus by incorporating dimensional reduction techniques the useful information from the data can be extracted. Similar assistance can also be obtained by feature selection techniques in machine learning.

Feature selection techniques are classified as *filter approach, wrapper approach* and *embedded approach*. A *filter approach* selects feature subsets without using any learning algorithm. It estimates the relevance score such as *Pearson correlation* and *mutual information* among features, which is used to select the best features. While a *wrapper approach* uses predictive accuracy of predetermined learning algorithms. For example, in a classification problem, a wrapper approach for feature selection tests a subset of the features of the classification problem and the subset that gives maximum classification accuracy is returned. An *embedded approach* allows com-

munication of different class of learning algorithms. The ensemble model based on dissimilar subsampling strategies, runs the learning algorithms on a number of sub-samples and the acquired features are united into a stable subset. For example, *decision tree* based classification of the data sample uses the inbuilt feature selection approach.

In machine learning perspective a new feature selection method named *Minimum Redundancy Maximal Relevance* (mRMR) [107] is proposed for selection of relevant features in medical imaging. The above said method includes two factor: one is relevancy among features and the target class and another one is redundancy between features. In such situation author proposed a probabilistic framework to minimize the redundancy and maximize the relevance at the same time. Besides mRMR, some other feature selection approaches used in machine learning are *chi-square test* [108], *correlation based feature selection* [109, 110], *branch-and-bound algorithm* [111].

In literature certain approaches for feature selection are proposed which is based on the concept of implication of *optimization algorithms*. *Optimization* algorithms help to find a subset of features that produces a higher accuracy with an optimal number of features. *Genetic algorithm* is such technique which is used to find the least number of features using optimization technique [112]. In medical science, the use of optimization also showed significant contribution when it is experimented to detect myocardial infarction using improved *Bat optimization* technique [113]. The use of the statistical measures with optimization approach is proposed for filtering relevant and informative features using rough set theory with *Bat optimization* [114]. With the advancement in the classification accuracy using reduced feature set based on evolutionary learning, a new methodology is proposed based on fulfilling multi-objectives for feature selection, i.e., *feature subset selection* and *feature weighting* [115]. Such approach has experimented with almost ten standard datasets of machine learning and found the significant contribution to machine learning society. The evolutionary learning algorithm when experimented with high dimensional dataset of gene expression data using *Binary Particle Swarm Optimization* (BPSO) is found significant for computer aided decision making [116, 117].

Besides the use of the various statistical and optimization approaches, there are certain approaches proposed in literature based on the concept of game theory. A game theoretic approach based feature selection is the revolution in the machine learning society for speed up the learning process and selecting the subset of fea-

tures in quick time. The game theoretic approach has experimented with the high dimensional dataset of hyperspectral images classification [118]. While for the low dimensional datasets of UCI machine learning data repository, the authors elaborate certain problems. The main problem of the most traditional feature selection methods are the ignorance of the some features which have strong classification ability as a group but is found weak as individuals. The cooperative game theory is claimed as the solution to such problem that gave significant results for feature subset selection when experimented with neighborhood entropy based cooperative game theory [119, 120]. The main task of cooperative game theory is to evaluate the power of each feature. The power can be served as a metric of the importance of each feature according to the complex and inherent interrelation among features.

2.3.2.3 Classification Models

In machine learning-based computer-aided decision-making systems, classification gives a significant contribution. It is the learning based decision making approach where the unknown sample is provided certain class label on basis of the past learning and decision making criteria. For example, distance is one such criteria which is used mostly in literature. The final output of the classifier can be found either a distinct value, one of the existing class label or a vector that reflect the likelihood to some specific class [121].

In classification based machine learning research, various models are proposed that are classified as *linear classification models, kernel learning based non-linear models, probabilistic models, artificial neural networks and ensemble learning models*. These models play significant role in machine learning based computer aided diagnosis of radiological applications. Each of the model has their own significance in medical image analysis. *Kernel based learning* e.g. *Support Vector Machine* (SVM) usually provides the best classification result of abnormality detection in radiological applications [122]. While the *probabilistic model* like Naive Bayes provides the theoretical aspects of medical imaging [123].

Some classification model estimates that there is a linear relationship exists between the input and output of the model. It is considered as the simplest approach for classification in machine learning based radiological applications. For example, *k-Nearest Neighbor* (KNN) and *Linear Discriminant Analysis* (LDA). *KNN* was introduced by Cover and Hart [124]. KNN uses k nearest samples for decision making. The working methodology of KNN classifier is based on an approximate

number of the nearest neighbors represented by the k. Among training sample the class instance associated with the majority of the k, be the class label of unidentified sample in dataset. In radiological applications, KNN was used to discriminate between the brain tumor types with various combinations of feature extraction and selection approaches [91, 98, 99]. For selecting the appropriate number of neighbors for classification, a model named *informative KNN* was proposed that uses the object informative theory, i.e., local and global to find the optimal nearest neighbors [125]. To increase the performance power of KNN, a weighted methodology is initiated that assigned the weightage to nearest data samples as compared to rest other data samples [126]. The weighted factor has lot of variations and techniques through which any data sample of the dataset has assigned weight e.g. uniform weighted mechanism and dual weighted approach [127].

Another linear model which has gained its importance in machine learning is *Linear Discriminant Analysis* (LDA) that is introduced in machine learning society in 1995 for mammographic images [128]. LDA works on the principle of transforming the data in such a way which maximizes the separation of mean class data and minimizes total within class scatterness of the given data in transform space. Using such projection space for diagnosis of the human brain tumors with proton magnetic resonance spectroscopy, the use of LDA is proposed with leave-one-out test paradigm [129]. In medical image classification, for tumor diagnosis, LDA also shows significant results when experimented with various feature extraction and selection mechanisms [85, 99]. Another approach that is found closely allied to linear discriminant analysis is quadratic discriminant analysis that tries to fetch the quadratic relationship between the independent and dependent variables [15]. It provided more powerful discriminant ability as compared with the linear separation of two classes learned by LDA.

In the radiological application of brain tumor analysis, ANN give its significant contribution in decision making. ANN is a methodology that is inspired by the brain and the way it learns and processes information. In literature various type of neural networks are discussed which is used in machine learning based applications. These networks can be classified on the basis of their structures. The basic ANN model is the perceptron model that has single layer and hence is also considered as linear classifier [130]. Extending the idea of the perceptron, another family member of the ANN is introduced in 1969 that has multiple layers called *Back Propagation Neural Network* (BPNN) for training the neural model [131]. In practice it is identified that among the multiple layers neural network, three layer based ANNs can be strong

enough to learn any complex function.

In the later years, some other neural network structures are introduced that have variety of variation. *Hopfield neural network* and *Boltzmann machine* are examples of that neural structures. *Hopfield network* is designed in respect that it consists of only single layer with neurons and all neurons are connected with each other [132] while *Boltzmann machine* is the stochastic extension of *Hopfield network* that solves complex combinatorial problems [133]. In the same duration when *Hopfield network* introduced, one network model is introduced called *Kohonen SOM* [134]. SOM is a unique network to other ones as it conducts unsupervised learning while other networks are supervised. It is noticed that the final network that is learned by SOM can express various characteristics of input signal. It is found useful for many medical applications that are not limited to dimension reduction, visualization of high dimensional data, clustering and tumor classification [83]. Besides the above discussed ANNs there are certain other important neural network that exist in machine learning and that are used for computer-aided decision-making, like *Radial basis function* (RBF) [135] and *Probabilistic neural network* [136].

In machine learning based problem-solving applications, kernel learning based model proved itself as a powerful model for data analysis in the variety of real-time applications. The most known kernel-based model is the *SVM*. *SVM* uses kernel based supervised learning methods that is used for data classification [137]. The definition of kernel in machine learning indicates a matrix that encodes similarities between the data samples. The encoded value is evaluated by certain weighting functions in integral equation that is used to calculate similarity among samples.

The main working concept of SVM is to minimize the observed classification error and maximize the geometric margin concurrently on the training set that leads to high simplification ability on the new data samples. Using the above discussed methodology, a feature selection approach is proposed for SVM model where the relevant features are identified by minimization of the bounds on leave-one-out error finding [138]. Besides feature selection, SVM is used widely in radiological applications where distinguishing between samples are required. In MRI-based tumor diagnosis, the use of SVM with radial basis kernel function is proposed for discriminating between the normal and abnormal brain images [85, 86, 88, 90, 93, 94].

In machine learning, *ensemble learning model* is known as collaborative or collective model. In the ensemble model, two or more models of various working

methodologies can participate in the decision making. In general, it is assumed that the learning performed by an ensemble of the various classifiers is very effective in decision-making [139]. It includes a collection of methods that learn certain target function in training which is defined initially and finally made the decision after combining all the ensemble models prediction.

The most widely known ensemble method is *bagging and boosting* algorithm for experimentation [140]. In such case the base classifiers are build up using bootstrap samples. Each bootstrap sample is formed by sampling the training set with uniformity and replacement. In *ensemble classifiers* the accuracy can be enhanced through introducing several versions of the base classifier when unstable learning algorithms are used *e.g. decision tree*. Another algorithm is the *AdaBoost algorithm* that calls the base classifier repeatedly and maintains the distribution of the weights on the training set [141]. In learning process, the weight of the wrongly classified samples are increased so that the weak learner is forced to focus on the firm examples in the training set. In medical applications the use of the ensemble learning is seen in many literatures for classification of the brain tumors based on ANN ensemble [82, 83, 91], KNN and SVM ensemble [98, 99].

2.4 Summary of chapter

This chapter introduces the basic concepts about the human brain, its structure, and the brain tumors. Later in this chapter, some primary basics have discussed the types of the brain tumor, its symptoms, and the possible causes through which brain tumor spreads in the human body. Finally, at the end of the chapter, the basics of the MRI is discussed in the regards of the image formation followed by the use of the MRI imaging for brain tumor diagnosis.

Further, an interconnection between the machine learning system and the radiology is discussed. The chapter focuses on the various phases of the machine learning systems i.e. the segmentation of an image, feature extraction mechanisms followed by feature selection methodology, and the classification model. Later, in the chapter, the contribution of the machine learning system in radiological applications is discussed. The subsection discusses the various machine learning approaches that are used to solve various applications of the radiology.

Later in the chapter, we presented most of the state-of-art techniques employed

for the classification of the brain tumors in MR images. The presented related work explored three main phases of the machine learning system in regards to the identification and classification of the brain tumors in the MR images. These three phases are *segmentation of the tumor regions, feature extraction, feature selection, and classification*. In each of the phase of machine learning system, various approaches are discussed that was proposed in past to solve the radiological problem of brain tumor analysis. These machine learning approaches produce higher accuracy in the discrimination of the brain tumor types in computer-aided diagnosis. It is because the machine learning model classification accuracy depends on the training of the dataset. A well-trained model can predict the class label of the unknown sample by its learning using the relevant and informative features extracted from MR image.

In the next chapter, the main contribution of the thesis is presented for tumor segmentation. The chapter gives a detailed description of the proposed biclustering approach based on the hybrid algorithm. Later, a hierarchical algorithm is proposed for the segmentation of the tumor region from the MR image using probabilistic mutual information.

Chapter 3

Machine learning based Tumor segmentation

Machine learning-based computer-assisted diagnosis in radiological application especially in the detection of the tumors in brain *Magnetic Resonance* (MR) slices has gained importance. The primary goal of the brain tumor MR image analysis is to gather the patient specific critical clinical information and features which are used for diagnosis. Such information is found embedded in the three-dimensional voxels of MR images that have to be extracted for monitoring the tumor affected regions in the MR slices. *Segmentation* is one such machine learning processes which is used to extract the regions from images based on learning algorithms.

In brain MR slices, the *segmentation* of the tumor consists in separating various tumor cells from different regions of the brain. The segmentation process includes extrication of solid lesions, edema and necrosis from different parts of the brain e.g. gray matter, white matter and cerebrospinal fluid. As the characteristic of the tumor is that it do not have any specific shape, size, and location thus a strong machine learning algorithm is always desired that can be categorized and differentiate the normal and abnormal brain cells with high precision. However, perfect segmentation of abnormalities is not straightforward.

In the past, various segmentation algorithms were proposed in radiological applications for abnormality detection. These algorithms adopt one of the following approaches *threshold based image segmentation, region splitting and merging based image segmentation and hierarchical theory based image segmentation*. But these algorithms have certain limitations, for example, (i) *Threshold based image segmentation* induces the formation of a binary image i.e. either a pixel belongs to

a certain group based on the threshold (white region) or the pixel is not of interest (black region). Also, the selection of the threshold value and variation in results with a slight change in threshold is problematic. (ii) *Region splitting and merging* based image segmentation mainly relies on the assumption that the neighboring pixels within one region have similar value. The region based algorithms are highly influenced by noise as they are dependent on the selection of the seed point from where the region grows. (iii) *Hierarchical theory based image segmentation* are mainly adopted for datasets having numeric attributes. The algorithms use the iterative process to find the similarity indexes between two pixels. The selection of the similarity criteria is found significant and thus results are influenced by noise. Besides this, the algorithm has the limitation of forming cascading of clusters in each iteration that make the algorithms non-supportive for images.

In this chapter, new algorithms are proposed to segment the tumor from MR images which reduce the limitations found in existing algorithms. The proposed algorithms focus on improving the segmented results by using binarization algorithms and use the hierarchical graph-based model to segment the tumor from MR images, which is not done in the past. The proposed algorithms are capable of segmenting the tumor region with accuracy and high precision. Finally, the performance comparison of the proposed algorithms with state-of-art algorithms is defined.

The rest of the chapter is organized as follows: In section 3.1, the proposed algorithms for segmentation of the tumor from MRI is given. In section 3.2, the description of the experimental dataset is given followed by the performance evaluation metrics used for validation of the results in section 3.3. The results and discussions over proposed algorithms are given in section 3.4. In section 3.5, the comparison of the proposed algorithms with existing algorithms are given. Finally, in section 3.6, the overall summary of the chapter is presented.

3.1 Proposed algorithms

In this chapter, for the segmentation of the tumor region from the brain MR image, two algorithms are proposed. The first algorithm is the hybrid algorithm which is based on the effective combination of two standard approaches named *OTSU method* and *Region growing method*. Both these algorithms were used independently in the past for image segmentation but the proposed hybrid algorithm is used to overcome the limitations of these algorithms and gain accuracy in segmentation. The second proposed algorithm is the *hierarchical approach* that is used to segment

the tumor region from the MR image with a high accuracy rate. The hierarchical approach is based on the concept of graph which results in the formation of clusters using *independent trees* and *mutual information*. Here the word hierarchical doesn't relate with the existing dendogram approach. But it uses the merging of the trees at each iterative level of the computation.

The motivation behind introducing the hierarchical concept in proposed approach is by examine the certain factors which were associated with it. Hierarchical clustering approaches can give different partitions depending on the level of resolution we are looking. Thus, it doesn't need the number of clusters to be specified. In addition, for fixed number of records, hierarchical algorithms performance increases and took less amount of time. The detailed discussion of the proposed hierarchical algorithm is given in its respective sub-section.

3.1.1 Hybrid algorithm for tumor segmentation

In this hybrid algorithm for tumor segmentation in brain MR images, threshold based image binarization approaches are used i.e. *OTSU* and *region growing*. *OTSU approach* is used to segment the image and get the binary image with a global threshold value that minimizes the within-class variance. The resultant binary image is further used by the *region growing* approach for initialization of the seed point and gets the final *n* clustered resultant segmented image that shows the tumor region with high precision. The basic working of the OTSU approach is described as:

OTSU algorithm: This algorithm uses the variance as the measure of finding homogeneity between the regions i.e. the regions that have high homogeneity they will have a low variance between them. OTSU approach uses the threshold value as a filter that separates the regions. The pixel value of an image having the value greater than or equal to set threshold will be considered as the member of the segmented region otherwise the pixel is considered as an outlier. The selection of the threshold is based on minimizing the within-class variance of two regions separated by the threshold operator [21]

Let the image has L gray levels and for each gray level value i, p(i) be its probability in the image. Let the initialized threshold is T. Then the pixels of an image be

classified as outlier (p_o) or region of interest (p_r) is given by the following equation:

$$p_0(T) = \sum_{i=1}^{T} p(i), \qquad p_r(T) = \sum_{i=T+1}^{L} p(i) \tag{3.1}$$

The mean gray level values of outlier and region are given by:

$$\mu_o(T) = \frac{\sum_{i=1}^{T} i.p(i)}{p_0(T)}, \qquad \mu_r(T) = \frac{\sum_{i=T+1}^{L} i.p(i)}{p_r(T)} \tag{3.2}$$

While the mean value of the whole image is given by:

$$\mu = \mu_o + \mu_r = \sum_{i=1}^{L} i.p(i) \;\; as \;\; p_0(T) + p_r(T) = 1 \tag{3.3}$$

The variance of the outlier and region are computed by:

$$\sigma_o^2(T) = \frac{\sum_{i=1}^{T}(i-\mu_o)^2.p(i)}{\sum_{i=1}^{T} p(i)} = \frac{\sum_{i=1}^{T}(i-\mu_o)^2.p(i)}{p_o(T)} \tag{3.4}$$

$$\sigma_r^2(T) = \frac{\sum_{i=T+1}^{L}(i-\mu_r)^2.p(i)}{\sum_{i=T+1}^{L} p(i)} = \frac{\sum_{i=T+1}^{L}(i-\mu_r)^2.p(i)}{p_r(T)} \tag{3.5}$$

Similarly the variance of whole image is given by:

$$\sigma^2 = \sum_{i=1}^{L} (i-\mu)^2.p(i) \tag{3.6}$$

The computation of the within class variance ($\sigma W^2(T)$) and between class variance ($\sigma B^2(T)$) is given by the summation of the two variances that is multiplied by the weights associated with them:

$$\begin{aligned}\sigma^2 &= p_o(T)\sigma_o^2(T) + p_r(T)\sigma_r^2(T) + p_o(T)(\mu_o(T)-\mu)^2 + p_r(T)(\mu_r(T)-\mu)^2 \\ &= \sigma W^2(T) + \sigma B^2(T)\end{aligned} \tag{3.7}$$

For the determination of the final threshold, it is required to compute the threshold that minimizes the within-class variance. As it has shown that total variance of the image is not dependent on T,thus the value of T that maximizes the between class variance will be the value that also minimizes the within class variance. Now let us consider the maximization of the σB^2, the equation for σB^2 be rewritten as

[21]:

$$\sigma B^2 = \frac{[\mu_o(T) - \mu.p_o(T)]^2}{p_o(T)p_r(T)} \tag{3.8}$$

where $\mu(T) = \sum_{i=1}^{T} i.p(i)$

For each of the value of the gray level, we have to test the possibility that the respective gray level value will be considered as a threshold value T if it maximizes σB^2.

Even then the method has certain limitations i.e. the process assumes that the image is bimodal only. Also, if the two classes have formed unequal then the method generates two maxima's and thus the selected maxima is not necessarily be the global one. Also, the method does not perform well when the image has low illumination, for example, medical images datasets of CT scans and MRI.

To overcome the limitation of the OTSU algorithm, in the proposed methodology a hybrid concept is introduced that uses the *region growing* algorithm on the extracted results by OTSU. The reason for using a such hybrid approach is that region based approach is the region growing methodology that uses the initialized seed point as the source to grow regions. Also, it relies on the assumption that the nearby or neighboring pixels within the single region have similar values. But, the selection of the seed point is again the limitation of the region-based algorithm. Since the algorithm uses a predefined threshold value to segment the regions based on the neighborhood criteria between seed point and nearby pixels, thus optimal threshold value is also the requirement of the algorithm.

In the proposed algorithm, initially the OTSU approach is applied to the low illuminated MR images, and the threshold value T is obtained. By using the threshold T, the image is segmented into two parts i.e. desired region and outlier region. Then the desired region is further used by the region growing method to initialize the seed point and grow the region. The threshold obtained by using OTSU method is used to filter the region again. Thus by using the region growing approach on OTSU, the problem of two maxima is resolved. Also, the limitation of the bimodality is resolved by positioning a number of seed points which segments the image into n clusters.

The pseudo code of the proposed hybrid algorithm is given as algorithm 1

Algorithm 1: Hybrid Algorithm for Tumor Segmentation

Input: Brain MR image (I) having tumor
Output: Tumor segmented image

1. T := Initialize the threshold

//Compute the histogram of all the gray label values (i) of an image
2. $P(i) := Hist\ (i)$

//Segment the input image in two regions based on threshold
3. $[p_o,\ p_r] := partition\ (I)$

//Compute the mean gray level of two regions
4. $\mu_o(T) := mean_{gl}\ (p_o)$, $\mu_r(T) := mean_{gl}\ (p_r)$

//Compute mean of whole image
5. $\mu := mean_{gl}\ (\mu_o(T),\ \mu_r(T))$

//Compute the variance gray level of two regions
6. $\sigma_o^2(T) := Variance_{gl}\ (p_o,\mu_o(T))$
$\sigma_r^2(\mathrm{T}) := Variance_{gl}(p_r,\mu_r(\mathrm{T}))$

// Compute variance of whole image
7. $\sigma^2 := Variance_{gl}(\sigma_o^2,\sigma_r^2)$

8. Obtain within class and between class variance σW^2, σB^2

// Find gray level that maximize σB^2 & set as new T
9. $T' := maximize\ (\sigma B^2)$

// assigning 3 seed points for 3 clusters
// Initialize any gray level as seed point in respective regions
10. $C1 := Seed\ (p_r)$, $C2,\ C3 := Seed\ (p_o)$

11. let n_i be the neighboring pixel in respective regions then,

// Compute distance between seed point & neighboring pixels
12. $D_i := Distance\ (C_i,\ n_i)$

13. if D_i T' then, // if distance is less than threshold then

14. n_i C_i, // assign neighboring pixel to respective cluster

15. Recompute the boundary of cluster, C_i and set the value as new seed point

16. Repeat step 12 to 15 until all pixels of an image allocate to certain cluster

17. Return the clusters C1, C2,C3 as an image having segmented tumor

3.1.2 Hierarchical algorithm for tumor segmentation

In this proposed algorithm, a hierarchical approach is used for segmentation of the tumor from brain MR images. The proposed algorithm uses the input image for the generation of the unique trees from the input matrix representing pixels. The main idea of the algorithm is to use a 8-bit DICOM Brain MR Image as an input and initialize all its pixels in individual clusters. This result in the generation of multiple clusters having similar value. Further, these clusters will iteratively merged in a

bottom-up fashion. After the first iteration of the proposed algorithm, the individual clusters having similar pixel values are merged to form a single tree. Thus, at the end of the first iteration, there exists only the limited number of trees having unique face value. The pseudo code of the proposed disjoint tree based algorithm is given by algorithm 2.

Algorithm 2: Disjoint Tree based algorithm for Tumor Segmentation

Input: I ← Brain MR Image
Output: out ← Clustered Image

Step 1: Preprocessing
f ← Filter−Image (I) // Filter the image to remove noise
gr ← Gray−image (f) // Convert the 3D voxel to gray scale

Step 2: Generation of Disjoint trees
V ← pi ; pi ∈ gr // Assign each pixel to a vertex
for each vi ∈ V
for each vk ∈ V && vi == vk
E ← Find−edge (vi ,vk) // Join the vertices having same vertex values
T ← Find−Tree (V, E) // Form a tree between same vertex values
V ← Update−Set (V) // Update the Vertex set
end for
end for
return (T)

Step 3: Tree Merging
k ← no of clusters // Initialize number of clusters required
while (k)
for each ti ∈ T
d ← Find−Distance (ti , tj) //Compute the distance between trees
m ← minimum (d) // Fetch the pair having minimum distance
c ← Minimum−Distance−Cluster(ti , tj) // Form the cluster
T ← Update−Set (T) // Update Tree set
end for
end while out ← cluster−set(c) // return the cluster set having k clusters

In the proposed algorithm there are three steps. The first step is the preprocessing step where the input MR image is filtered using the mean Gaussian filter of dimension 3x3 to remove the noise if it exists. During the image acquisition process, there may be a chance of obtaining the noise due to physical factors while forming the image, so noise filters are used. After filtering process, the formed image is the three bands noise free image which is converted to gray level. In gray level conversion, the bands of an image, i.e., *Red, Green, Blue* (R, G, B) are de-

composed to give the single band intensity image, also called gray level image. The formed Gray level image is used as an input for further steps in proposed method.

Trees generation: The second step of the proposed algorithm after the preprocessing is the generation of the trees. The proposed algorithm uses the preprocessed image pixel matrix to generate the pixel trees. The tree generation starts by initializing each of the pixels in the image matrix as the vertex. This results in the formation of a large number of trees having the single node (vertex). For example, if the size of the 8-bit image is 256 x 256 then the image pixel matrix consists of 65536 values. This results in the formation of the 65536 trees after initial step and each tree is considered as a cluster. This collection of trees are also called as forest. Further, these trees are processed to grow its structure based on the similarity parameter. The proposed algorithm uses the vertex value as the similarity metric using which the trees will grow. The concept of the tree formation is shown in Figure 3.1 with an example.

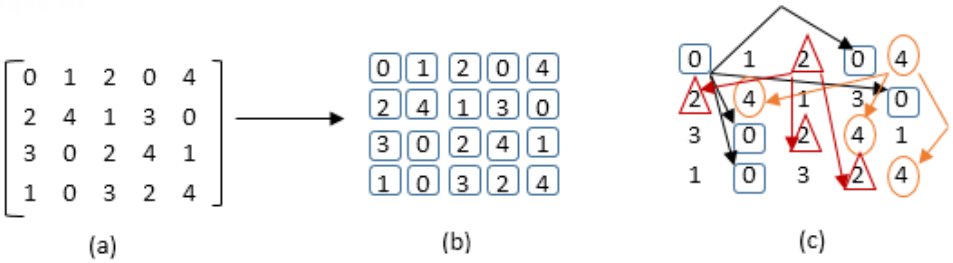


Figure 3.1: The method of Disjoint tree generation. (a) Sample image matrix. (b) Pixel trees having the single node. (c) Growing of the trees having same vertex value.

In the above Figure 3.1, for representation purpose, only growing of trees having node value 0, 2 and 4 are shown. Similarly, nodes having value 1 and 3 will also grow. When the tree will grow it is observed that each tree consists of the nodes which have the same vertex value. All such trees has unique value with them i.e. no two trees are found that has same vertex values after completion of step 2 of proposed method. Another issue which exists in step 2 is the generation of a large number of clusters. But at the end of step 2, these large number of clusters are reduced to a small number of clusters. This reduction in the number of clusters is explained as:

Suppose that the input gray image is of 8-bit format and for an 8-bit image the range of pixel intensity lies in between 0 (a minimum value representing dark

region) to the maximum gray level value 255 (representing bright region). While the pixel intensity range in between 1 to 254 indicate gray region in an image. This gray level value is found independent of the size of an image. For a 2D image, if the size of the image is 256 x 256, then the image matrix consists of 65536 values which are in the range of $0 \leq p \leq 255$, where p is the image pixel value. It is observed that initially there are 65536 pixel trees, or clusters, iff each pixel is assigned to a single cluster. When the tree starts growing, then these clusters is reduced to the maximum of 256 in number, as the pixels with same face value get merged to a single cluster. At any point of time there is a maximum of 256 unique pixels in the image matrix. This results in formation of maximum of 256 pixel trees after the completion of step 2 as shown in Figure 3.1 (c), the isolated tree clusters of node value 0 are merged to single tree cluster which holds all these isolated trees of 0 and so on.

Tree Merging: The last step of the proposed algorithm is the merging of the unique trees that exist in the forest obtained by using step 2. The trees are merged on the ground of the minimum distance between the trees. The proposed algorithm in step 3, starts by initializing the number of clusters required by the application. The choice of the number of clusters is user specific. After the initialization of required clusters, the output of step 2 which is in the form of pixel trees is computed in respect of finding the distance between trees. But the issue generates about how to find the distance between trees. To solve this problem, the proposed algorithm computes the mean value of each tree. The distance between the trees is computed using the mean values of the tree. The concept is explained by using an example in Figure 3.2. In the above Figure 3.2, (a) Represents the generation of the final

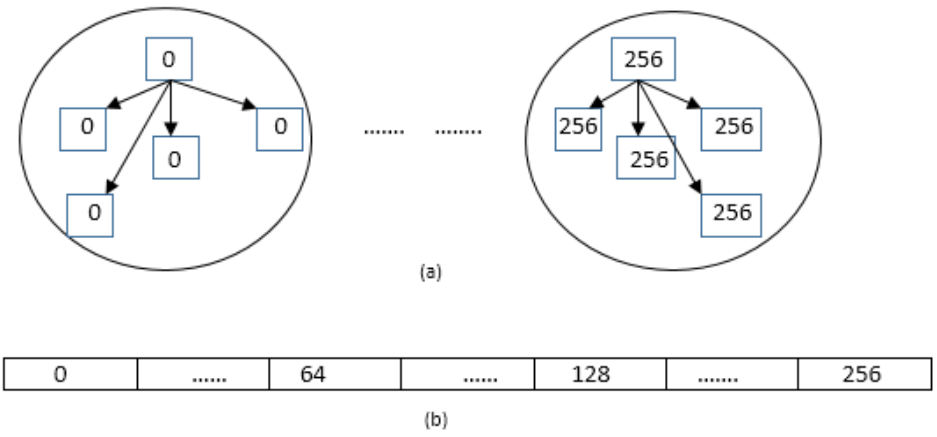


Figure 3.2: Process of tree generation

trees based on the unique gray level value in an input image. For an 8 bit image, the

maximum possible gray value is 256. While in Figure 3.2 (b), The generation of the 1D vector of size 256 indicating the mean value of the unique trees is presented.

After generation of the 1D vector representing the mean value of the each unique tree, the distance is computed between the trees. Graphically, it is represented with the help of the complete weighted graph in which every node of the graph is the mean value of the tree as indicated by 1D vector. The weight between any two nodes of the graph is the distance between the nodes. Then after the pair of nodes having minimum distance among all distances will be merged to form a joined tree and the vertex values will be updated to the maximum of the value among the trees to be joined. Similarly, the mean value of the tree is updated and the process iteratively executes till required number of clusters formed. The idea for replacing the value of new cluster with the maximum of the value joining clusters is for maximizing the inter-cluster distance as much as possible and minimizing intra-cluster distance. Thus the formed cluster is differentiated to each other as much as possible.

The illustration of the concept is shown by an example. Let us suppose that there are five unique trees of gray values 0, 14, 125, 201, 254 having 5, 3, 8, 4, 5 internal instances respectively in each unique tree that are formed after step 2. The mean value of these trees is same as their face values. Let us suppose that the required number of clusters are 3. The generation of the complete weighted graph and the process of tree merging is shown in Figure 3.3. At each iteration of the algorithm, the nodes of the trees having minimum distance is merged and the corresponding count of the internal nodes of the merged tree is the sum of both the internal nodes of trees that are merged as shown in Figure 3.3 (e). In the above Figure 3.3, an example of the tree merging is given using step 3 of the proposed algorithm. (a) Shows the complete weighted graph of unique trees, (b) First row represents the mean value of the tree and in the second row, the corresponding internal nodes are represented. (c) Distance matrix between the nodes. The value are marked with bold is the minimum distance value between consecutive nodes. (d)-(g) represents the iterative steps for merging of trees based on the pair having minimum distance with updated in respective mean and instance matrix and distance matrix. (h) Represents the final clusters, i.e. 3 with the number of instances as internal nodes. e

The proposed algorithm also presents the extension of the Disjoint tree algorithm in respect of merging the trees. The extended algorithm uses the probabilistic mutual information based on the joint probabilities to merge the trees. This extended method replaces the distance based tree merging, the concept of the previ-

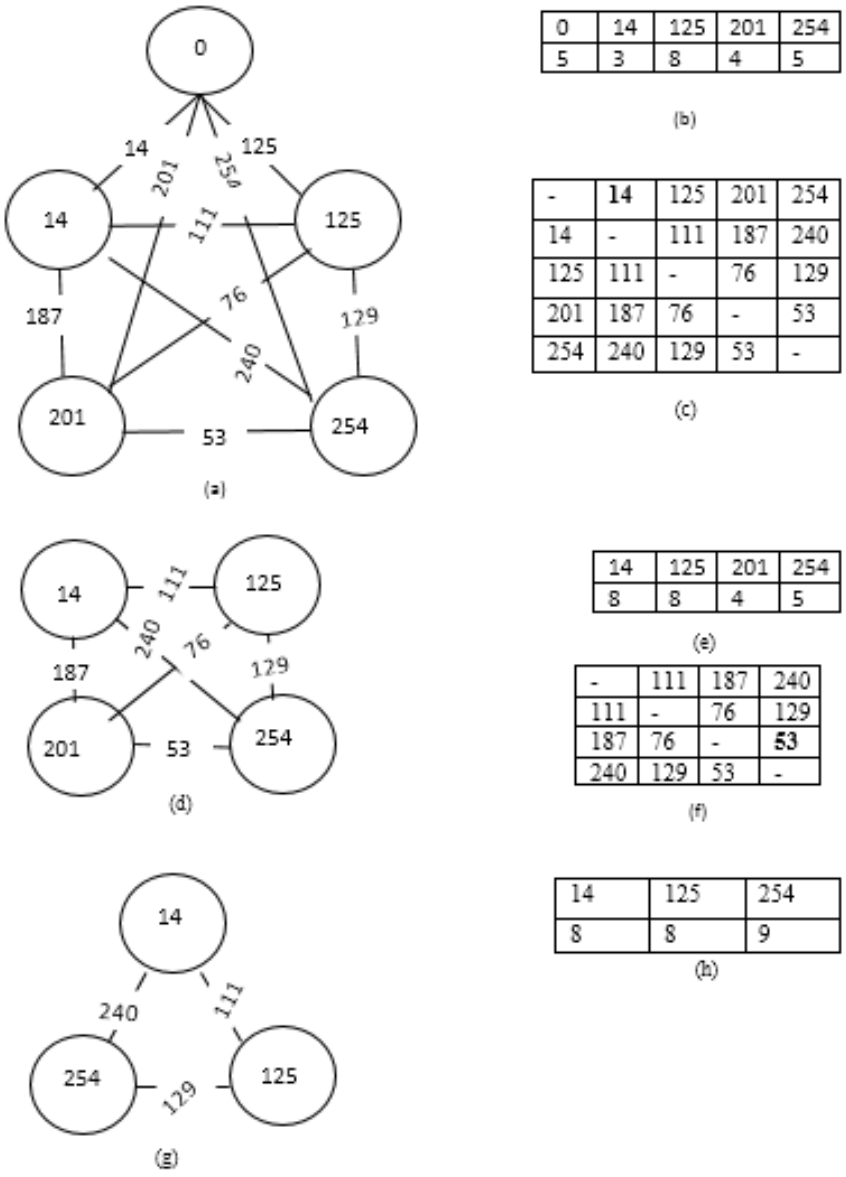


Figure 3.3: Example of tree merging

ous approach while rest of the algorithm remain same. The extended algorithm for merging the trees is given in algorithm 3.

The extended algorithm of merging trees using joint probability has certain advantages over previous distance based method. As in previous distance-based method the trees are merged using the difference in between the mean values of the trees. Also, in distance based method the formation and generation of trees are independent of the internal nodes of the tree. Besides this, in low illuminated images, there exist several trees on single iteration which are separated by unit distances

Algorithm 3: Merging of trees based on Probabilistic Mutual Information

Initialization: $k \leftarrow$ Number of clusters
$P \leftarrow$ *Probability-Distribution (T)* // prob. distribution of every unique tree
loop
while (k)
for each node *ni , nj* $\in T$ && *ni != nj*
$d \leftarrow$ *joint-Probability (ni , nj, P)* // joint probability between 2 nodes
end for
$mx \leftarrow$ *maximum (d)* // identify nodes with maximum joint probability
$out \leftarrow$ *Cluster(ni , nj, mx)* // merging of nodes having max joint probability
$T \leftarrow$ *Update-Set (T)* // Updation of the probability vaues of Tree set
end loop
$C \leftarrow$ *cluster-set(out)* //Final clustered set

only. These certain problems are resolved using the joint probability distribution method. The extended method uses the probability distribution of the unique tree among the forest. The probability distribution is computed on the number of nodes present in each unique tree among the forest. It is also noted that at any point of time the overall probability of the forest cant exceed to 1 i.e.

$$P1 + P2 + P3 + ... + Pn = 1 \tag{3.9}$$

Where *P1, P2, P3,...,Pn* are the respective probability distribution of trees in the forest. After computing the probability distribution, a complete weighted graph is generated in which the nodes of the graph is represented by two values which are represented by:

$$\forall ni \in G, ni = [Vi, Pi] \tag{3.10}$$

where *ni* is the node in graph *G* having vertex value *Vi* and probability distribution *Pi*. The weighted value of the graph is computed on basis of the joint probability. This joint probability is used as an edge weight in a graph formed between the tree nodes. Since each tree vector is independent of each other thus the joint probability between 2 independent variables is given by:

$$g(x1, x2) = g'_{x1}(X1) * g'_{x2}(X2) \tag{3.11}$$

where *x1* and *x2* are the two independent variables and g'_x1(X1), g'_x2(X2) are their probability distribution values.

The value of the computed joint probability distribution is further used to merge the trees. The pair of trees having maximum joint probability is used to merge and form the cluster. At the point of cluster formation, the tree set is updated with probability values. The method iteratively executes till required number of clusters formed. The proposed extended method is described by an example given for former method. For given five unique gray labels 0, 14, 125, 201, 254 having 5, 3, 8, 4, 5 as internal instances respectively, the complete graph of five nodes with probability distribution table is shown in Figure 3.4. In the above Figure 3.4, an example of

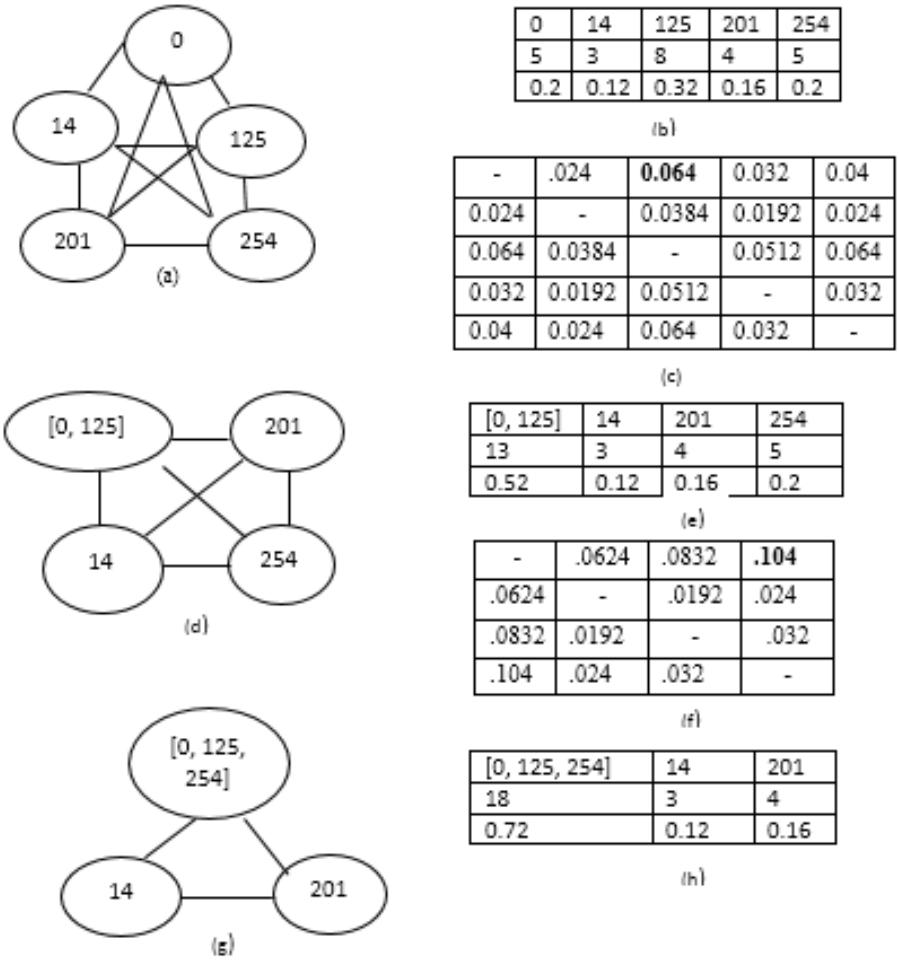


Figure 3.4: An example of extended proposed approach using Probability

the tree merging is given using step 3 of extended proposed method. (a) Shows the complete graph of unique trees (b) First row represent the unique tree gray value, Second row represent the corresponding internal nodes and the third row represents the probability distribution value. (c) Joint probability distribution matrix between the nodes. (d)-(g) represents the intermediate steps. (h) Represents the final clusters

3.2 Dataset Description

In this thesis, the proposed algorithms for image segmentation are experimented on brain MR images having abnormalities. The dataset of malignant brain tumors having five classes has experimented with proposed methodology. The experimental dataset is acquired by using 3.0 T GE scanner *(General Electronics Company, Milwaukee, WI)* from Department of Radiology, *Sawai Man Singh* (SMS) Medical College Jaipur, Rajasthan, INDIA. The MR images dataset used for segmentation consists of 650 weighted images of *T1, T2* and *Fluid Attenuated Inversion Recovery* (FLAIR) modality of three slices of head i.e. *Axial, Sagittal, and Coronal*.

All the patients images that are used for dataset generation are imaged using same imaging system and environment variables. The obtained voxels by any of the head slice have following specifications: 3D weighted voxels of *T1, T2, eT1, eT2* and *Fluid Attenuated Inversion Recovery* (FLAIR) each of having size 256x256x3 in DICOM format. In this thesis, 760 post contrast-enhanced weighted axial imaging modalities are used for examination purpose having the dimensionality of 256x256x3. The sample images of different MR modalities of the axial plane is shown in Figure 3.5.

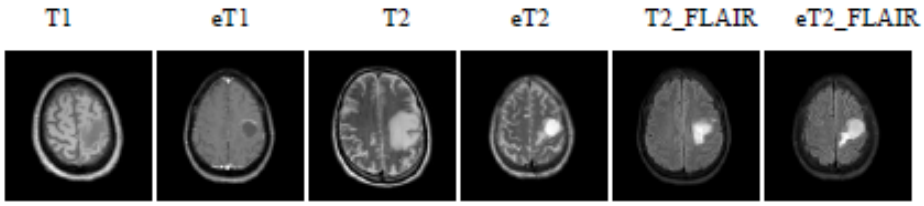


Figure 3.5: Axial MR images of different modalities.

3.3 Performance Evaluation Metrics

To evaluate the performance of the segmentation algorithms, there are certain parameters through which segmentation results are validated. These parameters are based on the two approaches depending upon the resultant output of the segmentation algorithm i.e., qualitative approach and quantitative approach.

The qualitative approach includes the validation of the segmented image results on basis of the quality analysis of the segmented output by the domain experts. For

example, in medical imaging, the segmented output is verified by the proficient radiologists. The qualitative validation of the machine learning-based problems does not include only the standalone domain expert consult but it includes the series of decisions by various experts to predict the quality of the segmented results. Thus qualitative approach based result validations are based on the visual parameters rather than any numerical value. While on the other hand, Quantitative approach deals with the numerical attributes for evaluating the performance of the segmentation algorithms. By the segmented results, quantitative approach is described by two evaluation methods i.e. Internal evaluation and External evaluation.

Some of the internal evaluation parameters are:

A *DB Index:* DB index is also called *Davies - Bouldin index* and is given by:

$$DB = \frac{1}{N_c}\sum_{i=1}^{n} max(\frac{D_i + D_j}{d(C_i, C_j)}) \ \ \& \ i! = j \tag{3.12}$$

where, N_c is the number of clusters, C_x is the cluster x centroid, $d(C_i, C_j)$ is the Euclidean distance between centroids, D_x is the average distance of every pixel to cluster centroid C_k. The segmentation algorithm which consists of smallest *DB Index* is considered as the best algorithm for segmentation.

B *Dunn Index:*It is defined as the ratio between the minimum inter-cluster distance to maximum intra-cluster distance for each partition and given by:

$$Dunn = min_{i\in[1,n]}(min_{j\in[1,n] \ \ \& \ i!=j}(\frac{d(i,j)}{max_{k\in[1,n]}d'(k)})) \tag{3.13}$$

where, *d(i,j)*is the distance between clusters i and j. *d'(k)* is the maximum distance between any pair of pixel elements in cluster k. The algorithm that produces high *Dunn Index* is considered as the best segmentation algorithm.

C *Silhouette Coefficient:*It is defined as the ratio of the average distance of the pixels within the cluster to the average distance between the clusters. It is given by:

$$SC = max_{i,j\in[1,n]}(\frac{D(i,j)}{D(C_i, C_j)}) \tag{3.14}$$

where, *D(i,j)*is the average distance between the pixels with the cluster, C_x is the cluster centroid.

On the other hand, the external evaluation performance measure depends upon calculating the degree of closeness of the clustering algorithm to the predetermined benchmarks of the classes. Some of the performance measures using external evaluation criteria are given below:

A *Rand Measure:*It depend on the ratio which calculates the degree of similarity to its benchmarks classes. It is given by:

$$RM = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.15}$$

where, $TP + FP + TN + FN$ are true positive, false positive, true negative, false negative respectively.

B *F- Measures:*It is used to balance the ratio of false negatives using a weighting parameter β. It is given by:

$$F = \frac{(1+\beta)^2.P.R}{\beta^2.P + R} \tag{3.16}$$

where, β is a weighting factor and, P, R are *Precision* and *Recall* and is given by:

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN} \tag{3.17}$$

C *Jaccard Index:*It is used when the similarity computation is required to find between the two datasets. Jaccard index is like a probability computation where the value 1 indicated full similarity between datasets and 0 indicated totally dissimilar datasets. it is given by:

$$J(X_i, X_j) = \frac{TP}{TP + FP + FN} \ \ or \ \ J(X_i, X_j) = \frac{(X_i \cap X_j)}{(X_i \cup X_j)} \tag{3.18}$$

D *Fowlkes- Mallows Index:*It also computes the similarity index between the segmented result obtained by the algorithm and benchmarks classes. The

higher value of Fowlkes- Mallows index represents the most similarity between the two. Mathematically it is given by the geometric mean of precision and recall, and is given by:

$$FMI = \sqrt{\frac{TP}{TP+FP} \cdot \frac{TP}{TP+FN}} \tag{3.19}$$

3.4 Result and Discussions

To segment the tumor region from MR images various algorithms are presented in the literature. These methods segment the input MR image into the two groups on the ground of the selected threshold value or group into multiple clusters by finding central distribution of the data with distance metric function. But in regards of the simplicity of these algorithms, they have certain limitations as well. In this thesis, two algorithms are proposed for the segmentation of the tumor region from MR image. The first algorithm is the hybrid algorithm that uses the OTSU based threshold value followed by the region growing algorithm for segmentation of the tumor region from MR image. The experimentation results of hybrid algorithm with various brain tumor MR images dataset are listed in Table 3.1. These results are qualitatively verified by an expert radiologist of *Sawai Man Singh* (SMS) Medical College Jaipur.

The second algorithm is based on the hierarchical method for the image segmentation. In this thesis, two hierarchical algorithms are proposed for the segmentation of the tumor region from MR images. The proposed algorithms are based on the concept of tree formation where the forest of unique gray label values is formed having the pixel tree of each distributed gray label. These pixel trees are further merged to form the final cluster. The two proposed hierarchical algorithms are different in terms of merging the pixel trees. The initial hierarchical algorithm uses the distance metric between the trees for merging, whose results are presented in Table 3.2.

The second hierarchical algorithm uses the probability distribution of the gray label value of pixel tree for tree merging. The probabilistic mutual information is used as a metric by which two trees are merged. The experimentation results of the

Table 3.1: Experimentation results for the proposed hybrid algorithm

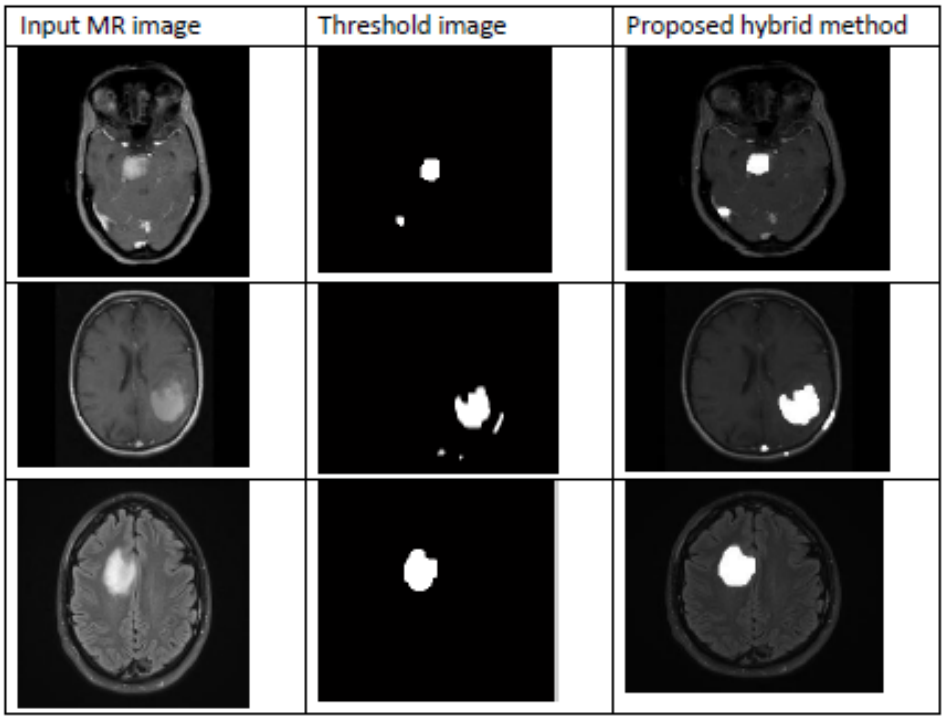
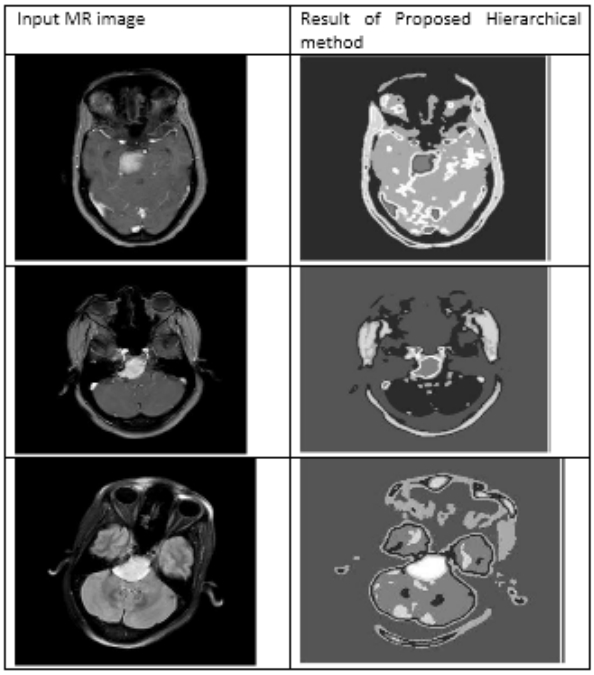
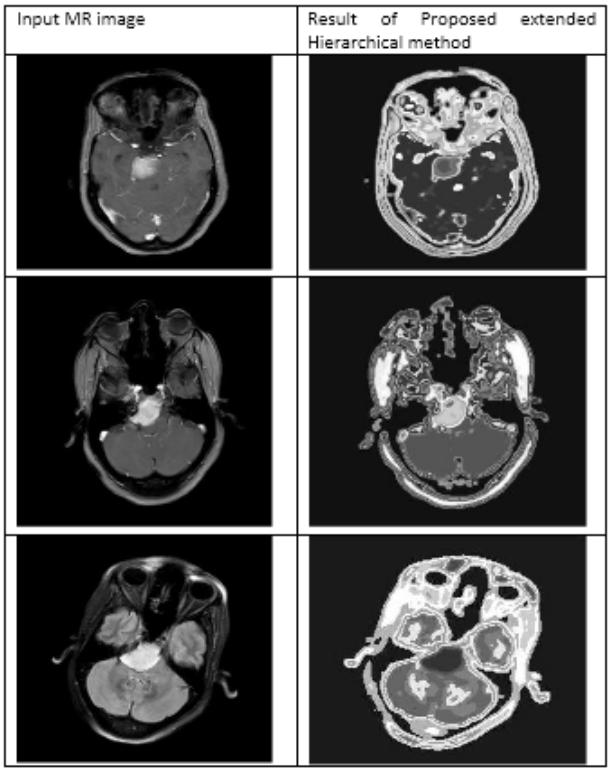


Table 3.2: Experimentation results of the Hierarchical algorithm for image segmentation



extended hierarchical approach are presented in Table 3.3.

Table 3.3: Experimentation results of the extended Hierarchical algorithm for image segmentation



3.5 Comparison with existing algorithms

For brain tumor dataset, formerly several researchers have reported various machine learning algorithms for segmentation of the tumor region inside MR images. Previous literature has reported various segmentation algorithms based on threshold-based approach, region growing approach and iterative approach. Some of the existing performances of state-of-art algorithms with proposed methods for tumor segmentation on basis of the qualitative analysis are represented in Table 3.4. These results are verified by the expert radiologists from the department of Radiology, *Sawai Man Singh* (SMS) Medical College Jaipur.

Some of the other experimental results based on the quantitative measures are presented in Table 3.5. For validation of the proposed algorithms with state-of-art algorithms that are present in literature is performed by the internal quantitative measure using DB index and Dunn index. The segmented results are validated with the help of numeric value obtained by computation. The listed values in Table 3.5 is the average values of the 650 patients MRI which are used for experimentation.

Table 3.4: Qualitative performance comparison of various algorithms on brain tumor dataset

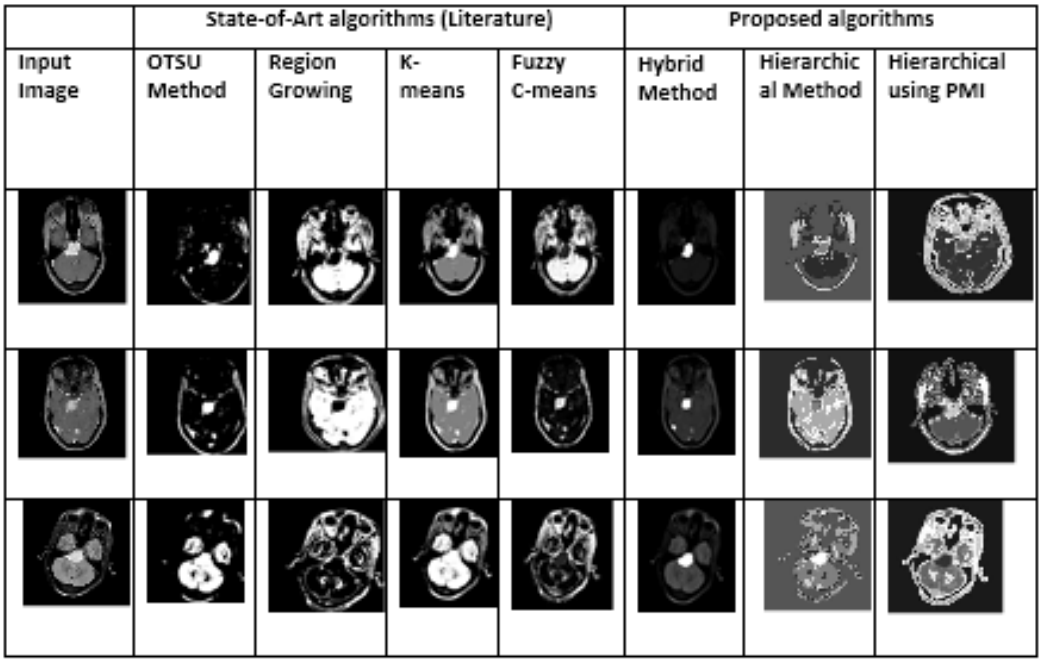


Table 3.5: Quantitative performance comparison of various algorithms on brain tumor dataset

Algorithm	DB Index	Dunn Index
OTSU	0.42	6.38
Region Growing	0.38	6.24
K-Means	0.21	7.75
Fuzzy C-Means	0.17	11.37
Hybrid algorithm	0.29	7.16
Hierarchical algorithm	0.14	12.36
extended Hierarchical algorithm	0.13	13.49

3.6 Summary of Chapter

Many researchers, proposed various algorithms as reported in Chapter 2, section 2.3.1, for segmentation of the tumor region from MR images. The literature analysis found certain limitations on state-of-art algorithms which are addressed in this thesis.

The experiments are done on the MR images of high-grade malignant brain tumors of variable shape, size, and locations. The qualitative experimental results

are shown in Table 3.4 for the algorithms including state-of-art algorithms and proposed algorithms. The verified results (by the radiologist i.e. qualitative validation) of the proposed algorithms outperform the state-of-art methods.

The performance of the segmentation algorithm is also analyzed by the quantitative statistical measures. These statistical measures help in finding how well the segmentation algorithm performs on the ground of the expected benchmarks. For validation of the proposed algorithms quantitatively, two indexes have been used, (i) DB index and (ii) Dunn index. The lower value of DB index while higher value of Dunn index represent the good segmentation results.

The analysis of Table 3.5 shows that the bi-clustering algorithms, Hybrid algorithm performs well as compared with OTSU and region growing algorithm. The reason is former algorithms have lower DB index and higher Dunn index than later algorithms. Among multi-cluster algorithms, two algorithms based on hierarchical algorithm without probabilistic mutual information and with probabilistic mutual information are proposed. It is interesting to find that hierarchical based algorithms give best results among all described algorithms. The hierarchical algorithm with probabilistic mutual information gets lowest DB index and highest Dunn index among overall algorithms and thus considered as the better segmentation algorithm for tumor extraction in MR images.

In the next chapter, the various proposed feature extraction algorithm are discussed for extracting relevant information from MR image. The proposed algorithm is used to extract the textural information from the spatial domain. Later, the spectral feature extraction mechanism is proposed to fetch the hybrid set of features that is used for MR classification.

Chapter 4

Classification of Malignant Brain Tumor MRI using Textural Features

Features are the important characteristics of machine learning. Using features, the hidden information present in the images are encoded. In MRI, the segmented ROI's are used to extract the information using feature extraction algorithms. The extracted features are used to represent the image in a digitized format which is used by the machine learning. In this thesis, two categories of features are extracted from segmented ROI's to discriminate among the malignant brain tumors. These are: i) Spatial domain features and ii) Spectral domain features. Both the type of features provide the textural information present in the ROI for classification.

The sections present in this thesis is organized as follows: In section 4.1, the proposed algorithms are given for the textural features extraction in spatial domain followed by the spectral domain features in section 4.2. In section 4.3, the proposed algorithms are given for the selection of the relevant features among extracted features. The experimental dataset which is used for the classification is presented in section 4.4. In section 4.5, the performance evaluation metric is given for evaluating the accuracy of the proposed algorithms. The experimental results and their discussions are presented in section 4.6 followed by the statistical validations of the proposed algorithms in section 4.7. Finally in section 4.8, the overall summary of the chapter is given.

4.1 Textural features from spatial domain

4.1.1 *Counting Label Occurrence Matrix* (CLOM)

Texture is one of an informative measure for feature extraction. To extract the textural information from MR images, a new algorithm is proposed named *Counting Label Occurrence Matrix* (CLOM). CLOM explores the frequency of the unique gray label values in a grid of 3x3 over the input MR image ROI. To extract the spatial distributed information, four occurrence sub-matrices are calculated across the four different orientations i.e. $\theta = 0^0, 45^0, 90^0, 135^0$. These four matrices are used to extract the textural features for segmented abnormal ROIs.

CLOM is a statistical algorithm for extracting the textural features that consider the spatial relationship of the pixels. CLOM characterizes the texture of an image by calculating how often a pixel with specific values occurs in different dimensions of an image. These dimensions are based on different orientations and represented by angular value (θ). The four orientations values of θ results in four occurrences matrices from which texture features are extracted. The generation of the occurrence matrix is illustrative in the following example: For an instance, consider the input MR image as shown in Figure 4.1 (a) with a grid of 4x4 having four gray level values as shown in Figure 4.1 (b).

Using grid value information three parameter values *M, D, R* are fetched as shown in Figure 4.1 (c). These values are used to construct the four occurrence matrices based on four orientations. For an orientation $\theta = 0^0$ and 90^0 the formed occurrence matrix is of dimension *R x (M+1)* while for 45^0 and 135^0 , the size of occurrence matrix is *D x (M+1)* as shown in Figure 4.1 (d)-(g).

The values of the occurrence matrix element at any index location are equal to the frequency of the appearance of gray value in respective row or column. As seen in the given example 4.1 (b), the frequency values of gray level 0, 1, 2, and 3 in first row are one, zero, two, and one respectively. Similarly, other values are computed for each row, each column, clockwise and anti-clockwise diagonals. The final four occurrence matrices are shown in Figure 4.1 (d)-(g).

The pseudo code of CLOM algorithm is given below.

Algorithm 4: Counting Label Occurrence Matrix (CLOM)

Input: Square image (I) of size M*M
Output: CLOM Matrix (C)

Procedure:(CLOM)
r, c ← size(I)
Grid ← splitting of an image (I) into a grid of predefined size
g ← no. of gray levels in input image (Grid)
for each $\theta \leftarrow 0^0$,45^0, 90^0, 135^0 **do**
if $\theta \leftarrow 0^0$ then
for each i ← 1 : r do
C ← frequency count of each gray level 'g' in row (i)
return (**C[r,g]**)
end **for**
end **if**
if $\theta \leftarrow 90^0$ then
for each i ← 1 : c do
C ← frequency count of each gray level 'g' in column (i)
return (**C[c,g]**)
end **for**
end **if**
if $\theta \leftarrow 45^0$ then
for each i ← 1 : 2^r-1 do
C ← frequency count of each gray level 'g' in right diagonal(i)
return (**C[2^r-1, g]**)
end **for**
end **if**
if $\theta \leftarrow 135^0$ then
for each i ← 1 : 2^r-1 do
C ← frequency count of each gray level 'g' in left diagonal(i)
return (**C[2^r-1,g]**)
end **for**
end **if**
end **for**
end **procedure**

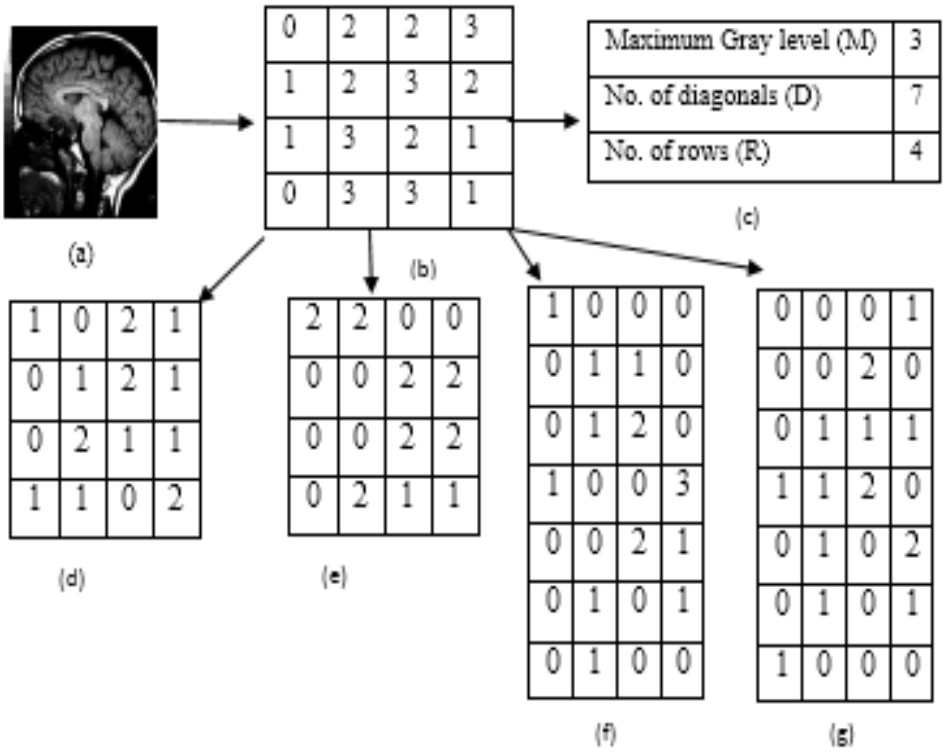


Figure 4.1: Example of calculating CLOM.

In this thesis, eight features are considered for extracting textural information of segmented brain tumor ROIs. The following notations are used to extract the features from CLOM. Let *P(i,j)* is the pixel at position *(i,j)* and *N* represents the total number of pixels in ROI. These features are considered as per formula given below [142]:

1. *Mean:* It is defined as the sum of the intensity values of pixels divided by the total number of pixels in the ROI of an image. It is given as:

$$m = \frac{1}{N} \sum_{i,j=0}^{g-1} p(i,j) \tag{4.1}$$

2. *Standard Deviation:* It describes the distribution of gray level value (intensity value) around the mean.

$$std = \sqrt{\frac{\sum_{i,j=0}^{g-1} (p(i,j) - m)^2}{N}} \tag{4.2}$$

3. *Entropy:* It represents the amount of randomness in intensity distribution of

an image. It is the measure of the disorder in an image.

$$Entropy = -\sum_{i,j=0}^{g-1} p(i,j)log(p(i,j)) \tag{4.3}$$

4. *Contrast:* It is defined as a measure of sudden change in intensity values of an image.

$$Contrast = \sum_{i,j=0}^{g-1} (|i-j|)^2 . p(i,j) \tag{4.4}$$

5. *Correlation:* a measure of correlation of a pixel to its neighbor pixel within the selected ROI is given as:

$$Corr = \frac{\sum_{i,j=0}^{g-1}(i*j)p(i,j) - (\mu_x.\mu_y)}{\sigma_x.\sigma_y} \tag{4.5}$$

$$where, \mu_x = \frac{1}{N}\sum_{i=0}^{g-1} i \sum_{j=0}^{g-1} p(i,j)$$

$$\mu_y = \frac{1}{N}\sum_{j=0}^{g-1} j \sum_{i=0}^{g-1} p(i,j)$$

$$\sigma_x = \sum_{i=0}^{g-1}(i-\mu_x) \sum_{j=0}^{g-1} p(i,j)$$

$$\sigma_y = \sum_{j=0}^{g-1}(j-\mu_y) \sum_{i=0}^{g-1} p(i,j)$$

6. *Homogeneity:* It measures the closeness of the distribution of elements in the CLOM.

$$H = \sum_{i,j=0}^{g-1} \frac{p(i,j)}{(1+|i-j|)} \tag{4.6}$$

7. *Energy:* It measures the textural uniformity of an image. It gives the sum of squared elements in the CLOM. The more distributed is the CLOM, lower is its energy.

$$E = \sum_{i,j=0}^{g-1} p(i,j)^2 \tag{4.7}$$

8. *Sum of square: Variance* It is given as the

$$S\,var = \sum_{i,j=0}^{g-1} (i-\mu)^2 . p(i,j) \tag{4.8}$$

4.1.2 *Texture Occurrence Matrix* (TOM)

The second proposed feature extraction algorithm in the thesis is *Texture Occurrence Matrix* (TOM). TOM is used to extract the texture information based on the nine texture extraction objects. These nine texture objects are shown in Figure 4.2.

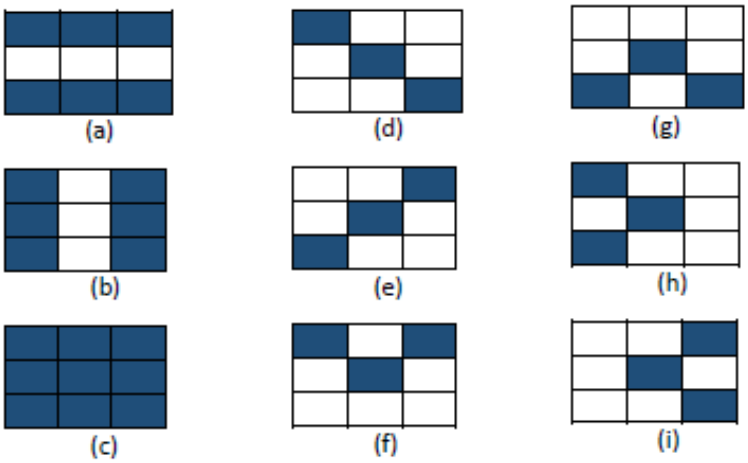


Figure 4.2: Nine texture structure objects used for texture analysis. (a) Horizontal extractor (b) Vertical extractor (c) Full block extractor (d) Anti-diagonal extractor (e) Diagonal extractor (f) Up block extractor (g) Down block extractor (h) Left block extractor (i) Right block extractor.

These texture extraction objects help to extract the texture information from the images. The block diagram of the proposed algorithm is shown in Figure 4.3.

In TOM algorithm, MR image is pre-processed to segment the ROI using low pass and high pass filters. Low pass filter truncates the high-intensity components of ROI and pass low-intensity values that smooth the ROI image. The tumor part in the resultant image is smooth enough to extract the soft tissues. On the other hand, high pass filter allows the high-intensity values to pass and results in sharpened ROI image.

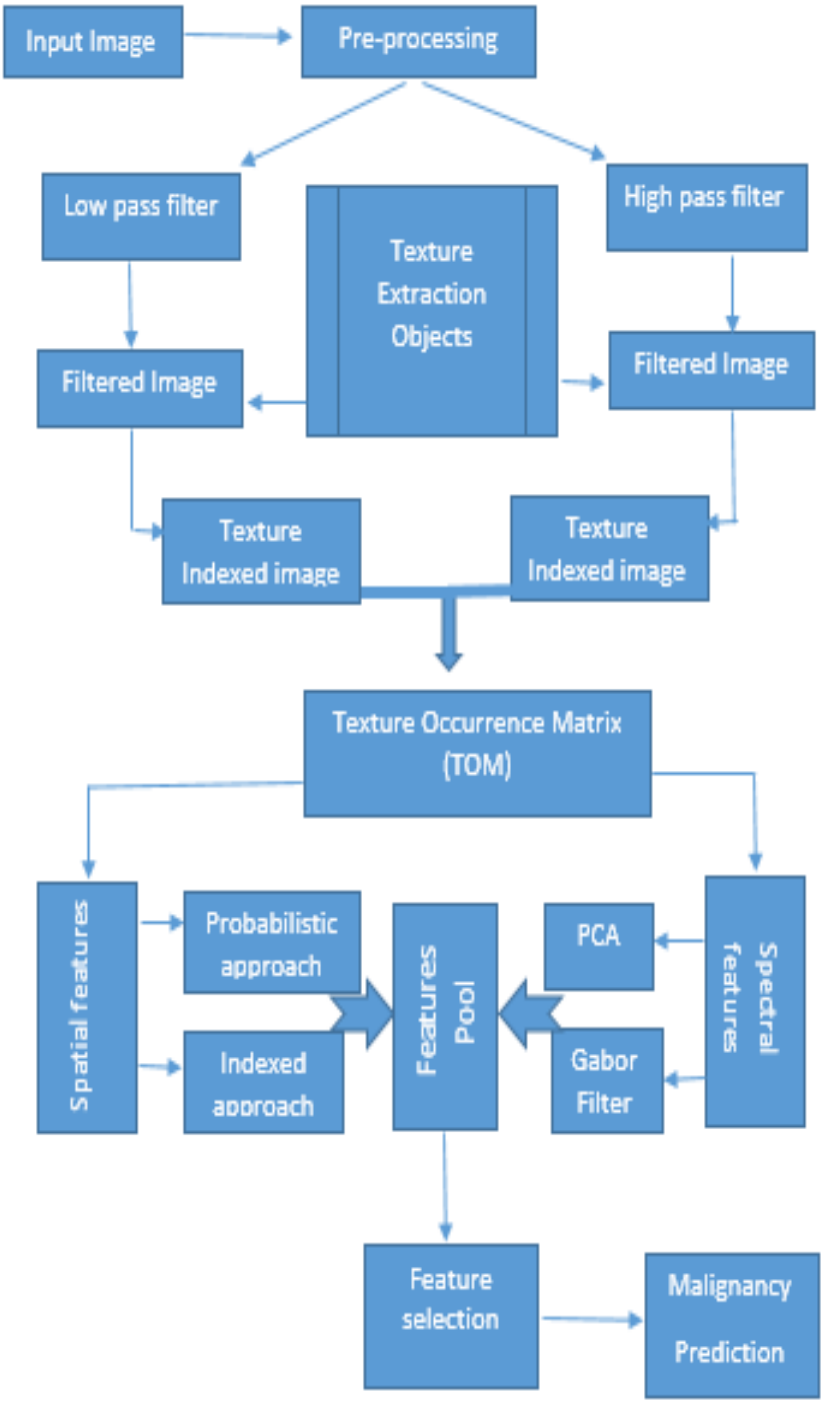


Figure 4.3: Basic building architecture of proposed TOM algorithm.

The resultant sharp image has better visualization of minute edges of and around tumor regions. These two variant images of ROI are processed simultaneously to get the common texture information present in both images and generate TOM. On these filtered images, texture extraction objects are masked to get the desired texture index value for the formation of TOM. In order to perform the masking between

filtered images and texture objects, the filtered image is divided into the blocks of 3x3 as of the size of texture objects. All the texture objects are sequentially rolled over filter image blocks to perform the masking which extracts the index of the texture object. The extracted texture position is stored in a matrix whose size is determined by the size of texture objects i.e. if the size of the original image matrix *(M)* is defined by $M = [..]_{RxC}$ and texture objects (*TO*) by $TO = [..]_{R'xC'}$ then the size of the positional matrix is $[..]_{\frac{R}{R'}x\frac{C}{C'}}$.

After extracting and representing the textural position in positional matrix, the positions where no textures are formed is indicated by 0. Next, the texture indexed matrix is formed by replacing the texture indexed by its value. The generation of the indexed matrix is illustrative by the following example as shown in Figure 4.4.

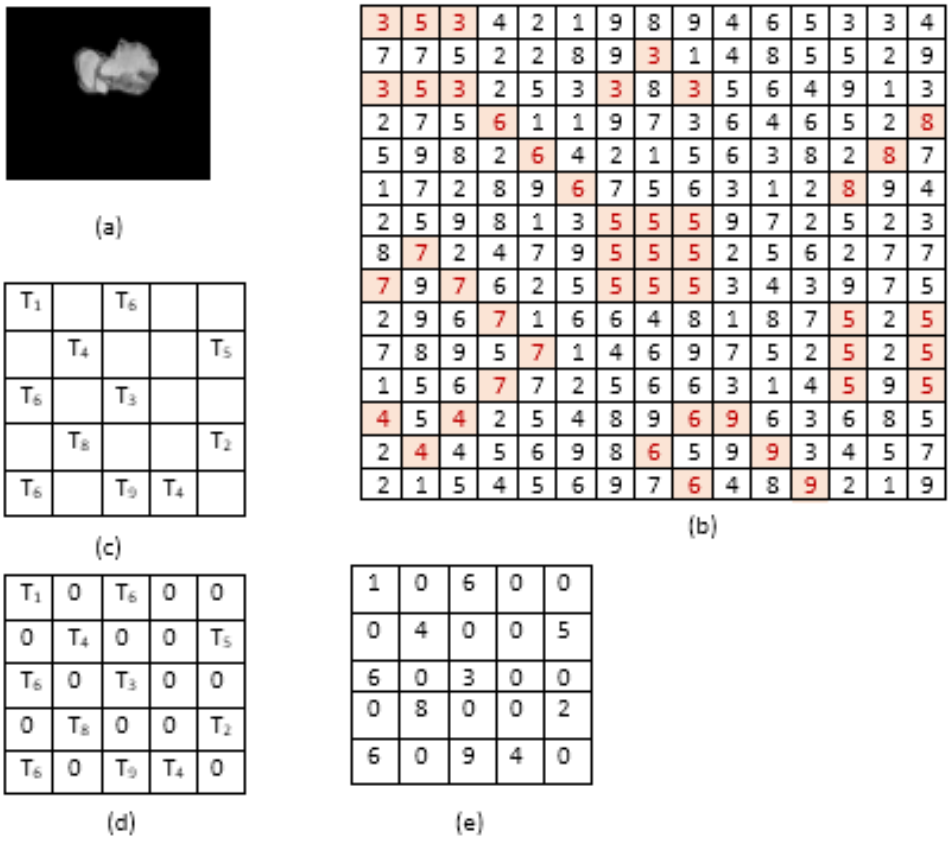


Figure 4.4: Working model for representation of extracted textures (a) Input ROI image (b) Result of extracted textures from input image in spatial domain (c) Texture extracted positional matrix (d) Intermediate Texture represented matrix (e) Final texture indexed matrix

Let us suppose that the two indexed matrices formed using high pass and low pass filter is shown in Figure 4.5 (a) and 4.5 (b). These formed indexed matrices are

used further to generate TOM as shown in Figure 4.5 (c)-(e).

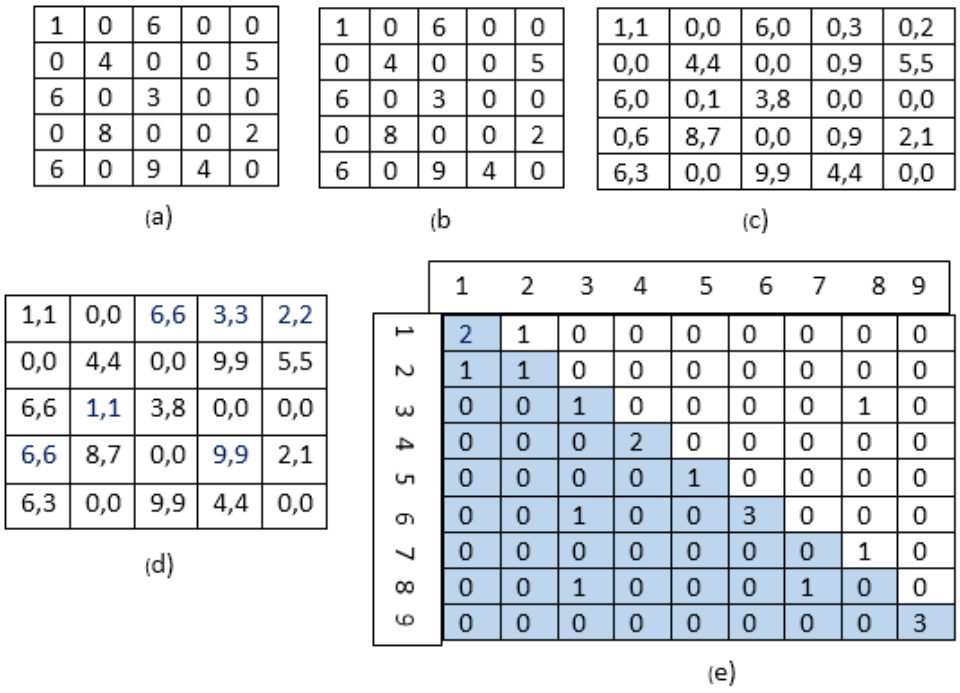


Figure 4.5: Formation of TOM (a) and (b) Matrices obtained using ROI preprocessing via texture objects. (c) An intermediate texture matrix. (d) Texture formulation matrix. (e) Final texture Occurrence Matrix.

An important thing in TOM is the formation of the set having an empty texture like 0, 6. It represents that the texture number 6 is only present in joined matrices. Similarly, the set formed by similar texture objects like 6, 6 represents the similar meaning as above. These issues are resolved by replacing such sets by uniformity. The uniformity is maintained by replacing the formed set 0, 6 to 6, 6 (shown in figure 4.5 (d)) as both means the same with respect to texture object. Finally, the TOM is generated using the texture formulation matrix. TOM is a square matrix of size equal to the number of texture objects, except null texture, which consists the values equal to the number of times a particular combination of texture objects encountered in formulation matrix. It is given by:

$$TOM(i, j) = N_{(i,j)} \tag{4.9}$$

Where, i and j are one of the nine texture objects, $N_{(i,j)}$ represents the count of the occurrence of the texture object $\{i, j\}$ in the formulation matrix.

To extract the textural information from TOM, several features are extracted

from the spatial and spectral domain. These features are discussed as:

4.1.2.1 Probabilistic Sum (PS) algorithm

To extract the spatial features from TOM, a new method named *Probabilistic Sum* (PS) algorithm is proposed. PS use the probability theory to find the probability value of each texture object from TOM. These probability values are finally added up to get the probabilistic sum value for each texture object. The concept is given by:

$$TextureSum(k) = \sum_{i=k}^{9} V(i,k); \quad wherek \in [1,2,...,9] \tag{4.10}$$

$$PS(i) = \frac{TextureSum(i)}{\sum_i TextureSum}$$

where, *V(i,k)* is the value indexed in TOM at position *(i,k)*. It is noticed that in proposed approach the size of *N* is the number of elements in symmetric matrix about its diagonal. The idea of PS approach is shown in Figure 4.6. The number of features added to feature set pool are same as the number of texture objects taken into consideration. In proposed approach, the number of features is nine.

2	1	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0
0	0	1	0	0	1	0	1	0
0	0	0	2	0	0	0	0	0
0	0	0	0	1	0	0	0	0
0	0	1	0	0	3	0	0	0
0	0	0	0	0	0	0	1	0
0	0	1	0	0	0	1	0	0
0	0	0	0	0	0	0	0	3

Texture count	Probabilistic Sum
3	0.1765
1	0.0588
3	0.1765
2	0.1176
1	0.0588
3	0.1765
1	0.0588
0	0.00
3	0.1765

Figure 4.6: Probabilistic Sum algorithm generating nine probabilistic features.

4.1.2.2 Indexed Approach

TOM indicates the spatial information of an image with the help of texture objects indexed value. Using the above matrix, eight textural features are extracted based

on the information present in TOM. Let the value of the TOM at specific index is represented by *V(i,j)* and *N* be the number of texture objects used in TOM. The $'\mu'$ is the mean value of *V*, extracted features are shown in Table 4.1.

Table 4.1: Number of features extracted from texture indexed matrix

Feature Name	Feature Definition
Second Order Angular Moment	$\sum_{i,j=1}^{N}(V(i,j))^2$
Contrast	$\sum_{k=1}^{N} k^2\{\sum_{i=1}^{N}\sum_{j=1}^{N} V(i,j)\}, \lvert i-j\rvert = k$
Inverse Difference Moment	$\sum_{i=1}^{N}\sum_{j=1}^{N}(\frac{1}{1+(i-j)^2}).V(i,j)$
Entropy	$-\sum_{i=1}^{N}\sum_{j=1}^{N} V(i,j) * \log(V(i,j))$
Variance	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i-\mu)^2 * V(i,j)$
Inertia	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2 * V(i,j)$
Cluster Shape	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i+j-\mu_x-\mu_y)^2 * V(i,j)$ Where, $\mu_x = \sum_{i=1}^{N} i \sum_{j=1}^{N} V(i,j)$ $\mu_y = \sum_{i=1}^{N}\sum_{j=1}^{N} j * V(i,j)$

4.1.2.3 Spectral features using Principal Component Analysis

Another domain of extracting features from images is the spectral domain. In the spectral domain, the spatial information of the image is not used directly to extract the features but the mapping of spatial information to some vector space is perceived. This vector space is used to represent the image information in terms of vectors, which is used as a feature for image classification. One such technique is *Principal Component Analysis* (PCA) [101] that transforms the multi dimensional data vector to one dimensional vector.

In the proposed algorithm, PCA is applied on TOM to extract spectral features. PCA maps the spatial information of TOM to vector space where information is represented by Eigen values and Eigen vectors. These vectors are considered as features which represent the image information. The size of the feature space is based on the number of Eigen-vector used to represent information. The dependency of Eigen-vectors is on the size of the matrix which is mapped from spatial domain to spectral domain. In proposed approach, the size of the TOM is based on the number of texture objects used to represent information. Thus using proposed approach with PCA, nine spectral features are extracted and added to feature space.

4.1.2.4 Gabor filter based spectral features

In the spectral domain, Gabor filter-based feature extraction is considered as an informative feature extraction mechanism [143]. It is believed that the frequency and orientation plays a significant role in representing textural information and its discrimination. In image analysis, an image is processed by Gabor filter which results in the generation of Gabor descriptors for that image. The Gabor descriptor is defined by:

$$G_{\theta,\lambda,\sigma,\varphi}(x,y) = \exp(-\frac{A^2+\gamma^2B^2}{2\sigma^2})\cos(\frac{2\pi}{\lambda}A+\varphi) \tag{4.11}$$

where, $A = x\cos(\theta) + y\sin(\theta)$, $B = -x\sin(\theta) + y\cos(\theta)$

In the above equation, λ represents the wavelength of the sinusoidal factor, θ represents the orientation of the Gabor function, φ is the phase offset, σ is the standard deviation of the Gaussian factor and γ is the spatial aspect ratio to specify the ellipticity of Gabor function.

To calculate the texture feature from an image the outputs of the symmetric (φ=0) and anti-symmetric ($\varphi=\frac{\pi}{2}$) Gabor kernel are combined using the distance metric and 2D linear convolution. The mathematical model of such is given by:

$$G_{\lambda,\theta}(x,y) = \sqrt{(I(x,y)\otimes g_{\lambda,\theta,\varphi=0}(x,y))^2 + (I(x,y)\otimes g_{\lambda,\theta,\varphi=\frac{\pi}{2}}(x,y))^2} \tag{4.12}$$

Textural features of an image are calculated by using a set of Gabor filters with different frequencies and orientations. Here five wavelengths i.e. $\lambda = \{2\sqrt{2}, 4, 4\sqrt{2}, 8, 8\sqrt{2}\}$ and four orientations in $[0,\pi)$ i.e. $\theta = \{0^0, 45^0, 90^0, 135^0\}$ which are taken at an equal interval of are used in proposed approach. The special aspect ratio (γ =1) is selected for computation. Thus for a single wavelength factor, four textural features for an image are extracted. Hence the total feature set obtained is of size 20. In proposed approach, two iterative values of standard deviation i.e. $\sigma = \{1.5, 2.5\}$ are used and finally the obtained number of features are 40 in total.

4.1.2.5 Summary of extracted features

The summary of overall extracted features from MR images are shown in Table 4.2. In total, there are 65 features which are present in our feature set to represent an image.

Table 4.2: Summary of extracted features

Feature Extraction approach	Number of extracted features
Probabilistic Sum approach	9
Indexed approach	7
Principal Component Analysis	9
Gabor filter	20*2 = 40
Total =	65

4.2 Features from spectral domain

Frequency or signal strength is another format to store the image in digital systems. The signal values are used to store and reconstruction of the image that makes signal processing as an important factor for feature analysis. At the first step, the input malignant image is mapped to signal domain using *Discrete Wavelet Transform* (DWT) [144]. DWT provides sufficient information both for analysis and synthesis of the original signal, with a significant reduction in the computation time. To use DWT for brain MR images, 2D variant of analysis and synthesis filter bank is implemented that results in forming an image in four bands i.e., LL, LH, HL, HH as shown below:

To extract the features from the 2 level DWT decomposition, the DWT coefficients of all four bands are taken into consideration. The nine spectral features are extracted from each of the four frequency bands of the DWT coefficient. The extracted features are mean, variance, average of energy, average of frequency, max of amplitude, min of amplitude, max of energy, min of energy and half of energy. These features are computed using formula given below [142, 145]:

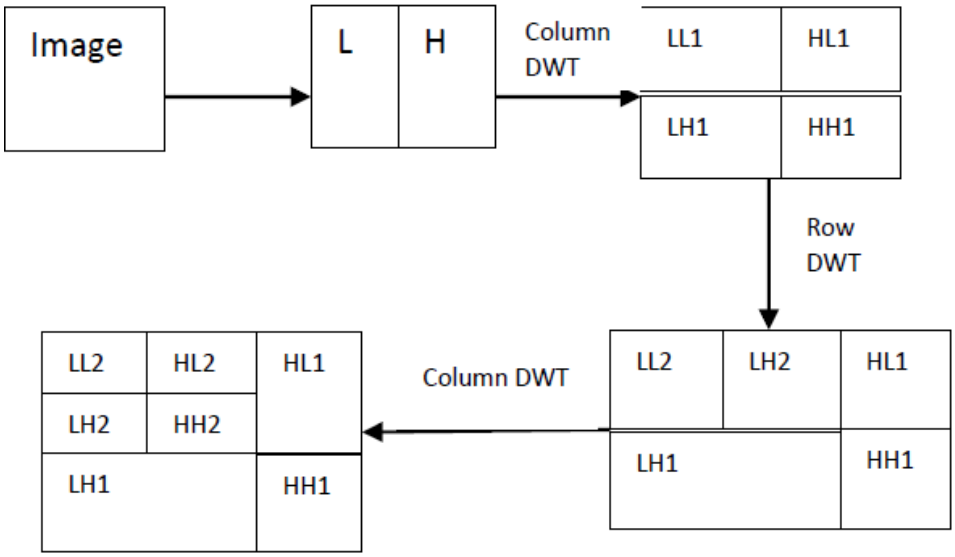


Figure 4.7: 2 Level DWT Block Diagram

1. *Mean:* It is the average of the value computed by applying DWT:

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4.13}$$

where, x_i is the DWT coefficient and 'n' is total number of DWT coefficients

2. *Variance:* It is defined as the sum of squared difference of each term in the distribution to the mean, divided by the number of terms in the distribution:

$$var = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m)^2 \tag{4.14}$$

3. *Average of energy:* It is defined as the mean value of the energy:

$$Avg_E = \frac{1}{n}\sum_{i=1}^{n}(x_i)^2 \tag{4.15}$$

4. *Average of frequency:* It is defined as the mean value of the frequency:

$$Avg_f = \frac{\sum_{i=1}^{n} f_i.x.p_i}{\sum_{i=1}^{n} p_i} \tag{4.16}$$

where, p = Power spectral density, f = Frequency vector [146]

5. *Max of amplitude:* It is defined as the highest amplitude value of the signal.

6. *Min of amplitude:* It is defined as the lowest amplitude value of the signal.
7. *Max of energy:* It is defined as the maximum value of the energy in the signal.
8. *Min of energy:* It is defined as the minimum value of the energy in the signal.
9. *Half of energy:* It is defined as the frequency that partition the spectrum into two regions having the same area.

4.3 Feature Selection

Feature selection, also known as *variable selection, subset selection or attribute selection. Feature selection* is defined as a process of selecting a relevant set of features or attributes from the large feature set vector which is used for building a machine learning model. A general feature set can have some features those are irrelevant or redundant in nature. These features have very less importance in machine learning models. Although it is a major concern to identify the features those are irrelevant in machine learning.

To filter out the redundant and irrelevant features, feature selection algorithms are divided into two parts i.e. (i) *feature ranking methods* and (ii) *feature subset selection method* [147].

4.3.1 *Feature subset selection using Cumulative Variance Method* (CVM)

To select the relevant and informative features subset, an approach is proposed namely *Cumulative Variance Method* (CVM). It is a statistic based approach that initially transformed the original features value to new values. The transformed value are used to compute the variance and cumulative variance for subset selection. The algorithm of CVM for subset selection is given below in algorithm 5:

CVM retrieve the relevant subset of the features which has the maximum contribution to decision making, on basis of statistical T-test. CVM finds the variance among the extracted Eigen-vectors for individual features and the subset is selected

Algorithm 5: Feature subset selection with CVM

Input: A feature vector having n features
Output: A subset of features having m features and m ≤ n

Procedure:
1. **for** each vector $v_i \in$ V **do**
2. Compute the mean of each feature vector, $\mu_i = \frac{(\sum_{j=1}^{n} v_i)}{n}$
3. Find the deviation of each feature instance from mean, $D_i = v_i - \mu_i$
4. end **for**
5. Compute the matrix, $C_{[}pxp] = D^T.D$; where,p = no.of features
6. Transformed the data from lower order to higher order using Eigen vector, [C - λI][u] = 0
7. Find the variance by using, $var_{(}\mu_{i)} = \frac{\sum_{i=1}^{n}(\mu_i - \bar{\mu_i})^2}{n-1}$
8. Sort the features on basis of variance in decreasing order of magnitude
9. Initialize variable Cumvar to zero i.e, Cumvar = 0
10. **for** each feature i = 1 to n do
11. Cumvar = Cumvar + var_i
12. end **for**
13. Apply T-Test to filter the subset having confidence interval of 99%.
14. Return the selected Subset
15. end **Procedure**

on the basis of the cumulative variance. The selected confidence level is the 99% [101] which is used as the threshold.

4.3.2 Hybrid algorithm for feature selection

The hybrid algorithm of feature selection uses both the univariate and multivariate feature selection algorithms. It is a two step procedure in which dimensions of the features are reduced by the univariate algorithm based feature selection. Further, the reduced feature set is optimized using the multivariate method. In this thesis, two such algorithms are used named *Fisher Discriminant Ratio* (FDR) [148], a univariate algorithm and *minimum Redundancy Maximum Relevance* (mRMR) [149], a multivariate algorithm.

Fisher Discriminant is an effective algorithm for dimension reduction in machine learning. The key concept behind the *Fisher Discriminant* is to find a line in the sample space, which separates the projection of the sample points by some point on the line. Mathematically it is defined as, better be the projection line at any point in time the value of the square of the difference between the means of sample

points is larger and the within-class scatter values is smaller. It is given by:

$$J_F(W) = \frac{(\bar{m_1} - \bar{m_2})^2}{(\bar{S_1})^2 + (\bar{S_2})^2} \tag{4.17}$$

where W is the direction projection, $\bar{m_i}$ and $(\bar{S_i})^2$ are mean and *within class scatter matrix* of class i where $i = 1, 2$. Thus it is desired to find the value of W at which $J_F(W)$ is maximum.

In the case where there are more than two classes, the analysis used in the derivation of the Fisher discriminant can be extended to find a subspace which appears to contain all of the class variability. Let us suppose that each of C classes has a mean μ_i and the same covariance Σ . Then the scatter between class variability may be defined by the sample covariance of the class means

$$\Sigma_b = \frac{1}{C}\sum_{i=1}^{C}(\mu_i - \mu)(\mu_i - \mu)^T \tag{4.18}$$

$$\Sigma_w = \Sigma_1 + \Sigma_2 + ... + \Sigma_c$$

where μ is the mean of the class and Σ_i is the covariance of class i. The equation (4.17) can be rewritten as:

$$J_F(W) = \frac{\vec{w}^T \Sigma_b \vec{w}}{\vec{w}^T \Sigma_w \vec{w}} \tag{4.19}$$

This means that when $\vec{w}$ is an eigenvector of $\Sigma^{-1}\Sigma_b$ the separation will be equal to the corresponding eigenvalue.

On the other hand, mRMR is a multivariate approach that focuses on minimizing the redundancy among features and select the features that are mutually maximally dissimilar to other features.

Let s denote the subset of features than the average minimum redundancy is given by:

$$min = \frac{1}{|s|^2}\sum_{f_i, f_j \in s} I(f_i, f_j) \tag{4.20}$$

where $I(f_i, f_j)$ presents the mutual information between the i^{th} and j^{th} features and $|s|$ is the number of features in S.

To compute the maximum relevance condition, it is to maximize the average relevance of all features in and it is given by:

$$max = \frac{1}{|s|} \sum_{f_i \in s} I(H_i, f_j) \tag{4.21}$$

where $I(H_i, f_j)$ is the mutual information between the target class H_i and features f_j.

The above two conditions suggested that the redundancy among the features must be minimum while the relevancy is to be the maximum. The combination of these two conditions are suggested as a single dimension reduction technique named mRMR as given by the following equation:

$$mRMR = max\{I(H_i, f_j) - \frac{1}{|s|} \sum_{f_i, f_j \in s} I(f_i, f_j)\} \tag{4.22}$$

4.4 Dataset Description

In this thesis, the proposed feature extraction algorithms experiment with brain MR images having abnormalities. The dataset of malignant brain tumors having five classes are experimented. The description of the dataset is presented in Table 4.3. The experimental dataset is acquired by using 3.0 T GE scanner (*General Electronics Company, Milwaukee, WI*) from Department of Radiology, *Sawai Man Singh* (SMS) Medical College Jaipur, Rajasthan, INDIA. All the patients images that are used for dataset are scanned using same GE scanner and environment variables. The obtained voxels of the head slice have following specifications: 3D weighted voxels of *T1, T2, eT1, eT2* and *Fluid Attenuated Inversion Recovery* (FLAIR) each of having size 256x256x3 in 8-bit DICOM format. We have used T1-weighted post contrast-enhanced axial images with dimension of 256x256x3.

Table 4.3: Dataset description of malignant tumors types and their sample images

Tumor Type	Number of samples
Central Neuro Cytoma (CNC)	133
Glioblastoma Multiforme (GBM)	160
Gliomas (GLI)	155
Intra Ventricular Malignant Mass (IVMM)	152
Metastasis (MTS)	160

4.5 Performance Evaluation Metrics

To evaluate the performance of the classifier there are certain parameters through which classification results are validated. These parameters are based on the two parameters i.e. the pre known class label and the class label predicted by the classifier. These measures are shown in Table 4.4.

Table 4.4: Measures for predicting Classification performance

True Positive (TP)	*False Positive* (FP)	***Precision*** $= \frac{TP}{(TP+FP)}$
False Negative (FN)	*True Negative* (TN)	***Negative Predict value*** $= \frac{TN}{(FN+TN)}$
Sensitivity$= \frac{TP}{(TP+FN)}$	***Specificity*** $= \frac{TN}{(FP+TN)}$	***Accuracy*** $= \frac{(TP+TN)}{Total}$

4.6 Results and Discussion

Experiment 1: The experimentation of the proposed CLOM feature extraction algorithm is validated with the help of two classifiers in machine learning. The dimension reduction of the extracted features using CLOM is computed by using the CVM algorithm. Two classifiers, *K-Nearest Neighbor* (KNN) and *Support Vector Machine using Radial Basis Kernel function* (SVM-RBK) are used to find the accuracy of the system. The experimentation results of the malignant tumor dataset with CLOM, CVM and KNN are presented in Table 4.5. The results are shown at k = 7 nearest neighbor. The reason is, during experimentation with KNN, several values of k, i.e., *k = (1, 3, 5, 7, 9)* are tested but the highest classification accuracy gained is at k=7. The results of the proposed method are compared with state-of-art feature selection methods such as *GLCM* and *Run-length matrix*. It is observed that the average accuracy of the CLOM for malignant tumor dataset is about 86.71% which is better than the *GLCM* (85.68%) by a factor of 1.20% and *Run-length* (79.31%) by a factor of 9.33%. The results are presented in Table 4.5.

The next experimentation is performed with the SVM classifier with radial basis kernel function. The experimentation result of the SVM-RBK is presented in Table 4.6. The result shows that the proposed CLOM algorithm with CVM method as feature selection gives classification accuracy of 89.94% as compared with the state-of-art *GLCM* (87.64%) and *Run-length* (84.48%). The experimental results show

Table 4.5: Classification accuracy of CLOM with KNN classifier at K = 7

MR image	CLOM(%)	GLCM(%)	Run Length(%)
Central Neuro Cytoma	85.2	84.67	80.10
Glioblastoma Multiforme	87.77	86.87	79.97
Gliomas	89.25	86.38	76.73
Intra Ventricular Malignant Mass	84.8	85.15	79.24
Metastasis	86.54	85.34	80.55
Avg. Accuracy	86.71	85.68	79.31

that CLOM is performing better than GLCM and Run-length by 2.62% and 6.46%.

Table 4.6: Classification accuracy of CLOM with SVM classifier

MR image	CLOM(%)	GLCM(%)	Run Length(%)
Central Neuro Cytoma	89.2	86.7	83.1
Glioblastoma Multiforme	88.7	86.8	82
Gliomas	91.3	88.3	86.8
Intra Ventricular Malignant Mass	90	87.6	84
Metastasis	90.5	88.8	86.5
Avg. Accuracy	89.94	87.64	84.48

The experimental results conclude that the proposed CLOM algorithm provides more textural information as compared with other feature extraction algorithms. The KNN classifier and kernel based SVM classifier provides significant classification accuracy for distinguishing different classes of malignant brain tumors.

Experiment 2:The second experiment is with the proposed feature extraction algorithm named TOM. In the experiment, textural features are extracted from the TOM and rest all the parameters remains same as experiment 1. The classification results of the TOM with CVM as feature selection and KNN as the classifier is presented in Table 4.7. While the results with SVM-RBK is presented in Table 4.8.

In the experimentation, it is found that the TOM based textural features provides much more textural information as the average classification accuracy gained is about 91% with KNN and 93% with SVM. In the comparative analysis, it is

Table 4.7: Classification accuracy of TOM with KNN classifier at k=7

MR image	TOM(%)	GLCM(%)	Run Length(%)
Central Neuro Cytoma	90.97	84.67	80.10
Glioblastoma Multiforme	91.25	86.87	79.97
Gliomas	90.32	86.38	76.73
Intra Ventricular Malignant Mass	90.78	85.15	79.24
Metastasis	91.87	85.34	80.55
Avg. Accuracy	91.03	85.68	79.31

Table 4.8: Classification accuracy of TOM with SVM classifier

MR image	TOM(%)	GLCM(%)	Run Length(%)
Central Neuro Cytoma	92.48	86.7	83.1
Glioblastoma Multiforme	93.12	86.8	82
Gliomas	92.25	88.3	86.8
Intra Ventricular Malignant Mass	93.42	87.6	84
Metastasis	93.75	88.8	86.5
Avg. Accuracy	93	87.64	84.48

observed that the TOM produces better results as compared with GLCM and Run-length matrix. The increase in the classification accuracy for TOM with state-of-art approaches is about 6.24% and 14.77%.

Experiment 3: The third experiment is conducted with the DWT based features and the fusion of the features from the spatial and spectral domain. In this experiment, the relevant features are selected from the two algorithms named CVM and hybrid algorithm (FDR + mRMR). All the experimental variables remain the same as in the previous ones. The performance of the classifier's using the variable features set is shown in Table 4.9 concludes that the highest accuracy achieved is 97.28% when the feature set having TOM based textural features are taken into consideration with hybrid feature selection approach and SVM-RBK classifier. The experimental results of proposed approaches are found satisfactory as compared with other feature extraction approaches and with both classifiers.

Experiment 4: In this experiment, the performance of the CVM and hybrid feature selection approach is compared with the state-of-art *Genetic algorithm* and *Independent Component Analysis* (ICA). For experimentation, the features are extracted using TOM by which the highest classification accuracy is achieved in the

Table 4.9: Average classification accuracy of different feature extraction algorithms with proposed dimension reduction techniques

	KNN		SVM	
	CVM (%)	Hybrid (%)	CVM (%)	Hybrid (%)
DWT	76.27	78.66	79.12	82.85
Run-length	81.67	83.33	85.50	87.66
GLCM	87.58	89.97	89.93	91.71
CLOM	88.33	89.97	91.64	93.43
TOM	91.03	93.34	93.00	97.28

previous experiment. The experimental results are presented in Table 4.10. The summarized results of Table 4.10 conclude that the hybrid algorithm of feature selection provides the satisfactory results with both KNN and SVM-RBK classifiers. While the other approach i.e. CVM also provides improved classification accuracies as compared with the Genetic algorithm and Independent Component Analysis algorithm.

Table 4.10: Classification accuracy of TOM features with different dimensional reduction techniques and classifiers

Dimension Reduction	**KNN (%)**	**SVM (%)**
Hybrid method	93.34	97.28
CVM	91.03	93.00
Genetic Algorithm	85.07	90.52
Independent Component Analysis	82.87	87.88

Comparative Analysis of CLOM and TOM

The comparative analysis of all the experiments is presented in Table 4.11. The experimental result suggests that among the two proposed feature extraction algorithms, TOM provides the better description of the texture information as compared with CLOM. The reason of outperformance of TOM is, it uses the nine texture filter objects to generate the texture matrix. Using texture matrix, 65 features are extracted from spatial and spectral domain. While using CLOM, only 32 features are extracted based on four different orientations. The experimental results shown in Table 4.11 concludes that using the triplet combination of TOM as feature extraction, hybrid algorithm as dimension reduction and SVM as classifier gains the highest accuracy of 97.28% for classification of the malignant tumor dataset.

Table 4.11: Overall comparative classification accuracy analysis of proposed algorithms

	KNN		SVM	
	CVM %	**Hybrid %**	**CVM %**	**Hybrid %**
CLOM	86.71	89.97	89.94	93.43
TOM	91.03	93.34	93	**97.28**

4.7 Statistical Validation

To validate the proposed feature extraction algorithm CLOM and TOM, a statistical validation test named *T-test* is used [142, 145]. *T-test* in statistics, is based on the set up a hypothesis to test whether a given sample mean is close enough to a population mean or not. *T-test* determines, test and concludes the difference between the sample mean and the population mean is significant enough to make decision. *T-test* is used when only the sample standard deviation is known. The *T-test* includes four main steps.

1. Define the null and alternate hypotheses,
2. Calculate the t-statistic for the data,
3. Compare t_{calc} to the tabulated t_{value}, for the appropriate significance level and degree of freedom.
4. If $t_{calc} > t_{value}$, we reject the null hypothesis and accept the alternate hypothesis. Otherwise, we accept the null hypothesis.

The mathematical description of *T-test* for two sample having unequal variance is given as,

$$t = \frac{\bar{\mu_1} - \bar{\mu_2}}{\sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}} \tag{4.23}$$

Where, μ_i is the mean of i^{th} sample, σ_i is the standard deviation for i^{th} sample, and n_i is number of observations in i^{th} sample.

Let the Null Hypothesis is H_0: $\mu = \mu_0$ i.e. both algorithms have same performance and alternate hypothesis H_A: $\mu > \mu_0$ i.e. algorithm A has better performance than algorithm B. In the statistical validation test of proposed algorithm,

Validation Test-1: CLOM with GLCM and Run legth

H_0: Assumed that CLOM and GLCM/Run length have same mean.

H_A: Assumed that CLOM has higher mean than GLCM/Run length.
The statistical validation with T-test of TOM and GLCM/Run length is given below in Table 4.12, 4.13:

Table 4.12: T-test result for CLOM and GLCM

	CLOM	**GLCM**
Mean	86.71	85.68
Variance	3.38307	0.83067
Observations	5	
Hypothesized Mean Difference	0	
degree of freedom	4	
t_{Stat}	2.776465	
t Critical two-tail	**1.94899**	

Table 4.13: T-test result for CLOM and *Run length*

	CLOM	**Run length**
Mean	86.71	79.31
Variance	1.189167	5.1225
Observations	5	
Hypothesized Mean Difference	0	
df	4	
t_{Stat}	8.397549	
t Critical two-tail	**3.182446**	

Validation Test-2: TOM with GLCM and Run legth H_0: Assumed that TOM and GLCM/Run length have same mean.
H_A: Assumed that TOM has higher mean than GLCM/Run length.
The statistical validation with T-test of TOM and GLCM/Run length is given below in Table 4.14, 4.15:

Table 4.14: T-test result for TOM and GLCM

	TOM	**GLCM**
Mean	91.038	85.682
Variance	0.33097	0.83067
Observations	5	
Hypothesized Mean Difference	0	
df	4	
t_{Stat}	11.11193754	
t Critical two-tail	**2.364624252**	

Table 4.15: T-test result for TOM and *Run length*

	TOM	**Run length**
Mean	91.038	79.318
Variance	0.33097	2.31457
Observations	5	
Hypothesized Mean Difference	0	
df	4	
t_{Stat}	16.11222239	
t Critical two-tail	**2.570581836**	

The validation test concludes that for a two-tail test inequality, if $t_{Stat} < -t_{Critical}$ two-tail or $t_{Stat} > t_{Critical}$ two-tail, then the null hypothesis is rejected and alternate hypothesis is selected.
In the experiment of T-test validation for proposed algorithms with both state-of-art algorithms, it is found that $t_{Stat} > t_{Critical}$, so we reject the H_0 and accept the H_A and conclude that CLOM and TOM has better classification performance than GLCM and Run length algorithms.

Validation test-3: CLOM and TOM algorithms

To test and validate the two proposed algorithms mutually, T-test is applied in between TOM and CLOM. The initialization of the hypothesis is given as: H_0: Assumed that CLOM and TOM have same mean.
H_A: Assumed that CLOM has higher mean than TOM.
The statistical validation with T-test of TOM and CLOM is given in Table 4.16: The

Table 4.16: T-test result between CLOM and TOM

	CLOM	**TOM**
Mean	86.71	91.03
Variance	1.189167	0.4143
Observations	5	
Hypothesized Mean Difference	0	
df	4	
t_{Stat}	-4.10325	
t Critical two-tail	**3.182446**	

validation result proved that the $t_{Stat} < -t_{Critical}$, which proved that the null hypothesis is rejected and alternate hypothesis is accepted. Subsequently, the negative value of t_{Stat} suggest that CLOM algorithm has lesser performance than TOM algorithm.

4.8 Summary of the chapter

The aim of the thesis is to design, implement, and evaluate a machine learning algorithm for discriminating between different classes of malignant brain tumors. This chapter introduces two new feature extraction algorithms named CLOM and TOM to extract the textural features from the segmented ROI. Later, two algorithms are discussed to select the relevant features subset and reduce the dimension of the feature vector. These algorithms are CVM and the hybrid algorithm that includes FDR and mRMR approaches. The experimentations are performed with the help of two classifiers i.e. KNN and SVM-RBK, a non-linear classifier. The experimentation results suggest that the fusion of the features from the spatial domain and spectral domain provides much more required information about the abnormality as compared with state-of-art feature extraction methods.

In the experiment 1, proposed CLOM based features are experimented with KNN and SVM classifiers. Using the eight textural features at each orientation, result it into thirty two features in total at 4 different orientation values. Further, these features are filtered using CVM algorithm and the selected features are feed into the classifiers as input. The experimentation results show the accuracy of 86.71% with KNN and 89.94% with SVM.

The second experiment is for TOM based features having 66 textural features in total. Using similar tune of the parameter setting, the classification accuracy of TOM is 91.03% with KNN and 93% with SVM. The overall increase in the comparative accuracy between CLOM and TOM is around 4% to 6%. The primary reason of such is that, TOM uses the filter objects based on texture information for generating texture matrix. Thus it holds more textural information as compared with CLOM.

The third experiment is a hybrid experiment where the features from two domains are concatenated to get hybrid textural information. Similarly, the two domains of feature selection algorithms, FDR and mRMR are merged to build hybrid dimensional reduction algorithm. This results in the enhancement of the classification accuracy for both the classifiers around 3% to 4%. The reason of such is, hybrid selection is a two phase process algorithm. In the first phase, univariate feature selection algorithm is applied to filter the extracted features followed by the multivariate feature selection algorithm. Thus the hybrid algorithm works similar to feature optimization.

Also, the proposed feature selection algorithms provide satisfactory results when

compared with other dimension reduction approaches in experiment 4. The highest average classification accuracy achieved is 97.28% with SVM classifier and 93.34% with KNN classifier.

The proposed feature extraction algorithms are also statistically validated by using the T-test method. Based on the validated results of T-test presented in Table 4.12, 4.13, 4.14 and 4.15, it is concluded that the proposed CLOM and TOM algorithms are better than the state-of-art feature extraction algorithms. Consequently, the validated result between two proposed methods, CLOM and TOM, it is found that the TOM algorithm is superior than CLOM algorithm based on T-test values as proved in Table 4.16.

In the next chapter, the advancement in the linear classification algorithm is proposed. The proposed algorithms helps in finding the optimal number of nearest neighbors without using any initial predefined threshold. The proposed algorithm is based on sample space reduction mechanism.

Chapter 5

Malignant Brain Tumor Classification with Variant of Nearest Neighbor Algorithm

Classifier is a machine learning algorithm that classifies examples into given set of classes. The gained classifier output can be in the form of a distinct value that indicates the one among the predefined class. The performance of the classifier is based on the stages of machine learning those are feature extraction and feature selection. To maximize the classifier performance, it is required to fine tune all the parameters of every stage of such system before the parameters are feed into the classifier. The performance of any classifier includes certain parametric stages like the *learning stage* of the classifier, *testing stage* of classifier and the evaluation of the classifier accuracy.

In the *learning phase*, the classifier is trained to recognize the label training examples based on the feature associated with sample and its class label. Using the feature vector, the classifier generates the prime description of category samples. The samples which are used in the learning phase are called *training set*. At the end of the learning, the classifier is prepared to predict the class label of the new data sample that still is undefined.

In the testing phase of the classifier, there exist some data samples which are still undefined. Based on the learning, the classifier predicts the class label of undefined samples. In the presence of the features extracted from the unknown samples, the classifier matches the closely related sample having similar or nearby similar characteristics. Thus the class label of the similar found sample is considered as the

class label to unidentified sample.

A linear classification approach based on the outcome of the closely related sample towards query sample is proposed in [125]. The approach uses the most likely sample as the reference, to make the prediction of the class label. Further, the concept is extended towards finding the number of samples having close resemblance to query sample. These closely found samples based on the distance similarity, helps in predicting the nominated class label of the new and unidentified test case. In turn, the decision about the test sample class is based on the maximum number of the samples having same class label in extracted closest samples. Subsequently, this approach resulted into the formation of an open research problem for finding the optimal number of close samples to query.

To find the optimal numeral of closely related samples, a firsthand mathematical algorithm is recommended that works on the iterative elimination of the samples. Rest of the chapter is organized as follows: In section 5.1, the proposed voting based mathematical algorithm is given with the description about its working. In section 5.2, the experimental results and discussion over the experimental work is presented. In the section 5.3, the statistical validation of the proposed algorithm is given using McNemar's test. Finally at the end in section 5.4, the overall summary of the chapter is presented.

5.1 A Voting based Novel Mathematical Algorithm

The proposed algorithm is used to find the ideal number of nearest samples from the whole dataset. The proposed algorithm identifies the number of close samples to query sample, by recursively eliminating the samples from the sample space.

In machine learning, a voting based novel mathematical algorithm is proposed which removes the barrier of choice of the number of closely link samples for classification. In literature several methods were presented that classifies the data instances by the pre initialized number of nearby neighbors. The main contribution of the proposed algorithm that differentiates it from existing algorithms is to provide the mechanism that finds the optimal number of nearby neighbors and also provides the weighted voting value that classifies the data instance. Proposed algorithm uses the space reduction approach for iteratively filter out the samples which are far enough to make any judgment in classification. Besides this, iterative dy-

namic threshold value is also computed by given algorithm based on its updated reduced sample space.

The algorithm of proposed mathematical algorithm is given below:

5.1.1 Description of the proposed algorithm

Proposed algorithm is reflex enough in finding the selection of the number of nearby samples to query sample. In addition, the proposed algorithm provides the strength by recursively eliminating the sample space based on the dynamic generation of the threshold value. The descriptive analysis of the proposed algorithm is given below:

Consider an sample space S having N data samples with their respective class labels and a query sample Q with unknown class as shown in Figure 5.1.

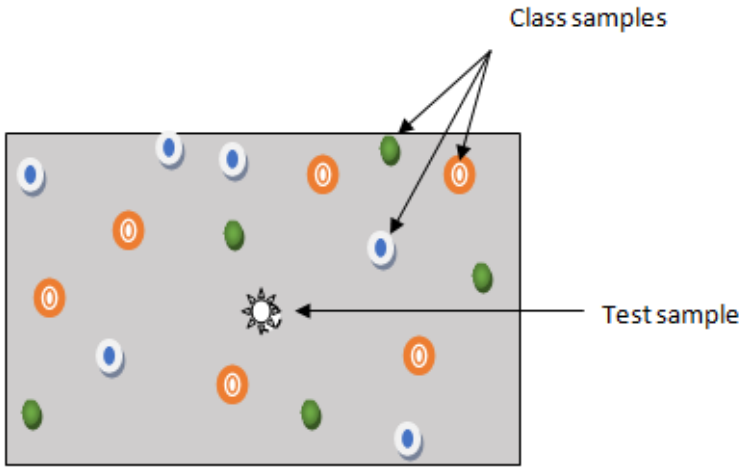


Figure 5.1: Original sample space S

In the above figure, it is shown that the sample space S has three known class samples data in the sample space with a test query sample. The goal is to find the class label of the query on behalf of the number of samples closely related to the query sample. The close resemblance between the samples and query is identified by distance computation. The minimum the distance formed between the samples, more similar resemblance samples they considered.

In the proposed algorithm, each of the samples is associated with the certain features and its respective class label as given by equation 5.1.

$$S = \{f_1, f_2, ..., f_n | c_i\} \tag{5.1}$$

Algorithm 6: A Voting based Novel Mathematical Rule

Input: TR: Training sample space, $\{S_i, c\}$, Where, i = $\{1, 2, 3, , N\}$; $S_i \in$ Training Samples; c ∈ class label; N ∈ no. of training samples;
x : Query vector/Test sample.

Step 1: Calculate sum of sample and query vector
for i = 1 to N
$SSum_i = \sum_k f_k$; where, $f_k \in$ feature in i[th] sample
//$SSum_i \in$ Summation of features for i[th] sample
end
QSum = $\sum_k f_k$; // Sum of query vector features

Step 2: Compute distance between all samples to query vector
for i = 1 to N
$dist_i = |SSum_i QSum|$
end

Step 3: Ranking closely related samples
$Rank_i$ = sort (dist); // sort in increasing order
// where, dist = $\{dist_1, dist_2, , dist_N\}$

Step 4: Feature Distance computation
$F_min = Rank_1$; // Fetch I[st] minimum sample
for i = 1 to k // for each k features
$SFmean_i$ = mean (f_k) // where, $SFmean_i$ is the mean value of i[th] feature for all samples
end

Step 5: Sample Space Reduction
// Generating threshold
While (! conserve or no elimination of samples) do
T = ($\sum SFmean_i$) - QSum
for i = 1 to N
if ($dist_i > T$)
eliminate S_i; // eliminate sample having distance ¿ T
// Where $S_i \in$ S; S is a sample set
update S;
end if
end for
goto: step 4
end while

Step 6: Fetching number of closest samples to Query
k = size (S) // leftover samples in sample set

Step 7: Weight Initialization
D_1 = dist($Rank_1$)
D_k = dist($Rank_k$)
for i = 1 to N // where N is no. of samples in updated set S'
$W_i = \frac{(D_k - D_i)}{(D_k - D_1)}$ // Assigning weight to each sample
end

Step 8: Weighted neighbor computation
for i = 1 to N
WS = $\sum W_{(ic)}$; // Where, c is sample class label
// fetching total weight for respective class
end

Step 9: Class label assignment
if (size (S) ==1)
Q_c = c(S) // Query class label = sample class label
else
Q_c = argmax ($\sum WS_k$) //WS_k is total weighted sum for class k

$$q = \{\bar{f}_1, \bar{f}_2, ..., \bar{f}_n\}$$

Where, for each sample, f_k is the associated feature value with its class label c_i. Furthermore, the proposed algorithm computes the feature sum value for one and all samples and query vector, as given by equation 5.2. Using the computed summation values of all features, the distance measurement is computed between every sample to query using equation 7.3.

$$SSum_i = \sum_k f_k \ and \ Qsum = \sum_k f_k \tag{5.2}$$

$$dist_i = |SSum_i \quad Qsum| \tag{5.3}$$

Based on the computed distance value, the proposed algorithm ranks the samples in accumulative order as given below:

$$Rank = sort(dist) \tag{5.4}$$

After samples ranking, the proposed algorithm identifies the dynamic threshold T that reduces the sample space by eliminating the samples from the sample space. The elimination of the samples is based on their extracted feature sum value. If any sample violates the algorithm discussed in step 4 of the given algorithm, then sample is deleted from space.

Subsequently, to find the threshold T, the average value of each feature in sample space is computed as given in step 4 of the algorithm. The computation of the threshold is given by the equation 5.5.

$$T = (\sum SFmean) - QSum_j \tag{5.5}$$

The computed threshold is further used to reduce the size of the original sample space, from S to S, as shown in Figure 5.2.

By using the iteratively recursive elimination of the samples from sample space, the size of the sample is reduced to S. At every recursive call of the proposed algorithm, the threshold is dynamically updated as elimination of the samples from space changes the feature sum mean value. The proposed algorithm iteratively changes the threshold, and as the threshold value converges, the elimination of the samples stops. Consequently, the first nearby sample to query in the reduced sample

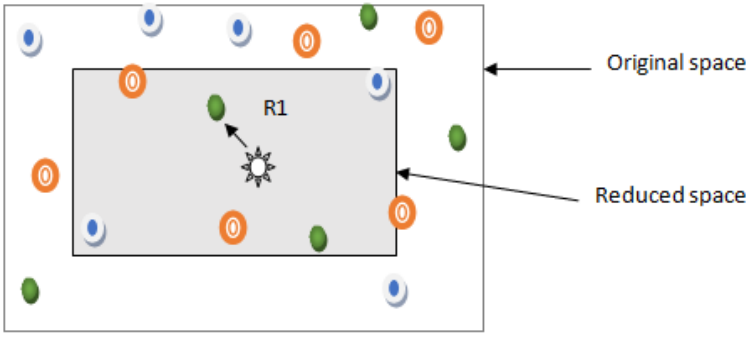


Figure 5.2: Reduced Sample space from original space

space is mapped with the ranked one as R1 as shown in Figure 5.2.

Subsequently, in the state of algorithm convergence, the number of sample points that remain the sample space S be the k nearby samples to the query sample. The reduced space S' consists of the required number of associated samples towards query vector. In turn, the proposed algorithm overcome the limitation of the pre-initialization required for the value of k as previously seen in classification models, KNN.

Next step of the proposed algorithm is to find the weighted voting for predicting the class label for unknown query sample Q. The weighted concept in the proposed algorithm is in such form that the most nearby sample towards query sample gets the more weightage while the farthest sample get least weight. As the most nearby sample has close resemblance to query, thus nearby sample get maximum possible weight.

In the reduced sample space there exists 'k' samples. The distance from query sample Q to the 'kth' sample is denoted by D_k while the most nearby sample to query sample is denoted by D_1. The weighting factor for each of the sample in sample space is given by equation 5.6.

$$W_i = \frac{(D_k - D_i)}{(D_k - D_1)} \tag{5.6}$$

It is noticed that the value of W_i is found as:

$$W_i = \begin{Bmatrix} 1; & i = 1 \\ 0 < W < 1; & 1 < i < k \\ 0 & i = k \end{Bmatrix} \tag{5.7}$$

Further, in mathematical computation, it is found that the most nearby sample get the maximum weight of W = 1. In addition, the sample found at the farthest k distance will get the least weight of W=0. While, the samples which are formed in the inequality intermediate range, get the weight in between 0 to 1. These all samples will participate in the voting with their associated weight value. The maximum of the weighted class label be considered as the winner and its class label is used as the predicted class for query sample.

5.2 Results and Discussion

To test the performance of the proposed mathematical algorithm, the experimentation is performed with the malignant brain tumors MR images dataset. The dataset consists of five types of Grade IV tumors. The description of the dataset is given in Table 4.3. Initially, the dataset is pre-processed before it is experimented for classification. The preprocessing of the dataset includes extraction of the texture features for all the images of the dataset. In addition, the relevant and the informative features among the overall extracted features are filtered using the discussed feature selection approach in previous chapter. Finally, the selected features are experimented with the proposed algorithm to find the classification accuracy. To test the new mathematical algorithm, the best combination result of TOM as feature extraction and Hybrid approach as dimension reduction gained in previous chapter 4 in Table 4.11 is used for experimentation. The experimental result with new mathematical algorithm is shown in Table 5.1.

Table 5.1: Experimentation results of proposed algorithm with tumor dataset

Dataset	KNN	WKNN	UWKNN	DWKNN	Proposed Rule
Central Neuro Cytoma	93.41	90.97	90.32	91.56	93.66
Glioblastoma Multiforme	94.13	91.25	91.25	93.07	95.33
Gliomas	93.20	86.89	90.97	91.28	94.66
Intra Ventricular Malignant Mass	92.78	90.32	90.78	92.84	95.70
Metastasis	93.17	91.87	91.87	93.95	93.95
Average Accuracy	93.34	90.26	91.03	92.54	94.66

5.3 Statistical measures and validation - McNemar's Test

To validate the proposed algorithm and to find the significance between classification algorithms, a statistical test is performed which discovers whether the classification algorithm is significant or not [150]. To validate and finding the significance of the two classification algorithms, an approximate statistical test is used named *McNemar's test* [151]. This test is used to determine experimentally the probability of incorrectly identifying the Type I error. Type I error is the incorrect rejection of a true null hypothesis, or False Positive. It identifies an effect in machine learning that is not present. Thus to evaluate the performance of two classification algorithms, McNemar's test is preferred in this chapter of the thesis.

McNemar's test is based on the χ^2 distribution value at specific degree of freedom and trust level. To apply *McNemar's test* in machine learning, the dataset is divided into two parts i.e. training set, R and testing set T. On the training set apply the two classification algorithms, f_A and f_B to learn the system individually. After the training, these algorithms are tested using the testing dataset T. For every sample, in testing set, $s_T \in T$, a 2x2 contingency table is maintained that represent the comparative classification performance of the algorithms f_A and f_B. The cell values of the formed contingency table are the 'pairs' not the individual value of any algorithm. The sample table is shown as:

Where $N = (M_{00}) + (M_{01}) + (M_{10}) + (M_{11})$ is total number of the samples in the test set T.

The cells M_{01} and M_{10}, in *McNemar's test* are called *discordant cells* because

Table 5.2: Contingency Table for McNemars Statistical Analysis

Number of samples misclassified by both f_A and f_B (M_{00})	Number of samples misclassified by f_A but not by f_B (M_{01})
Number of samples misclassified by f_B but not by f_A (M_{10})	Number of samples misclassified by neither f_A nor f_B (M_{11})

these cells value represent the pairs with difference in observations by algorithm f_A and f_B. While the cells M_{00} and M_{11}, are called *concordant cells* because of providing similarity observation results by the algorithms. As the concordant calls don't provide any difference in algorithms pair, so these can't be used in finding the *McNemar's test* statistic. *McNemar's test* is based on χ^2 (χ^2) distribution for goodness of fit. Under the condition of null hypothesis (H_0), it is expected that the two algorithms should have the same error rate that indicates(M_{01}) = (M_{10}). In any of the case, if the null hypothesis is found true, then the following inequality will always be find true:

$H_0 = \frac{M_{01}}{M_{01}+M_{10}} = \frac{M_{10}}{M_{01}+M_{10}} = 0.5$

The McNemar's χ^2 value is computed using the pair difference counts of discordant cells as:

$$M_c = \frac{(|M_{01} - M_{10}| - 1)^2}{M_{01} + M_{10}} \tag{5.8}$$

As the McNemar's sample distribution is a χ^2 distribution, thus for the test at α = 0.05 and df =1, the expected value of M_c at $\chi^2_{1,0.95}$ = 3.841. Based on the observed M_c value experimentally, it is found that if the observed M_c value < 3.84 then the null hypothesis is accepted otherwise the alternate hypothesis is accepted for M_c value > 3.84.

In this thesis, to test and validate the proposed classification algorithm with state-of-art KNN and it's variants, McNemar's test is applied. Let the null hypothesis (H_0) is defined as:
(H_0): Proposed algorithm and state-of-art algorithms have same performance and
(H_1): Proposed algorithm and state-of-art algorithms have different performance.

Based on such, the proposed algorithm is statistically verified whose results are shown in Table 5.3.

Table 5.3: Experimental result of McNemars test of proposed algorithm vs. state-of-art algorithms

Dataset	KNN	WKNN	UWKNN	DWKNN
Central Neuro Cytoma	6.89	6.89	4.58	4.53
Glioblastoma Multiforme	3.87	3.87	4.22	4.06
Gliomas	5.58	6.84	5.89	5.89
Intra Ventricular Malignant Mass	7.51	5.87	7.51	7.97
Metastasis	7.34	6.34	6.34	6.12
Average	**6.238**	**5.962**	**5.708**	**5.714**

In the experimentation result with McNemar's algorithm, it is found from different dataset that the proposed method AVNM is statistically significant different from existing algorithms (*KNN, WKNN, UWKNN, DWKNN*). Finally, the validation algorithm suggest that proposed algorithm is better in classification than state-of-art algorithms.

5.4 Summary of chapter

In this chapter, a new statistical algorithm based approach is introduced that provides the enhancement in the classification model. Proposed algorithm is totally dynamic in nature that provides the optimal number of nearby samples. These samples are associated with the query sample on basis of the distance similarity. The more closely the existing sample to query sample, more resemblance they have in nature.

The proposed algorithm consists of two parts namely threshold computation and space reduction. In first part, the algorithm computes recursively and finds dynamic threshold value T using existing samples in space. In second part, considering the threshold T, the algorithm reduces the original sample space to new space that has less number of samples.

Using the proposed algorithm, the sample space recursively reduces till the threshold value becomes unchangeable or constant. At this state, where no change in the sample space is encountered, then the sample space is considered as converged state. Subsequently, it holds optimal number of samples that lies inside the final reduced sample space. These samples are considered as the sufficient number of required nearby samples to the test query sample and participate in the voting for

predicting the class label of test sample.

Further, among the n samples, the algorithm uses the weighted approach for assigning maximum weight to most nearby sample and minimum weight to the farthest sample. Later, the assigned weight is used as the multiplicative factor with sample class label to get the voting for test sample's class label. The winner class label among the voting is used as the predicted class label to test query sample.

Proposed algorithm experimented with the dataset of malignant brain tumors MR images. The experimented results are presented in Table 5.1. It is noticed that the proposed algorithm gives satisfactory results for the classification of malignant brain tumors. The results are compared with the KNN and its variants based on weighted mechanism and neighborhood selection. In the next chapter, overall conclusions of the thesis are presented with the important finding based on the proposed approaches in machine learning. Finally, the future scope is presented in detail.

Chapter 6

Conclusions and Future Scope

6.1 Conclusions

Brain tumor is one among the most harmful diseases in the dome of medical science. The process of brain tumor type classification serves as the assessment for tumor treatment. The machine learning-based diagnosis helps the radiologist to make proper diagnosis and patient management. The radiological diagnosis requires a moderately sophisticated assessment of the various *Magnetic Resonance* (MR) imaging, *Computer Tomography* (CT) imaging and *Positron Emission Tomography* (PET) imaging. The assessment is basically performed by experienced radiologists using heuristic learning and digital systems. There are several kinds of literature exists that point out the requirement of tumor cells classification. As the every tumor type has different characteristics and treatment protocols, thus the proper classification of tumor type is a major concern. Literature gives the various ideas about the tumor cells segmentation and their classification but has certain limitations with them.

The main problems with state-of-art existing systems are as follows:

1. Most of the existing machine learning algorithms for medical image segmentation suffer from the selection of the predefined variables like threshold value, the number of required clusters and initial starting point. Those algorithms depend on the size of the image and thus the performance of those algorithms has found to be limited.

2. Machine learning works on the set of the extracted features from the segmented regions. These features are used to describe the image in machine

learning models. To select the appropriate feature extraction method is an issue in machine learning system.

3. Machine learning models may perform well if an efficient feature selection technique is used to eliminate the noisy and irrelevant features to reduce the size the high dimensional feature vector. To investigate an effective feature selection technique is a challenge.

4. In classification problems, the selection of the number of nearest neighbor for decision making is still an open issue. This nearest neighbor selection is based on the choice of the user and it varies from dataset to dataset. The selection of an optimal value for nearest neighbors is another open issue in machine learning problems.

The work in this thesis is divided into three major parts. These are *Magnetic Resonance* (MR) image segmentation, Classification of the malignant brain tumor MRI using textural features, and Malignant brain tumor classification with variant of nearest neighbor algorithm. In every part, several algorithms are proposed to fulfil the research objective of the thesis. The proposed solutions of research questions are discussed in the chapters 3,4,5. In every chapter, conclusions are drawn from the respective methodologies developed based on the experimental results. The consolidate result of proposed algorithms provide an overall picture of the contributions of the thesis.

These overall contributions are summarized as follows:

1. Novel image segmentation algorithms are proposed in the machine learning system that extract the tumor area from the T1-weighted post contrast axial brain MRI. The proposed segmentation algorithm is the hybrid Bi-clustering algorithm that clusters the MR image into two groups and hierarchical multi-cluster algorithm that segments the abnormality regions iteratively fast as compared with existing algorithms. The quantitative validation of the given algorithms using DB index and Dunn index is given in Table 3.5. The DB Index value of *Hybrid algorithm* is 0.29 which is much lower that state-of-art *OTSU* and *Region Growing* algorithms. While the DB Index of extended *Hierarchical algorithm* 0.13 is lowest among all state-of-art algorithms.

2. Two new feature extraction mechanism are proposed named *Counting label Occurrence Matrix* (CLOM) and *Texture Occurrence Matrix* (TOM) for ex-

tracting textural features in the spatial and spectral domain. The average accuracy gained by CLOM is 86.71% using KNN classifier and 89.94% with SVM classifier. The average increase in the performance of the classifier is found to be in the range of 1.20% to 9.33% as compared with state-of-art algorithms. While with TOM the average accuracy gained is 91% and 93% with KNN and SVM classifiers respectively. The gained in the performance with state-of-art using TOM is around 14.77%.

3. A new feature selection algorithm is proposed that is based on the statistical computations named *Cumulative Variance Method* (CVM). Additionally, a hybrid algorithm based on univariate and multivariate algorithms is proposed. This enhances the performance of the machine learning system in terms of tumor classification accuracy. The experimentation with CVM results in the gain of highest accuracy (93.67%) with TOM based features. Further, the extension towards hybrid algorithm gained the classification accuracy of 97.28% with SVM as the classifier.

4. A new mathematical algorithm is proposed that automates the selection of a number of nearest neighbors for decision making like KNN. Automation in the selection of nearby neighbors is computed using proposed algorithm based on sample space reduction. Proposed algorithm resolves the dependency of the initial selection of the number of nearest neighbor, or the kernel function implementation for hyperplane generation in accuracy computation. The experimentation results conclude that using proposed rule the accuracy increase to 94.66% which is initially 91% using KNN classifier.

All the experimental results are performed using the clinical dataset that consists of six hundred sixty (660) malignant brain tumor MRI. All the images are of T1-weighted post-contrast axial modality taken from 3.0 T GE MR Scanner. The dataset includes five classes of malignant brain tumor. These classes are *Central Neuro Cytoma* (CNC), *Glioblastoma Multiforme* (GBM), *Gliomas* (GLI), *Intra Ventricular Malignant Mass* (IVMM), and *Metastasis* (MTS) obtained from the Department of Radiology, *Sawai Man singh* (SMS) Medical College Jaipur, Rajasthan, India.

6.2 Future Work

A fascinating extension of this thesis would be to explore and compute the features mined from other modalities like CT, PET and fMRI and fuse them with the features employed in this thesis. This assembly of multimodal classification systems may improve the malignant brain tumor classification accuracy. Another interesting idea is to incorporate clinical features along with the statistical features derived from the MRI. The adoption of such hybrid features in the malignant tumor classification may improved accuracy.

In future, the proposed AVNM classification algorithm will also be experimented with some state-of-art classifiers named Artificial Neural Network, SVM and Decision Tree. Some of the experiments will also be conducted to find the ensemble classification model that improves the performance of the classification accuracy.

Finally, the reduction of algorithmic processing time is another important issue. Since for real-time clinical applications, processing should not exceed the order of minutes. The algorithms developed in this thesis require broad computations and, thus, proposed algorithms should be parallelized for faster implementation. The role and use of the big data analytics in the medical domain may provide the fast real time analysis. This helps to run proposed algorithms in a network of computers that will run in parallel mode.

Bibliography

[1] This is your brain and this is how it works. http://slideplayer.com/slide/9003301/slide2.htm. Accessed: 20-08-2016.

[2] Kenneth L Becker. *Principles and practice of endocrinology and metabolism*. Lippincott Williams & Wilkins, 2001.

[3] Anssi Auvinen, Maila Hietanen, Ritva Luukkonen, and Riitta-Sisko Koskela. Brain tumors and salivary gland cancers among cellular telephone users. *Epidemiology*, 13(3):356–359, 2002.

[4] Fred H Hochberg, George Miller, Robert T Schooley, Martin S Hirsch, Paul Feorino, and Werner Henle. Central-nervous-system lymphoma related to epstein–barr virus. *New England Journal of Medicine*, 309(13):745–748, 1983.

[5] Cancer research uk. http://www.cancerresearchuk.org/about-cancer/type/brain-tumour/about/brain-tumour-risks-and-causes.htm. Accessed: 20-08-2016.

[6] Brain tumor symptoms and signs. http://www.cancer.net/cancer-types/brain-tumor/symptoms-and-signs.htm. Accessed: 20-08-2016.

[7] Brain tumor symptoms and causes. http://www.mayoclinic.org/diseasesconditions/brain-tumor/symptoms-causes/dxc-20117134.htm. Accessed: 20-08-2016.

[8] William R Shapiro and Joan Rankin Shapiro. Biology and treatment of malignant glioma. *Oncology (Williston Park, NY)*, 12(2):233–40, 1998.

[9] F Dhermain, D Ducreux, F Bidault, A Bruna, F Parker, T Roujeau, A Beaudre, JP Armand, and C Haie-Meder. [use of the functional imaging modalities in radiation therapy treatment planning in patients with glioblastoma]. *Bulletin du cancer*, 92(4):333–342, 2005.

[10] Paul Kleihues and Leslie H Sobin. World health organization classification of tumors. *Cancer*, 88(12):2887–2887, 2000.

[11] Robert W Brown, Y-C Norman Cheng, E Mark Haacke, Michael R Thompson, and Ramesh Venkatesan. *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons, 2014.

[12] Paul Tofts. *Quantitative MRI of the brain: measuring changes caused by disease*. John Wiley & Sons, 2005.

[13] Directions and planes of section. https://faculty.ishington.edu/chudler/slice.html. Accessed: 20-08-2016.

[14] Michael Buckland and Fredric Gey. The relationship between recall and precision. *Journal of the American society for information science*, 45(1):12, 1994.

[15] ZQ John Lu. The elements of statistical learning: data mining, inference, and prediction. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(3):693–694, 2010.

[16] Shijun Wang and Ronald M Summers. Machine learning and radiology. *Medical image analysis*, 16(5):933–951, 2012.

[17] Radiology today magazine. http://www.radiologytoday.net/archive/050409p16.shtml. Accessed: 20-08-2016.

[18] Sarunas J Raudys and Anil K Jain. Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Transactions on pattern analysis and machine intelligence*, 13(3):252–264, 1991.

[19] Yaser S Abu-Mostafa. The vapnik-chervonenkis dimension: Information versus complexity in learning. *Neural Computation*, 1(3):312–317, 1989.

[20] Michael J Kearns. *The computational complexity of machine learning*. MIT press, 1990.

[21] Zhiwei Tang and Yixuan Wu. One image segmentation method based on otsu and fuzzy theory seeking image segment threshold. In *International Conference on Electronics, Communications and Control (ICECC)*, pages 2170–2173. IEEE, 2011.

[22] Yi Wan, Jiangchang Wang, Xingbo Sun, and Ming Hao. A modified otsu image segment method based on the rayleigh distribution. In *3rd International Conference on Computer Science and Information Technology (ICCSIT)*, volume 5, pages 281–285. IEEE, 2010.

[23] P Natarajan, N Krishnan, Natasha Sandeep Kenkre, Shraiya Nancy, and Bhuvanesh Pratap Singh. Tumor detection using threshold operation in mri brain images. In *International Conference on Computational Intelligence & Computing Research (ICCIC)*, pages 1–4. IEEE, 2012.

[24] Anil K Jain. Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8):651–666, 2010.

[25] Ming-Ni Wu, Chia-Chen Lin, and Chin-Chen Chang. Brain tumor detection using color-based k-means clustering segmentation. In *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP)*, volume 2, pages 245–250. IEEE, 2007.

[26] Eman Abdel-Maksoud, Mohammed Elmogy, and Rashid Al-Awadi. Brain tumor segmentation based on a hybrid clustering technique. *Egyptian Informatics Journal*, 16(1):71–81, 2015.

[27] Joseph C Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. 1973.

[28] James C Bezdek. *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.

[29] D Selvathi, J Anitha, et al. Effective fuzzy clustering algorithm for abnormal mr brain image segmentation. In *International Advance Computing Conference (IACC)*, pages 609–614. IEEE, 2009.

[30] R Preetha and GR Suresh. Performance analysis of fuzzy c means algorithm in automated detection of brain tumor. In *World Congress on Computing and Communication Technologies (WCCCT)*, pages 30–33. IEEE, 2014.

[31] Heena Hooda, Om Prakash Verma, and Tripti Singhal. Brain tumor segmentation: A performance analysis using k-means, fuzzy c-means and region growing algorithm. In *International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, pages 1621–1626. IEEE, 2014.

[32] Vignesh Rajesh, Bharathan Venkat, Vikesh Karan, and M Poonkodi. Brain tumor segmentation and its area calculation in brain mr images using k-mean clustering and fuzzy c-mean algorithm. *Journal of Fuzzy Systems*, 7(4):103–107, 2015.

[33] Sookpotharom Supot, Chaichana Thanapong, Pintavirooj Chuchart, and Sangworasil Manas. Segmentation of magnetic resonance images using discrete curve evolution and fuzzy clustering. In *International Conference on Integration Technology (ICIT)*, pages 697–700. IEEE, 2007.

[34] Amritpal Singh et al. Detection of brain tumor in mri images, using combination of fuzzy c-means and svm. In *2nd International Conference on Signal Processing and Integrated Networks (SPIN)*, pages 98–102. IEEE, 2015.

[35] M Karnan and N Nandha Gopal. Hybrid markov random field with parallel ant colony optimization and fuzzy c means for mri brain image segmentation. In *International Conference on Computational Intelligence and Computing Research (ICCIC)*, pages 1–4. IEEE, 2010.

[36] Abbas Biniaz, Ataollah Abbassi, Mousa Shamsi, and Afshin Ebrahimi. Fuzzy c-means clustering based on gaussian spatial information for brain mr image segmentation. In *19th Iranian Conference of Biomedical Engineering (ICBME)*, pages 154–158. IEEE, 2012.

[37] N Nandha Gopal and M Karnan. Diagnose brain tumor through mri using image processing clustering algorithms such as fuzzy c means along with intelligent optimization techniques. In *International Conference on Computational Intelligence and Computing Research (ICCIC)*, pages 1–4. IEEE, 2010.

[38] WE Phillips, RP Velthuizen, S Phuphanich, LO Hall, LP Clarke, and ML Silbiger. Application of fuzzy c-means segmentation technique for tissue differentiation in mr images of a hemorrhagic glioblastoma multiforme. *Magnetic Resonance Imaging*, 13(2):277–290, 1995.

[39] Matthew C Clark, Lawrence O Hall, Dmitry B Goldgof, Robert Velthuizen, F Reed Murtagh, and Martin S. Silbiger. Automatic tumor segmentation using knowledge-based techniques. *IEEE Transactions on Medical Imaging*, 17(2):187–201, 1998.

[40] Lynn M Fletcher-Heath, Lawrence O Hall, Dmitry B Goldgof, and F Reed Murtagh. Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. *Artificial intelligence in medicine*, 21(1):43–63, 2001.

[41] Mohammed Yakoob Siyal and Lin Yu. An intelligent modified fuzzy c-means based algorithm for bias estimation and segmentation of brain mri. *Pattern recognition letters*, 26(13):2052–2062, 2005.

[42] SR Kannan. A new segmentation system for brain mr images based on fuzzy techniques. *Applied Soft Computing*, 8(4):1599–1606, 2008.

[43] László Szilágyi, Sándor M Szilágyi, and Zoltán Benyó. A modified fuzzy c-means algorithm for mr brain image segmentation. In *International Conference on Image Analysis and Recognition*, pages 866–877. Springer, 2007.

[44] Mohamad Forouzanfar, Nosratallah Forghani, and Mohammad Teshnehlab. Parameter optimization of improved fuzzy c-means clustering algorithm for brain mr image segmentation. *Engineering Applications of Artificial Intelligence*, 23(2):160–168, 2010.

[45] Weiling Cai, Songcan Chen, and Daoqiang Zhang. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. *Journal of Pattern Recognition*, 40(3):825–838, 2007.

[46] László Szilagyi, Zoltán Benyo, Sándor M Szilágyi, and HS Adam. Mr brain image segmentation using an enhanced fuzzy c-means algorithm. In *25th Annual International Conference of Engineering in Medicine and Biology Society*, volume 1, pages 724–726. IEEE, 2003.

[47] Nelly Gordillo, Eduard Montseny, and Pilar Sobrevilla. State of the art survey on mri brain tumor segmentation. *Magnetic Resonance Imaging*, 31(8):1426–1438, 2013.

[48] Jeffrey William Prescott. Quantitative imaging biomarkers: the application of advanced image processing and analysis to clinical and preclinical decision making. *Journal of digital imaging*, 26(1):97–108, 2013.

[49] Jung Leng Foo. A survey of user interaction and automation in medical image segmentation methods. *Human Computer Interaction Technical Report,Iowa State University, ISU-HCI-2006-02*, 2006.

[50] Sílvia Delgado Olabarriaga and Arnold WM Smeulders. Interaction in the segmentation of medical images: A survey. *Medical image analysis*, 5(2):127–142, 2001.

[51] Chi-Hoon Lee, Mark Schmidt, Albert Murtha, Aalo Bistritz, Jöerg Sander, and Russell Greiner. Segmenting brain tumors with conditional random fields and support vector machines. In *International Workshop on Computer Vision for Biomedical Image Applications*, pages 469–478. Springer, 2005.

[52] Weibei Dou, Su Ruan, Yanping Chen, Daniel Bloyet, and Jean-Marc Constans. A framework of fuzzy information fusion for the segmentation of brain tumor tissues on mr images. *Image and vision Computing*, 25(2):164–171, 2007.

[53] Ying Wang, Zhi Xian Lin, Jian Guo Cao, and Mao Qing Li. Automatic mri brain tumor segmentation system based on localizing active contour models. In *Advanced Materials Research*, volume 219, pages 1342–1346. Trans Tech Publ, 2011.

[54] Dipak KumarKole and Amiya Halder. Automatic brain tumor detection and isolation of tumor cells from mri images. *International Journal of Computer Applications*, 39(16):26–30, 2012.

[55] Khan M Iftekharuddin, Jing Zheng, Mohammad A Islam, and Robert J Ogg. Fractal-based brain tumor detection in multimodal mri. *Applied Mathematics and Computation*, 207(1):23–41, 2009.

[56] Sérgio Pereira, Adriano Pinto, Victor Alves, and Carlos A Silva. Brain tumor segmentation using convolutional neural networks in mri images. *IEEE Transactions on Medical Imaging*, 35(5):1240–1251, 2016.

[57] Jianhua Yao. Image processing in tumor imaging. *New Techniques in Oncologic Imaging*, pages 79–102, 2006.

[58] Tony F Chan and Luminita A Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2012.

[59] Suhuai Luo, Rongxin Li, and Sébastien Ourselin. A new deformable model using dynamic gradient vector flow and adaptive balloon forces. In *APRS Workshop on Digital Image Computing*, pages 9–14, 2003.

[60] Suhuai Luo. Automated medical image segmentation using a new deformable surface model. *International Journal of Computer Science and Network Security*, 6(5A):109–115, 2006.

[61] Vicent Caselles, Francine Catté, Tomeu Coll, and Françoise Dibos. A geometric model for active contours in image processing. *Numerische mathematik*, 66(1):1–31, 1993.

[62] Ravi Malladi, James A Sethian, and Baba C Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175, 1995.

[63] Stanley Osher and James A Sethian. Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.

[64] Satyanad Kichenassamy, Arun Kumar, Peter Olver, Allen Tannenbaum, and Anthony Yezzi. Gradient flows and geometric active contour models. In *Fifth International Conference on Computer Vision*, pages 810–815. IEEE, 1995.

[65] Anthony Yezzi, Satyanad Kichenassamy, Arun Kumar, Peter Olver, and Allen Tannenbaum. A geometric snake model for segmentation of medical imagery. *IEEE Transactions on Medical Imaging*, 16(2):199–209, 1997.

[66] Kaleem Siddiqi, Yves Bérubé Lauziere, Allen Tannenbaum, and Steven W Zucker. Area and length minimizing flows for shape segmentation. *IEEE Transactions on Image Processing*, 7(3):433–443, 1998.

[67] Dorothy M Greig, Bruce T Porteous, and Allan H Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 271–279, 1989.

[68] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[69] Yuri Boykov and Marie-Pierre Jolly. Interactive organ segmentation using graph cuts. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 276–286. Springer, 2000.

[70] Daniel García-Lorenzo, Jeremy Lecoeur, Douglas L Arnold, D Louis Collins, and Christian Barillot. Multiple sclerosis lesion segmentation using an automatic multimodal graph cuts. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 584–591. Springer, 2009.

[71] Vladimir Estivill-Castro. Why so many clustering algorithms: a position paper. *ACM SIGKDD explorations newsletter*, 4(1):65–75, 2002.

[72] Sreekumar K Athira Krishnan. A survey on image segmentation and feature extraction methods for acute myelogenous leukemia detection in blood microscopic images. *Athira Krishnan et al,/(IJCSIT) International Journal of Computer Science and Information Technologies*, 5(6):7877–7879, 2014.

[73] Kenneth Falconer. *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons, 2004.

[74] Atiq Islam, Syed MS Reza, and Khan M Iftekharuddin. Multifractal texture estimation for detection and segmentation of brain tumors. *IEEE transactions on Biomedical Engineering*, 60(11):3204–3215, 2013.

[75] Mu Zhou, Lawrence O Hall, Dmitry B Goldgof, Robert A Gatenby, and Robert J Gillies. A texture feature ranking model for predicting survival time of brain tumor patients. In *IEEE International Conference on Systems, Man, and Cybernetics*, pages 4533–4538. IEEE, 2013.

[76] Nathalie Just. Improving tumour heterogeneity mri assessment with histograms. *British Journal of Cancer*, 111(12):2205–2213, 2014.

[77] Robert M Haralick, Karthikeyan Shanmugam, et al. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, (6):610–621, 1973.

[78] Gang Zhang, Zong-min Ma, Li-guo Deng, and Zhi-ping Cai. Orientational filter based on dominant directions in content-based image retrieval. *Journal of North Eastern University of Natural Science*, 28(7):978, 2007.

[79] Suchita Goswami and Lalit Kumar P Bhaiya. A hybrid neuro-fuzzy approach for brain abnormality detection using glcm based feature extraction. In *International Conference on Emerging Trends in Communication, Control, Signal Processing & Computing Applications (C2SPCA)*, pages 1–7. IEEE, 2013.

[80] SN Deepa and B Aruna Devi. Artificial neural networks design for classification of brain tumour. In *International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–6. IEEE, 2012.

[81] Vikas Gupta and Kaustubh S Sagale. Implementation of classification system for brain cancer using backpropagation network and mri. In *Nirma University International Conference on Engineering (NUiCONE)*, pages 1–4. IEEE, 2012.

[82] SN Deepa and B Aruna Devi. Neural networks and smo based classification for brain tumor. In *World Congress on Information and Communication Technologies (WICT)*, pages 1032–1037. IEEE, 2011.

[83] D Jude Hemanth, C Kezi Selva Vijila, A Immanuel Selvakumar, and J Anitha. Performance improved iteration-free artificial neural networks for abnormal magnetic resonance brain image classification. *Journal of Neurocomputing*, 130:98–107, 2014.

[84] Shaheen Ahmed, Khan M Iftekharuddin, and Arastoo Vossough. Efficacy of texture, shape, and intensity feature fusion for posterior-fossa tumor segmentation in mri. *IEEE Transactions on Information Technology in Biomedicine*, 15(2):206–213, 2011.

[85] VP Rathi and S Palani. Brain tumor mri image classification with feature selection and extraction using linear discriminant analysis. *arXiv preprint arXiv:1208.2128*, 2012.

[86] Pantelis Georgiadis, Dionisis Cavouras, Ioannis Kalatzis, Antonis Daskalakis, George C Kagadis, Koralia Sifaki, Menelaos Malamas, George Nikiforidis, and Ekaterini Solomou. Improving brain tumor characterization on mri by probabilistic neural networks and non-linear transformation of textural features. *Computer Methods and Programs in Biomedicine*, 89(1):24–32, 2008.

[87] Umer Javed, Muhammad Mohsin Riaz, Abdul Ghafoor, and Tanveer Ahmed Cheema. Mri brain classification using texture features, fuzzy weighting and support vector machine. *Progress In Electromagnetics Research B*, 53:73–88, 2013.

[88] Evangelia I Zacharaki, Sumei Wang, Sanjeev Chawla, Dong Soo Yoo, Ronald Wolf, Elias R Melhem, and Christos Davatzikos. Mri-based classification of brain tumor type and grade using svm-rfe. In *International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1035–1038. IEEE, 2009.

[89] Rafael C Gonzalez and Richard E Woods. Digital image processing, 2008.

[90] G Farias, M Santos, and V Lopez. Brain tumour diagnosis with wavelets and support vector machines. In *3rd International Conference on Intelligent System and Knowledge Engineering*, volume 1, pages 1453–1459. IEEE, 2008.

[91] El-Sayed Ahmed El-Dahshan, Tamer Hosny, and Abdel-Badeeh M Salem. Hybrid intelligent techniques for mri brain images classification. *Digital Signal Processing*, 20(2):433–441, 2010.

[92] Heba Mohsen, El-Sayed Ahmed El-Dahshan, and Abdel-Badeeh M Salem. A machine learning technique for mri brain images. In *8th International Conference on Informatics and Systems (INFOS)*, pages BIO–161. IEEE, 2012.

[93] Mohd Fauzi Bin Othman, Noramalina Bt Abdullah, and Nurul Fazrena Bt Kamal. Mri brain classification using support vector machine. In *4th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO)*, pages 1–4. IEEE, 2011.

[94] Noramalina Abdullah, Lee Wee Chuen, Umi Kalthum Ngah, and Khairul Azman Ahmad. Improvement of mri brain classification using principal component analysis. In *IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, pages 557–561. IEEE, 2011.

[95] D Sridhar and IV Murali Krishna. Brain tumor classification using discrete cosine transform and probabilistic neural network. In *International Conference on Signal Processing Image Processing & Pattern Recognition (ICSIPR)*, pages 92–96. IEEE, 2013.

[96] G Preethi and V Sornagopal. Mri image classification using glcm texture features. In *International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)*, pages 1–6. IEEE, 2014.

[97] Salim Lahmiri and Mounir Boukadoum. Classification of brain mri using the lh and hl wavelet transform sub-bands. In *IEEE International Symposium of Circuits and Systems (ISCAS)*, pages 1025–1028. IEEE, 2011.

[98] Atiq ur Rehman, Aasia Khanum, and Arslan Shaukat. Hybrid feature selection and tumor identification in brain mri using swarm intelligence. In *11th International Conference on Frontiers of Information Technology (FIT)*, pages 49–54. IEEE, 2013.

[99] Hashem Kalbkhani, Mahrokh G Shayesteh, and Behrooz Zali-Vargahan. Robust algorithm for brain magnetic resonance image (mri) classification based on garch variances series. *Journal of Biomedical Signal Processing and Control*, 8(6):909–919, 2013.

[100] M Saritha, K Paul Joseph, and Abraham T Mathew. Classification of mri brain images using combined wavelet entropy based spider web plots and probabilistic neural network. *Journal of Pattern Recognition Letters*, 34(16):2151–2156, 2013.

[101] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.

[102] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *Journal of Science*, 290(5500):2323–2326, 2000.

[103] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Journal of Neural Computation*, 15(6):1373–1396, 2003.

[104] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Journal of Science*, 290(5500):2319–2323, 2000.

[105] Ronald R Coifman, Stephane Lafon, Ann B Lee, Mauro Maggioni, Boaz Nadler, Frederick Warner, and Steven W Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *National Academy of Sciences of the United States of America*, 102(21):7426–7431, 2005.

[106] Pierre Comon. Independent component analysis, a new concept? *Journal of Signal processing*, 36(3):287–314, 1994.

[107] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.

[108] Huan Liu and Rudy Setiono. Chi2: Feature selection and discretization of numeric attributes. In *ICTAI*, pages 388–391, 1995.

[109] Mark A Hall and Lloyd A Smith. Feature subset selection: a correlation based filter approach. 1997.

[110] Lei Yu and Huan Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *ICML*, volume 3, pages 856–863, 2003.

[111] Matthias Ring and Bjoern M Eskofier. Optimal feature selection for nonlinear data using branch-and-bound in kernel space. *Journal of Pattern Recognition Letters*, 68:56–62, 2015.

[112] Cheng-Lung Huang and Chieh-Jen Wang. A ga-based feature selection and parameters optimizationfor support vector machines. *Journal of Expert Systems with Applications*, 31(2):231–240, 2006.

[113] Padmavathi Kora and Sri Ramakrishna Kalva. Improved bat algorithm for the detection of myocardial infarction. *Journal of SpringerPlus*, 4(1):1, 2015.

[114] E Emary, Waleed Yamany, and Aboul Ella Hassanien. New approach for feature selection based on rough set and bat algorithm. In *9th International Conference on Computer Engineering & Systems (ICCES)*, pages 346–353. IEEE, 2014.

[115] Sujoy Paul and Swagatam Das. Simultaneous feature selection and weighting–an evolutionary multi-objective optimization approach. *Journal of Pattern Recognition Letters*, 65:51–59, 2015.

[116] Suresh Dara and Haider Banka. An elitist binary pso algorithm for selecting features in high dimensional data. In *Advanced Computing, Networking and Informatics*, volume 1, pages 679–686. Springer, 2014.

[117] Haider Banka and Suresh Dara. Feature selection and classification for gene expression data using evolutionary computation. In *23rd International Workshop on Database and Expert Systems Applications*, pages 185–189. IEEE, 2012.

[118] Prudhvi Gurram and Heesung Kwon. Coalition game theory based feature subset selection for hyperspectral image classification. In *Geoscience and Remote Sensing Symposium*, pages 3446–3449. IEEE, 2014.

[119] Kai Zeng, Kun She, and Xinzheng Niu. Feature selection with neighborhood entropy-based cooperative game theory. *Computational Intelligence and Neuroscience*, 2014:11, 2014.

[120] Xin Sun, Yanheng Liu, Jin Li, Jianqi Zhu, Huiling Chen, and Xuejie Liu. Feature evaluation and selection with cooperative game theory. *Journal of Pattern Recognition*, 45(8):2992–3002, 2012.

[121] Anil K Jain, Robert P. W. Duin, and Jianchang Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

[122] Issam El-Naqa, Yongyi Yang, Miles N Wernick, Nikolas P Galatsanos, and Robert M Nishikawa. A support vector machine approach for detection of microcalcifications. *IEEE Transactions on Medical Imaging*, 21(12):1552–1563, 2002.

[123] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Journal of Image and Vision Computing*, 21(11):977–1000, 2003.

[124] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

[125] Yang Song, Jian Huang, Ding Zhou, Hongyuan Zha, and C Lee Giles. Iknn: Informative k-nearest neighbor pattern classification. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 248–264. Springer, 2007.

[126] Sahibsingh A Dudani. The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics*, (4):325–327, 1976.

[127] Jianping Gou, Taisong Xiong, and Yin Kuang. A novel weighted voting for k-nearest neighbor rule. *Journal of Computers*, 6(5):833–840, 2011.

[128] Heang-Ping Chan, Datong Wei, Mark A Helvie, Berkman Sahiner, Dorit D Adler, Mitchel M Goodsitt, and Nicolas Petrick. Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space. *Journal of Physics in Medicine and Biology*, 40(5):857, 1995.

[129] RICHARD Leblanc, A Otivier, RONALD Pokrupa, and DOUGLAS L Arnold. Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy. *Journal of Nature Medicine*, 2(3), 1996.

[130] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Journal of Psychological Review*, 65(6):386, 1958.

[131] Arthur Earl Bryson. *Applied optimal control: optimization, estimation and control*. CRC Press, 1975.

[132] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *National Academy of Sciences*, 79(8):2554–2558, 1982.

[133] David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. A learning algorithm for boltzmann machines. *Journal of Cognitive Science*, 9(1):147–169, 1985.

[134] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Journal of Biological Cybernetics*, 43(1):59–69, 1982.

[135] John Moody and Christian J Darken. Fast learning in networks of locally-tuned processing units. *Journal of Neural Computation*, 1(2):281–294, 1989.

[136] Donald F Specht. Probabilistic neural networks. *Journal of Neural Networks*, 3(1):109–118, 1990.

[137] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[138] Jason Weston, Sayan Mukherjee, Olivier Chapelle, Massimiliano Pontil, Tomaso Poggio, and Vladimir Vapnik. Feature selection for svms. 2000.

[139] TG Ditterrich. Machine learning research: four current direction. *Journal of Artificial Intelligence Magzine*, 4:97–136, 1997.

[140] Petronio L Braga, Adriano LI Oliveira, Gustavo HT Ribeiro, and Silvio RL Meira. Bagging predictors for estimation of software project effort. In *International Joint Conference on Neural Networks*, pages 1595–1600. IEEE, 2007.

[141] Yoav Freund and Robert E Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *European Conference on Computational Learning Theory*, pages 23–37. Springer, 1995.

[142] William W Hines, Douglas C Montgomery, and David M Goldman Connie M Borror. *Probability and statistics in engineering*. John Wiley & Sons, 2008.

[143] Sundaram Arivazhagan, L Ganesan, and S Padam Priyal. Texture classification using gabor wavelets based rotation invariant features. *Pattern recognition letters*, 27(16):1976–1982, 2006.

[144] Ke Huang and Selin Aviyente. Wavelet feature selection for image classification. *IEEE Transactions on Image Processing*, 17(9):1709–1720, 2008.

[145] Atindra Mohan Goon, Bhagabat Dasgupta, and Milankumar Gupta. *Fundamentals of statistics*. World Press Private Limited, 1963.

[146] Wavelet power spectrum. http://northstar-www.dartmouth.edu/doc/idl/html_6.2/Wavelet_Power_Spectrum.html. Accessed: 20-08-2016.

[147] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(Mar):1157–1182, 2003.

[148] Seung-Jean Kim, Alessandro Magnani, and Stephen Boyd. Robust fisher discriminant analysis. In *Advances in Neural Information Processing Systems*, pages 659–666, 2005.

[149] Yudong Cai, Tao Huang, Lele Hu, Xiaohe Shi, Lu Xie, and Yixue Li. Prediction of lysine ubiquitination with mrmr feature selection and analysis. *Journal of Amino acids*, 42(4):1387–1395, 2012.

[150] Cross valiadated. http://stats.stackexchange.com/questions/45851/how-to-statistically-compare-the-performance-of-machine-learning- Accessed: 20-08-2016.

[151] Alan Agresti and Maria Kateri. *Categorical Data Analysis*. Springer, 2011.

APPENDIX I

List of Publications

Journals:

J1. A. Vidyarthi, N. Mittal, *Texture Based Feature Extraction Method for Classification of Brain Tumor MRI*, Journal of Intelligent and Fuzzy Systems, Reprint DOI:10.3233/JIFS-169223, pp. 1 -12, 2017 (SCIE Indexed, IF: 1.004).

J2. A. Vidyarthi, N. Mittal, *AVNM: A Voting based Novel Mathematical Rule for Image Classification*, Journal of Computer Methods and Programs in Biomedicine Elsevier, Vol. 137, pp. 195201, 2016 (SCI Indexed, IF: 1.862)

J3. A. Vidyarthi, N. Mittal, *Utilization of Shape and Texture features with Statistical Feature Selection Mechanism for Classification of Malignant Tumors in MR Images*, Journal of Biomedical Engineering, De Gruyter, Vol. 59, pp. 155-159, 2014 (SCIE Indexed, IF: 1.650)

Conferences:

C1. A. Vidyarthi, N. Mittal, *Performance Analysis of Gabor-Wavelet based Features in Classification of High Grade Malignant Brain Tumors*, In the proceedings of 39th National System Conference (NSC), PP. 1-6, 2015 (IEEE).

C2. A. Vidyarthi, N. Mittal, *Brain tumor Segmentation Approaches: Review, Analysis and Anticipated Solutions in Machine Learning*, In the proceedings of 39th National System Conference (NSC), PP. 21-26, 2015 (IEEE).

C3. A. Vidyarthi, J. Nagpal, N. Mittal, *CLOM: Counting Label Occurrence Matrix for Feature Extraction in MR Images*, In the proceedings of 2nd International Conference on Signal Processing and Communication (ICSC), pp. 111 115, 2015(IEEE).

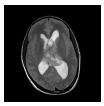
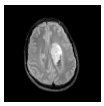
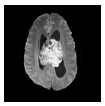
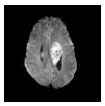
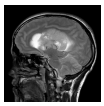
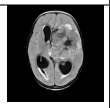
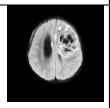
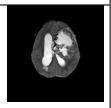
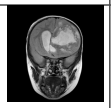
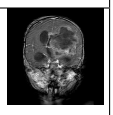
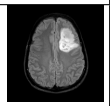
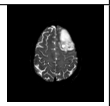
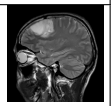
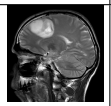
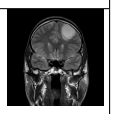
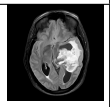
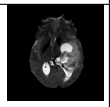
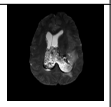
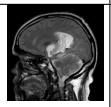
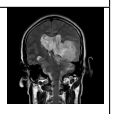
C4. A. Vidyarthi, N. Mittal,*Disjoint Tree based Clustering and Merging for Brain tumor Extraction*, In the Proceedings of 2nd International Conference on Advanced Computing, Networking and Informatics (ICACNI), pp. 445- 452, 2014 (Springer).

C5. A. Vidyarthi, N. Mittal,*A Hybrid Model for extraction of Brain Tumor in MR Images*, In the Proceedings of International Conference on Signal Processing and Communication (ICSC), pp. 202- 206, 2013 (IEEE).

C6. A. Vidyarthi, N. Mittal,*Comparative Study of Brain Tumor Classification in MR Images*, In the Proceedings of 3rd International Conference on Soft Computing for Problem Solving (SocPros), pp. 889- 897, 2013 (Springer).

APPENDIX II

Dataset

Central Neuro Cytoma					
Glioblastoma Multiforme					
Gliomas					
Intra Ventricular Malignant Mass					
Metastasis