

**A FRAMEWORK FOR EFFICIENT DETECTION OF BRAIN
TUMOR**

**Thesis submitted in fulfillment of the requirement for the award of
the degree of**

**DOCTOR OF PHILOSOPHY
IN
COMPUTER SCIENCE AND ENGINEERING**

Submitted By

**Prabhjot Kaur Chahal
(Registration No: 901403028)**

Under the guidance of

Dr. Shreelekha Pandey

Assistant Professor

Department of Computer Science & Engineering

Thapar Institute of Engineering and Technology, Patiala-147004



**COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY,
PATIALA – 147004**

September 2022

CERTIFICATE

I, Prabhjot Kaur Chahal, Regn. No. 901403028, hereby declare that the thesis entitled **“A framework for efficient detection of brain tumor”** submitted to the Computer Science and Engineering Department at Thapar Institute of Engineering and Technology, Patiala, Punjab, India is an authenticated record of my own work for the award of degree of “Doctor of Philosophy” under the supervision of Dr. Shreelekha Pandey. This report has not been submitted to any other institution for award of any other degree.


Prabhjot Kaur Chahal
Regn. No. 90143028

Place: Patiala

Date: 26.09.2022

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Verified by:


Dr. Shreelekha Pandey

Assistant Professor

Computer Science & Engineering Department

Thapar institute of Engineering & Technology, Patiala

ABSTRACT

One of the most crucial tasks in any brain tumor detection system is the isolation of abnormal tissues from normal brain tissues. Interestingly, domain of brain tumor analysis has effectively utilized the concepts of medical image processing, particularly on MR images, to automate the core steps, i.e. extraction, segmentation, classification for proximate detection of tumor. Research is more inclined towards MR for its non-invasive imaging properties. Human brain tumor detection and classification are time-consuming however vital tasks for any medical expert. Assistance via computer aided diagnosis or detection systems is commonly used to enhance diagnosis capabilities attaining maximum detection accuracy. But such systems are becoming challenging and are still an open problem due to variability in shapes, areas, and sizes of tumor. The past works of many researchers under medical image processing and soft computing have made noteworthy review analysis on automatic brain tumor detection techniques focusing segmentation as well as classification and their combinations. Despite significant research, brain tumor segmentation is still an open challenge due to variability in image modality, contrast, tumor type, and other factors. Many great works ranging from manual, semiautomatic, or fully automatic tumor segmentation with MR brain images are available, however, still creating a space for developing efficient and accurate approaches in this domain.

In this research work, various brain tumor detection techniques for MR images are reviewed along with the strengths and difficulties encountered in each to detect various brain tumor types. The current segmentation, classification, and detection techniques are also conferred emphasizing on the pros and cons of the medical imaging approaches in each modality. The survey presented aims to help researchers to

derive essential characteristics of brain tumor types and identifies various segmentation/classification techniques which are successful for detection of a range of brain diseases. An attempt to summarize the current state-of-art with respect to different tumor types would help researchers in exploring future directions.

The research proposes a hybrid weighted fuzzy k-means (WFKM) brain tumor segmentation approach using MR images to retrieve more meaningful clusters. It is based on fuzzification of weights which works on spatial context with illumination penalize membership approach which helps in settling issues with pixel's multiple memberships as well as exponential increase in number of iterations. The segmented image is further utilized for successful tumor type identification as benign or malignant by means of SVM. Experimentation performed on MR images using Digital Imaging and Communications in Medicine (DICOM) dataset shows that fusion of proposed WFKM and SVM outperforms many existing approaches. Further, performance evaluation parameters show that the proposal produces better overall accuracy. Results on variety of images further prove applicability of the proposal in detecting ranges and shapes of brain tumor. The proposed approach excels qualitatively as well as quantitatively reporting an average accuracy of 97% on DICOM dataset with total number of images varying from 100 to 1000.

Furthermore, the proposed WFKM is integrated with Hadoop-MapReduce and MDCS platforms to introduce scalability as well as parallelism. The generation of voluminous MR data, however has raised the need to handle massive MR data in significant time. The work thus focuses on analyzing the performance of WFKM in a scalable, distributed, and parallel processing environment. The performance of the collaboration is explored on variable sized datasets i.e. 215MB, 1.2GB and 7.3GB using single node (standalone), 2-node and 3-node configurations. With an increase in the size of MR image data, the processing-time decreases as the number of nodes gets increased, however, authenticate the objective of proposed work. On the other hand, the read and write operation times elevate as the data size increases irrespective to multiple nodes in the configuration. The respective classification

accuracy for 215MB, 1.2GB and 7.3GB are reported as 97.07%, 97.01%, and 96.30%. However, reduction in 90% processing-time is observed for 3-node configuration as compared to standalone on a dataset of 215MB. Similarly, 89% and 88.7% decrease in the processing time is noted with 1.2GB and 7.3GB of data, respectively.

ACKNOWLEDGMENTS

I deem it my duty to express a word of sincere gratitude to all those helping hands that came forward to undertake my research thesis. First of all, I wish to acknowledge the benevolence of omnipotent Almighty who gave me opportunity, courage, ability, and patience to overcome all hurdles.

With profound sense of gratitude and heartiest regard, I express my sincere feelings of indebtedness to my supervisor Dr. Shreelekha Pandey, Assistant Professor, Computer Science Engineering Department, Thapar Institute of Engineering and Technology for her valuable advice, motivation, encouragement, and moral support throughout this journey. I strongly admire the delightful ambiance for learning provided by her that made this thesis possible. It has been a wonderful experience to work under her shield of guidance. I really want to thank Dr. Shivani Goel (former guide), Bennett University, Noida from bottom of my heart as she helped me to get into this field and explore new things. Her trust and positive attitude always boosted me to get out even in intense situations.

I am grateful to Head of Department, Dr. Maninder Singh, who made my study a knowledgeable experience during my stay in the department. I am much beholden to the Director, Dean (RSP) and the Management of Thapar Institute of Engineering and Technology for providing all the necessary resources and encouragement to produce results. I am thankful to my doctoral committee members Dr. Prashant Singh Rana, Dr. Sanmeet Kaur, Dr. Sushma Jain and Dr. Mandeep Singh for their constructive suggestions and ensuring correct pace of the progress of my research work. I sincerely thank the faculty and support staff of Computer Science and Engineering department for their constant motivation.

I would like to pay deep gratitude to my parents S. Sarbjit Singh Chahal, Mrs. Arwinder Kaur Chahal and Mrs. Rajinder Kaur (mother-in-law) for their unconditional love, affection, and encouragement through all the good and bad times. I would like to shower hearty gratitude, feeling of indebtedness, and endless love to my husband S. Amritpal Singh for his moral support, and sacrifices which helped me to fulfill my dream. I would also like to acknowledge my sister Mehak Chahal and Navpreet Kaur (sister-in-law) for encouragement and much needed support during this journey. I want to pay humble gratitude to my sisterly friend Manvi Sharma for being my strength at every moment.

Finally, I want to bestow all my love and gratitude to my little world, Avitaaz Singh who came into my life during this journey and motivated me to strive my goals with positiveness.

(Prabhjot Kaur Chahal)

Contents

Certificate	i
Abstract	iii
Acknowledgment	v
Abbreviations	xv
NOMENCLATURE	xxii
1 Introduction	1
1.1 Brain Tumor - Characteristics	1
1.2 Brain Tumor - Techniques	2
1.3 Motivation	3
1.4 Imaging Modalities and Contrast Parameters	4
1.5 Tumor Types and their Grades	9
1.6 Generic Brain Tumor Detection System	10
1.6.1 Integration of Hadoop with MDCS for MR images	11
1.7 Objectives and Scope of the Work	13
1.8 Organization of the Thesis	14
2 Literature Review	17
2.1 A Few Statistics - MR Image Modalities and Data Sources	18
2.2 Traditional and Emerging Techniques	21

2.2.1	Thresholding Based Techniques	22
2.2.2	Region and Contour Based Techniques	24
2.2.3	Statistical Based Method	30
2.2.4	Machine Learning Techniques	34
2.2.5	Deep Learning Techniques	39
2.2.6	Auto-encoder Techniques	44
2.2.7	Hybrid Techniques	48
2.3	Medical Imaging Systems for Efficient Handling of Big Data	52
2.4	Discussions	54
2.5	Summary	58
3	The Proposed Segmentation Approach	63
3.1	Background	64
3.1.1	k-means	64
3.1.2	Fuzzy c-means	65
3.2	The Proposed Approach	66
3.2.1	Preprocessing	67
3.2.2	The Proposed Weighted Fuzzy k-means Approach	68
3.2.3	Feature Extraction and Reduction	71
3.2.4	Classification	72
3.3	Experimental Setup and Data Acquisition	72
3.4	Results and Discussions	73
3.5	Summary	76
4	Detection Framework using Big Data	79
4.1	Hadoop-MapReduce System Model	80
4.2	The Hadoop and MDCS Based Architecture for MR images	81
4.2.1	WFKM on Hadoop with MDCS Platform	83
4.3	Experimental setup and results	85
4.3.1	Initial configuration and Dataset	85

4.3.2	Results and Discussions	87
4.4	Summary	89
5	Results and Discussions	91
5.1	Comparative Results for WFKM	91
5.1.1	Segmentation Results	92
5.1.2	Classification Results	98
5.2	Comparative Results of WFKM Integrated with Hadoop-MDCS . . .	103
6	Conclusions and Future Scope	107
6.1	Conclusions	107
6.2	Future Scope	109
	Appendices	111
A		111
A	Feature vectors	111
B		113
A	Segmentation using FKM	114
B	FKM with Hadoop-MapReduce	115
	List of Publications	146

List of Figures

1.1	Normal brain tissues after segmentation a. White matter (WM), b. Grey matter (GM), and c. Cerebrospinal fluid (CSF).	5
1.2	Different categories of imaging modalities used for the diagnosis of brain tumor.	7
1.3	Brain anatomy visualized in three planes using T2-W MR image (a) axial, (b) sagittal, and (c) coronal [1].	7
1.4	The MR image types (a) T1-W, (b) T2-W, (c) CE-T1W (d) FLAIR, (e) PD-W, (f) ADC, and (g) DWI [1].	8
1.5	Grading classification of brain tumor types.	10
1.6	Generic process flow of brain tumor detection system.	12
1.7	The general segregation of Hadoop-MDCS architecture for brain tumor detection.	13
2.1	Usage (in percentage) of different MR image modalities in brain tumor research domain based on studies considered in this research work. . .	19
2.2	Popularity (in percentage) of various data sources in brain tumor research domain based on studies considered in this research work. .	21
2.3	Classification of brain tumor segmentation techniques used in MR images.	22
2.4	(i) Manual segmentation by four different experts to detect glioma in MR image. (ii) Automatic Segmentation of glioma using improved GVF snake [2].	55

2.5	Difference between manual and semi-automatic segmentations. Images in sequence are original image, manual segmentation, semi-automatic segmentation, and difference between manual segmentation [3].	55
2.6	Satisfying segmentation of the tumour, regardless motion artefacts (top and middle), One of the worst cases of over-segmentation observed (bottom) [4].	56
3.1	The proposed system model for brain tumor detection using MR image.	66
3.2	Visual segmentation results on T1W MR image using Proposed hybrid WFKM for a. benign tumor and b. malignant tumor.	73
3.3	Performance evaluation of proposed WFKM using four different SVM classifier kernels.	75
3.4	Performance evaluation of WFKM in terms of accuracy with varying number of images.	76
4.1	Medical Image processing system architecture using Hadoop and MDCS platform.	81
4.2	Proposed framework of brain tumor detection using WFKM integrating Hadoop and MDCS platform.	83
4.3	Plot of system accuracy on train and validation datasets	86
4.4	Initial configuration of distributed parallel processing framework . . .	87
4.5	Performance evaluation of processing time on different nodes	88
5.1	Visual segmentation results of benign tumor on T1W MR image using a. k-means, b. FCM, c. Canny edge detection, d. FKM and e. Proposed hybrid WFKM.	93
5.2	Visual segmentation results of malignant tumor on T1W MR image using a. k-means, b. FCM, c. Canny edge detection, d. FKM and e. Proposed hybrid WFKM.	94

5.3	Visual segmentation results of soft fuzzy rough c-means (SFRCM) [5] and proposed hybrid WFKM.	95
5.4	Visual Segmentation results of High grade gliomas (HGG) using MR Images. a. k-means b. FCM, c. Canny edge detection d. FKM, and e. Proposed hybrid WFKM	96
5.5	Visual Segmentation results of Low grade gliomas (LGG) using MR Images. a. k-means b. FCM, c. Canny edge detection d. FKM, and e. Proposed hybrid WFKM	97
5.6	The segmentation of MR brain tumors using WFKM a) Meningioma, b) Astrocytoma, c) Glioblastoma Multiforme and d) Pilocytic Astro- cytoma	98
5.7	Linear classification accuracy of proposed hybrid WFKM compared with existing approaches	99
5.8	Quadratic classification accuracy of proposed hybrid WFKM com- pared with existing approaches	99
5.9	RBF classification accuracy of proposed hybrid WFKM compared with existing approaches	100
5.10	Poly classification accuracy of proposed hybrid WFKM compared with existing approaches	100
5.11	Performance Evaluation of standalone and multi-node system with respect to (a) Read time (b) Processing time and (c) Write time . . .	105
5.12	Performance comparison of processing time on different nodes	105
B.1	Proposed fusion based approach for brain tumor detection using Hadoop- MapReduce mechanism	113

List of Tables

1.1	Intrinsic Parameters	5
1.2	Extrinsic Parameters	6
1.3	Echo and Repetition times along with color palette of normal and abnormal brain tissues for each MR image type.	9
2.1	Datasets commonly used in brain tumor segmentation approaches. . .	20
2.2	Summary of a few studies utilized traditional approaches for brain tumor detection.	40
2.3	Summary of a few studies utilized deep learning approaches for brain tumor detection.	41
2.4	Summary of a few studies utilized Auto-Encoders for brain tumor detection.	46
2.5	Summary of a few studies utilized hybrid segmentation approaches for brain tumor detection.	51
2.6	Segmentation approaches and their values for various parameters iden- tified during the survey.	57
2.7	Summary of performances reported by approaches under each of the considered segmentation category grouped on tumor type.	60
4.1	Initial configuration table	86
5.1	Performance comparison with the existing works based on accuracy using same dataset.	102

5.2	Comparison of specificity, sensitivity, and dice coefficient index of the proposed scheme	103
A.1	Summarization of feature vectors	112

LIST OF ABBREVIATIONS

Acronym	Meaning
AAE	Adversarial auto-encoders
AANN	Adaptive Artificial Neural Network
AAS	Anaplastic astrocytoma
ACPSO-FNN	Adaptive chaotic particle swarm optimization forward neural network
ACC	Accuracy
ADC	Apparent diffusion coefficient
AFBNN	Adaptive firefly back propagation neural network
AMOSA	Archived multi-objective simulated annealing
ANFIS	Anisotropic diffused filtering with fuzzy logic
ANN	Artificial neural networks
AS	Astrocytoma
AS-FLGMM	Adaptive scale fuzzy local Gaussian mixture model
AT	Active tumor
AUC	Area under the curve
BAMM	Bayesian bounded asymmetric mixture model
BFO	Bacteria foraging optimization
BGGMM	Bounded generalized gaussian mixture model
BPNN	Back propagation neural network
BRATS	Brain tumor segmentation challenge
CA	Cellular automata
CAAE	Constrained adversarial auto-encoders
CARISI	Convolutional autoencoder-based inter-slice interpolation

Acronym	Meaning
CAD	Computer aided detection
CBAC	Content based active contour
CBMRS	Content based medical image retrieval system
CCA	Canonical correlation analysis
CED	Canny edge detection
CE-MRI	Contrast enhanced magnetic resonance imaging
CGMM	Constrained gaussian mixture model
CMLS	Cognition based modified level set
CNN	Convolutional neural networks
CRF	Conditional random fields
CSF	Cerebrospinal fluid
CT	Compute tomography
CVAEs	Condtonal variational auto-encoder
DAEs	Denoising auto-encoders
DCNNs	Deep convolutional neural networks
DF	deformable model
DICOM	Digital imaging and communications in medicine
DNN	Deep neural networks
DSC	Dice similarity coefficient
DS1	Data source 1
DS2	Data source 2
DS3	Data source 3
DTI	Diffusion tensor image
DWI	Diffusion weighted imaging
DWT	Discrete wavelet transform
EBP	Evidence-based practices
EBS	Edge based segmentation

Acronym	Meaning
ECNN	Enhanced convolutional neural network
EM	Expectation maximization
EPFCM	Enhanced possibilistic fuzzy c-means
ETM	Simple modified expectation maximization
FCM	Fuzzy c-means
FCMM	Modified fuzzy c-means
FEM	Finite element method
FKM	Fuzzy c-means and k-means
FLAIR	Fluid attenuated inversion recovery
FN	False negative rate
FP	False positive rate
fMRI	Functional magnetic imaging
FPANN	Feed forward back propagation artificial neural network
FPCNN	Feedback pulse-coupled neural network
FSVM	Fuzzy support vector machine
GA	Genetic algorithm
GAN	Generative adversarial net
GBM	Glioblastoma multiforme
GFS	Google file system
GM	Grey matter
GMM	Gaussian mixture model
GN	Granuloma
GS	Gold standard
GVF	Gradient vector field
HCP	Human connectome project
HDFS	Hadoop distributed file system
HEM	Hybrid expectation maximization

Acronym	Meaning
HGA	High grade astrocytoma
HGG	High grade glioma
HH	High-high intensity region
HL	High-low intensity region
HLS	Hybrid level set
HMMs	Hidden markov models
HMRF	Hidden markov random field
IBSR	Internet brain segmentation repository
ICA	independent component analysis
ICM	Iterated conditional modes
IDM	Inverse difference moment
ISLES	Ischemic stroke lesion segmentation
KIFCM	k-means fuzzy c-means
JC	Jaccard Coefficient
KNN	K-nearest neighbor
LGG	Low grade gliomas
LH	Low-high intensity region
LL	Low-low intensity region
LSFT	Least squares features transformation
LVQ	Learning vector quantization
MBC	metastatic bronchogenic carcinoma
MCBE	Multi-center based encoding
MCWS	Marker-controller watershed segmentation
MDCS	MATLAB distributed computing server
MEN	Meningioma
MET	Metastatic
MFS	Magnetic field strength

Acronym	Meaning
MFKM	Modified fuzzy k-means
MOO	Multiobjective optimization
MJS	Matlab job scheduler
MRA	Magnetic resonance angiography
MRF	Markov random field
MRI	Magnetic resonance imaging
MSE	Mean square error
NBIA	National biomedical imaging archive
NBTR	National brain tumor foundation
NEM	Neighborhood expectation maximization
NITRC	Neuroimaging tools and resources collaboratory
NON	Normal dataset
OASIS	Open access series of imaging studies
PC	Personal computer
PCA	Principal component analysis
PCC	Pearson's Correlation Coefficient
PCNN	Pulse-coupled neural network
PCT	Parallel computing toolbox
PD	Proton density
PD-W	Proton density weighted
PGA	Parallel genetic approach
PICE	Prior information constraint evolution
PNN	Probabilistic neural network
PSNR	Peak signal-to-noise ratio
PT	Pituitary tumors
RAF	Radiofrequency
RBF	Radial basis function

Acronym	Meaning
RET	Rough entropy based thresholding
RF	Random forest
RG	Region grow
ROI	Region of interest
SA	Simulated annealing
SCGAGMM	Asymmetric gaussian mixture model with spatial constraint
SCGM-EM	Spatially constrained generative model expectation maximization
SDAEs	Stacked denoising auto-encoders
SEN	Sensitivity
SFRCM	Soft fuzzy rough c-means
SGLDM	Spatial gray level dependence method
SI	Similarity Index
SIFT	Scale invariant feature transform
SOM	Self-organized maps
SPEC	Specificity
SVM	Support vector machine
TCIA	The cancer imaging archive
TC	Tumor core
TE	Echo time
TI	Inversion time
TN	True negative rate
TP	True positive rate
TR	Repetition time
T1-W	T1-weighted
T2-W	T2-weighted
URL	Uniform resource locator
VAE-GAN	Variational auto-encoder-generative adversarial network

Acronym	Meaning
WFKM	Weighted fuzzy k-means
WHO	World health organization
WM	White matter
WMMFCM	Multi-resolution wavelet morphological pyramid fuzzy c-means
WNDCHRM	Weighted neighbor distance using compound hierarchy of algorithms representing morphology
WT	Whole tumor

Nomenclature

Notations	Abbreviation
c	cluster center
d	Euclidean distance
i or k	Cluster index
j	Image number/Pixel index
l	Illumination point.
m	Fuzziness exponent.
m_i	Number of objects in the cluster
n	Total number of pixels
p	Probability
r or $\mathbb{R}$	Region.
s_k	Cumulative probability
th	Threshold to check the fitness of a cluster.
u_{ik}	Represents degree of membership of pixel I_i in cluster C_k .
w	Weight of the pixel
z	Iteration number
μ	Mean intensity
ρ	Pixel intensity
σ_B^2	Class variance of the image

Notations	Abbreviation
C_k	Represents centre of k^{th} cluster.
C_p	Cluster pool
F	Array of $c \times 1$ to store fitness of each cluster
G	Number of grey levels in the image
HH	High-High intensity region.
HL	High-Low intensity region.
I	Image
I_i	i^{th} pixel of an image I.
I_i^k	Image's i^{th} pixel belongs to k^{th} cluster
LH	Low-High intensity region.
LL	Low-Low intensity region.
P_{file}	Pixel file
U	Represents u_{ik} matrix store values of pixels of each cluster.
W	Array of $c \times 1$ to store weights of each cluster.
X	Data set of different data points of k^{th} cluster.
$\mathbb{C}$	Cluster
$\mathbb{J}$	Objective function
$\mathbb{M}$	Distance form hyper-plane to origin
$\mathbb{P}$	Pixel
$\mathbb{U}_{mem}$	Membership function
$\mathbb{W}$	Weights
$\mathbb{X}$	Decision boundary for classes
$\mathfrak{C}$	Classification classes

Chapter 1

Introduction

Brain tumor and its analysis is of great interest due to growing technology in medical image processing. According to the survey conducted by National Brain Tumor Foundation (NBTF), the development of brain tumor amongst masses and the death rate due to brain tumor is increasing comparatively across the globe. Every five year brain tumor patient's survival statistics is conducted (recently conducted in January 2020) [6]. Also in past few decades several works have proposed frameworks or approaches to highlight brain tumor area which may or may not be followed by phases like classification, treatment planning, and outcome predictions. Segmentation in medical images is generally governed by factors like poor contrast, noise, and missing boundaries [7] [8] [9]. For effective control of these factors, diagnosing imaging processes like Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) scan, are employed. These imaging processes are helpful in detecting different types of diseases [10]. Usage of harmless magnetic fields and radio waves makes MR images very popular in successfully diagnosing and treating brain tumors [1].

1.1 Brain Tumor - Characteristics

Life threatening brain tumor compresses/displaces brain tissues to make room for the growing mass. Tumor can occur in any body part hampering the normal working

of the system but is fatal when arise in brain. Different types of brain tumors exist varying in shapes, sizes, and/or positions which have to be dealt accordingly [11]. The abnormal tissues can be labeled as primary and secondary brain tumors that, based on their characteristics even get streamlined further. Primary tumors are formed when cancer tissues initiates and stays at the same place as well as position. The secondary (or metastatic) tumors are those which spread to other body parts from where they are emerged. As per the terminology of radiologists, primary tumor is either benign (slow growing, less aggressive, and noncancerous) or malignant (rapid growing, life threatening, and cancerous). In addition to these primary information, the treatment planning depends on grade and tissue cell information which is usually collected during a process called biopsy. Tissue cell type identifies those cells that had given rise to the tumor and tumor grade estimates its aggressiveness. Tumors often have different grades at various stages of their growth [12] [13] [14].

1.2 Brain Tumor - Techniques

Based on human interactivity, MR images can be processed using manual [3] [15] [2], semi-automatic [16] [17] [18], and fully automatic techniques [19] [20] [21] [22] [23]. In medical image processing, segmentation should be accurate which is thus commonly performed by experts manually and hence time consuming. On the same hand, designing a fully automatic and efficient segmentation approach is yet far from reality. Moreover, such systems always demand for a second opinion as the main object here is human life [24]. Also in the absence of experts, performance of automatic techniques depends solely on the knowledge bases. Researchers have proposed several methods to improve such knowledge bases and thus the capability of tumor detection systems. The clinical and pathological acceptance of segmentation techniques solely rely on the computation ease and the degree of user supervision [25]. Manual methods require complete user supervision so an urge for better solutions rises. The existing semi-automatic or interactive methods definitely dominate in practice

currently as well as in coming time. Semi-automatic segmentation methods are popularly adapted in clinical applications by radiologist to replace time consuming and challenging tasks of manual segmentation process [26] [27]. User initialization is required in almost all semi-automatic methods and to assure accuracy repeated user interaction is mandatory. Under such circumstances, the work presented in this thesis also attempts to visualize the acceptance and growth of such automatic/semi-automatic systems in medical as well as research domain for various tumor types.

This chapter initiates with aim to introduce the background related to brain tumor and its characteristics in Section 1.1. Section 1.2 discusses current trends in the brain tumor detection techniques especially for MR images. Section 1.3 highlights the motivation of proposed research work. Section 1.4 describes commonly used MR image modalities and the popular contrast parameters. Section 1.5 briefs the grading and classification of brain tumor types/diseases based on WHO guidelines followed by a generic architecture of MR image based brain tumor detection system, covering all the basic modules in Section 1.6. It also describes the integration of Hadoop and MDCS for brain tumor detection using MR images. Objectives and scope of the work are given in Section 1.7. Lastly, Sections 1.8 briefs organization of rest of the thesis chapters.

1.3 Motivation

One of the preliminary phases in most of the medical image analysis tools is segmentation which has also become a challenging and dynamic research field in the domain of image processing. A variety of state-of-the-art methodologies with good prior knowledge for brain MR image segmentation prevails. However, segmentation using MR brain images has to be as efficient as possible and hence there still exists scope for improved accuracy and/or computation speed of existing systems. Furthermore, brain tumor segmentation is challenging due to variations in tumor shape, appearance, size, and position. An expanding tumor often deforms normal tissue

thereby giving an abnormal geometry. In addition, intensity values of tumor overlap with those of normal tissues. Researchers have extensively explored segmentation in brain MR images to properly deal with the said issues and the work presented in this research is in-lined with the same.

Segmentation is followed by the tasks of extraction and then detection of tumor type. These are important for proper clinical examination as well as treatment. For the brain tumor detection, various traditional and emerging techniques have been proposed [28] [29] [27] [30] [31] [32] [33]. This research targets few popular types of brain tumors and the methodologies applied for their detection. The correlation between tumor data source, tumor type, and detection technique is proximate to control the efficiency. This further leads to increase in knowledge about the relationship between various tumor disorders with anatomical deviation. Overall segmentation is a prior phase in tumor detection as well as analysis tools and thus needs the highest precision. An attempt to develop a hybrid approach combining the concepts of clustering with fuzzification is made to automate the delicate process of segmentation aiming to achieve better efficiency. The proposed algorithm is also utilized in designing frameworks incorporating innovative detection and classification techniques for brain tumor MR images.

1.4 Imaging Modalities and Contrast Parameters

Segmentation of MR images needs to be very efficient for proper analysis of tumor. A range of works have utilized the capabilities of image segmentation techniques to extract the meaningful content from medical images (i.e., tumor) which thus helps in proper analysis of brain tumor. But, the variability of tumor shapes and the presence of cerebrospinal fluid (CSF) particularly make a difficult brain tumor detection task more complex. No matter which segmentation technique is employed, quality of segmentation depends greatly on contrast of medical images, amount of noise, and lastly incomplete boundaries. Particularly, in medical images good con-

Table 1.1: Intrinsic Parameters

Proton density (PD) measures relative concentration of protons in the concerned tissue to that in water at same temperature and volume.

Longitudinal relaxation time constant (T1) is defined as the time when 63% of the longitudinal magnetization has recovered. It measures the re-growth of longitudinal magnetization and is also known as *spin – lattice relaxation*.

Transverse relaxation time constant (T2) is defined as the time when 63% of the transverse magnetization has decayed. It measures the decay of transverse magnetization and is also known as *spin – spin relaxation*.

Chemical shift arises due to the differences in resonance frequencies of fat and water. This MR image artifact happens because the water and the fat structure shift with different frequencies in the organ.

Susceptibility indicates whether a tissue attracts or repels out in the magnetic field. It may also be defined as the extent to which a tissue gets magnetized when placed in an external magnetic field.

trast supersedes all other basic requirements, as an abnormal structure is identified completely on the basis of certain contrast characteristics. Contrast is defined as a function of tissue density in MR images [20] [34]. In other words, good image contrast is a native property of brain MR images and thus their generation do not need any kind of contrast enhancing agents. This fact is also responsible for the popular usage of these images in brain tumor detection. However, image contrast is influenced by intrinsic (Table 1.1) and extrinsic (Table 1.2) parameters.

Proper tuning of these parameters helps to change the contrast behavior and

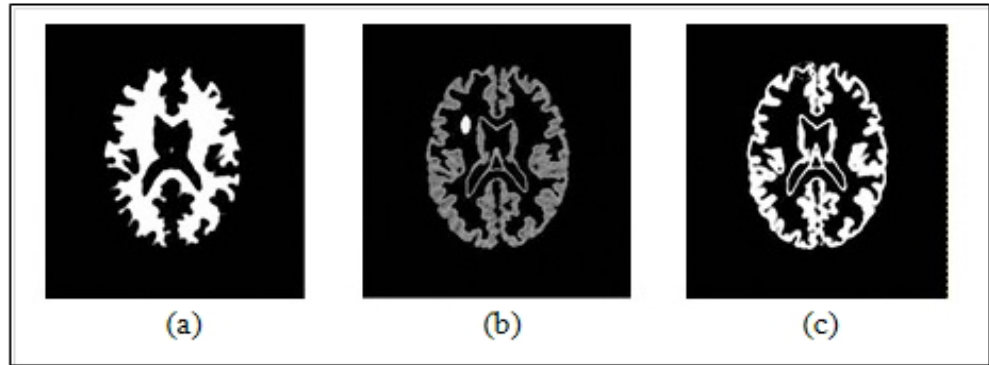


Figure 1.1: Normal brain tissues after segmentation a. White matter (WM), b. Grey matter (GM), and c. Cerebrospinal fluid (CSF).

Table 1.2: Extrinsic Parameters

Magnetic field strength (MFS) should be both strong and uniform MFS to draw proper conclusions. Its measuring unit is Tesla (T) and most of the systems work at 1.5T.

Radio frequency (RAF) is an electromagnetic wave with frequency range similar to that of radio and television transmissions. The prone tissue gets heated and highlighted by absorbing the pulses of the applied radio-frequency energy. The preferred range for RF pulses is 1 MHz -100 MHz.

Repetition time (TR) is the time that exists between succeeding pulse sequences of the same curve. It determines the longitudinal magnetization between each pulse. Image contrast properties are controlled by TR variations.

Echo time (TE) refers to the time between the radio-frequency excitation pulse and the signal peak induced in the coil.

Inversion time (TI) is the time period between the initial 180° RAF pulse (to invert the magnetization) and the 90° RF pulse (to get the residual longitudinal magnetization into the transverse plane). This pulse sequence is the core component of an inversion recovery pulse sequence.

RF pulse amplitude (flip angle) is the angle that rotates the net magnetization relative to the main magnetic field direction by 120°, applying the RF excitation pulse.

b-value measures the strength and timing of the gradients applied to generate the diffusion-weighting images. It is used to control the effect of diffusion, for example, higher b-value is needed for stronger diffusion effects to sense the slow moving water molecules.

highlight the regions of interest, i.e. differentiate tumor tissues from the normal (inner) brain tissues (fat, gray matter - GM, white matter - WM, and cerebrospinal fluid - CSF) [12]. The visual representation of normal brain tissues is depicted in Fig. 1.1. Anatomy of MR image modalities can be visualized in three different planes namely, axial, sagittal, and coronal as depicted in Fig. 1.3.

Imaging modalities used for the diagnosis of brain tumor can be categorized broadly as structural and functional magnetic resonance imaging as shown in Fig.1.2. In structural, a non-invasive process is applied to examine anatomical structure of brain. The brain image modalities, viz. T1-weighted (T1-W), T2-weighted (T2-W), fluid attenuated inversion recovery (FLAIR), proton density (PD) and magnetic resonance angiography (MRA) cover the structural technique. MRA is based on the

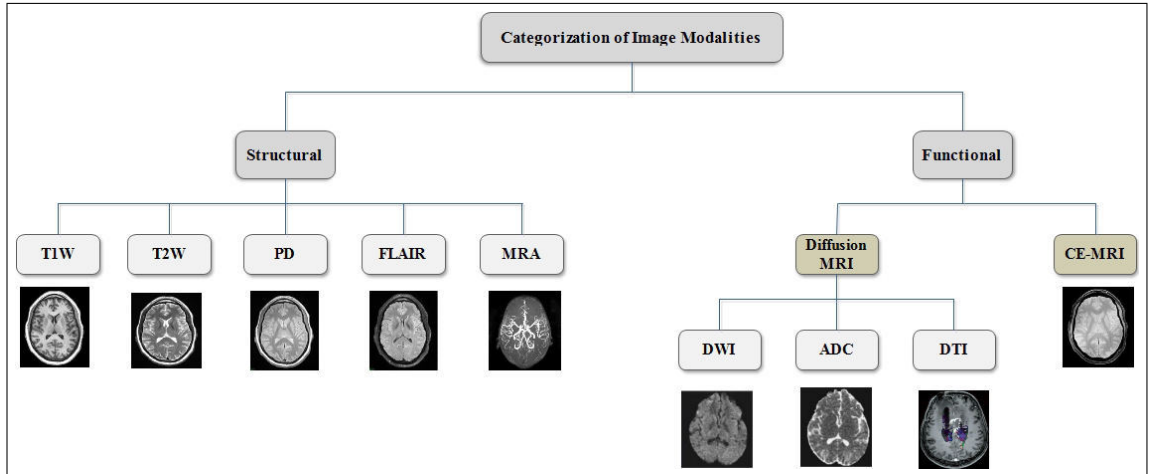


Figure 1.2: Different categories of imaging modalities used for the diagnosis of brain tumor.



Figure 1.3: Brain anatomy visualized in three planes using T2-W MR image (a) axial, (b) sagittal, and (c) coronal [1].

vascular anatomy imaging examination and shown in categorization only, however, its extended description is not considered in this research work. The second category i.e. functional MR imaging (fMRI) on the other hand measures human brain function or metabolic activity associated with blood flow change in the brain. Diffusion MRI and contrast enhanced MRI (CE-MRI) are the two sub-categories that lie under functional. Further, diffusion MRI is categorized as diffusion weighted image (DWI), apparent diffusion coefficient (ADC), and diffusion tensor image (DTI).

Differences in the compositions of tissue types result in varying MR image signals. Accordingly, different types of pulse sequences are used to emphasize various aspects of normal and tumor tissues. Variable sequences have different completion time and also they provide diverse valuable information. For the purpose of clarity, a few

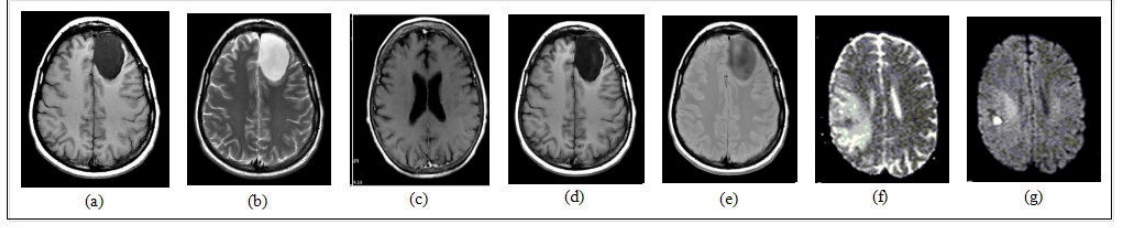


Figure 1.4: The MR image types (a) T1-W, (b) T2-W, (c) CE-T1W (d) FLAIR, (e) PD-W, (f) ADC, and (g) DWI [1].

images from each of the MR image types considered in this research are shown in Fig. 1.4. Depending upon the pulse sequence used an MR image can be T1-W or T2-W. Sometimes to get better results a special pulse sequence is employed to obtain FLAIR) or PD-W. Apart from these some diffusion based MR images like DWI and ADC are also popularly used. DWI and ADC are particularly useful in early diagnosis and longitudinal evaluations. During the study, it is observed that a variation of T1-W obtained using a contrast agent Gadolinium is also popular in the community. It is known as CE-T1W and is similar to T1-W with enhanced clarity as is shown in Fig 1.4.

Proper tuning of MR imaging parameters helps in differentiating normal and abnormal brain tissues based on color. Table 1.3 summarizes values of a few contrast parameters (echo time - TE and repetition time - TR) for all the seven MR image modalities and colors of normal (fat, GM, WM, and CSF) along with abnormal brain tissues. Specific values of these parameters are available for all MR image types, except DWI and ADC. Instead of using single TE and TR values, most of the existing works vary these parameters within certain limits based on the protocol being followed while working with DWI and ADC MR images. The focus of any system is to effectively detect the tumor tissue by separating it from other brain tissues and which are commonly differentiated based on colors. A color palette is thus included for better visualization of colors associated with normal and tumor tissues in each MR image type.

Table 1.3: Echo and Repetition times along with color palette of normal and abnormal brain tissues for each MR image type.



	White	Light Grey	Medium Grey	Grey	Medium Black	Black
MR Image	T1-W	T2-W	FLAIR	PD-W	DWI	ADC
TE	<30mm	>80mm	>80mm	<30mm	User Dependent	
TR	<80mm	>2000mm	>3000mm	>1000mm	User Dependent	
FAT	White	Medium Black	Grey	Light Grey	Medium Black	Medium Black
White matter	Medium Grey	Grey	Medium Grey	Medium Grey	Grey	Light Grey
Grey matter	Grey	Medium Grey	Light Grey	Light Grey	Medium Grey	Medium Grey
CSF	Black	White	Black	Grey	Black	White
Tumor tissue	Medium Black	Medium Grey	White	White	White	Black

1.5 Tumor Types and their Grades

A comprehensive study of multi-grade brain tumor detection approaches specifically for MR images is undertaken in this research. It has been observed that the performance of segmentation or classification approach follows a particular trend depending on tumor category contained in the collected dataset or acquired brain images. The World Health Organization (WHO) guidelines for brain tumor grading categorization are totally histopathological, which however hinders the medical imaging diagnosis and detection approaches [35]. In addition to these tumor types, Alzheimer's disease [36] [37] [38], and Lesions [39] [4] [40] are also considered based on their data sources and detection techniques. The categorization of brain tumor types is shown in Fig. 1.5

Keeping this constrict in mind an attempt to develop computer aided detection (CAD) is gaining insight by many researchers. The effectiveness of detection system

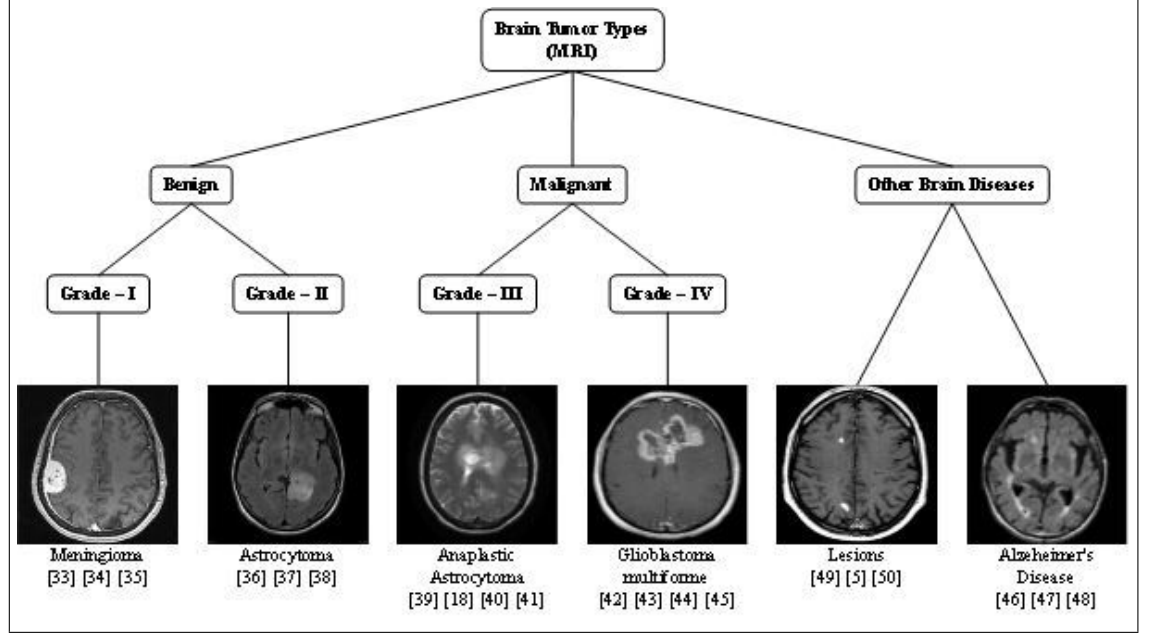


Figure 1.5: Grading classification of brain tumor types.

rely on background knowledge related to MR imaging contrast and its modality, tumor type/disease and data source used in developing a system. Therefore, all the studies covering said aspects have been taken into considerations. The performance evaluation of existing approaches corresponding to various tumor types highlighting the effectiveness of overall framework is summarized. This research work focuses on popularly existing tumor categories, namely Meningioma (MEN) [41] [42] [43], Astrocytoma (AS) [44] [45] [46], Anaplastic Astrocytoma (AAS) [47] [21] [48] [49], and Glioblastoma multiforme (GBM) [50] [51] [52] [53] with prospective to each of the four grades. The study targets towards providing wider options to the experts, radiologists, surgeons, and researchers for the analysis of detection techniques best suited for multi-grade tumor types classes.

1.6 Generic Brain Tumor Detection System

The starting module is collection and acquisition of images which are to be analyzed in later phases. Standard brain tumor databases, namely BRATS, Harvard medical school, BrainWeb, and Internet Brain Segmentation Repository (IBSR),

are commonly used by the researchers. Moreover, studies have validated their approaches on data repositories collected from various clinical and pathology labs. The pre-processing module is meant to improve image visual quality by filtering all the high noise levels. Presence of such disturbances makes abnormal/normal tissue discrimination relatively difficult and thus the precise interpretations. Next is the segmentation module that highlights region of interest and guides all the proceeding modules, especially analysis. Although segmentation is important but it may not be present sometimes when direct classification supersedes other objectives. The feature extraction module follows segmentation. A variety of standard texture features along with other information like intensity and edge, are commonly used. Robust learning models are built by picking a sub-class of relevant features and exempting the irrelevant ones. This is successfully achieved by utilizing standard dimension reduction techniques such as principle component analysis (PCA), independent component analysis (ICA), canonical correlation analysis (CCA), etc. Finally, the resultant features with reduced dimension are used for classifier training to differentiate normal brain tissues from abnormal ones. After training, performance of the system is evaluated on several parameters (such as accuracy, similarity, sensitivity, dice coefficient, visual comparison, etc.). Once output is retrieved, the system is set to formulate and interpret the desired diagnosis result. To recapitulate, an expert advice is always preferred to have a second opinion for prefect diagnosis. Systems reporting improved performances, i.e. refined results assist medical specialists in making better interpretations (Fig. 1.6).

1.6.1 Integration of Hadoop with MDCS for MR images

In addition to higher accuracy rates, another important aspect in CAD systems is lower processing time. Several studies have brought together different traditional as well as emerging approaches to achieve an optimum output with respect to both, i.e. precision and time. Availability of numerous diagnostic centers, laboratories, hos-

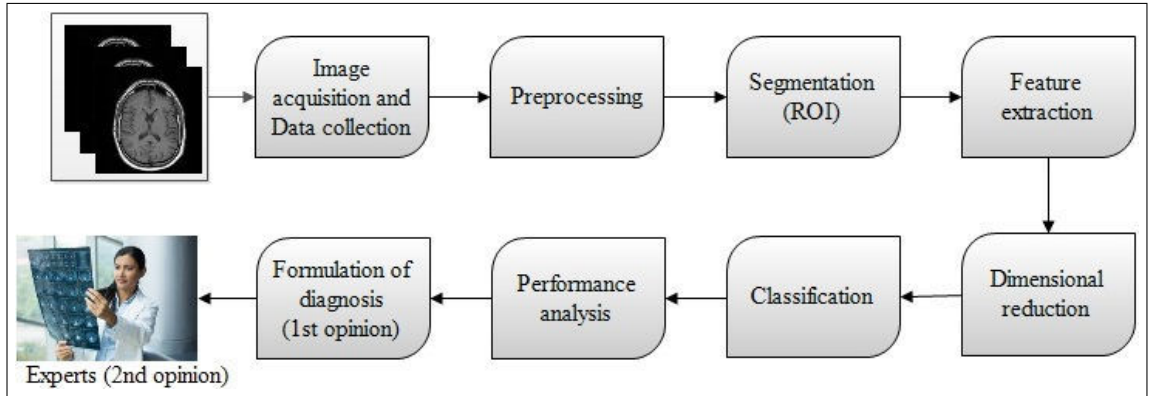


Figure 1.6: Generic process flow of brain tumor detection system.

pitals, or online data repositories for storage, testing, experimentation, and scalable analysis, makes the rate of growth of MR brain image data exponential. Swift handling of this massive data in an organised and parallel way is realized via multiple cloud and network based platforms [54]. The era of cloud is characterized as excessive computational speed along with efficient data storage [55] [56]. Popular in the category is Hadoop, a multi-node, parallel, and distributed framework that increases scalability by managing data in portions and distributing it among the DataNode systems. Hadoop distributed file system (HDFS) exhibits worthwhile features for scalable storage and aids all the data files intact. The programming paradigm, MapReduce manages similar behavioral jobs based applications using parallel and distributed mechanisms [57]. The MapReduce first forms numerous chunks of data, and then processes them through concurrent map function, further the retrieved output is sorted and forwarded to the next function, that is, the reducer. The resultant is always preserved in the file storage named HDFS.

Popularity of real-time applications augmented such environments which in turn increased storage capabilities to process big data on large scale [58] [59]. The environment such as Cloud, Matlab distributed computing server (MDCS), and Hadoop enhances computation of brain tumor MR image data [60] [61]. The distributed data on HDFS is processed via configuring MDCS containing NameNodes and slave nodes is shown in Fig. 1.7.

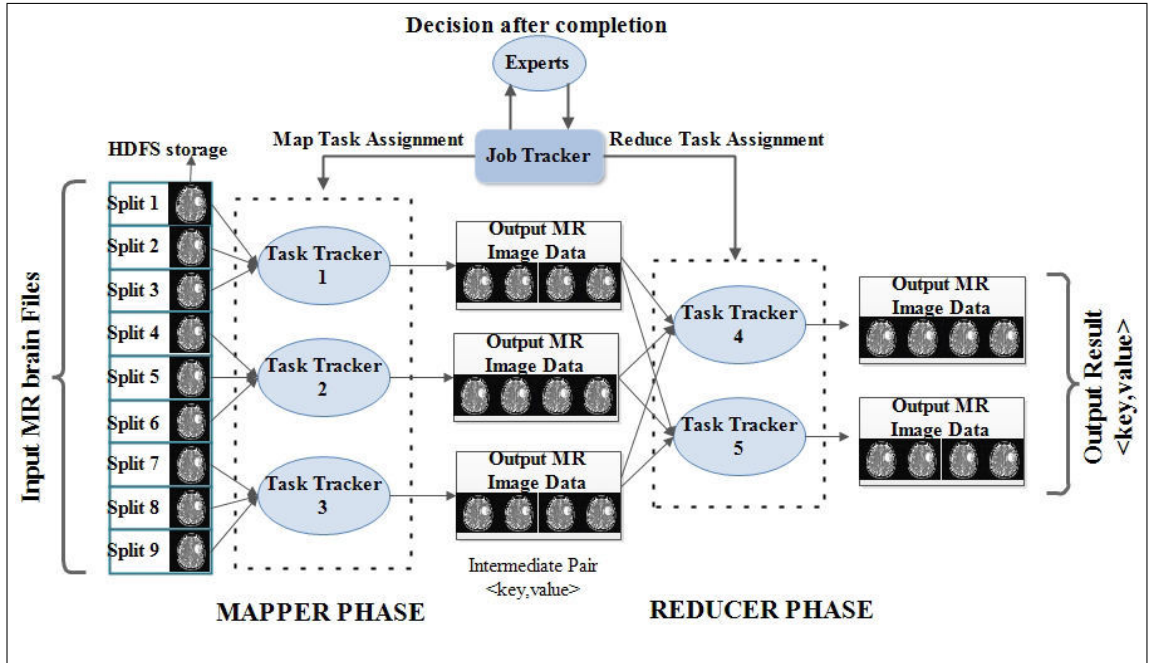


Figure 1.7: The general segregation of Hadoop-MDCS architecture for brain tumor detection.

1.7 Objectives and Scope of the Work

The research work aims to design and implement a semi-automated framework for the detection and classification of brain tumor using MR images. Performance improvements, as underlined in the existing literature Chapter 2, via introducing fuzziness is the main motivation behind the work presented in this thesis. The objectives of this work to strive the effectiveness of the proposals have been illustrated as below.

To study and analyze various existing techniques to detect brain tumor. One of the most crucial tasks in any brain tumor detection system is the isolation of abnormal tissues from normal brain tissues. Therefore, a comprehensive review of traditional as well as emerging machine learning and soft computing approaches in the field of medical image processing, especially brain tumors, studying their pros and cons is essential. Also, studying the related existing works helps in understanding the current research scenario, specifically techniques utilized so far in brain tumor detection and thus to improvise the methods that would assist in the

design and development of efficient frameworks.

To propose an efficient framework for the detection of brain tumor. Existing systems for tumor detection lack in aspects of spatial context and illumination due to multiple memberships, that degrade segmentation and hence the whole performance. Objective is thus to automate the delicate process of segmentation in MR images aiming to achieve a framework having better efficiency in terms of accuracy as well as computation time. It is to be noted that irrespective of the performances reported for either of the two parameters by any tumor detection system, a second opinion is still required for better diagnosis in any of the cases.

To validate the proposed framework. The proposals are tested on freely available brain tumor image dataset namely, The Cancer Imaging Archive (TCIA) obtained in the format of Digital Imaging and Communications in Medicine (DICOM). Along with this the proposals, particularly for glioma tumor type (HGG and LGG), are also validated and tested on different datasets collected from Google image search engine.

1.8 Organization of the Thesis

The research work presented in the thesis is organized in 6 chapters. A brief narration of all the chapters is as follows:

Chapter 1 introduces life threatening brain tumor and absolute need for its early detection. Predominance of MR imaging contrast, modalities, and its variability is also defined together with the tuning tissue parameters influencing the type of MR image. Further, this chapter highlights brain tumor types, grading, and classification as well for proper tumor detection. It presents a generic architecture of an image based brain tumor detection system and its incorporation with Hadoop-MDCS framework. Specifying objectives and extent of the work, the chapter ends describing the thesis organization. The work accomplished in Chapter 1 covers the Objective 1 partially.

Chapter 2 presents a discussion on various segmentation/classification techniques explored by the researchers for an accurate identification of brain tumor. Complete coverage of all categories of tumor detection techniques has been done to understand the performance differences. Based on the observations made, an attempt to identify the most popular MR image modality and commonly used brain image dataset is done. Performance of state-of-the-art techniques are analyzed to identify those that work well across various brain tumor categories. Chapter 2 has been published as a research article titled ‘A survey on brain tumor detection techniques for MR images’ in *Multimedia Tools and Applications*, Springer. The work done in this chapter covers Objective 1 of the thesis work.

Chapter 3 proposes a novel hybrid segmentation algorithm based on the concept of weights, termed as weighted fuzzy k-means for the efficient detection of brain tumor and covers partially Objective 2 and Objective 3 of the research. The presented approach excels standard approaches in the process of segmentation as well as classification of brain tumor MR images. This is however achieved by improving the clustering method of segmentation via assigning quantization weights to the pixel values. This chapter (Chapter 3) has been published as a research article titled ‘A Hybrid Weighted Fuzzy Approach for Brain Tumor Segmentation using MR Image’ in *Neural Computing and Applications*, Springer.

Chapter 4 focuses on efficient handling of massive MR image data, i.e. high accuracy and low execution time, which is one of the significant concerns in any framework. It analyzes the performance of varying sized datasets by executing the proposed WFKM on a combination of Hadoop-MDCS architecture that effectively deals with huge MR brain image data by utilizing clusters of commodity computers. To recapitulate, the traditional segmentation approaches cover either of the performance parameters and this collaboration heads to target both of them (accuracy and execution time). This chapter (Chapter 4) has been published as a research article titled ‘An efficient Hadoop-based brain tumor detection framework using big data analytics’ in *Software: Practice and Experience*, Wiley. The work done in this

chapter covers Objective 2 and Objective 3 partially of the research work.

Chapter 5 discusses the experimental results to further validate the frameworks proposed in Chapter 3 and Chapter 4. A comparative performance analysis in terms of accuracy as well as computation time with the traditional existing techniques is provided. This chapter (Chapter 5) is collectively covered in both the technical papers.

Chapter 6 concludes purpose of the research work summarizing the highlights and main contributions of the thesis. A deliberation with prospective to future research directions is also drawn along with the summarization of objectives accomplished.

Chapter 2

Literature Review

The existing benefits of MR imaging over other diagnostic imaging techniques, as stated in Chapter 1 drives the basis of this research on MR image based brain tumor detection techniques. This chapter thus aims in reviewing brain tumor detection trends applied after data collection and image acquisition. A comprehensive review of traditional as well as current approaches in the field of medical image processing, especially brain tumors, highlighting their pros, cons, and future challenges/trends is presented [21, 44]. Emerging machine learning and soft computing approaches have created a remarkable impact in medical imaging too [72, 79]. As highlighted in one of the recent reviews, deep learning based brain tumor segmentation is in trend [7]. Thus, summarization together with a discussion of recent studies that have explored deep learning architecture, its mechanism, performance and quickness from the perspective of their quantitative utility in MR image based tumor identification systems is also included. Along with a detailed assessment of current scenario, highlights to future directions are also provided. Classification is driven by segmentation and there are many works that target on efficient segmentation which may or may not be followed by classification. Thus, the chapter discusses approaches individually from both the phases and their combinations too, i.e. hybrid systems. Section 2.1 of the chapter attempt of this study is to highlight MR image modalities that are frequently used in the research domain of brain tumor

detection. Segmentation is a significant step in automating the process of brain tumor detection. Researchers have extensively explored various existing techniques and tried to introduce fuzziness at several levels for better accuracy and is covered in Section 2.2. The subsection of techniques discuss the performance of different deep learning classifiers as well with respect to various tumors. Section 2.3 briefs about the architecture of Hadoop-MapReduce incorporated with medical imaging segmentation approaches. Observations extracted after survey of various systems based on human intervention has been covered in Section 2.4. Further, Section 2.5 summarizes the chapter based on the traditional and emerging techniques applied to detect brain tumor using different MR modalities and data sources.

2.1 A Few Statistics - MR Image Modalities and Data Sources

Before looking at the existing literature, let's first have a look on the information related to MR image modalities and data sources utilized in various studies within the domain of brain tumor detection. The different MR modalities vary from each other on the basis of various parameters, i.e. intrinsic and extrinsic (discussed in Chapter 1, Table 1.2 and 1.2). The easily available and commonly worked MR image sequences are T1-W and T2-W. The researchers working on traditional techniques retrieve high accuracy rates with T2-W MR sequences even more than 90% but steadily this trend has changed and the demand of systems handling all types emerge. The same can be observed in Table 2.3 and Table 2.4 where T1-W, T2-W, CE-T1W, and FLAIR are mostly utilized. Fig. 2.1 depicts the usage trend (in percentage) of various MR images and is prepared based on the papers reviewed in this research. No doubt the proportion of DWI is very small, however based on the observations made it can be said that in coming scenario the diffusion based will be preferred due to its sensitivity in detecting even minute infarcts as compared to T1-W and

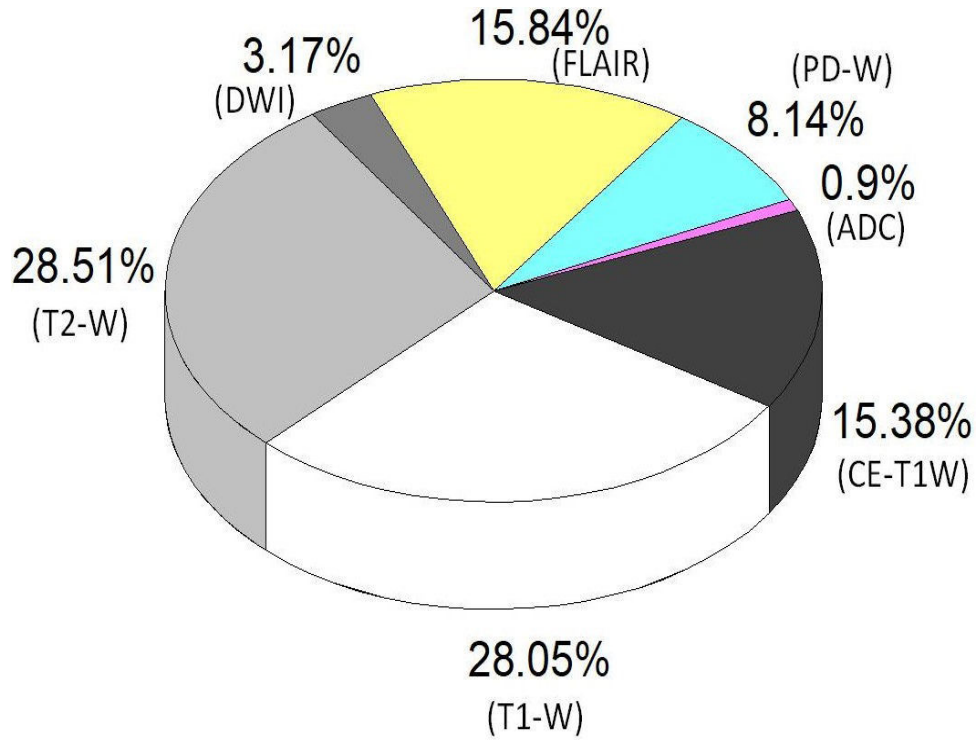


Figure 2.1: Usage (in percentage) of different MR image modalities in brain tumor research domain based on studies considered in this research work.

T2-W, which are the current winners.

Though the research is focused on tumor detection, but the scarcity of freely available medical data due to security reasons sometimes limit the important task of experimentation. Thus, based on the papers reviewed, a list of data sources having various MR images with modalities is presented in Table (2.1). In addition, the popularity of these data sources are also identified and the same is shown in Fig. 2.2. Another popular category is ‘private datasets’ which represents real data collected from clinical laboratories or pathology or hospitals. Clearly, the mostly used one is the real/private datasets followed by BRATS. The reason behind is an urge to prove the acceptability of the proposed approach in real-time scenario.

Table 2.1: Datasets commonly used in brain tumor segmentation approaches.

Name	Description	Web Source
BRATS	Brain Tumor Segmentation Challenge (BRATS) has the dataset available from 2012 to 2018.	https://www.smir.ch/BRATS/Start2012
BrainWeb	Simulated MRIs of the human brain.	http://www.bic.mni.mcgill.ca/brainweb/
IBSR	MGH CMA Internet Brain Segmentation Repository (IBSR). Repository with raw and segmented (GM/WM/CSF) MRI volumes.	http://neuro-www.mgh.harvard.edu/cma/ibsr/
OASIS	Open Access Series of Imaging Studies. Structural MR from 416 subjects.	http://www.oasis-brains.org
Harvard medical school	Download includes the collection of electronic resources of image data, slides, presentation and software provided by NCIGT.	http://ncigt.org/downloads
NBIA	National Biomedical Imaging Archive (NBIA) is a searchable repository of vivo images providing biomedical research and academia with access to image archives.	https://imaging.nci.nih.gov/ncia/login.jsf
TCIA	The Cancer Imaging Archive provides access of public MRI/CT scan image data.	https://wiki.cancerimagingarchive.net
ISLES	Ischemic Stroke Lesion segmentation is challenge of medical image. In ISLES 2016 publicly available cases 35 training and 40 testing.	https://www.smir.ch/ISLES/Start2016 .

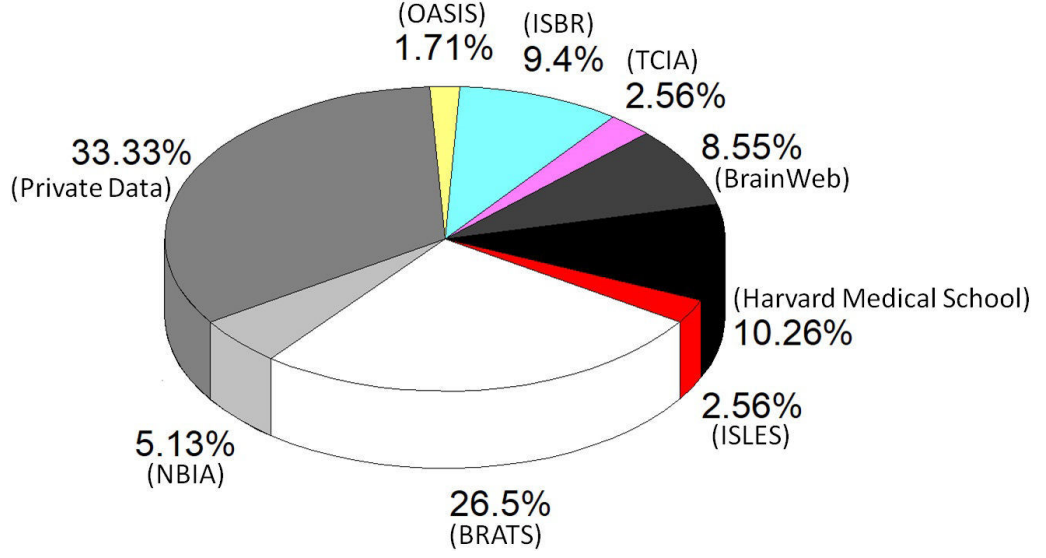


Figure 2.2: Popularity (in percentage) of various data sources in brain tumor research domain based on studies considered in this research work.

2.2 Traditional and Emerging Techniques

Among all the modules in a brain tumor detection system, the work presented in this research focuses in reviewing trends applied after data collection and image acquisition. The existing benefits of MR imaging over other diagnostic imaging techniques, as stated in Chapter 1, constrain the focus on MR image based brain tumor detection techniques. Based on the range of studies included for review, the six categories as shown in Fig.2.3 are observed. Each paper can be considered under one of these categories. The five standard categories identified are thresholding, region-, contour- (or shape-), statistical-, and machine-learning based. In addition to these traditional ones, the emerging categories, viz. deep learning and hybrid are also included. The following subsections discuss some of the prominent studies in each of these categories. Section 2.2.1 to Section 2.2.3 focus on segmentation approaches, Section 2.2.4 to Section 2.2.6 target the classification methods and lastly Section 2.2.6 covers the hybrid techniques.

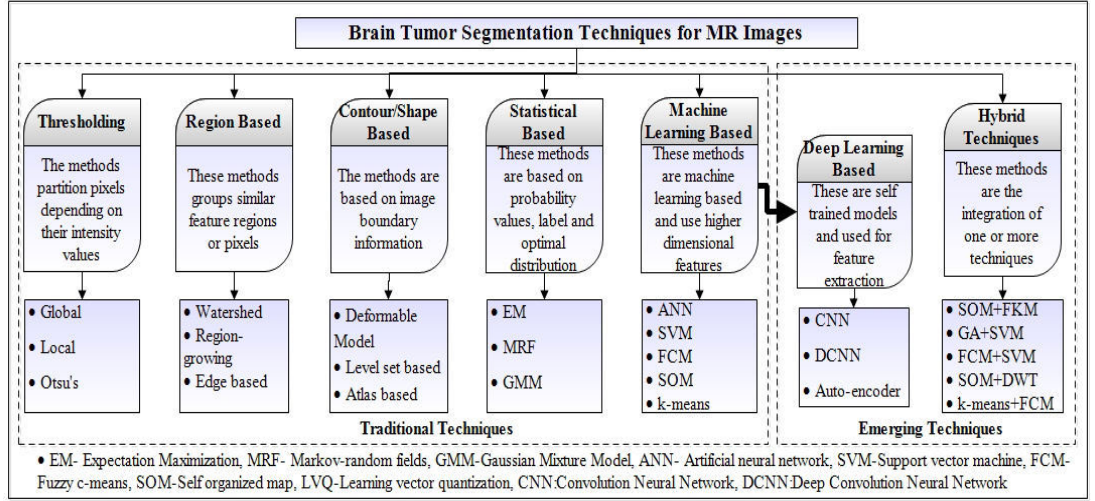


Figure 2.3: Classification of brain tumor segmentation techniques used in MR images.

2.2.1 Thresholding Based Techniques

Thresholding is a simple and aged procedure for image segmentation [19] [62], based on the images with different intensities. In this method image is partitioned into different regions by comparing their intensities with one or more predefined intensity value(s). These intensity values are known as threshold, which actually separates the image into different classes. Threshold values are generally estimated by prior knowledge related to local statistical properties like mean intensities. For example, studies efficiently computed local threshold values for T1-W images [25] [63]. There are several thresholding techniques established such as: global, local, histogram, and Otsu's. Global, the simplest one, computes single threshold for entire image. It works well for images with bimodal histogram, otherwise local thresholding or histogram thresholding is preferred. Local thresholding divides image into sub-images and generates threshold of each region. It is thus computationally slower than global. On the other hand, histogram thresholding is based on the number of peaks (representing different modes or regions) in a histogram [19] [62].

A novel framework utilizing global thresholding to enhance binary image followed by morphological operations for tumor segmentation is presented [64]. Similarly, another work considered partial volumes of every region in PD-W and T2-W MR

images to compute thresholds for segmentation. The segmented volumes of brain tissues (GM, WM, and CSF) are measured with an aim to determine the existence of neurodegenerative diseases (Alzheimer's and cerebral palsy) [65]. Furthermore, thresholding based segmentation in either case, local or global, is commonly observed as an initial phase in segmentation process because it minimally deteriorates the information provided by MR images. Working on same line, a semi-automatic fast approach calculates tumorous area from post CE-T1W MR tumorous image volumes by combining the concept of intensity threshold with morphological segmentation [66]. Thresholding is applied to selected region of interest manually. Visual results comparing set of patient data suffering from intracerebral glioma with traditional data thresholding method are presented and found to be similar. The technique is also popular in current scenario and sometimes used in combinations with other techniques to uplift the accuracy rate.

An automatic pathological area extraction approach using multi-resolution symmetry analysis and Otsu's threshold technique is also designed [67]. The approach is evaluated on T2-W and FLAIR images from BRATS dataset with 357 axial and 443 coronal images. The approach is evaluated in two phases based on accuracy values, detection (91.15%) and extraction (95%). Although the system is proposed for 2D MR, it is suitable for tumor detection in 3D as well. In recent works an effective combination of rough entropy based threshold and random forest (RF) to identify, extract, and classify four different tumor types (Glioblastoma multiforme - GBM, metastasis - MET, meningioma - MG, and granuloma - GN) is developed [68]. The study adopted thresholding to appropriately delineate tumor area in CE-T1W images, followed by morphological post processing, and lastly the RF classification. Experimental results prove that the proposed approach effectively improves diagnostic power of computer assisstive system.

An efficient threshold based segmentation approach which is easier to implement than existing threshold methods is designed [69]. The study computes threshold considering summation of unique pixel values, omitting black pixels, and then di-

viding by the count of unique pixels. Using the obtained threshold, grayscale image gets converted to a binary image. Experimental results are obtained on 100 MR images from TCIA. Overall efficiency in terms of accuracy, sensitivity, and specificity is reported as 96%, 94.28%, and 100%, respectively. In comparison Otsu's method results in 84.72% accuracy [70].

2.2.2 Region and Contour Based Techniques

Region based methods search for group of pixels having similar features to get connected regions [71]. These methods are simple and also resistant to noise. An effective seed pixel based region-growing segmentation uses similarity criteria to verify and add neighboring pixels to a region [72] [73] [74] [75]. The process iterates till no further pixel satisfies the criteria. On the other hand, seed-based segmentation considers a set of seeds denoting the objects to be segmented. A pixel is assigned to a region whose mean is the closest to the pixel's intensity. Systematic watershed segmentation is eminent in identifying background and foreground of an image. The seed point selection is really difficult in such methods but several iterations lead to the attainment of minimum values for foreground and background locations.

Edge based segmentation (EBS) methods observe hasty intensity changes near edges and thus are suitable for simple noise free images. EBS work in two-folds to balance between accuracy and noise immunity. If accuracy is higher, then noise can make image irrational; similarly higher degree of noise results in undetected image portions and incorrect object positions. Gradient based methods take difference of neighboring pixel intensities and works well for little noisy images with rapidly changing near edge intensity. Canny edge detection (CED) first filters noise, separates edge pixels, and then calculates image threshold. Sobel edge detection aids image smoothening and emphasizes high spatial frequency regions using a two dimensional spatial gradient measurements. Contour based segmentation uses a collection of control points and traces object's boundaries by evolving curves. The

linear interpolation gives a piecewise linear curve. The active contour model is an effective way of segmentation, in which the object's boundaries are noticed by evolving curves. This method is capable of detecting sharp image edges.

Watershed approaches have also been extensively used in brain tumor segmentation. One study proposes a watershed based segmentation algorithm for tumor detection both in 2D and 3D brain modalities [76]. Highlight is the block method employed to divide image into regions, which is then analyzed on multiple parametric values namely, edges, gray values, and local contrast of pixels. Focusing on the fact that proper parameter tuning is the key to success for watershed segmentation methods, a study explores several versions of watershed [77]. Working on the same lines, another work uses markers and floods to design marker-controlled watershed segmentation (MCWS) [78]. In place of regional minima, markers are adopted at the start of image gradient. Experiments show better accuracy values for MCWS than other segmentation methods. Moving ahead the over segmentation problem in an efficient watershed algorithm is resolved by developing a marker based watershed algorithm [79]. Internal and external markers are detected by applying atlas based approach utilizing prior image information. The algorithm uses different feature combinations such as color, edge, orientation, and texture to extract tumor region from MR images. Experiments are performed on 45 brain tumor axial orientation diffusion weighted images (DWI) provided by Medical College hospital. It is evident from results that color and orientation features cannot be considered for watershed segmentation. The concept of traditional watershed segmentation is effectively combined with Scale-Invariant Feature Transform (SIFT) to design a novel watershed matching approach [80]. Images are preprocessed with median and trilateral filters for artifacts removal and hence an improved segmentation quality. Watershed followed by SIFT matching between segmented and original image is done to compute tumor region. Experimental results report an accuracy of 98.5% operating on 100 T1-W images from Montreal neurological institute. In addition, other studies have worked on issues of over segmentation by means of pre- and post- processing to elim-

inate noise or by merging watershed with other concepts to report better results. Researchers used region and Fuzzy c-means clustering to eradicate this above stated disadvantage [81] [82] [83]. The study of existing literature clearly states that the popularity of watershed is still prevalent due its advantages over other conventional methods.

Sobel, Prewitt and Roberts are commonly covered under edge detection to enhance image areas containing edges or significant transitions. Performance comparison of three edge-based detection techniques based on these operators is presented using a dataset of 13 patients among which 8 having tumor [84]. The aim is to retrieve tumor as well as brain structure edge pattern. Sobel is observed to report the smallest standard deviation and thus the most suitable operator in detecting brain tumor for the considered dataset. Limitations like discontinuity in detected edges, improper edge detection, false or thick edges, no edge detection in noisy images etc. usually occur with traditional edge based methods. Thus, combination of such approaches with other well known approaches is explored by many researches to effectively deal with these drawbacks. Like a combination with cellular automata (CA) is proposed [85]. Similarly, an improved Sobel edge detection algorithm utilizing the concept of image dependent thresholding is designed [86]. It first identifies closed contours with lesser false edges and then uses their intensity information for tumor extraction. Experiments are conducted on seven case studies, each with 20 – 30 images and are compared with conventional algorithms based on gray level uniformity measure, relative ultimate measurement accuracy, and Q-parameter. Improved Sobel algorithm effectively extracts tumor in all the cases. The algorithm can further be improved by decreasing boundary lines thickness and thus increasing region area. Edge detection sometimes fails to achieve the desired results, a study is introduced which combines Fuzzy inference system with edge detection to design an improved segmentation approach [87]. Initially k-means approach is applied to form the clustering groups which act as an input to Mamdani fuzzy inference system to generate various threshold parameters. The classical sobel edge detector

is used to feed these parameters to exactly locate the tumor position. A modified version of contour based method termed as content-based active contour (CBAC) is presented [88]. The study combines intensity and texture information within active contour to overcome issues that prevail due to the presence of diffused and false edges in input images. Efficacy of CBAC is tested on two real datasets containing more than 600 images of pre- and post-contrast T1-W and T2-W MR image modalities. Experimental results show remarkable performance in segmenting different types of homogeneous, heterogeneous, and diffused tumors as well as synthetic images.

Unlike the discussed segmentation techniques, deformable model based segmentation builds a connected model using object's prior knowledge and information like shape, texture, orientation, location etc. [89]. A few incorporates statistical information drawn from the group of training datasets too [25]. The internal and external influence weighted forces parametrically defines curves/surfaces that makes deformable models smooth. Existing deformable models are categorized as geometric and parametric model [90]. Parametric deformable models can significantly match and track anatomic structures of images over time [91]. Out of all the region based segmentation methods, deformable model based ones are the most extensively explored in addition to level set segmentation methods. Level-set segmentation segments curves or surfaces while dealing properly with issues of corner and breaking curve.

Such models are extensively incorporated in brain tumor detection system. Like a modified deformable region model is presented using 2D MR image modalities to extract and retrieve boundary of the brain tumor [92]. It effectively utilizes point sampling to lower the count of boundary points to be processed. The corresponding effect in the form of reduced extraction time of tumor boundary is observed when compared with traditional active contour models. However, the model is not a general one and performs poorly in some applications. It requires manual input to set the ideal boundary limits for acceptable results in those applications. A framework to segment internal brain structures is proposed by integrating spatial relations

into deformable model based on initial tumor segmentation [93]. Segmentation and recognition of surrounding brain structures are based on prior knowledge of their spatial arrangement. The approach is validated on 20 cases of CE-T1W MR image. The experimental results report sensitivity of 84%. Further, a novel, accurate, and robust pipeline of region-based fuzzy clustering method called Enhanced Possibilistic Fuzzy C-Means (EPFCM), which is used for the initial segmentation of the tumor, is combined with one of the contour-based methods called parametric deformable model with gradient vector field (GVF) is introduced [94]. EPFCM output provides initial contour for deformable models to deal with the issues of initialization and poor boundary capturing. Tumor segmentation results are presented on a dataset having 10 CE-T1W and 5 FLAIR MR images. The reported average similarity and Jaccard index are 95.3% and 82.1%, respectively.

Geometric deformable model, also known as level set method helps to overcome problems like corner fabrication and curve breaking. Segmentation of 3D volumetric images are commonly done using this model as it automatically handles topological changes during splitting and merging of contours. A variational level set approach based on existing well-known method [95] is introduced for automatic bias correction and segmentation of images altered by intensity variations [96]. The presented region-based active contour model is flexible in initialization and ensures smoothness of the computed bias field by the data term while variational formulation. It uses a local intensity fitting term that attracts contour towards object boundaries and preserves regularization term to ensure accuracy. Results are compared based on coefficient of variance for two simulated images (one corrupted with bias field without noise and another with both bias and noise) from BrainWeb. The obtained lower variance values indicate that images bias corrected with the presented approach are more homogeneous. Similar form of variational methods are presented by other researchers where knowledge is preserved in the form of template guide outlines and atlases of normal brain structures [97] [98]. For accurate tumor matching, an improved semi-automatic version of level-set segmentation employing a constraint

termed as PICE (prior information constrained evolution) is presented [18]. It combines image gradient and intensity information in level-set formulation. PICE is evaluated and compared with traditional level set segmentation using dataset of T1-W MR images and has reported an outstanding accuracy (nearly 98%) which is 10% more than the other existing approaches. Level-set segmentation is explored in combination with region and boundary driven methods known as hybrid level set (HLS) to achieve an accurate performance [99]. The idea is to employ region method for propagation and boundary as a stopping function so as to assist user in better region of interest (ROI) selection. Experiments are performed on 246 axial tumor slices of 10 different patients. So far, semi-automatic detection methods are considered the best as human interactivity is required to initialize the segmentation process [100]. A semi-automatic level set method is used to initialize the tumor region for segmentation. A repeated segmentation process occurs by tuning the parameters of level-set. Moreover, to handle the issue of slow computing of 3D medical images, a tool that depends on level-set models is proposed.

Segmentation of images with unclear pixels – region’s intensities relationship is commonly done using atlas based segmentation which allows transfer of labels as well as segmentation. It is prominently employed to detect object’s shape and patient’s group morphologically in computer aided design. It’s generic aspects are detailed beautifully that helps to set the forum for brain tumor segmentation [101]. An expectation-maximization based automated system is introduced aiming to segment brain tumor as accurately as manually [102]. In addition, it segments normal brain tissue classes, namely WM, GM, and CSF along with edema. Spatial probabilistic atlas containing expert prior knowledge of the brain structures guides the whole process. Contrast enhancement based atlas is computed for each subject under consideration for better results. Experimental results obtained for T1-W, T2-W, and CE-T1W MR images proof applicability of the system. Similarly, a tool based on mesh-free method is designed to segment brain tissues (WM, GM, CSF, fats, edema, and tumor) [81]. Study aims on improvement in the registra-

tion accuracy by manually seeding atlas with the prior knowledge of tumor position and mass effect. The initial source of work is presented in [82] and is further validated on synthetic cases using a finite element method (FEM) based deformation. The experimentation results show good segmentation performance. Another mesh free based, non-parametric, automatic system is introduced to fragment brain tissues from volumetric tumor images. The tumor development model contrive as a mesh-free Markov Random Field energy minimization issue that guarantees correspondence among atlas and patient's image, preceding to registration phase.

2.2.3 Statistical Based Method

Statistical based segmentation is simpler to implement and further, the associated statistical probabilities make acceptance of such methods in medical image processing quite natural [19]. These methods work with parameters like mean, median, standard deviation, probability distribution, and amplitudes to statistically describe an image. They are maily classified into three, namely Markov random field (MRF), Gaussian mixture model (GMM) and expectation maximization (EM). MRF is non-deterministic and considers spatial information via dependency concepts of neighbors and cliques. One can embed prior knowledge into MRF through clique possibilities. GMM in comparison has effortless implementation and requires few parameters. Lastly, EM uses other methods, like k-means for initialization of matrices such as mean and covariance. Each iteration recomputes these matrices based on changing membership of training information till the convergence criteria is reached. In general, an effective statistical simple segmentation model is designed and progressively extended for the multi-scale environment.

A novel 3D segmentation algorithm combining concept of simulated annealing (SA) with an iterated conditional version of Markov random field (MRF) is designed [83]. The algorithm segments entire 3D MR volume and all MR images acquired with MR sequence automatically, just after the training of typical echo

intensities and MRF parameter setting as per the expected irregularity. It unites nonparametric intensity distributions, neighborhood correlations, and irregularities effectively. As a result, the algorithm is accurate as well as robust against noise, irregularities, and structure thickness. Also, SA improves the performance as compared to iterated conditional modes (ICM) but reports longer execution times. A novel two stage fully automated approach is proposed to separate tumorous and non-tumorous region (inclusive of edema) from multi spectral MR images [103]. The integration of Bayesian classification of Gabor decomposition with MRF classification is performed. Bayesian extracts texture features of tumor/edema which are further refined by MRF. The performance evaluation depicts DSC of 0.668 and 0.56 for tumor and edema, respectively. The work is further extended to propose an iterative MRF framework based on pixel- as well as region-information to classify a brain tumor image [104]. The main objective of applying iterative model is to overcome the issue of over-segmentation and mis-characterization of the tissue types. Previously the aim is to compute the tumor area with high probability, neglecting the sub classification of different tissue types. An automatic, non-parametric, simple, fast, and mesh free method to segment brain tissues from volumetric MR image brain tumor images is presented [105]. It combines non-rigid registration of an average atlas with biomechanically justified tumor growth model to simulate soft-tissue deformations caused by tumor mass-effect. The tumor growth model is formulated as a mesh-free MRF energy minimization problem that ensures correspondence between atlas and patient image, prior to registration step. MRF efficiently computes finite differences in the neighborhood of each voxel. In addition to tumor induced deformation, it can model soft-tissue deformations and is thus clinically-oriented. Experiments are conducted on T1-W MR images of four patients database from ContraCancrum project and results are compared using DSC. For real patient cases DSC values of WM, GM, and CSF ranges from 70% – 82%. The segmentation results maintain similar accuracy as that of the existing methods.

Combining the concept of GMM and fuzzy clustering, a segmentation model

termed as AS-FLGMM (adaptive scale fuzzy local GMM) is designed for accurate and robust tumor detection [106]. A novel way to adaptively determine variance in every distribution is employed based on local scale assessment technique assuming that local information within image area pursue GMM. It completely overlooks distinction of variance among Gaussian segments. Working on the same assumptions, another work proposes FLGMM (fuzzy local GMM) to automate brain tumor MR image segmentation approach [107]. Summation of weighted GMM energy function is defined as the objective function covering the whole image. The computational evaluation is performed using 40 T1-W images taken from BrainWeb and IBSR datasets which proves efficacy of the combination. Similarly, a framework based on the fusion of multiple number of Gaussians is used to apprehend the spatial features of complex tissues [108]. The proposed constrained GMM (CGMM) is composed of four-dimensional Gaussians, where three represent spatial features and one is the intensity attribute per region. Satisfactory experimental results are conducted using 36 T1W real MR images derived from IBSR dataset. A study introduces a generative asymmetric Gaussian mixture model with spatial constraint (SCGAGMM) to overcome drawbacks, such as robustness and over-smoothness, of existing EM based segmentation models [109]. SCGAGMM successfully reduces the drawbacks and has flexibility to cover different tumor shapes. Segmentation performance of SCGAGMM is compared with four state-of-the-art algorithms namely, spatially constrained generative model and EM (SCGM-EM), fast and robust spatially constrained GMM (FRSCGMM), Bayesian bounded asymmetric mixture model (BAMM), and bounded generalized GMM (BGGMM). Experimental datasets consists of T1W, T2W, and PD MR images collected from BrainWeb and IBSR. In addition, results are obtained with synthetic as well as natural color images. In all the cases, SCGAGMM reports the best DSC values. DSC of 86.91%, 86.78%, and 67.84% is respectively attained for GM, WM and CSF tissue classes. However, presented algorithm lacks in automatically determining the number of classes/components for different applications.

Aiming to segment brain tumor very close to manual procedure, an expectation-maximization (EM) scheme based automated system is introduced [102]. In addition to tumor, tissue classes namely WM, GM, and CSF are segmented along with edema. Segmentation is guided by spatial probabilistic atlas containing expert prior knowledge of brain structures. To achieve better results contrast enhancement based atlas is computed for each subject under consideration. Experimental results are generated for T1-W, T2-W, and CE-T1W MR images using Jaccard similarity measure. Five cases each with different tumor types are considered to compare results with manual and semi-automated methods. Compared to manual results, this system yields surface distances of approximately 1 – 4 mm. However, semi-automated method supersedes in all the considered cases evaluating Jaccard similarity between 49% - 71%. Segmentation algorithm, termed as SHE, is developed by effectively combining EM with the concept of spatial accuracy weighted hidden Markov random field [110]. SHE focuses on improved segmentation quality in clinical applications where low-resolution sequences (T2-W, FLAIR) are used together with high resolution T1-W images. Experiments are performed on a dataset of 15 patients and are compared with original hidden MRF (HMRF) with EM using sensitivity and specificity parameters. Improved tissue segmentation is reported by SHE, especially for all the normal tissues (WM, GM, and CSF). Comparison is done on multiple channels of MR images as well. For axial and coronal channels, the respective sensitivity and specificity values are 76.4% and 87.5%. SHE reports the corresponding improvements of 5.3% and 1.9% when all the three channels are utilized. Another work explores spatial relationships between pixels to estimate finite mixture model parameters with Gaussian distribution assumption and introduced a simple modified EM algorithm, termed as ETM [111]. The acceptance is proved by conducting experiments on a real dataset of 32 malignant T1-W MR images (with contrast agent gadolinium). Comparison with baseline approaches is performed using Gold Standard (GS) of the image dataset, a parameter provided by Carlos Van Buren Hospital of Valparaiso-Chile. Baseline approaches compared are EM, Neigh-

borhood EM (NEM), Hybrid EM (HEM), Fuzzy C-Means (FCM), Modified Fuzzy C-Means (FCMM), region growing (RG), and Genetic Algorithms (GA) with 4 regions or classes defined a priori (tissues types in the image) [112], [113], [114]. ETM reports the best segmentation results obtaining 95.13% accuracy on 32 T1-W MR images. Another combination utilized Hidden MRF (HMRF) with EM to efficiently segment tumor from other brain tissues [115]. It minimizes energy function during each iteration. Experiments are presented on 30 images of high grade glioma T1-W, T2-W, CE-T1W and FLAIR modalities from BRATS 2012 dataset. A comparative analysis with existing FCM algorithm proves applicability of the integration.

2.2.4 Machine Learning Techniques

Applicability of machine learning techniques in the domain of automatic brain tumor segmentation is extensively explored to efficiently proceed with analysis and diagnosis phases. It is observed that the introduction of such learning techniques eases work at the end of radiologists as well as experts in medical practice [116], [117]. These techniques understand relationship depth of patterns and complex data more accurately. They are broadly categorized as supervised (classification) and unsupervised (clustering) [118]. A few studies that have utilized learning for brain tumor segmentation are FCM [119], [120], [51], self-organized maps (SOM) [62], [121], [122], k-means [123], [124], support vector machine (SVM) [125], [126], [127], learning vector quantization (LVQ) [128], [129], artificial neural networks (ANN) [130], [131] etc. Subsequent sections detail utilization of such approaches.

Supervised techniques derive functional relationship between features and labels during training by utilizing the labeled information. Then in testing, labels are generated for unlabeled information based on the estimated feature [132]. A range of well-known classifiers is used in several studies for the identification of brain tumor. A distance and voting function based k-nearest neighbor (KNN) is the simplest of all the classification techniques. It is known to achieve higher accuracy

and stability for MR image data but is observed to take large running time. Artificial neural network (ANN) maps an image to network of neurons which are considered as pixels. ANN considers detection as an energy minimization problem and tries to establish not only the connection but weights between nodes too while training. Weights are modified by means of an error factor computed by comparing observed and desired output for the given input. One can extract features as well using ANN. Next is SVM which uses the concept of hyper planes to train itself in appropriately selecting margins and thus distinguishing between or among classes. SVM helps in overcoming issues related to local minima and neurons overhead as compared to ANN [133], [134].

Clarke is one of the first researchers who introduces an automatic segmentation system for MR images using ANN [130]. The presented classification approach segments and labels glioblastoma-multiforme (GBM) tumors using T1-W, T2-W, and PD-W MR images. Later several studies exploring various classifiers come into existence. A user friendly SVM based novel segmentation approach embedded with region analysis is presented to optimally extract flexible decision boundary across tumor [135]. The approach automatically learns parameters from non-linear distribution of image data without any additional prior knowledge. Final results are obtained after region analysis. Comparative results with fuzzy clustering on a five patient dataset having 24 pairs of T1-W and CE-T1-W images prove efficiency of the approach. A novel version of probabilistic neural network (PNN) integrating least squares features transformation (LSFT) of non-linear type is designed to differentiate metastatic and primary tumors [20]. The system is trained with 66 texture features of 67 T1-W post-contrast MR images. The system achieves a classification accuracy of 95.24% to distinguish metastatic and primary tumors whereas 93.48% accuracy to classify between gliomas and meningiomas. An automated diagnosis method combining an improved version of orthogonal DWT, i.e. Slantlet transform with intensity histograms is presented [136]. The main focus is to classify two-dimensional MR image as either a normal brain or one suffering from Alzheimer's disease. Feature

vector is created using a two step process: intensity histogram computation followed by Slantlet transform (magnitudes of outputs corresponding to six spatial positions are used). Experiments are performed on a dataset consisting of 75 T2-W images (normal brain – 39 and Alzheimer’s disease – 36) collected from Harvard Medical School. Results report excellent classification accuracy of 100% and an improvement in time localization too.

An automatic seed point selection based system termed as adaptive neuro fuzzy interface system (ANFIS) is presented for detection and diagnosis of brain tumor [137]. The system effectively combines the concepts of neural networks and anisotropic diffused filtering with fuzzy logic. ANFIS classifier is trained using three different texture features, viz. gray level co-occurrence features (contrast and energy), Haar wavelet transform, and law’s energy features. Experimental results are generated using dataset consisting of 120 (80 abnormal + 40 normal) training and 80 (30 normal + 50 abnormal) testing images. Comparison with several conventional processes using parameters like accuracy, sensitivity, similarity index, extra fraction, overlap fraction, and positive prediction value proves applicability of the presented system. The respective reported values are 87%, 81%, 81.7%, 18.2%, 81.7%, and 81.7%. A study analyzing the affect of three dimension reduction techniques on the classification accuracy of SVM for glioma grade is presented [138]. It suggests that careful selection of features reduces number of features by up to 98% while maintaining desired classification accuracies comparable to those reported in literature. Also feature compactness helps in more streamlined work-flow with minimal computational needs. Predictive power of feature reduction techniques is valued on a dataset of 101 untreated glioma patients using T2-W and CE-T1W MR images. Experimental results report PCA as the best candidate that uses the smallest feature set in achieving equal classification accuracy. Specifically, PCA reports 85% classification accuracy with 89% sensitivity and 84% specificity for three principal components. In comparison, the respective values for Pearson’s correlation coefficients (PCC) are 82%, 89%, and 77% with two principal components; similarly

for independent component analysis (ICA) they are 79%, 87%, and 75% with nine principal components.

Detection of brain tissues in the presence of unlabeled image information is commonly done via clustering. These techniques submit complete arrangement of pixel vectors to cluster examination plot then find relevant groups algorithmically. Similar objects share same cluster and dissimilar ones are placed among different clusters. Clustering identifies brain tissues as regions sharing same label [132]. A few of the popular clustering approaches in tumor detection through MR images are k-means, self-organized map (SOM), fuzzy c-means, pulse-coupled neural network (PCNN), and feedback pulse-coupled neural network (FPCNN).

A segmentation approach to automatically track non-enhancing growth or shrinkage of brain tumor size over time in MR images is designed [47]. Segmentation is done effectively in two phases. Starting with unsupervised fuzzy clustering, domain knowledge is subsequently integrated then the final segmentation output is generated. Experimental results generated on weighted modalities of MR images, i.e. T1-W, T2-W, and PD, are found as satisfactory. A modified PNN combined with LVQ to classify specified ROI of tumor area is proposed [139]. Each ROI is weighted based on LVQ and then extracted ROI features are utilized to evaluate brain tumor from MRI. For experimentation 64 MR images are collected and comparison is done with conventional PNN. The results prove a classification accuracy of 100% along with that a reduction in the processing time by 79% approximately. The classification technique consisting of three stages, feature extraction, dimensionality reduction, and classification [140]. In classification stage, two classifiers have been developed, the first one is based on feed forward back propagation artificial neural network (FP-ANN) and the other on k-nearest neighbor (k-NN). These two classifier are used to identify the abnormal tissues and resulted with a success of 97% and 98% respectively. The method can be deployed for all type of MR images T1-W, T2-W and PD-W.

A fully automatic and unsupervised brain tumor segmentation method along

with a method to retrieve membership functions perfect MR image is designed [51]. The objective function is defined based on intensity values. Experimental results are obtained on CE-T1W MR images taken from three datasets: normal dataset (NON), glioblastoma multiforme (GBM), and meningioma (MEN). Results are better than the existing approaches. Specifically Jaccard similarity coefficient on NON is 97% whereas on GBM and MEN it is observed to be 83.61% and 76.71% respectively. A hybrid segmentation algorithm based on soft sets namely soft fuzzy rough c-means (SFRCM) is proposed to extract the white matter (WM), gray matter (GM) and the cerebrospinal (CSF) fluid from MR brain image with bias field correction [5]. The soft fuzzy rough sets are obtained by embedding the concept of soft sets, rough sets, and fuzzy set. The soft rough set defines the lower and upper approximations of cluster using the soft fuzzy set indicating the degree of belongingness. The approximations are defined for each pixel with fuzzified values rather than crisp values. The efficacy of the proposed algorithm SFRCM is evaluated on three brain databases: (i) simulated Brain Web database with 20 images (ii) IBSR database with 20 images and (iii) BRATS database with 10 images. T1-W, T2-W, and FLAIR MR images are used to detect high grade glioma. The visual results clearly differentiate between brain tissues and tumor. The accuracy of the tumor detection with Gaussian noise at 3% and salt and pepper at 5% turns to be 94.04% and 87.94% respectively.

Combining several computational methods at each intermediate stage, an efficient as well as automatic CAD system for brain tumor diagnosis using MR images is designed [141]. The system employs FPCNN for segmentation, DWT for feature extraction followed by dimension reduction using PCA, and lastly feed forward back propagation neural network (BPNN) classifies between normal and abnormal tissues. Experimental results are obtained on 101 MR images (14 normal and 87 tumorous) taken from the Harvard Medical School website. The classification accuracy on testing images is 99%, sensitivity is 100%, and specificity turns to be 92.8%. Neural network based approaches have complex architecture and demands enough training, an attempt to resolve these limitations is made by integrating SOM

with unsupervised LVQ [142]. The proposed SOM based clustering module does not utilized any additional network. The generated output is further processed using concepts of thresholding and skull stripping to obtain effective segmentation.

Summarization of some good works under each of the discussed traditional approaches are presented in a Table 2.2. Basis used to choose any study are the values of performance parameters with respect to the size of experimental dataset.

2.2.5 Deep Learning Techniques

Another emerging domain of machine learning is ‘deep learning’. It’s multiple layer architecture represents data with manifold layers of abstraction which helps to overcome several challenges that occur in traditional machine learning approaches. Additionally, it’s generalization and self-learning characteristics enable better quantitative analysis of imaging features and thus better detection of neurological disorders. Thus, deep learning based detection techniques are gaining acceptance in the field of medical imaging process [146] [147]. CAD systems designed with deep learning techniques are frequently observed in medical image analysis ranging from breast lesions and pulmonary nodules [148] [149] [150], chest and pulmonary tuberculosis [151] [152] to brain tumors [153].

Several deep learning techniques, namely deep neural networks (DNN) [163] [153], convolutional neural networks (CNNs) [164] [48] [165], deep convolutional neural networks (DCNNs) [166] [167], auto-encoders [168], stacked auto-encoders [22] [15], are developed for efficient brain tumor detection, segmentation, and classification using MR images. Researches in great pace continues to dig more and more deep learning approaches to achieve the best performance. Deep learning combined with other techniques is observed to enhance the classification accuracy. Working differently, one study focuses on proper treatment planning and presented a hybrid system that predicts 1p/19q status of low grade gliomas (LGG) [154]. ANTs open source software library for image registration is utilized to register multi-modal CE-

Table 2.2: Summary of a few studies utilized traditional approaches for brain tumor detection.

Reference	Technique	Modality	Size of Dataset	Tumor type	Performance(%)
[68]	RET	T1W, CE-T1W	100 cases	GBM, MET, MEN, Glioma	ACC GBM:84.6, MET:83.5 MEN:92.2, Glioma:85
[69]	New thresholding method	-	100 images	-	ACC:96, SEN=94.28, SPEC=100
[143]	Multi-level Otsu's thresholding	T2W	DS-I: 1100 images DS-II: 600 images	GBM, AST, EPD, ODG	ACC=98, SEN=100
[72]	Statistical classification via atlas prior	T1W,CE-T1W	20 cases	LGG, MEN	ACC: 99
[79]	Marker based watershed	DWI	45 images	-	DSC:93.98 *(for the 5 features) and DSC=73.93 for color and texture
[80]	Watershed matching	T1W	200 images	-	ACC=98.5
[111]	Modified EM	T1W, CE-T1W	12 synthetic and 32 real	-	Synthetic ACC: 99.96 Real ACC: 95.13
[105]	MRF	T1W	4 patient data (Four datasets)	-	DSC for tumor DS-I:93 DS-II:94, DS-III:95, DS-IV:87
[109]	SCGAGMM	T1W, T2W, PD	100 images	-	DSC GM:86.91, WM:86.78 CSF:87.84
[144]	ST	T2W	75 images	Alzheimer's disease	ACC:100
[119]	SVM	CE-T1W,T1W, FLAIR,ADC,DWI	22 images	GBM,ODG	*intra-patient ACC: 98.75 *Inter-patient ACC:98
[137]	ANFIS	ADI	120 training, 80 testing	-	ACC:87 SEN:81 SI:81.7
[51]	KBFC	CE-T1W	20 images	GBM,MEN	JC: Normal:97 GBM:83.61 MEN:76.71
[141]	FPCNN	T2W	101 images	GBM, MEN, Alzheimer's disease, glioma, sarcoma	ACC:99, SEN:100 SPEC:92.8
[145]	BCFCM	-	-	-	ACC:99.25

ACC: Accuracy, ADI :Anisotropic diffused image, ANFIS: Adaptive neuro fuzzy inference system, AST: Astrocytoma, BCFCM: Bias- field corrected fuzzy c-means, CSF: cerebrospinal, DS-I: First data source, DS-II: Second data source, DS-III: Third data source, DS-IV: Fourth data source, DSC: Dice similarity coefficient, EM: Expectation Maximization, EPD: Ependymoma, GBM: Glioblastoma multiforme, GM: Grey matter, JC: Jaccard Coefficient, KBFC: Knowledge based fuzzy clustering, KFCM: K-means integrated with Fuzzy c-means, LGG: Low grade gliomas, MET: Metastasis, MEN: Meningioma, MRF: Markov random field, ODG: Oligodendroglioma, RET: Rough entropy based thresholding, SCGAGMM: Spatially constrained generative asymmetric gaussian mixture model, SEN: Sensitivity, SI: Similarity index, SPEC: Specificity, ST: Slantlet transform, WM: White matter.

Table 2.3: Summary of a few studies utilized deep learning approaches for brain tumor detection.

Reference	Approach	Modality	User Interaction	Size of Dataset	Tumor type	Performance (%)
[154]	CNN	CE-T1W, T2W	Semi-automatic	159 images	LGG	ACC: 87.7 SEN: 93.3 SPEC:82.22
[155]	WMMFCM	T1W, T2W, FLAIR, CE-T1W	Automatic	DS-I: 81 DS-II:152	HGG, LGG, MS	ACC DS-I:95.853 DS-II:97.05
[156]	DNN-DWT	T2W	Automatic	66 images	GBM, MBC	ACC:96.97
[157]	Supervoxel learning	T1W, T2W, FLAIR, CE-T1W, DTI	Automatic	30 images	HGG, LGG, edema	BRATS SEN:96, DSC:89 Clinical SEN:86, DSC:0.84
[158]	EnsembleNet + Incremental (2CNet,3CNet)	T1W, T2W, FLAIR, CE-T1W	Automatic	HGG:210, LGG:75	GBM	DSC:88
[48]	CNN	T1W, T2W, FLAIR, CE-T1W	Automatic	65 images	Astrocytomas, AA, GBM, LGG, HGG	DSC WT:88, TC:83 AT: 77
[159]	DCNN	T1W, T2W, FLAIR, CE-T1W	Manual and Automatic	30 images	LGG, HGG, Edema	DSC WT:80, TC:67 AT:85
[160]	IntNet, SkipNet, SENet	T1W, T2W, FLAIR, CE-T1W	Automatic	Training:374 Testing:110	HGG, LGG, edema	Combined dataset IntNet DSC:90, SEN:88 SPEC:73
[161]	ECNN	-	Automatic	-	-	ACC:92
[162]	(FC-NNs) and (CRFs)	T1W, T2W, FLAIR, CE-T1W	Manual and Automatic	HGG:220, LGG:54	HGG, LGG, edema	BRATS DSC WT:84, TC:73, AT:62

AA: Anaplastic astrocytomas, ACC: Accuracy, AT: Active tumor, CNN: Convolutional neural network, CRF: Conditional Random Fields, DNN-DWT: Deep neural network discrete wavelet transform, DS-I: First data source, DS-II: Second data source, DSC: Dice similarity coefficient, DTI : Diffusion tensor imaging, ECNN: Enhanced convolutional neural network, EnsembleNet: End-to-End Incremental deep neural networks, FCNs: Fully Convolutional Networks, FC-NN: fully convolutional neural networks, FC-ResNets: Fully Convolutional Residual Networks, GBM: Glioblastoma multiforme, HGG: High grade gliomas, IntNet: Interpolated network, LGG: Low grade gliomas, MBC: Metastatic bronchogenic carcinoma, MS: multiple sclerosis, SENet: Squeeze-and-Excitation Networks, TC: Tumor core, 3CNet: Three convolutional layers, 2CNet: Two convolutional layers, WMMFCM: multi-resolution wavelet morphological pyramid fuzzy c-means, WT: Whole tumor, XCNet: number of convolutional layers.

T1W and T2-W images (in total 159 MR images of LGG with status non-deleted for 57 images and co-deleted for 102 images) [169], followed by tumor segmentation using semi-automatic software [170], and finally CNN classification is performed. Cross domain results reported 87.7% accuracy, 93.3% sensitivity, and 97.7% specificity.

A three-stage automatic segmentation approach, termed as WMMFCM is proposed aiming to overcome the limitations of FCM [155]. Techniques employed in respective three stages are multi-resolution wavelet (WM), morphological pyramid (M), and Fuzzy C-means (FCM) clustering. Performance is verified using two datasets: BrainWeb (152 MR images with T1-W, T2-W, and PD modalities) and BRATS (81 images from multimodal brain tumor segmentation having glioma with T1-W, T2-W, FLAIR, CE-T1W modalities). An accuracy of 97.05% is reported for BrainWeb and with BRATS 95.853% accuracy is achieved. In the consecutive year, another work utilized the concept of small kernels based CNN [48]. A novel way to deal with over fitting, provided fewer number of weights in the network is presented. Starting with uncommon intensity and patch normalization, the study proved effectiveness of combination together with data augmentation. Next training with patches is done by rotating them artificially. Lastly, defined threshold is placed to impose volumetric constraints, i.e. elimination of small clusters, which are basically the erroneous ones and may be classified as small tumors. In the brain tumor experimentation phase, the accuracy rate for a baseline network turned to be 84% whereas for U-net 88% accuracy is achieved.

Focusing on the same issue of over-fitting with sparse data, an automated system based on deep convolutional neural networks (DCNN) is designed [159]. It blends DCNN with an introduction of max-out and drop-out layers. Acceptability of the approach is evaluated on BRATS 2013 dataset having T1-W, T2-W, CE-T1W, and FLAIR MR image modality. Experiments are conducted on a system using training is to testing ratio of 80:20 based on three parameters namely, dice similarity coefficient (whole tumor - 80%, core - 67%, and enhancing - 85%), sensitivity (whole

tumor - 82%, core - 63%, and enhancing - 83%), and specificity (whole tumor - 85%, core - 82%, and enhancing - 88%). Several approaches are designed to improve and outdo CNN performance in terms of hardware specifications, accuracy, and processing time especially while handling large size images [156]. The approach uses Fuzzy c-means for the MR image segmentation which is followed by discrete wavelet transform (DWT) integrated with the Deep Neural Network (DNN) to classify 66 T2-W MR images into four classes: normal, glioblastoma, sarcoma and metastatic bronchogenic carcinoma tumor. The performance of the algorithm resulted in 96.97% classification rate.

In the same year extended version of the DCNN is presented to resolve the segmentation issues [171] [165] [172]. Another popular issue studied by a large group of researchers is the occurrence of multiple tumors [156] [48]. Tumor multiplicity demands more precision and thus increases complexity; in such a scenario MR image input type and its features matter a lot. A novel multimodal supervoxel based segmentation approach integrated with random forest (RF) is introduced to work with MR as well as with diffusion tensor imaging (DTI) [157]. For each supervoxel, a variety of Gabor features are extracted to train RF classifier. It classifies each supervoxel as healthy and tumorous (core or edema) using multimodal images from BRATS and clinical dataset (30 images). Performance is reported in terms of sensitivity and dice score. The respective values for clinical dataset are 86% and 0.84. In comparison better values are reported for BRATS, 96% and 0.89 respectively. One work has proposed three different network models, namely, Interpolated Network (IntNet), Skip-Net, and SE-Net for brain tumor segmentation [160]. Experiments on BRATS 2015 MR images with four modalities (T1-W, T2-W, CE-T1W and FLAIR) proved superiority of IntNet over SE-Net and Skip-Net. IntNet achieves highest values for all the three considered parameters dice coefficient (90%), sensitivity (88%), and specificity (73%) on a complete dataset. In the current scenario, enhancement of CNN is applied to resolve the extensive manual diagnosing process. An optimal automatic segmentation approach combining enhanced CNN (ECNN)

with BAT algorithm is developed [161]. BAT works on the function loss whereas small kernels characteristics of ECNN allow network with less weight assignment that controls over-fitting. The performance of ECNN is found to 3% more accurate in comparison to traditional CNN. For the segmentation of Glioblastomas (high as well as low), an end-to-end incremental DNN based model known as “EnsembleNet” is proposed [158]. It aggregates on parallel instances with incremental XCNet that generates CNNs model (2CNet and 3CNet) by using non-parametric fusion technique. The dice score turns out to be 0.88 on BRATS 2017 dataset.

2.2.6 Auto-encoder Techniques

Auto-encoders, a fully associated neural networks able to learn effectively in an unsupervised manner, are studied extensively in the past few years so as to lessen data dimensions (encoding). The reduced representation is then processed (decoding) to reconstruct the associated input. Several works in the field of medical imaging have proved better performance of auto-encoders in comparison to PCA.

Stacking denoising auto-encoders (SDAE) scheme for brainstem segmenting is found to perform well over traditional SVM [15]. Experimental results are obtained on T1-W modality using manual as well as automatic process. Shape similarity accessed via DSC ranges from 0.84 to 0.90 for four manual and three automatic annotations methods. DSC with SVM is observed as 0.88 whereas it has shown an improvement of 0.03 when SVM is embedded with the proposed scheme. Overall the proposed scheme is observed to perform better, reporting mean DSC of 0.92. Another study also utilized SDAE technique for accurate segmentation of gliomas [22]. Among several layers in the architecture three are meant for SDAE having 3000, 1000, and 500 neurons respectively. System is trained with fixed sized 3D patches 11x11x3 of T1-W, T2-W, CE-T1W, and FLAIR MR image modalities from BRATS having 220 HGG and 54 LGG patients. Experimentation results outperform for high grade glioma as compared to low grade with DSC of 0.81 for whole tumor,

0.68 for core tumor, and 0.64 for active tumor. Another reason for the popularity of SDAE is its ability to represent complex functions finitely [153]. In comparison, existing approaches require manual interference with prior knowledge whereas in case of automatic shallow learning is involved. Utilizing this fact a novel deep learning system is proposed. The pre-processing phase extract image patches which are processed by SDAE based classification network to retrieve high level features. Finally, morphological filters are applied to fetch the segmentation results. The approach has reported satisfactory performance on real patient's data collected from West China Hospital. Furthermore, other versions of denoising auto-encoders (DAEs) are also introduced for effective lesion detection and segmentation from multi sequence MR images [173]. The focus is to minimize false-positive (FP) for gliomas using a single layer DAE, which is referred as novelty detector (ND). Further the work modifies ND via cascading, i.e. cascaded ND (CND) to generate exclusive error distributions for different gliomas components. The study also demonstrates semi supervised learning by pretrained SDAEs using patches drawn from 20 patient volumes. Experiments are tested on BRATS-2013 and BRATS-2015 using T1-W, T2-W, CE-T1W, and FLAIR MR modalities. The performance is compared with previous submission [22] to BRATS 2015 challenge and resulted in an improvement of 8% in DSC considering tumor core for LGG volumes. Visual results are presented using ISLES 2015 challenge dataset for lesion detection of 28 patient volumes using ND and CND.

Auto-encoders are combined with other concepts too. One such study attempts to overcome several issues related to generative adversarial net (GAN) by integrating it with deep convolutional auto-encoders [174]. The system is termed as variational auto-encoder GAN (VAE-GAN) and is designed to encode complete brain MR slice context. Focus is to determine the benefits of mapping healthy anatomy to a well-structured, latent manifold with GAN. Experimental results are generated by training the system using 83 healthy cases and testing it on 49 multiple sclerosis (MS) lesions with FLAIR modality. Comparison done with deep auto-encoding

Table 2.4: Summary of a few studies utilized Auto-Encoders for brain tumor detection.

Reference	Technique	Modality	User Interaction	Size of Dataset	Tumor type	Performance(%)
[15]	SDAEs	T1W	Manual and Auto-matic	9 patients	Brainstem cancer	DSC: 92
[22]	SDAEs	T1W, T2W, FLAIR, CE-T1W	Automatic	HGG: 220, LGG:54	HGG, LGG	WT HGG:81, LGG:72 SEN:79, SPEC:84
[153]	SDAEs	-	Manual and auto-matic	10 images	-	ACC:98.04
[173]	DAEs and SDAEs	T1W, T2W, FLAIR, CE-T1W, DWI	Semi-automatic	HGG: 220, LGG:54	HGG, LGG, edema	DSC for HGG WT:85, TC:71 AT:75
[174]	VAE-GAN	FLAIR	Automatic	Training: 83 Testing: 49	MS lesion	DSC=60.50
[175]	CVAEs	-	Automatic	Training : 3000 Testing: 29550	GBM	ACC Baseline:84 U-net:88
[176]	CAAE	T2W	Automatic	Training:35 Testing:42	Lesions	AUC, AAE with $\lambda=0.5$ is 0.906 $\lambda=1.0$ is 0.923
[177]	GAN	T1W, T2W, FLAIR, CE-T1W	Automatic	Training:274 Testing:110	HGG, LGG	DSC:82

ACC:Accuracy, AT: Active tumor, AUC: Area under the curve, CAAE: Constrained adversarial auto-encoders, CARIS: Convolutional autoencoder-based inter-slice interpolation, CVAEs: Conditional variational auto-encoder, DAEs: Denoising auto-encoders, DCNN: Deep convolutional neural networks, DSC: Dice similarity coefficient, GAN: Generative adversarial nets, HGG: High grade gliomas, LGG: Low grade gliomas, MN: Meningiomas, PSNR: Peak signal-to-noise ratio, PT: Pituitary tumors, SDAEs: stacked denoising auto-encoders, TC: Tumor core, VAE-GAN: Variational auto-encoder-Generative adversarial network, WT: Whole tumor.

models proves that the performance of spatial VAE-GAN is the best with DSC of 0.6050 ± 0.1927 . Similarly, a conditional generative network based on VAE is designed that learns the input distribution by embedding label independence in latent space [175]. Another work proposes an adaptive augmentation algorithm to train discriminative model utilizing generative models [178]. The approach helps in controlling image characteristics by tuning the latent variables properly. Results also report performance improvement over other baseline and U-net architectures. For FLAIR MR images of BRATS dataset, respective accuracies of 84% and 88% are achieved for baseline and U-net network.

Convolutional autoencoder-based inter-slice interpolation (CARISI) framework is designed for brain tumor inter-slice interpolation using convolutional auto-encoders (CAEs) [179]. The main focus is to effectively deal with the gap that occurs due to rapid changes between two consecutive image slices. Experiments performed on dataset proves the ability of CARISI in slice generation that are more accurate, refined, and helps in exhaustive 3D shape reconstruction of the brain tumor by assisting medical domain at its best. In some interpolations, it is assumed that pixels change gradually and thus experimentation satisfies with the best possible results [180]. However, the said assumption gets violated in some cases [181]. A discussion on several other common issues faced in latent consistency while investigating VAE and adversarial auto-encoders (AAE) is presented deeply in one of the research studies [176]. In addition, lesion detection in an unsupervised learning-manner with improved accuracy and latent constraint is proposed. Experiments are executed by training the model with 35 T2-W MR images extracted from Human Connectome Project (HCP) data source. Though detection performance is evaluated on T2-W images of 42 subjects taken from BRATS 2015. The accuracy is depicted through area under the curve (AUC) which turns to be 0.923 for AAE considering $\lambda=1$. A thought to replace conditional random fields (CRF) with an automatic GAN is also explored with an aim to achieve better results compared to traditional state-of-art brain tumor segmentation methods [177]. The main reason behind is the

high-order smoothing which in turn affects the performance. Experimental results generated on BRATS 2015 database outperforms CNN based approaches reporting dice coefficient of 82% and segmentation time of 10.8 seconds.

2.2.7 Hybrid Techniques

Hybrid technique means integrating two or more techniques so as to attain better results compared to those reported by individual techniques. The objective of such integrations is to conquer loopholes of one technique by another technique. Based on the observations made during the literature survey the term ‘hybrid’ in respect to brain tumor detection systems has three variations, viz. segmentation-segmentation, segmentation-classification, and classification-classification. Some of the works from each of the three variations has been discussed so as to perceive their advantages as well as uses.

Integration is not new in the field of medical image processing, hybrid techniques were popular even in back years to fulfill the increasing demands of perfection. SVM is very popular in such scenarios and is aggregated not only with traditional but with emerging approaches too. A novel approach combining wavelets individually with self-organizing maps (SOM) and SVM is presented to classify brain MR images [182]. Dataset consisting of 52 axial, T2-W MR brain images (normal and abnormal, i.e. affected by Alzheimer’s disease) is utilized for performance evaluation. Effective classification results of more than 94% and 98%, reported respectively for SOM and SVM demonstrate utility of the proposed method. Another work integrates SVM with genetic algorithms (GA) and implements spatial gray level dependence method (SGLDM) to optimally extract texture features from normal and tumor regions [183]. Experiments are done on real brain dataset from Harvard Medical School website consisting of 83 images (29 - normal, 22 - malignant; suffering from low grade glioma and meningioma, and 32 - benign; suffering from bronchogenic carcinoma, glioblastoma multiform, sarcoma, and grade IV tumors). The approach

reports accuracy varying from 94.44% to 98.14% and sensitivity ranging from 91.9% to 97.3%. SVM classifier is even integrated with fuzzy c-means (FCM) clustering for tumor detection and classification [184]. Results are concluded using 120 patients bases and the performance is compared on four different datasets with ANN. Larger datasets report better performance with ANN but for smaller ones SVM is found more suitable. Overall it can be said that hybrid SVM produces satisfactory accuracy rate and lesser errors.

FCM is another popular candidate in the hybrid category. A novel region-based fuzzy clustering method namely, enhanced possibilistic fuzzy c-means (EPFCM) is designed to overcome initialization and weak boundaries constraints in region based approaches [94]. The combination is utilized to segment tumor in 10 CE-T1W and 5 FLAIR MR images and is observed to achieve an average similarity and Jaccard indices as 95.3% and 82.1%, respectively. An effective segmentation approach, KIFCM integrates k-means clustering with FCM [185]. k-means reduces execution time whereas FCM excels accuracy of brain detection. The study used three datasets, namely digital imaging and communications in medicine (DICOM) - 22 images, Brain Web dataset - 152 images, and BRATS - 81 images. Both KIFCM and FCM report similar accuracy for these datasets but KIFCM takes lesser processing time by reducing the number of iterations. Precisely, for the three datasets KIFCM respectively reports 90.5%, 100%, and 100% accuracy. Another work explores a novel combination of SOM with learning vector quantization (LVQ) to develop effective segmentation approach [142]. The study also proposes skull stripping algorithm. T1-W, T2-W, and FLAIR MR images of twenty patient's suffering from glial tumor is utilized to perform experiments. The skull stripping algorithm outperforms other well-known algorithms on IBSR dataset. However, performance on BRATS 2012 dataset reports an average dice similarity indices of 91% (WM), 87% (GM), 96% (CSF), 61% (tumor), and 77% (edema) which are better than other state-of-the-art techniques.

The concept of hybrid is extended further with the idea of blending more than

two approaches. FKM poorly handles issues related to large range of data. In an attempt to improve data handling capacities, FKM is integrated with SOM to design a tumor detection approach [21]. SOM assists in initial clustering and dimensionality reduction too. This combination obtains an accuracy of 96.18% and sensitivity of 87.18% for T1-W, T2-W and FLAIR MR images collected from Harvard Brain Repository. It is found to perform superior in comparison to existing segmentation techniques. A hybrid, efficient, fully-automatic brain tissue segmentation framework based on clustering techniques is presented [186]. The system starts by fixing pixel intensity values in order to enhance image contrast, then super-pixel algorithm is employed to join pixels with similar intensity into objects. The objects are further processed independently using three clustering techniques (k-means, FCM, and SOM) to produce the labels corresponding to each object. Lastly, neural network (NN) is trained with extracted object features and their labels generated by clustering techniques. Results obtained on T1-W MR images collected from internet brain segmentation repository (IBSR) are reported in terms of accuracy (98.10%), specificity (98.97%), and sensitivity (79.66%). These values are observed to outperform those achieved for other clustering approaches.

A novel segmentation algorithm, termed as SFRCM, i.e. soft fuzzy rough c-means is designed to precisely extract brain soft tissues, namely WM, GM, and CSF, with bias field correction [5]. Initially soft fuzzy rough sets are obtained by embedding the concepts of soft sets, rough sets, and fuzzy sets. They define lower and upper cluster approximations for each pixel using soft fuzzy sets indicating degree of belongingness. The efficacy of SFRCM is evaluated on simulated BrainWeb database (20 images), IBSR (20 images), BRATS (10 images) having T1-W, T2-W and FLAIR MR images to detect high grade glioma. Visual results clearly show differences between brain tissues and tumor. Quantitatively, SFRCM achieves tumor detection accuracy of 94.04% with Gaussian noise (3%) and 87.94% with salt and pepper (5%) noise. A novel integration of an optimization technique with FKM is introduced to process to MR images [191]. The bacteria foraging optimization

Table 2.5: Summary of a few studies utilized hybrid segmentation approaches for brain tumor detection.

Reference	Technique	Modality	User Interaction	Size of Datasets	Tumor type	Performance(%)
[182]	DWT+SVM DWT+SVM with radial basic function based kernel DWT+SOM GA+SVM	T2W	Automatic	52 images	Alzheimer's disease	ACC=96
[183]						
[187]	Hybrid ACPSO-FNN (DWT+PCA), EPFCM+ deformable model	T2W	Automatic	83 images	LGG, MEN, GBM	ACC= 94.44 - 98.14, SEN= 91.4 - 97.3
[94]						ACC=98.75
			Manual and Automatic	15 images	gliomas, MEN, Pick's dis- ease, Alzheimer's disease	JC=82.1 SI=95.3
[184]	FCM + SVM	-	-	120 images	-	Linear ACC=91.66,SEN=83.33, SPEC=100 Quadratic ACC=83.33, SEN=66.66, SPEC=100 Polynomial ACC=87.50,SEN=75, SPEC=100
[185]	K-means integrated with Fuzzy c-means	DS2= T1W, T2W, PD DS3= T1W, T2W, CE-T1W, FLAIR	Automatic	DS1=22 images, DS2=152 images, DS3=81 images	HGG, LGG	Accuracy DS1=90.5 DS2=100 DS3=100 DSC WM= 91 GM=96 CSF=61
[142]	SOM+LVQ	T1W, T2W, FLAIR	Manual and automatic	20 patient data	Glial tumor	
[21]	SOM+FKM	T1W, T2W, CE-T1W, FLAIR	Automatic	4 patient data	LGG, HGG, HGA, MBC	ACC=96.18 SEN= 87.18
[188]	Semi-MrMOO, Multi-center based encoding Soft fuzzy rough c-means clus- tering algorithm (SFRM)	AMOSA, T1W,T2W,PD T1W, T2W, FLAIR	Automatic	-	Alzheimer's disease	-
[5]			Automatic	100 images	HGG	Accuracy Gaussian(3%)=94.04, Salt&Pepper(5%)=87.94 ACC=98.10 SEN=79.66 SPEC=98.97 DSC=79.3
[186]	Super-pixel algorithm+ means, FCM, SOM)	(k- T1W	Automatic	28 patient data	-	ACC=98
[189]	CMLS+AANN	T1W, T2W, CE-T1W, FLAIR	Automatic	384 images	HGG, LGG	ACC=99.84
[190]	AFBNN	T2W	Automatic	81 images	-	SEN=97.24 SPEN=99.85

AANN: Adaptive Artificial Neural Network, ACC: Accuracy, ACPSO-FNN: Adaptive Chaotic Particle Swarm Optimization Forward Neural Network, AFBNN: Adaptive Firefly backpropagation neural network, AMOSA: Archived MultiObjective Simulated Annealing, MOO: Multiobjective optimization, CMLS: Cognition based modified level set, DF: deformable model, DSC: Dice similarity coefficient, DS1: Data Source 1, DS2: Data Source 2, DS3: Data Source 3, DWT: Discrete wavelet transform, EPFCM: Enhanced Possibilistic Fuzzy C-Means (EPFCM), FCM: Fuzzy c-means and k-means, GA: Genetic algorithm, GBM: Glioblastoma multiforme, HGA: High grade astrocytoma, JC: Jaccard Coefficient, MBC: metastatic bronchogenic carcinoma, MCBE: Multi-center based encoding, MEN: Meningiomas, SEN: Sensitivity, SFRM: Soft fuzzy rough c-means clustering algorithm, SI: SIMILARITY, SPEC: Specificity.

(BFO) based modified fuzzy k-means algorithm (MFKM) is designed and is proved to take minimum computational time in handling MR image sequences. MFKM initializes the segmentation process then BFO algorithm works to derive optimal threshold values. Experiments performed produce appreciable results in terms of sensitivity and specificity, which turns to be 97.14% and 93.94%, respectively.

The complete summarization of deep learning, auto-encoders, and hybrid techniques highlighting their relationship with MR image modalities, tumor disease, and data set affecting the performance evaluation is respectively given in Table 2.3, 2.4, and 2.5. However, based on this summarization it can be said that automatic method of segmentation is mostly used with all MR image modalities whereas traditional techniques mostly focus single modality.

2.3 Medical Imaging Systems for Efficient Handling of Big Data

The emerging trends have drifted the research from accuracy toward processing time so as to perform tumor segmentation in minimal time. In addition, another prospective is a shift from small MR tumor data to larger volume. Tremendous amount of MR medical data storage, processing, and distribution needs much beyond the traditional algorithms or hybrid frameworks. The need is to parallelly handle MR brain image data through multi-nodes thus increasing speed and efficiency of the system [192] [193]. In 2016, MR brain tumor segmentation is evaluated using parallel genetic approach (PGA) in Hadoop and Spark tools for varying data sizes [54]. The novel approach is in cloud environment, therefore, an effective lossless compression methodology for transferring of brain data is used to have communication with the cloud. For experimentation Connectome project dataset is taken in range of 2GB to 10GB which are processed through 1 to 5 cluster nodes. The execution time of Hadoop ranges from 12sec to 24sec whereas in case of Spark it reduces and ranges from 4sec to 7sec. However, in this work the network based platform (Spark)

for brain image processing supersedes the cloud based (Hadoop) platform. In the same year, 2D brain MR image analysis system has been proposed integrated with Hadoop MapReduce framework [194]. The objective is to reduce the retrieval time of the brain MR images. The support of core image processing approaches makes MapReduce capable of handling large image data than traditional content based medical image retrieval system (CBMRS). However, such image retrieval studies are related to evidence-based practices (EBP), which act as a guide to experts, doctors, and radiologists to undertake patients with similar cases. The two stage approach leads to a decision of surgery, if more number of retrieved images is tumor affected. Performance of the system is validated on 10 images with execution times of 29msec and 27msec for replication factors of 2 and 3, respectively.

In 2017, framework for automatic diagnosis of 3D brain MR images is proposed, which utilizes scale invariant feature transform (SIFT) to extract image features [195]. Further, SIFT feature vectors are mapped with each input image using MapReduce and are stored using Hadoop distributed file system (HDFS) within the system. MapReduce divides massive images across different clusters which in return reduces time. Evaluation is presented on different MR image datasets by running Hadoop in chunks of 20, 60, and 100 image classes; however the whole system accuracy turns to be 80%. Multimodal brain tumor MR image segmentation using one of the statistical segmentation techniques, known as Hidden Markov Models (HMMs) is presented employing MapReduce for training to increase system throughput [196]. The proposal is evaluated and validated on BRATS-2013 dataset, consisting of simulated and real patients details suffering from high grade glioma (HGG), in terms of dice similarity coefficient (DSC) and processing time. Average DSC of 89.2% (simulated) and 65.6% (real) is reported for tumorous cases. Further, MapReduce with 8 multi-node platform results in better processing time (approximately less than a minute) for a single patient. It is evident that messing with performance and time simultaneously with any external platform is impractical. However, incorporation of multiple cloud and network based platforms ensure organized as well as parallel

handling of massive MR image data.

Another study attempts mapping and prediction of healthy/effected MR brain images based on analysis of visual parameters, i.e. content based image retrieval (CBIR) approach, utilizing Hadoop platform for large scale image data [197]. The proposal further employed Lucene indexing technique for reduced search time thus faster performance and classifies tissue abnormality via support vector machine (SVM) for efficient predictions. Research on medical image retrieval is growing at fast pace to extract texture features with the assistance of Hadoop, HDFS and MapReduce paradigm. Working on similar lines, researchers designed an efficient medical image retrieval system using CBIR approach incorporated with Hadoop (Hadoop-CBIR) [198]. Experimental evaluations showing considerable accuracy at minimum retrieval time are presented with high number of input images. In particular, examination of 80 images (5GB) with replication factor (RF) as 5 measures a response time of 72msec.

2.4 Discussions

Studies covered in this research clearly presents the depth as well as the width of research in the field of brain tumor detection. One can thus easily observe that the employed techniques, MR image modalities, brain disease, data source, and number of cases definitely affect performance of the developed system. Lot many systems are handled solely as manual, semi-automatic, or fully automatic, but in some cases combination is also applied. To recapitulate automatic (without user interference) is considered the best from the researcher point of view. However, the same cannot be true when applied practically on any affected patient. A second opinion is always required to assure the system's decision. References in support of these observations are included in Fig. 2.4. Difference in observations made by four different practitioners while manually segmenting MR image to detect tumor (glioma) is clearly visible in Fig. 2.4(i). Similarly, Fig. 2.4(ii) shows automatic segmentation results generated using original as well as modified gradient vector flow snake (GVF).

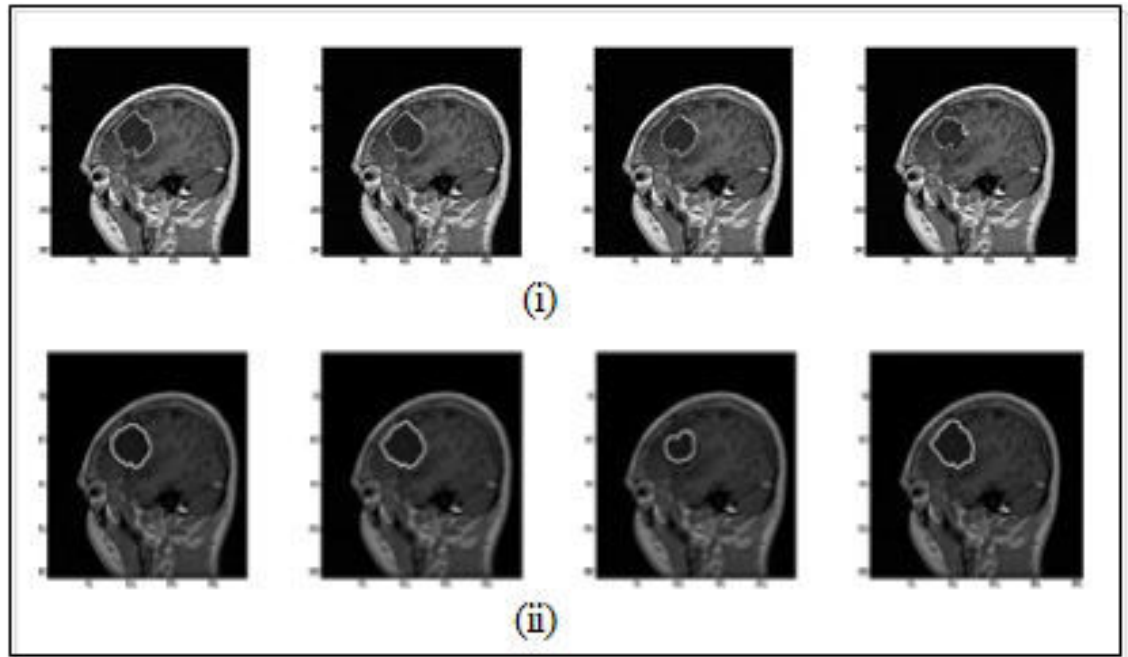


Figure 2.4: (i) Manual segmentation by four different experts to detect glioma in MR image. (ii) Automatic Segmentation of glioma using improved GVF snake [2].

Again differences in locating glioma via both versions of GVF is observed, however tumor extraction is still near to accurate. From this deliberation can be concluded that automatic is not always effective. Similarly, for [3] semi-automatic systems, Figure 2.5 is included that displays results generated using a system assisted follow-up in neuroimaging of therapeutic intervention (AFINITI) as well as manual annotation. It is visible that the calculated tumor sizes in terms of volume are generally smaller than those marked during manual segmentation.

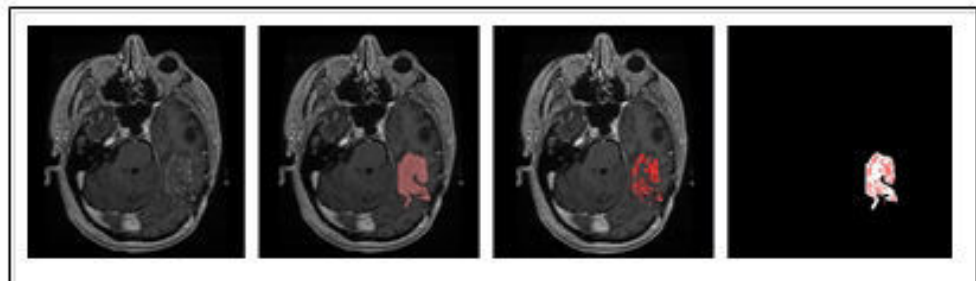


Figure 2.5: Difference between manual and semi-automatic segmentations. Images in sequence are original image, manual segmentation, semi-automatic segmentation, and difference between manual segmentation [3].

Though drift from traditional to emerging techniques will happen to keep various parameters (MRI sequence, tumor type, dataset etc.) synchronized. Most of studies signify need of semi-automatic as well as fully automatic detection, also proved the same with the best visual results. However, there are cases where even the latest deep learning approaches fail to segment the brain MR image effectively. In figure 2.6, colors differentiate different brain tumor areas. Here manual performs better and due to over segmentation automatic segmentation lacks behind.

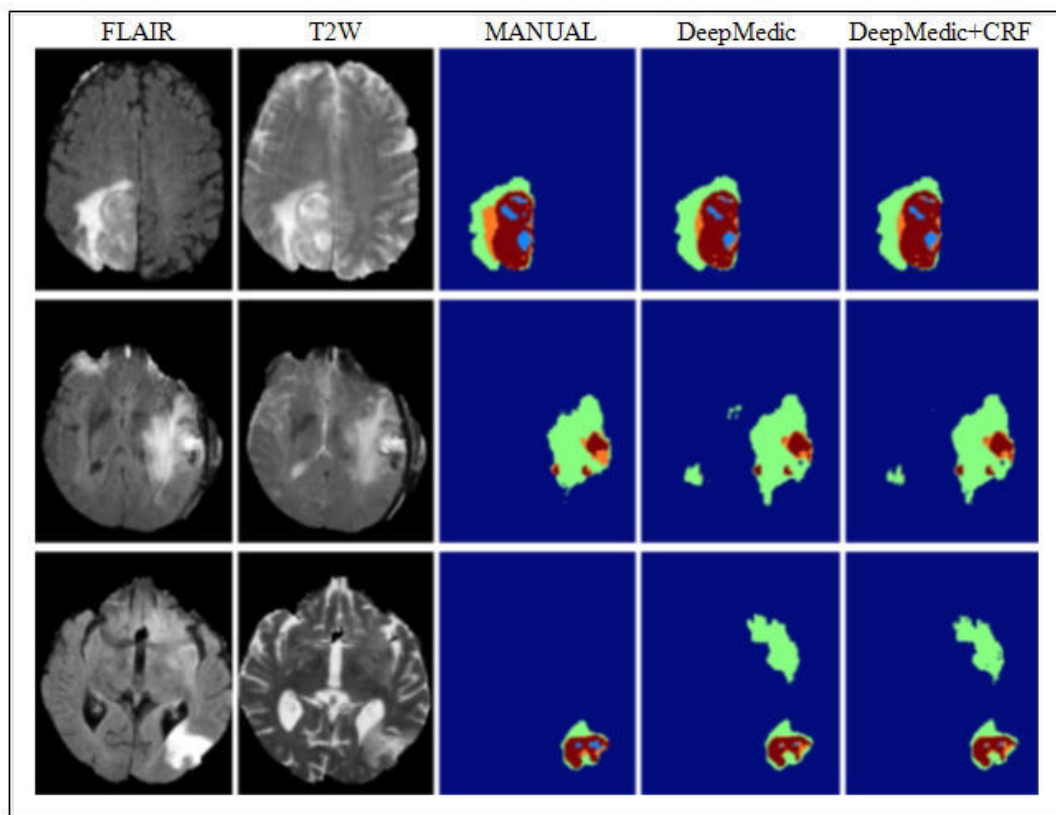


Figure 2.6: Satisfying segmentation of the tumour, regardless motion artefacts (top and middle), One of the worst cases of over-segmentation observed (bottom) [4].

Lastly, an attempt to compare all the segmentation approaches is made based on certain parameters which are figured while surveying the existing literature in the domain. Parameters identified are Computation Time (time required to generate the result after an input image is provided), Computational Simplicity (ease with which algorithm can be implemented), Dataset Size (ability to handle big data),

Table 2.6: Segmentation approaches and their values for various parameters identified during the survey.

Segmentation technique	Computation time	Computation simplicity	Dataset size	Noise invariance	Support to real dataset	Spatial information
Thresholding	Slow	Low	Small	Yes	No	No
Edge Based	Fast	Low	Small	Yes	No	No
Contour Based	Slow	Moderate	Small	Yes	No	Yes
Region Based	Slow	Low	Small	Yes	No	Yes
Atlas based	Slow	Moderate	Small	Yes	No	No
Level set	Slow	Moderate	Small	No	No	No
Deformable Model	Fast	Low	Large	Yes	No	No
EM	Fast	Low	Large	Yes	No	No
MRF	Fast	Moderate	Small	No	No	No
GMM	Fast	Low	Small	No	No	No
K-NN	Fast	Low	Large	Yes	Yes	Yes
SVM	Slow	Moderate	Large	Yes	No	Yes
NN	Fast	Moderate	Small	Yes	No	Yes
LVQ	Slow	Low	Small	Yes	No	Yes
K-means	Fast	Low	Large	Yes	No	Yes
FCM	Slow	Low	Large	Yes	No	Yes
SOM	Fast	Low	Large	Yes	No	Yes
CNN	Slow	High	Small	Yes	Yes	Yes
DCNN	Slow	High	Small	Yes	Yes	Yes
Auto-encoders	Slow	High	Small	Yes	Yes	Yes

Noise Invariance (capability to resist presence of noise in input image), Support to Real Dataset (cross-domain performance), and Spatial Information (is the algorithm utilizing spatial image data effectively). The range of allowed values for each identified parameters is as follows: Computation time (Slow or Fast), Computational simplicity (Low, Moderate, or High), Dataset size (Small or Large), noise invariance (Yes or No), Support to real dataset (Yes or No), Spatial information (Yes or No). Using the above information, segmentation approaches along with their values are presented in Table 2.6. Clearly, the traditional ones are better computational but provide no support to real datasets, are not invariant to noise, and does not even utilize spatial information at the desired extent. In comparison, new participants in the domain especially the CNN, DPNN, and auto-encoders supersedes in all aspects.

2.5 Summary

Statistical analysis of MR image modalities, data sources, as well as current scenario in each of the considered detection technique is presented. Utilization of the best modality based on the ease of availability is summarized considering the articles covered in this research. Based on the existing literature of brain tumor segmentation techniques several deliberations concerning performance improvements of the developed system can also be concluded. The most prominent is the automatic CAD system for detection. Firstly, the execution speed of segmentation approach can be improved by consolidating automation within it [79] [95] [138] [191]. Secondly, system efficiency/adaptability can be improved by refreshing prior knowledge bases having parameters ranging from intensity to tissues, like shape, location, area, symmetry, and normal anatomic variability. Studies have worked on filtering as well as better utilization of dataset for precise results [159] [171] [172]. Another possible improvement can be done by taking care of over-segmentation and noise filtering [22] [79] [80] [81] [82] [83] [154] [4]. Furthermore, CAD systems developed using fully or semi automatic unsupervised segmentation approach are preferred over patient-specific trained ones. Varying benefits of existing systems as well as details of the employed technique(s) - either traditional or emerging is presented in depth. Detection method produces open contour sometimes, and is relatively more delicate to threshold. One can think to resolve such intensity changes and threshold related issues.

The neural networks (NN) implementation and it's modified versions helped researchers with great deal. The self-learning capacity of emerging deep learning techniques completely changes the scene of brain tumor detection systems. Thus, looking at the literature of past few years, usage of such techniques can be highlighted as a current trend in this domain too. Now-a-days multi-modal images are gaining interest in brain tumor segmentation as reviewed from the articles of deep learning, auto-encoders and hybrid approaches. Similarly, concepts like GA

and optimization are strongly making their room. Undoubtedly, present era in the focused domain, i.e. brain tumor segmentation, can be characterized with huge data and high research, thus it demands for features like quickness along with accuracy. As this is somehow impossible for a single technique to fulfill, the idea of integrating two or more techniques is blindly employed. Combination may either be inter- or intra-phase. Based on the observations made, the study identifies following possibilities with respect to brain tumor systems. They are segmentation-segmentation, segmentation-classification, and classification-classification. It can also be segmentation-clustering and clustering-clustering. The said observation is supported by summarizing combinational techniques in Table 2.5. An attempt to present another perspective of the current research with respect to each tumor type is also made. Table 2.7 shows the best results, as far as the knowledge of authors are concerned, from each of the considered segmentation category grouped on tumor type. Clearly, there are still few combinations, i.e. tumor type and approach applied which can be worked on as they are either unexplored or has not yet reported acceptable accuracy. Several attempts to automate the segmentation process have already been made and are still being made, but the analysis of existing approaches clearly proves that the level of human interference can significantly improves system's performance and thus matters a lot. Therefore, instead of having a system either completely automatic or time consuming solely manual research will slowly reach a middle way by identifying a semi-automatic more efficient systems.

Furthermore, continuously increasing MR image data needs much beyond the traditional algorithms. Distributed and parallel processing of data from native storage requires a well-defined platform and thus several open source platform comes into picture, like Hadoop, MDCS, etc. The output is expected to be prioritized based on speed and accuracy of the system. The traditional segmentation approaches cover either of the performance parameters, accuracy or processing time. Therefore, research is getting accelerated incorporating Hadoop and Matlab distributed computing server (MDCS) with detection approaches to ease the processing time.

Table 2.7: Summary of performances reported by approaches under each of the considered segmentation category grouped on tumor type.

Tumor type	Technique	Reference	Performance Parameter				
			ACC	SEN	SPEC	JSC	DSC
GBM	Thresholding	[143]	98	100			
	Region/Contour	[18]		98			
	Statistical	[49]				72.3-82.5	
	Classifiers	[119]	98				
	Clustering	[141]	99	100	92.8	83.61	
	Hybrid	[34]	99	100	92	93	
	Deep learning	[48]	96.97			77-88	
Low grade	Thresholding	[67]					82.18-85.11
	Region/Contour	[72]	99				
	Statistical	[110]		91.7	89.4	75	
	Classifiers	[138]	85	89	84		
	Clustering	[62]		86	93		
	Hybrid	[185]	100				
	Deep learning	[155]	97.05				
High grade	Thresholding	[199]		99.89			
	Region/Contour	[93]		84			
	Statistical	[110]		91.7	89.4	75	
	Classifiers	[138]	85	89	84		
	Clustering	[62]		86	93		
	Hybrid	[5]	94.04				
	Deep learning	[155]	97.05				
Astrocytoma	Thresholding	[200]	98	100			
	Region/Contour	-					
	Statistical	[49]				76.9	
	Classifiers	[201]	85				
	Clustering	[143]	93				
	Hybrid	[21]	97.14	93.94			
	Deep learning	[48]					83
Meningioma	Thresholding	[68]	92.2				
	Region/Contour	[72]	99				
	Statistical	[102]				49-71	
	Classifiers	-					
	Clustering	[141]	99	100	92.8		
	Hybrid	[21]	97.14	93.94			
	Deep learning	[202]					81
Alzheimer's	Thresholding	[203]	80				
	Region/Contour	[204]	98	96.45	99.09		
	Statistical	[205]	91				
	Classifiers	[144]	100				
	Clustering	[141]	99	100	92.8		
	Hybrid	[34]	99	100	92		
	Deep learning	[206]	100	100	100		

Study of the existing literature has been carried out systematically and thus the brain tumor detection systems are analyzed from different prospectives like modalities, datasets, categorization of performance parameter, tumor types, etc. Intensity and illumination losses are commonly observed in almost all MR image modalities while undergoing segmentation, which there after effects the performance of systems. Fuzziness may enhance overall efficacy of any medical image analyses system if fuzzy values are appositely incorporated within each of its module, particularly segmentation. However, a colossal growth in this research domain has generated the possibilities of improved membership functions for better segmentation, i.e. illuminated clusters highlighting the tumorous area(s). In turn the research headway towards designing and implementing a fuzzified segmentation approach based on penalizing membership function along with the development of a framework for identifying brain tumor efficiently. Hence an attempt has been made to cover all the observations made during the literature survey.

Chapter 3

The Proposed Segmentation Approach

Segmentation is a significant step in automating the process of brain tumor detection. Researchers have extensively explored clustering methods, like k-means, fuzzy c-means (FCM), etc., along with SVM and have introduced fuzziness at several levels for better accuracy. In other words, introduction of fuzziness shows an improvement in the performance of a brain tumor detection system. In addition, a prolific modification in the existing techniques sometimes proves fruitful. These observations are inspirations for the work carried out in this chapter. An automated algorithm to segment abnormal tissues in MR brain images is designed and implemented. The proposed algorithm is termed as weighted fuzzy k-means (WFKM). It uniquely introduces the notion of weights in fuzzy c-means (FCM) to overcome a few of the inadequacies in the existing algorithms. WFKM incorporates the notion of k-means as well to effectively detect tumor infiltration levels. Essentially WFKM works on spatial context with illumination penalize membership approach, thus it is found helpful in settling issues related to pixel's multiple memberships as well as exponential increase in number of iterations. It is also observed to report better classification accuracy.

In this chapter the detailed overview of system framework is done. Initiating

and setting the background related to the k-means and fuzzy c-means segmentation approaches for brain tumor is done in Section 3.1. Followed by Section 3.2, where the overall detailed architecture of the proposed brain tumor detection system and gives an insight of all the basic modules involved. Section 3.3 discusses all the experimental setup and data acquisition required for the implementation of proposed approach. Section 3.4 summarizes as well as highlights the experimental observations from the prospective of segmentation effectiveness quantitatively (accuracy measures) and qualitatively (visually). Finally, a summarized deliberation of the chapter is presented in Section 3.5.

3.1 Background

3.1.1 k-means

k-means clustering aims to partition n data-points into c clusters. The objective is to minimize the total intra-cluster variance as defined in Eq. 3.1. Each cluster is identified via its centre and is computed using Eq. 3.2. During initialization some cluster centres are chosen and data-points are distributed among clusters. Each data-point belongs to only one cluster whose centre is nearest to it (Eq. 3.3). An iterative process of attributing the closest cluster to each data point and recalculating new cluster prototypes is then followed till the termination criteria is met. In the research work presented here data-points are pixels of an input MR image. According to the terminology mentioned in Nomenclature, standard k-means equations are defined here.

$$J(C) = \sum_{k=1}^c \sum_{\forall i, I_i \in C_k} \|I_i - C_k\|^2, 1 \leq i \leq n \quad (3.1)$$

$$C_k = \frac{1}{|C_k|} \sum_{\forall i, I_i \in C_k} I_i, 1 \leq i \leq n \text{ and } 1 \leq k \leq c \quad (3.2)$$

$$I_i \in C_k \iff \|I_i - C_k\| < \|I_i - C_j\|, 1 \leq i \leq n, 1 \leq k, j \leq c \text{ and } k \neq j \quad (3.3)$$

where $\|I_i - C_k\|$ is any distance measure used to compute similarity between a data-point and a cluster center.

3.1.2 Fuzzy c-means

Fuzzy c-means (FCM) introduced by Professor Jim Bezdek in 1981, is the basis of the proposed methodology [207]. FCM is similar to k-means but partitions n data-points into c fuzzy clusters, i.e. a data-point belongs to more than one cluster with some probability which is defined by a membership function. u_{ik} is the degree of membership of pixel I_i in k^{th} cluster and is based on the distance between I_i and C_k . Each cluster is identified by its centre. It works with a fuzziness exponent, m , to control the number of iterations. An objective function that FCM aims to minimize is given in Eq. 3.4. Similar to k-means, it's an iterative optimization algorithm that updates u_{ik} and C_k using Eq. 3.5 and Eq. 3.6, respectively, till the termination criteria (Eq. 3.7) is met. Considering image pixels as data-points and following the terminology mentioned in Nomenclature, standard fuzzy c-means equations are defined here.

$$J(C, U) = \sum_{i=1}^n \sum_{k=1}^c u_{ik}^m \|I_i - C_k\|^2, \quad 1 \leq m \leq \infty \quad (3.4)$$

$$U_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|I_i - C_k\|}{\|I_i - C_j\|} \right)^{\frac{2}{m-1}}}, \quad 1 \leq i \leq n \text{ and } 1 \leq k \leq c \quad (3.5)$$

$$C_k = \frac{\sum_{i=1}^n u_{ik}^m I_i}{\sum_{i=1}^n u_{ik}^m}, \quad 1 \leq k \leq c \quad (3.6)$$

$$\|U^{(z+1)} - U^{(z)}\| \leq \epsilon, \quad 0 \leq \epsilon \leq 1, \text{ and } z \text{ is an iteration step} \quad (3.7)$$

3.2 The Proposed Approach

Traditional and swift k-means can handle large datasets so is popularly utilized in computer aided detection systems to segment tumorous region from brain [208], however it lacks ROI completeness in case of malignant tumor [209]. In contrast, FCM retains more information and operates accurately on any tumor type but at the cost of execution time due to noise and outliers sensitivity. A few recent researches integrated these approaches to attain better results [185] [210]. Working on the same lines, the proposed hybrid approach named as weighted fuzzy k-means (WFKM) is designed to bring benefits of the specified algorithms together. The main aim is to achieve best possible clusters in lesser number of iterations and hence better execution time. The experimental results clarify that WFKM is efficient enough to detect tumor that cannot be detected by k-means, FCM, and their combination.

The process flow of the proposed approach is depicted in Fig. 3.1, it starts by taking original MR image as an input which undergoes pre-processing mainly for noise removal. In the current proposal Gaussian mask filter is applied. Next skull scripting is performed which however, is an important phase of brain segmentation process to separate out normal tissues like WM, GM, and CSF from tumor tissues. Thus, for improved and effective detection Otsu's thresholding is utilized in this work followed by segmentation using proposed hybrid WFKM approach. To accentuate tumor shape morphological filtering using two basic operations (dilation, i.e., shrinking the foreground and erosion, i.e., expanding the foreground) is done. After this phase features are achieved which in turn helps to retrieve the minimum maximum range of the values, however, affecting the performance of system. Lastly, the

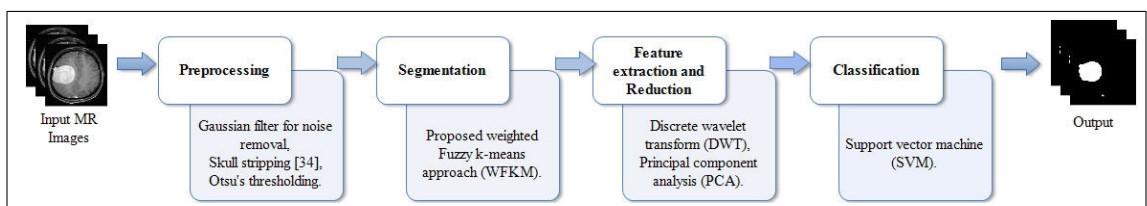


Figure 3.1: The proposed system model for brain tumor detection using MR image.

highlighted white region in the obtained image signifies tumor and the system finally classifies the image as tumorous accordingly along with the class (benign or malignant). All the main phases of the proposed system are detailed in the subsequent sections.

3.2.1 Preprocessing

MR brain images are prone to sensitivity but achieving maximum quality with minimum redundancy is at the priority. Thus several measures to enhance quality of raw input image are applied in pre-processing phase before proceeding further. Basically, signal-to-noise ratio gets improved by tasks like unwanted noise removal, background smoothing, inner regions refinement, edges and region preservation. In the developed brain tumor detection system, Gaussian filtering is employed by convolving an input image with Gaussian mask represented as a linear low-pass filter to eradicate noise. It brings each pixel value in closer concordance with neighboring pixel values which reduces edge blurring and helps in smoothing. Next pre-processing step is skull stripping to confiscate irregularities like skull, eyes, irrelevant anatomy structures, and scalp [211]. Lastly intensity-based Otsu's thresholding is utilized to extract salient features of brain tumor MR images. The original algorithm is meant for binarization by calculating an optimal threshold that minimizes the summation of spread of foreground and background pixels. The outcome thus highlights the tumorous region and improves processing speed due to less storage. A modified Otsu's thresholding is employed to obtain a gray image as an output, i.e. if an image is to be segmented into K clusters, then algorithm computes $K - 1$ optimal thresholds values. Mathematically, the between class variance (σ_B^2) as given in Eq. 3.8 is maximized over $K - 1$ possible thresholds.

$$\sigma_B^2 = \sum_{k=0}^{K-1} s_k \cdot (\mu_k - \mu_T)^2 \quad (3.8)$$

where,

$$\mu_T = \sum_{k=0}^{K-1} \mu_k \cdot s_k \quad (3.9)$$

$$\mu_k = \sum_i I_i^k \cdot \frac{p_i^k}{s_k} \quad k \in \{0, 1, \dots, K-1\} \quad (3.10)$$

$$s_k = \sum_i p_i^k \quad k \in \{0, 1, \dots, K-1\} \quad (3.11)$$

If G be the number of gray levels in an input MR image and g_j is the number of pixels at j^{th} gray level, then p_i^k , the probability associated with i^{th} image pixel lying in k^{th} cluster is defined as in Eq. 3.12.

$$p_i^k = p_j \quad \begin{cases} \text{if intensity of } I_i^k \text{ falls in } j^{th} \text{ gray level} \\ p_j = \frac{g_j}{\sum_l g_l} \end{cases} \quad (3.12)$$

3.2.2 The Proposed Weighted Fuzzy k-means Approach

The proposed weighted fuzzy k-means (WFKM) approach aims to fuzzify pixel values along with weights assignment to accurately identify tumorous region in the pre-processed brain MR image. WFKM blends fuzzy c-means with k-means utilizing the spatial context information for clustering. A penalized membership approach is designed to control variations on illumination and grey values of a pixel while clustering. In addition, noise proneness due to multiple pixel membership generations is also assisted. The approach basically supervises scattered clusters and unclear segments due to FCM during each of the iterations. WFKM works towards minimization of the objective function $J(C, U)$ as defined in Eq. 3.13. Here l_k is the luminosity (Eq.3.14) and w_k is the weight (Eq. 3.15) computed for cluster k . The membership of pixel I_i in k^{th} cluster is updated as in Eq. 3.16. The constant, α ,

controls the impact of illumination.

$$J(C, U) = \sum_{i=1}^n \sum_{k=1}^c u_{ik}^m \|I_i - C_k\|^2 - \frac{1}{2} \alpha |l_k - w_k|, \quad (3.13)$$

$$1 \leq m \leq \infty$$

where,

$$l_k = \frac{\sum_{i=1}^n n I_i^k}{\sum_{i=1}^n u_{ik}} \quad (3.14)$$

$$w_k = \frac{\sum_{i=1}^n \sum_{j=1}^n |I_i^k - I_j^k|}{\sum_{i=1}^n u_{ik} \times (\sum_{i=1}^n u_{ik} - 1)}, \quad 1 \leq k \leq c \quad (3.15)$$

$$u_{ik} = \frac{1}{\sum_{j=1}^c (|I_i - C_k| - |I_i - C_j|)^{\frac{2}{m-1}}}, \quad (3.16)$$

$$1 \leq i \leq n \text{ and } 1 \leq k \leq c$$

The steps constituting WFKM, a hybrid segmentation approach for brain tumor detection, are given in Algorithm 1. The process starts by partitioning an input image into four clusters based on the intensity values defined for *LL* (Low-Low), *LH* (Low-High), *HL* (High-Low), and *HH* (High-High). After that the membership of each pixel is recomputed to define the fuzzy cluster centres. Then the processing of cluster weight values, pixel illuminations, and dividing different intensity clusters starts to achieve improved segmentation of pixels. The loop control in the algorithm goes back after weighted FKM to fuzzification due to the minimal error as controlling factor. However, the cluster non-uniformity has a proportional relationship to MSE (mean square error), though a small MSE declination directly reflects the cluster differences. When the MSE threshold (defined as 10^{-7}) is achieved in the proposed algorithm the loop gets stopped, satisfying the criteria. Another stopping condition is based on the number of iteration, i.e. once 50k iterations are achieved the algorithm halts.

Inclusion of cluster weights. The pixel membership notion in FCM results in better clustering, specifically with overlapping large datasets. At the same time, FCM has some issues which demand refinements. WFKM attempts to establish the

Algorithm 1 WFKM

INPUT: A pre-processed input MR image I.**OUTPUT:** A segmented image.

- 1: Divide image into four clusters based on the intensity values defined for LL, LH, HL, and HH.
 - 2: Compute U

$$u_{ik} = \begin{cases} 1, & \text{If } I_i \text{ is in } C_k \\ 0, & \text{otherwise} \end{cases}, \forall i \text{ and } 1 \leq i \leq n; \forall k \text{ and } 1 \leq k \leq c$$
 - 3: Calculate cluster centers using Eq. 3.6.
 - 4: Update U using Eq. 3.16.
 - 5: Compute W .

$$W[k] = \frac{\sum_{i=1}^n \sum_{j=1}^n |I_i^k - I_j^k|}{\sum_{i=1}^n u_{ik} (\sum_{i=1}^n u_{ik} - 1)}, \forall k \text{ and } 1 \leq k \leq c.$$
 - 6: **while** threshold not achieved **do**
 - 7: Calculate illumination variation factor for each pixel.

$$I_i = \frac{\sum_{j=1}^n |I_i^k - I_j^k|}{\sum_{j=1}^n u_{jk}}, 1 \leq k \leq c; \forall i \text{ and } 1 \leq i \leq n.$$
 - 8: Compute fitness of each cluster.

$$F[k] = \frac{\sum_{i=1}^n |mean_k - I_i^k|}{mean_k}, mean_k = \frac{\sum_{j=1}^n I_j^k}{\sum_{j=1}^n u_{jk}}, \forall k \text{ and } 1 \leq k \leq c.$$
 - 9: **for** $k' \leftarrow 1$ to c **do**
 - 10: **while** ($F[k'] < th$) **do**
 - 11: Divide cluster k' into two clusters based on its intensity range
 - 12: Update cluster centers using Eq. 3.2
 - 13: Execute k -means using the updated cluster centers.
 - 14: Update U using Eq. 3.16.
 - 15: Update cluster centers using Eq. 3.6.
 - 16: Update W .

$$W[k] = \frac{\sum_{i=1}^n \sum_{j=1}^n |I_i^k - I_j^k|}{\sum_{i=1}^n u_{ik} (\sum_{i=1}^n u_{ik} - 1)}, \forall k \text{ and } 1 \leq k \leq c.$$
 - 17: Compute penalizing factor.

$$0.5 \times \alpha \times \left| \frac{\sum_{i=1}^n I_i^k}{\sum_{i=1}^n u_{ik}} - W[k] \right|, \forall k \text{ and } 1 \leq k \leq c.$$
 - 18: **end while**
 - 19: **end for**
 - 20: **end while**
-

same by incorporating the concept of weights. Literature highlights an exponential increase in number of iterations with a reduction in FCM termination criteria. Multiple pixel memberships, particularly with those which are either outliers or at boundary, tend towards noisy patterns. The proposal of weight facilitates effective handling of such concerns by removing outliers as well as by considering other imaging artifacts such as spatial information in context. The proposed penalize membership approach in combination with cluster weights throughout the clustering process updates illumination as well as gray value variations of a pixel. Penalizing distance based on illumination also helps in handling multiple pixel memberships and outliers.

3.2.3 Feature Extraction and Reduction

Features of WFKM segmented image is extracted by making use of a well-known mathematical tool, i.e. Discrete wavelet transform (DWT), in the domain of medical images. WFKM initiates more clarity in an MR image with respect to intensity, texture, size, and edges of tumor, which helps in better computation of wavelet coefficients. DWT through one to three levels of decomposition with Daubechies-4 filter banks is implemented in the developed brain tumor detection system. ¹13 ROI features namely, contrast, correlation, energy, homogeneity, mean, standard deviation, variance, entropy, root mean square, smoothness, kurtosis, skewness, and inverse difference moment (IDM), are calculated to form an MR image feature vector. The obtained features are reduced using principal component analysis (PCA). It's effective, well-structured, and transfigure input feature sets having several interlinked variables into a lower-dimensional space while variations intact. Reduced features are less complex and improve memory requirement as well as computation time. It further helps in an efficient and accurate classification of a tumor as benign or malignant.

¹Refer Appendix A for details of the feature vectors.

3.2.4 Classification

Classification of extracted tumor is done using a supervised support vector machine (SVM). Among several existing classifiers, literature observes SVM as the most suitable for brain tumor detection due to its computational efficiency and good generalization performance [46] [184] [212] [213]. Moreover, it is independent of any dimension and feature set. For tumor classification as malignant or benign, the proposed system undergoes training using features extracted after the WFKM segmentation followed by the testing. Initially, the collected dataset has been processed for proper analysis of the defined parameters. In training phase, the sensitive parameters have been selected for effective results. The epoch values are iterated till the required accuracy not achieved.

3.3 Experimental Setup and Data Acquisition

The proposed approach is implemented and executed successfully in MATLAB R2018b running under Windows-10 operating system with CPU 2.6 GHz, i7 processor and 8GB of memory (RAM). Experimental results are generated on images obtained from digital imaging and communications in medicine (DICOM) dataset considering all scan parameters, namely relaxation time (TR), spin echo (TE), and inversion time (T1) [214]. Images acquired through MRI scanner under the magnetic field strength of 1.5 Tesla have 2D structure on which the segmentation steps are executed. Images with 256 x 256 pixel resolution consisting of T1W, T2W, FLAIR, and CE-T1W modalities have been used to analyze performance of the proposed WFKM algorithm. Brain images from three different planes, namely axial, coronal, and sagittal are referred as their symmetry signifies normality of human brain whereas asymmetry strongly supports the abnormality. 5 mm thick slices with 7.50 mm spacing in between are utilized during experimentation. A total 300 MR images are taken from DICOM dataset, out of which 70% is used for training, 15% for result validation, and the rest for testing. Cross-validation is a technique applied

to evaluate expected performance by partitioning the dataset into splits of train, validation, and test. Training set to train the system, and a test set to evaluate it on all these splits. Training and validation done at using 10 fold cross validation for improving the accuracy. Experiments are also performed using more DICOM images ranging from 20 to 1000 images.

3.4 Results and Discussions

Performance of proposed WFKM approach is evaluated on brain MR images with and without tumors. WFKM has reported comparable segmentation results to the existing research works. Hybrid approach (WFKM) is mainly meant to improve segmentation, thus results are presented visually. Furthermore, for classification, tumor features are extracted and optimized to better train the SVM classifier.

Segmentation results: Fig. 3.2 illustrate phase by phase visual interpretation to detect tumor using the proposed WFKM segmentation approach. Fig. 3.2(a) presents the examination of MR image belonging to benign tumor class, whereas Fig. 3.2(b) verifies the same for malignant tumor class. The cluster formation in each phase is clearly evident due to intensity variation generated by inculcating the concept of weights and thus better convergence towards the tumor region.

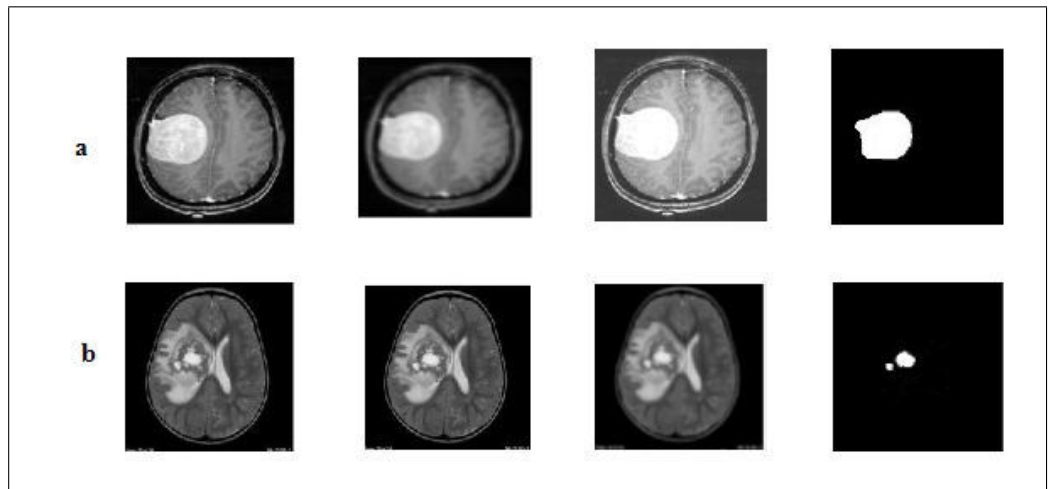


Figure 3.2: Visual segmentation results on T1W MR image using Proposed hybrid WFKM for a. benign tumor and b. malignant tumor.

The performance evaluation: The two tumor classes, benign and malignant are contained in the dataset. Classification accuracy shows how effectively an approach identifies the type of tumor from MR image scans. Evaluation parameter used to assess efficacy of the classification approach in this research work is accuracy Eq. 3.21. In addition, performance is evaluated in terms of sensitivity Eq. 3.22, specificity Eq. 3.23, and dice coefficient index Eq. 3.24. Confusion matrix parameters namely, true positive rate (TP), true negative rate (TN), false positive rate (FP), and false negative rate (FN) are also computed. These are respectively defined as in Eq. 3.17, Eq. 3.18, Eq. 3.19, and Eq. 3.20.

$$TP = \frac{\text{Number of tumorous MR brain images}}{\text{Total number of images in dataset}} \quad (3.17)$$

$$TN = \frac{\text{Number of non-tumorous MR brain images}}{\text{Total number of images in dataset}} \quad (3.18)$$

$$FP = \frac{\text{No. of non-tumorous MR images detected as tumorous}}{\text{Total number of images in dataset}} \quad (3.19)$$

$$FN = \frac{\text{No. of tumorous MR images detected as non-tumorous}}{\text{Total number of images in dataset}} \quad (3.20)$$

Therefore,

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} * 100 \quad (3.21)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} * 100 \quad (3.22)$$

$$\text{Specificity} = \frac{TN}{TN + FP} * 100 \quad (3.23)$$

$$\text{Average dice coefficient index} = \frac{2TP}{2TP + FP + FN} * 100 \quad (3.24)$$

SVM classifiers with four different kernels, namely linear, polynomial, Gaussian (RBF), and quadratic are used to compare classification performance after segmentation using the proposed WFKM . Accuracy graphs plotted with respect to each of the four kernels are shown in Fig. 3.3. Higher the classification accuracy better is

the approach. Transition is specified in the graph to get a clear picture of accuracy values correspond to each image. With linear kernel, the classification accuracy lies between 97% and 98%.

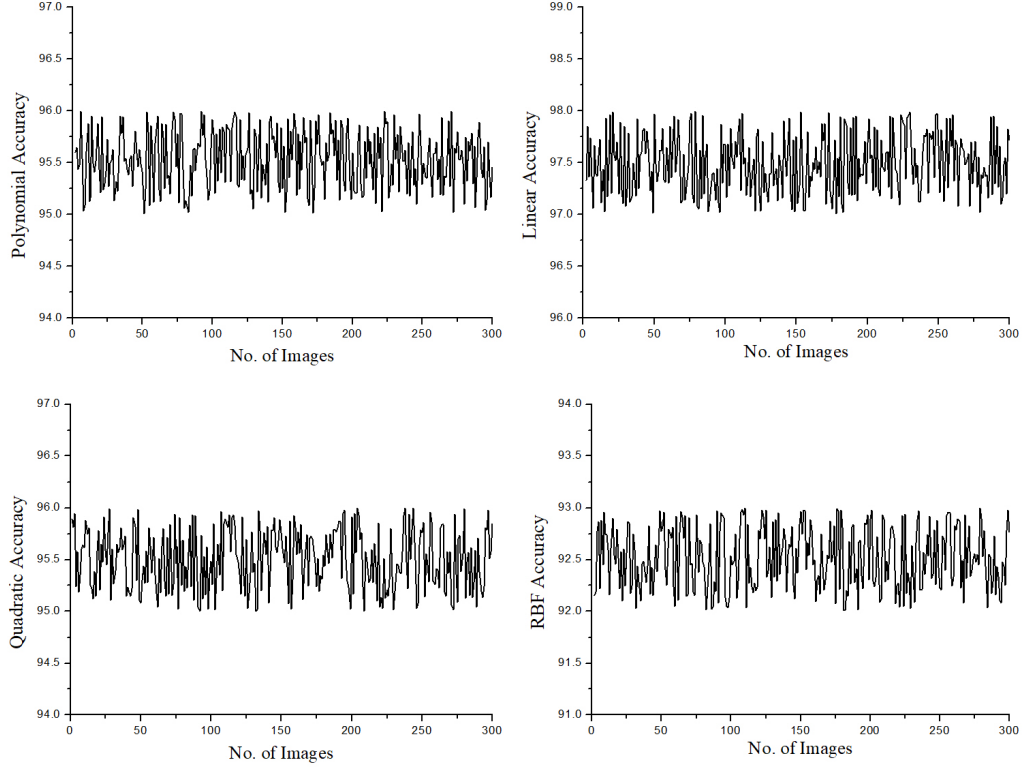


Figure 3.3: Performance evaluation of proposed WFKM using four different SVM classifier kernels.

However, in case of quadratic, poly, and RBF the average accuracy is reported as 95.5%, 96%, and 92.5%, respectively. The overall accuracy of the system is reported as 97.7% on 300 MR images from DICOM dataset. However, the change in the number of images do not affect the overall accuracy of the proposed system, the minute changes can be due to the feature spaces of the dataset or may be the smaller dataset is more linear to the system. The same can be depicted in the graph shown in Fig. 3.4 where the achieved average accuracy of the proposed system varies from 97.12% to 96.63% with the number of images starting from 25 till 1000.

Other than accuracy parameter, the proposed work has been evaluated and

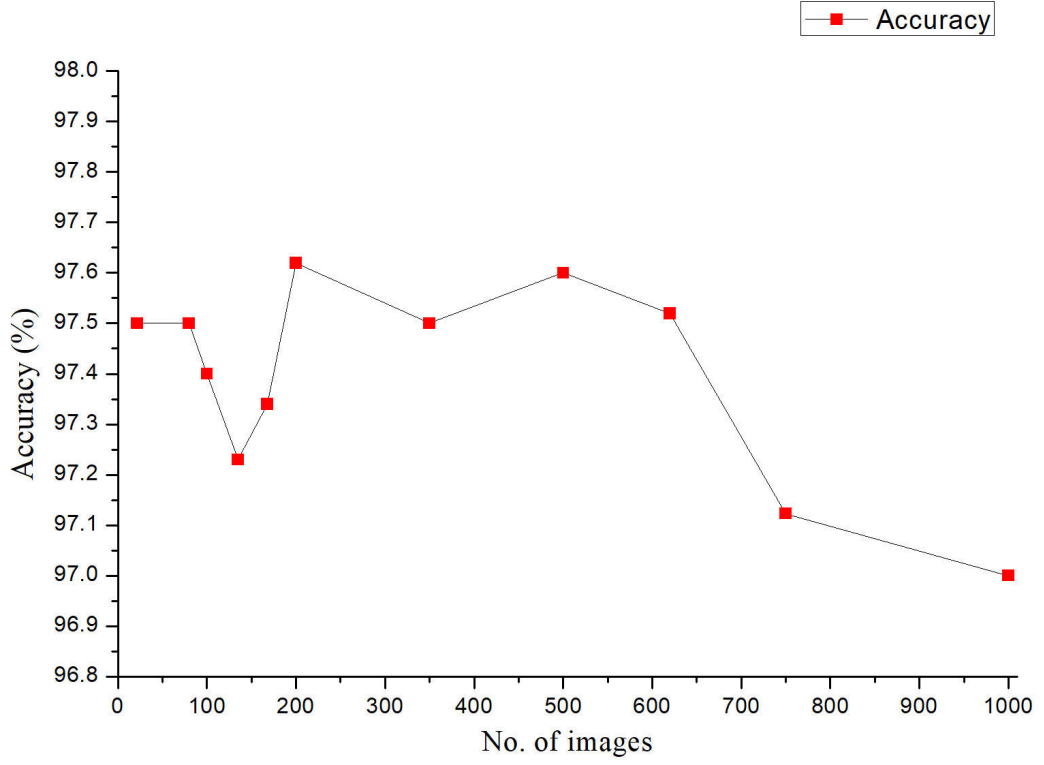


Figure 3.4: Performance evaluation of WFKM in terms of accuracy with varying number of images.

compared on the basis of sensitivity, specificity, and dice similarity index. These parameters demonstrate effectiveness of the proposed approach in spotting abnormal and normal brain tissues within MR brain images, except the dice similarity index which statistically validates the spatial overlap effectiveness of the proposal. WFKM has reported specificity, sensitivity, and average dice similarity coefficient as 94.2%, 94.6%, and 95.7%, respectively. Higher values of these specified parameters further supports better overall performance of the proposed detection system.

3.5 Summary

In this chapter, a novel hybrid weighted fuzzy based algorithm, termed as WFKM, has been proposed using MR images to retrieve more meaningful clusters. It is based on fuzzification of weights which works on spatial context with illumination penalize membership approach which helps in settling issues with pixel's multiple member-

ships as well as exponential increase in number of iterations. The segmented image is further utilized for successful tumor type identification as benign or malignant by means of SVM. Performance is validated both quantitatively (in terms of classification accuracy) as well as qualitatively (i.e., visual examination of segmentation outputs).

Experimentation performed on MR images using Digital Imaging and Communications in Medicine (DICOM) dataset shows that fusion of proposed WFKM and SVM outperforms both ways, i.e. qualitatively and quantitatively. Further, performance evaluation parameters show that the proposal produces better overall accuracy. Results on variety of images further prove applicability of the proposal in detecting ranges and shapes of brain tumor. Experiments showcased that WFKM is accurately detecting tumor along with its precise location in brain MR images. For completeness the presented system includes a classification module after WFKM segmentation. Features extracted after effective WFKM segmentation helps in better training and so is the reported accuracy in classifying brain tumor MR images. Overall accuracy of nearly 97.7% is achieved in linearly classifying brain MR images from DICOM dataset as normal and abnormal tumor classes with total number of images varying from 100 to 1000. Classification accuracy reported after WFKM segmentation clearly concludes that the proposed algorithm identifies the tumor region efficiently. This observation is further supported by values obtained using other considered performance parameters.

Chapter 4

Detection Framework using Big Data

The exponential increase of brain MR image data in the medical imaging field demands not only accurate but faster tumor segmentation as well. The computer aided detection (CAD) systems acting as a second option to experts, radiologists and surgeons needs to be swift enough to handle parallelism. However, handling of massive MR data for segmentation with high accuracy and low execution time is one of the significant concerns in any framework. The traditional segmentation approaches cover either of the performance parameters, accuracy or execution time. The proposed hybrid approach (WFKM) outperforms all traditional approaches in terms of accuracy. ²The next focus is to process WFKM on Hadoop platform and observe the execution time. Mainly performance of the proposed algorithm is analyzed on varying sized datasets using MapReduce on Hadoop. Observations are made to judge WFKM in dealing huge MR brain data across clusters of commodity computers. Results obtained clearly concludes that the said collaboration helps in achieving both the performance parameters, i.e. accuracy as well as execution time.

²The initial research work tested applicability of Hadoop-MapReduce by performing experiments using a combination termed as FKM here. FKM is a hybrid segmentation approach based on the concepts of fuzzy c-means and k-means algorithms. The background details and FKM algorithm is included as Appendix B of the thesis. Observation made further motivated the integration of proposed hybrid WFKM approach with Hadoop-MapReduce platform with an improvisation namely, MDCS.

The chapter is organized firstly to briefly explain the Hadoop-MapReduce system model and its implication on medical imaging in Section 4.1. Section 4.2 highlights the assimilation of proposed WFKM on Hadoop-MDCS architecture using MR images to improve the scalability. Section 4.3 summarizes as well as focuses on the experimental setups and results. It evaluates the results from the prospective of segmentation effectiveness along with processing-time. Finally, Section 4.4 winds up with a summary of the chapter.

4.1 Hadoop-MapReduce System Model

Hadoop is an open source well-defined platform that ensures efficient and swift processing of massive traffic in distributed environment. In today's world, data (so is the MR image data) is growing continuously thus its processing from the native storage becomes essential. The idea is thus to manage the processing of MR images via a legible platform, i.e. Hadoop for effectual data communication. The objectives ascertained by means of this integration are (i) proper storage platform maintaining consistency and scalability of bulky image data; (ii) platform reliability through distributed and parallel infrastructure; (iii) lastly, an efficient hybrid approach to extract the required output that further eases response time of the system.

Two main components of Hadoop are MapReduce (for data processing) and Hadoop Distributed File System – HDFS (for storing huge data) [215]. The sequential working of a generic Hadoop system model for MR brain images starts through inputting the links /URLs of the datasets or downloaded paths of image data. HDFS stores the gathered data in the form of contiguous blocks of 64 MB (default) size. These splitted image data chunks are stacked in each replicated racks with NameNode and DataNode. Daemons associated with NameNode and DataNode are JobTracker and TaskTracker respectively. JobTracker assigns jobs for further processing to the TaskTracker in Mapper phase. The Map function processes input and generates an intermediate output, both in the form of $\langle key, value \rangle$ pair. Further,

the Reducer phase agglomerates intermediate outcomes to generate the final output.

4.2 The Hadoop and MDCS Based Architecture for MR images

Hadoop is a platform designed to run applications in distributed environment. On the other hand, Matlab is a user-friendly and easily accessible tool meant for processing of enormous image datasets. MATLAB Distributed Computing Server (MDCS) framework eases immense image processing in clusters. Additionally, big data storage HDFS layer in Hadoop permits data storage and its access in multiple formats. Therefore, integration of MDCS with add-on libraries of Hadoop enhances performance of the developed platform for specific application which in this work is MR brain tumor image processing. The general integration model is shown in Fig. 4.1 and following is the layer-wise description of its four layers, namely Input, HDFS, MDCS, and Output layer.

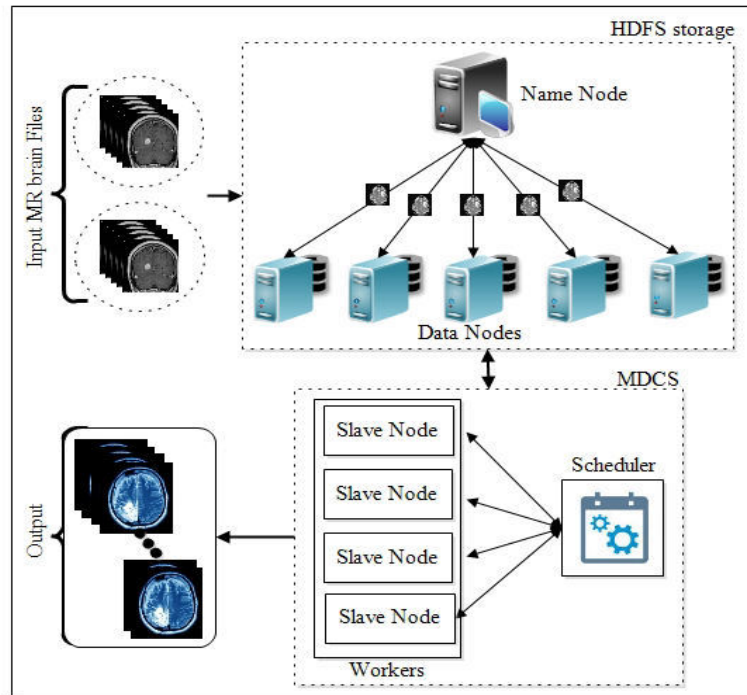


Figure 4.1: Medical Image processing system architecture using Hadoop and MDCS platform.

- **Input Layer** provides input to Hadoop configured platform in the form a text file containing image links/URLs or the downloaded dataset path. Hadoop's ripper module extracts images from the Google defined URLs or from the dataset path as contained in the text file. In case data is to be downloaded through URLs then for smooth processing NameNode distribute load among the DataNodes. Shuffled downloaded images are sorted by DataNodes and are stored in blocks of size 128 MB. If any file overflows from the selected block then a new block is created accordingly to store the files.
- **HDFS Layer** is responsible to store the data effectively for further processing. HDFS works with Google File System (GFS) to make use of distributed environment for processing on various distributed nodes [216] [217]. Block approach is utilized for data storage on distributed drives with block sizes of 64MB, 128MB, and 256MB. The incoming images get divided into number of blocks to manage the workload among cluster nodes. For well-organized execution of the task various daemons are processing in the background of HDFS [218]. NameNode is the root node of the system that handles the incoming data. DataNode handles application data (in the form of blocks) on different servers. HDFS layer forwards data in the same form to Mapper-Reducer Layer for processing.
- **MDCS Layer** is a configurable layer in the MATLAB environment that provides scalability, recoverability, and speed. MDCS platform helps in faster data processing by dividing load among available slaves in the cluster. One has to configure MATLAB Job Scheduler (MJS) to provide a communication link between different nodes as well as to divide the incoming load among available slave nodes in the cluster.
- **Output Layer** is the last layer. It stores the obtained output in the form of blocks on HDFS platform. The extracted output can now be moved to the local drive.

4.2.1 WFKM on Hadoop with MDCS Platform

For correct diagnosis, a huge number of MR images are commonly produced by smart machines in the medical domain. The proposed segmentation algorithm, WFKM has shown to effectively detect brain tumor in MR images and Hadoop with MDCS is known to handle huge data in all intents and purposes. Thus the framework discussed in Chapter 3 consisting of proposed WFKM, DWT, PCA, and SVM is assimilated on Hadoop with MDCS platform to store MR image dataset in distributed manner as well as to process it using the clustering mechanism. Fig. 4.2 elaborates the proposed integration.

MDCS permits the proposed WFKM segmentation approach to execute non-interactively on clusters containing nodes. MATLAB client node with parallel computing toolbox (PCT) in MDCS composes and forwards task to cluster nodes using parallel scheduling system for processing [219]. Further, MATLAB job scheduler configured on MATLAB client machine distributes jobs, i.e. incoming tasks, on the

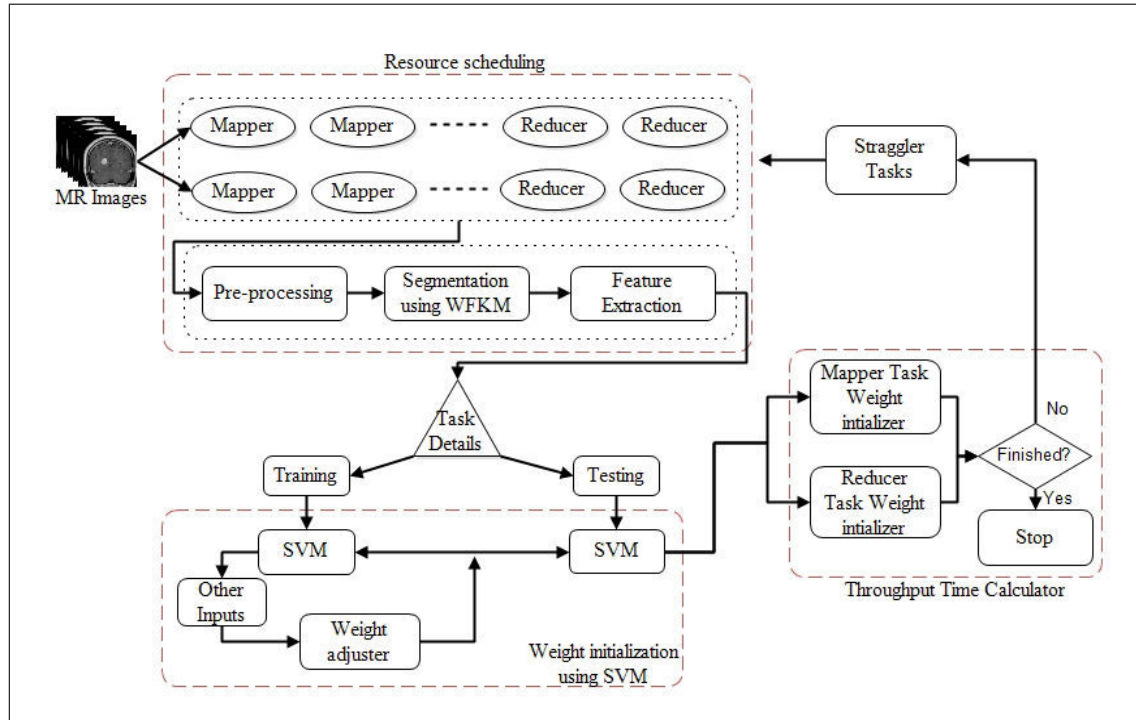


Figure 4.2: Proposed framework of brain tumor detection using WFKM integrating Hadoop and MDCS platform.

worker nodes (personal computer accessible to perform computation) that are live in the cluster for parallel execution.

HDFS first divides large datasets into number of blocks and stores them on the distributed machines. These blocks are then forwarded to MDCS that distributes the workload among DataNodes for smooth processing. The same sequence of steps is now executed. MR Images, in blocks, are pre-processed for smoothness and removal of unwanted noise. Segmentation using proposed *WFKM* approach is completed. Based on the extracted features of the segmented region, the process of classifier training and testing is carried out using the selected set of dataset images. For classification purpose, test images are also forwarded to the MDCS framework which maps them with the trained model and store them into the relevant class. After classification, the retrieved result is stored on the HDFS in the form of blocks, the left over tasks are considered as Straggler task and are re-processed for classification of the image.

The working of the proposed integration, i.e. *WFKM* on Hadoop with MDCS platform, is discussed in Algo. 2. As identified earlier, Hadoop using HDFS stores clusters of MR Images in defined sized blocks for distributed storage as well as efficient processing of allocated tasks. The execution starts with a call to an input function, (**WFKM_INPUT(key, value)**) where key is the directory path storing image files and value is the image number. The image, as given in the value parameter, is forwarded to the proposed *WFKM* segmentation algorithm (**Algo. 1**) and the extracted feature of the segmented region is stored in the file using **file.write(key, S)** function for further processing. Later, features are mapped via call to an output function (**WFKM_OUTPUT(key, value)**). The process starts by loading all images from blocks and matching image name with the segmented image, if matched then the corresponding matching image is processed using **WFKM_OUTPUT()** function. A block is generated on HDFS containing image file name and pixels information. The width and height of the indexed image is extracted and updated. If the image is active, then information is stored in the generated block on HDFS and

Algorithm 2 Hadoop implementation and MDCS configuration for WFKM segmentation algorithm

INPUT:- Block of I images from HDFS

OUTPUT:- Report file of selected I images from workers

```

1: Start
2: FUNCTION WFKM_INPUT(key, value) ▷ key: Directory index, value:
   Image index
3: Segmented Image  $S \leftarrow$  Algo. I
4: if  $S \neq null$  then
5:   file.Write(key, $S$ )
6: end if
7: END FUNCTION
8: FUNCTION WFKM_OUTPUT(key, value) ▷ key: HDFS path, value:
   Image index
9: for ( $R_{img} \in \mathbb{S}$ ) do
10:   $P_{file} = \text{HDFS.generateNewFile}(R_{img}.\text{pullName}() + ".\text{pixels}")$ 
11:  for ( $m=0$  to  $m \leq R_{img}.\text{pullWidth}()$ ) do
12:    for ( $n=0$  to  $n \leq R_{img}.\text{pullHeight}()$ ) do
13:       $P_{val} \leftarrow \mathbb{S}.\text{pullPixel}(m,n)$ 
14:      if  $P_{val}$  is active then
15:         $P_{file}.\text{Write}(m,n)$ 
16:      end if
17:    end for
18:  end for
19: end for
20: END FUNCTION

```

the extracted values are used for processing the MR Images.

4.3 Experimental setup and results

This section starts with a discussion on the dataset utilized and the initial configuration of the developed platform. Afterwards, obtained results are evaluated to authenticate the proposed integration of WFKM on Hadoop with MDCS.

4.3.1 Initial configuration and Dataset

Experiments are performed on a digital imaging and communications in medicine (DICOM) dataset collected from TCIA. The dataset is of $\approx 8\text{GB}$ in size and has

Table 4.1: Initial configuration table

Parameter	Value
Hadoop installation type	Fully distributed (3.1)
Matlab installation type	Matlab distributed computing server 2014a
Number of nodes	3
Configuration of slave nodes	4GB, 8GB RAM
Configuration of master node	8GB RAM

images with dimensions 256×256 , i.e. a total of 65,536 pixels in a image.

An instance of ripper interface named as FlickrRipper is available in Hadoop that can extract the dataset by working on the URLs of searched MR Images. These URLs are provided in the form of a file. FlickrRipper can be used for all sets of results in any of the formats. Here as well images are pre-processed using the Gaussian filter to enhance the smoothness by removing unwanted noise. To train the SVM classifier the complete dataset is divided into three portions, 70% images are used for training, 15% for result validation, and the remaining for testing. Fig. 4.3 shows training and validation done at different epochs, clearly the maximum accuracy is achieved at 30th epoch.

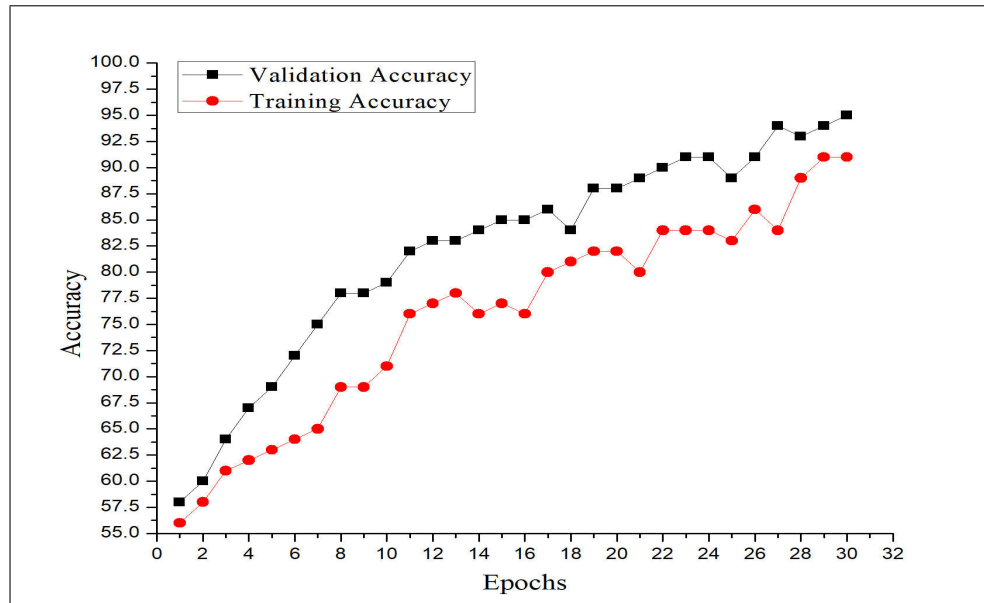


Figure 4.3: Plot of system accuracy on train and validation datasets

For optimal results, a compatible number of nodes/machines are elemental and for better performance it's essential that these machines are properly tuned. For experimental purpose, 3-nodes are configured to surge the processing speed of the proposed integration. Table 4.1 presents the node configuration details utilized to perform experimentation for the evaluation of the proposal. In this research, the proposed framework is developed on physical machines instead of virtual environment to attain the desired results. The initial configuration of MATLAB for parallel processing of MR Images dataset is shown in Fig. 4.4. It's a master-slave model.

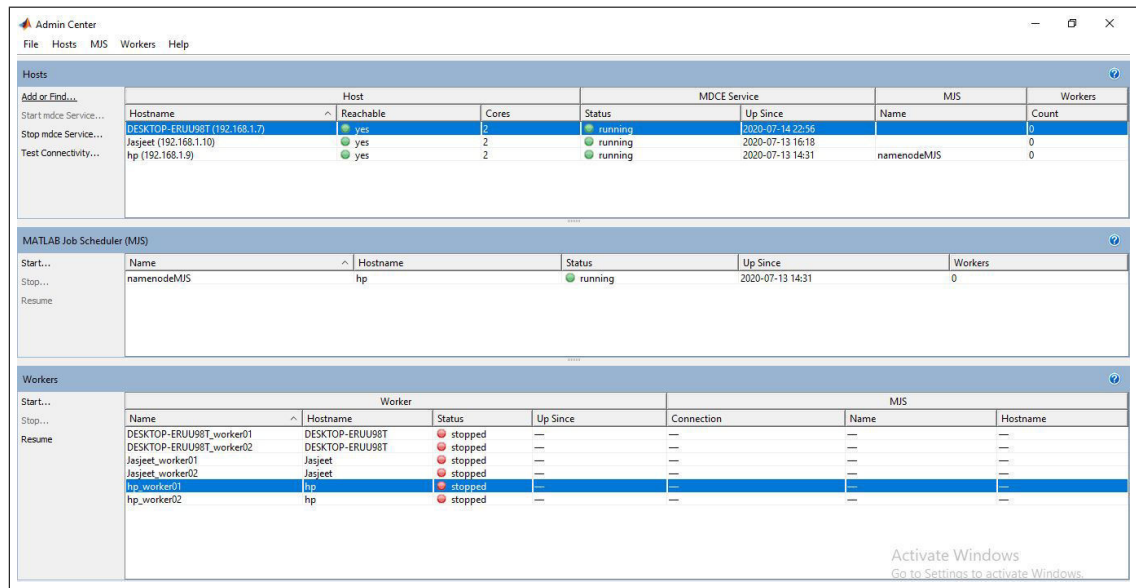


Figure 4.4: Initial configuration of distributed parallel processing framework

In all three nodes are configured, one acts as a master and others as slave nodes. Two or more cores are available with each node and as per the workload on a machine workers can be initialized to process the incoming data.

4.3.2 Results and Discussions

The aim of integrating the proposed WFKM approach with Hadoop and MDCS platforms is to handle bulkier datasets quickly via parallel processing of the stored images. To validate the integrated framework, processing-time, read-time and write-time, calculated on different sizes of datasets namely, 125MB, 1.2GB, and 7.3GB

are shown in Fig. 4.5.

HDFS architecture works on master-slave model. The master node manages file access and slave nodes are responsible for storing (or extracting) the data on (or from) various DataNodes. Master node actually forwards index of the destination node to slaves for storing or extracting the required data. Involvement of master and slave nodes to write or read the data at various connected nodes increases the write-time as well as the read-time of the configured topology. The same can be observed in Fig. 4.5, i.e. write-time and read-time for 7.3GB is the highest and they are the lowest for 125MB. The extracted images are forwarded to Jobtracker which parallelly processes them and results in faster execution. Therefore, the processing-time reported in each of the three cases is better then either read- or write-time for the 3-node configuration. Also, classification accuracy of the proposed framework is reported as 97.07%, 97.01%, and 96.30% respectively for 215MB, 1.2GB and 7.3GB datasets.

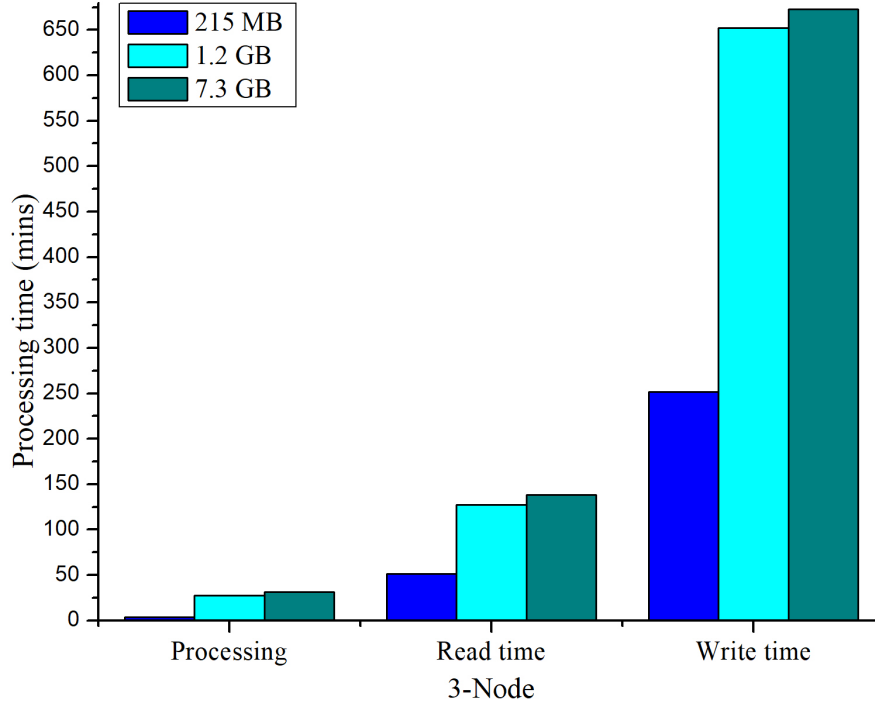


Figure 4.5: Performance evaluation of processing time on different nodes

4.4 Summary

In the era of industrialization, a hinge is required to process brain tumor MR Images at big data health-care system. In the proposed framework, parallel processing support to the WFKM segmented output is incorporated to ease overall time of the system. During experimentations variable sized datasets (215MB, 1.2GB, and 7.3GB) are considered. Timings for three different phases, namely read, processing, and write are observed for 3-node configuration. The more bulky the dataset more are the read and write timings for the system due to the additional manager operating on the tasks. However, the main computation part falls in the processing time category and the 3-node configuration has reported desired results in each of the three cases. Additionally, all the reported classification accuracy excel the traditional frameworks.

Chapter 5

Results and Discussions

This chapter compares the proposals discussed in the last two chapters with the existing works in literature. The novel weighted fuzzy k-means (WFKM) approach is first combined with a simple classifier and is then integrated with Hadoop-MDCS. The analysis is thus presented qualitatively for segmentation to showcase the improved segmentation reported using WFKM. Furthermore, classification accuracy based comparison with traditional systems is followed. It was observed that the extracted tumor features after WFKM processing are optimized and better train the classifier. Lastly, the processing time variations are compared using variable sized datasets.

Section 5.1 discusses the comparative results of segmentation and classification of the proposed hybrid WFKM with the existing approaches. The comparative results of WFKM integrated with Hadoop-MDCS are highlighted in Section 5.2.

5.1 Comparative Results for WFKM

The foremost objective of the research is to propose an efficient framework to segment the tumorous part of the MR brain image. For experimentation and validation of the proposed approach, the MR images are taken from online available data sources. Apart from DICOM dataset, MR brain tumor images are retrieved from

NITRC [209], kaggle, and Google search engine. The comparative results of the proposed WFKM for segmentation and classification are included in the following subsections.

5.1.1 Segmentation Results

The proposed WFKM segmentation approach is compared with existing approaches based upon visual examination and the same is discussed in this section. The phase by phase visual interpretation is shown so as to compare the performance of five segmentation approaches (k-means, FCM, canny edge detection, FKM, and WFKM) to detect tumor. The input medium is similar in all the cases i.e T1-W MR brain image taken from DICOM. Therefore, in every visual comparison the first column of all the rows is same followed by pre-processing and Otsu's thresholding. Performance is evaluated and compared not only using brain MR images with and without tumors, but with respect to a broader tumor category termed as glioma (tumor initiated in the glial cells) as well. Almost 30% of all the brain tumors are comprised of glioma, moreover 80% of them are malignant in nature based on the survey conducted by National Brain Tumor Society. Glioma is further subdivided into two types namely, high grade glioma (HGG) and low grade glioma (LGG). The results are included for both the glioma types. In addition, WFKM has reported comparable segmentation results to the existing research works termed as SFRCM. Lastly, fine segmentation results on different tumor types are included to further validate the proposal.

Segmentation results in Fig. 5.1 presents the examination of MR image belonging to benign tumor class whereas Fig. 5.2 verifies the same for malignant tumor class. The end result of k-means depicted in Fig. 5.1(a) and Fig. 5.2(a) are far from the tumorous region; they are the worst performer. The clusters formed are far away from the actual tumor which lead to an inaccurate identification. Furthermore, Fig. 5.1(b) and Fig. 5.2(b) depicting the segmentation of FCM is though better than the k-means. Also somehow the result is able to separate the tumorous intensity values

k-means. Also somehow the result is able to separate the tumorous intensity values as shown in the intermediate section of Fig. 5.1(b) and Fig. 5.2(b) but final output highlights all high intensity brain regions rather than the crucial tumor region. Clearly, in Fig. 5.1(a)(b) and Fig. 5.2(a)(b) the intensity ranges to extreme level that the tumor area is not visible at all. The next approach canny edge detection is able to detect the outlier boundaries of the tumor very effectively as seen in Fig. 5.1(c) and Fig. 5.2(c), however, still lack in accurate tumor identification. Contrary to previous results, FKM is able to retrieve the tumorous area quite significantly shown in Fig. 5.1(d) and Fig. 5.2(d). The output of hybrid approach WFKM in both the tumor classes as is shown in Fig. 5.1(e) and Fig. 5.2(e) gives better convergence towards the tumor region. The cluster formation in each phase is clearly evident due to intensity variation generated by inculcating the concept of weights.

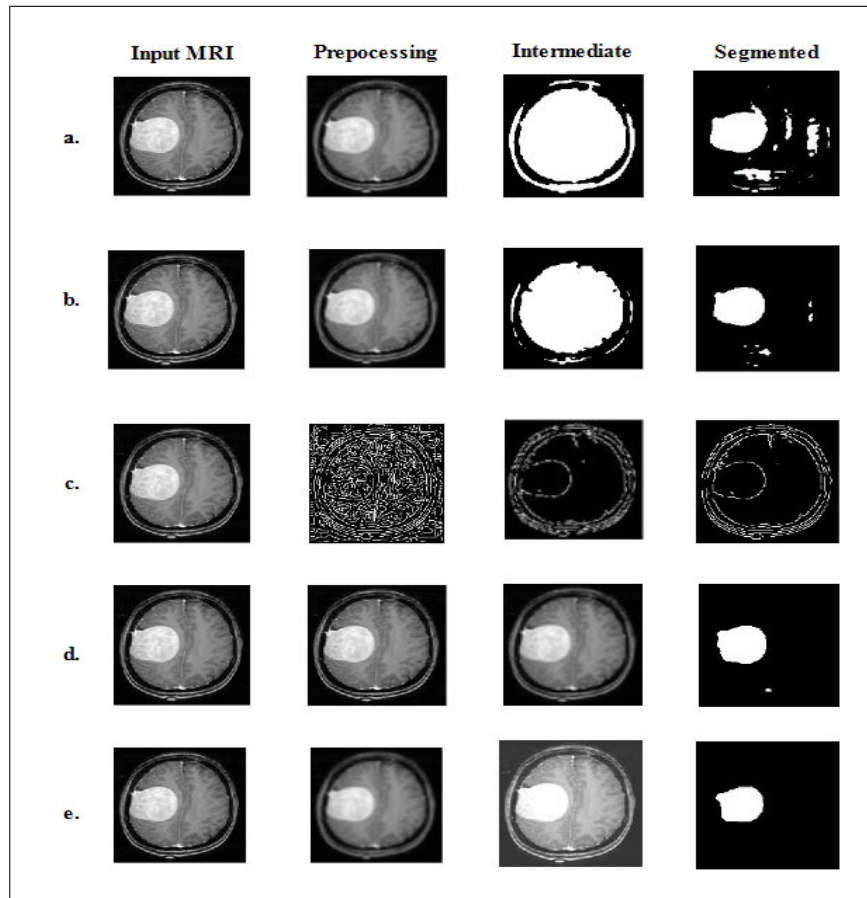


Figure 5.1: Visual segmentation results of benign tumor on T1W MR image using a. k-means, b. FCM, c. Canny edge detection, d. FKM and e. Proposed hybrid WFKM.

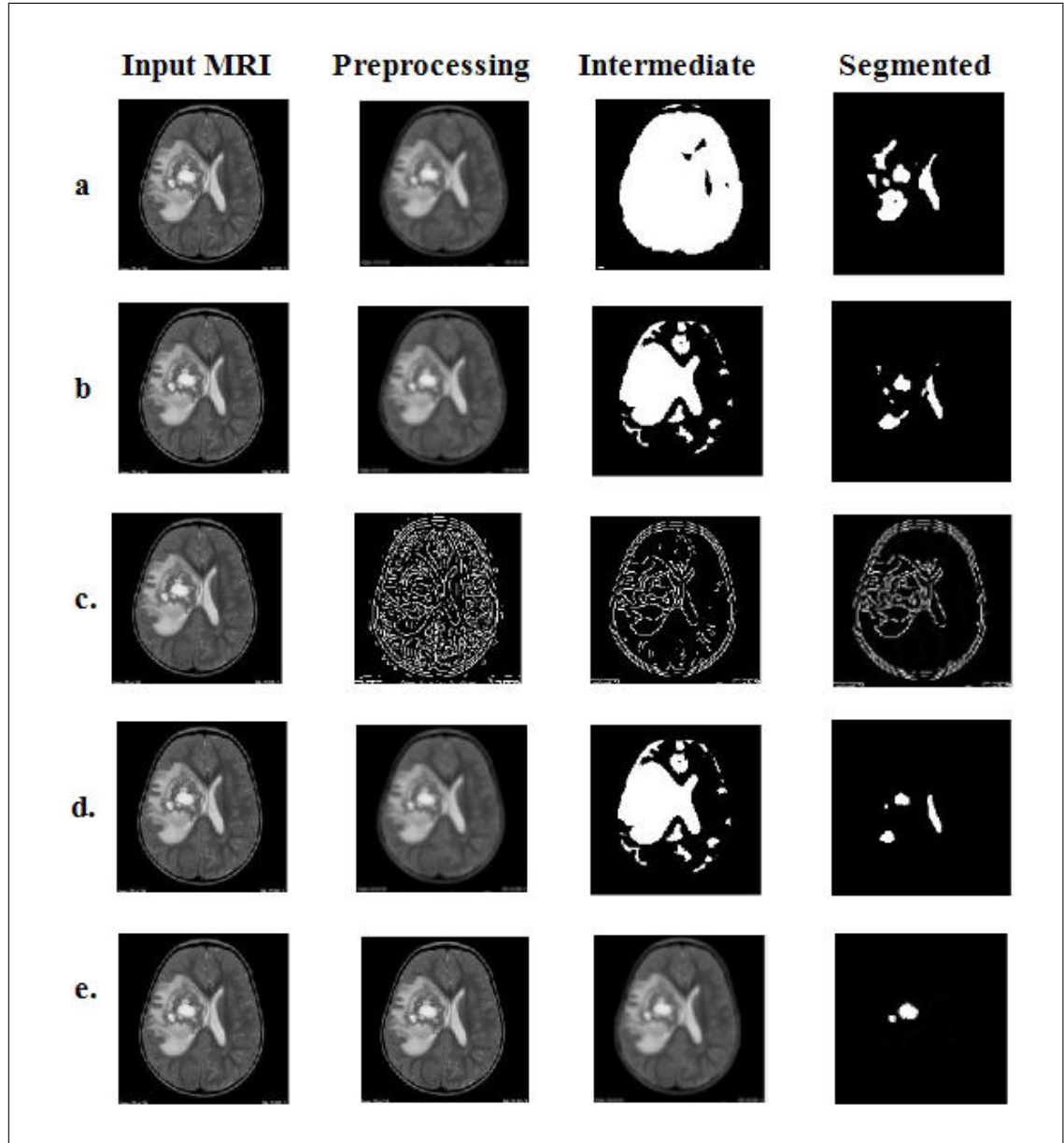


Figure 5.2: Visual segmentation results of malignant tumor on T1W MR image using a. k-means, b. FCM, c. Canny edge detection, d. FKM and e. Proposed hybrid WFKM.

The segmentation accuracy of the proposed hybrid WFKM is also compared with results given in the research article [5] visually for the two cases. In Fig. 5.3, the same input MR image is taken for the comparison of the existing soft fuzzy rough c-means (SFRCM) with the proposed hybrid WFKM. However, the visual representation of the segmentation results depicts that the proposed hybrid WFKM segments the tumor region almost the same as done by SFRCM. This also concludes

that the developed segmentation approach excels to efficiently identify the tumor region in MR images.

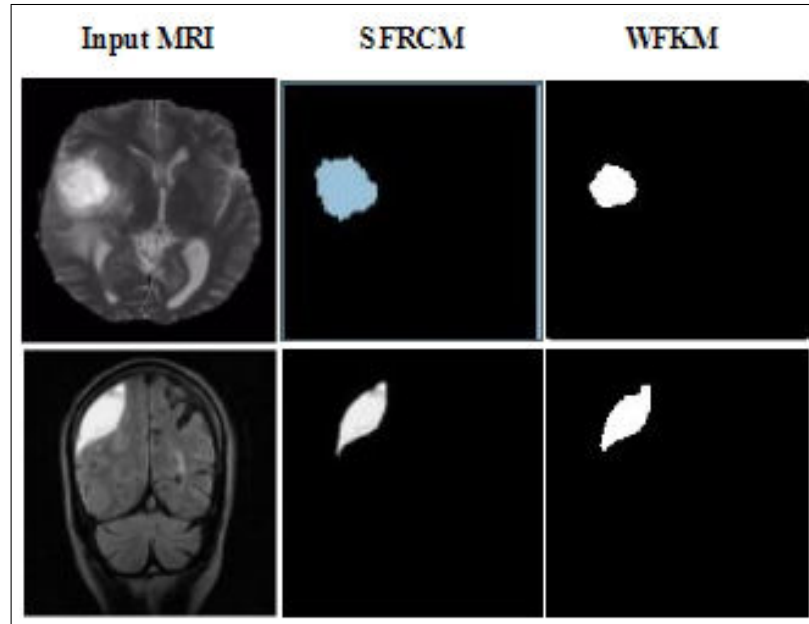


Figure 5.3: Visual segmentation results of soft fuzzy rough c-means (SFRCM) [5] and proposed hybrid WFKM.

The visual results attained using the proposed WFKM segmentation approach on both glioma types, i.e. HGG and LGG, are compared with those obtained by traditional approaches. Respective visual comparison results for HGG and LGG are shown in Fig. 5.4 and Fig. 5.5. Observations made for all the traditional approaches, i.e. k-means (Fig. 5.4(a) and Fig. 5.5(a)), FCM (Fig. 5.4(b) and Fig. 5.5(b)), canny edge detection (Fig. 5.4(c) and Fig. 5.5(c)), and FKM (Fig. 5.4(d) and Fig. 5.5(d)), for malignant and benign cases also exist while experimenting with HGG and LGG categories. k-means leads to inaccurate identification of tumorous area and FCM, though bit better than k-means but illuminates irrelevant pixels in addition to tumor tissues. Similarly, canny edge detection and FKM separate the tumorous intensity values effectively but lack in accurate tumor identification. The output of hybrid novel WFKM in both the glioma categories gives better convergence with improved illumination factor as shown in Fig. 5.4(e) and Fig. 5.5(e). The area with intense intensity pixels in case of WFKM is crisp and less prone to multiple

memberships, however, the cluster formation in each phase is clearly evident due to intensity variation generated by inculcating the concept of weights.

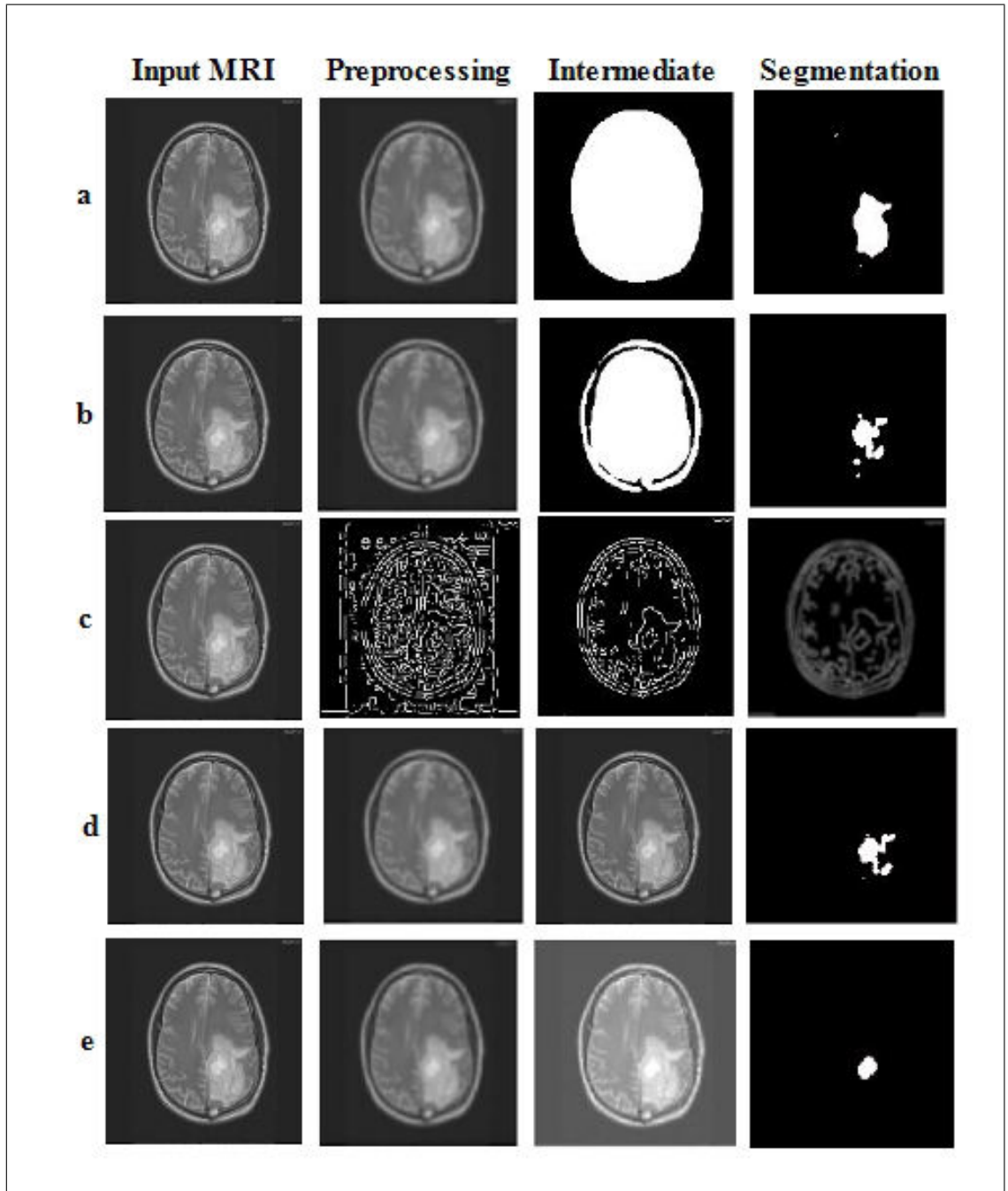


Figure 5.4: Visual Segmentation results of High grade gliomas (HGG) using MR Images. a. k-means b. FCM, c. Canny edge detection d. FKM, and e. Proposed hybrid WFKM

In addition, WFKM segmentation results on four different tumor types namely Meningioma, Astrocytoma, Glioblastoma Multiforme, and Pilocytic Astrocytoma are compared with the classical approaches. The visual representation of obtained

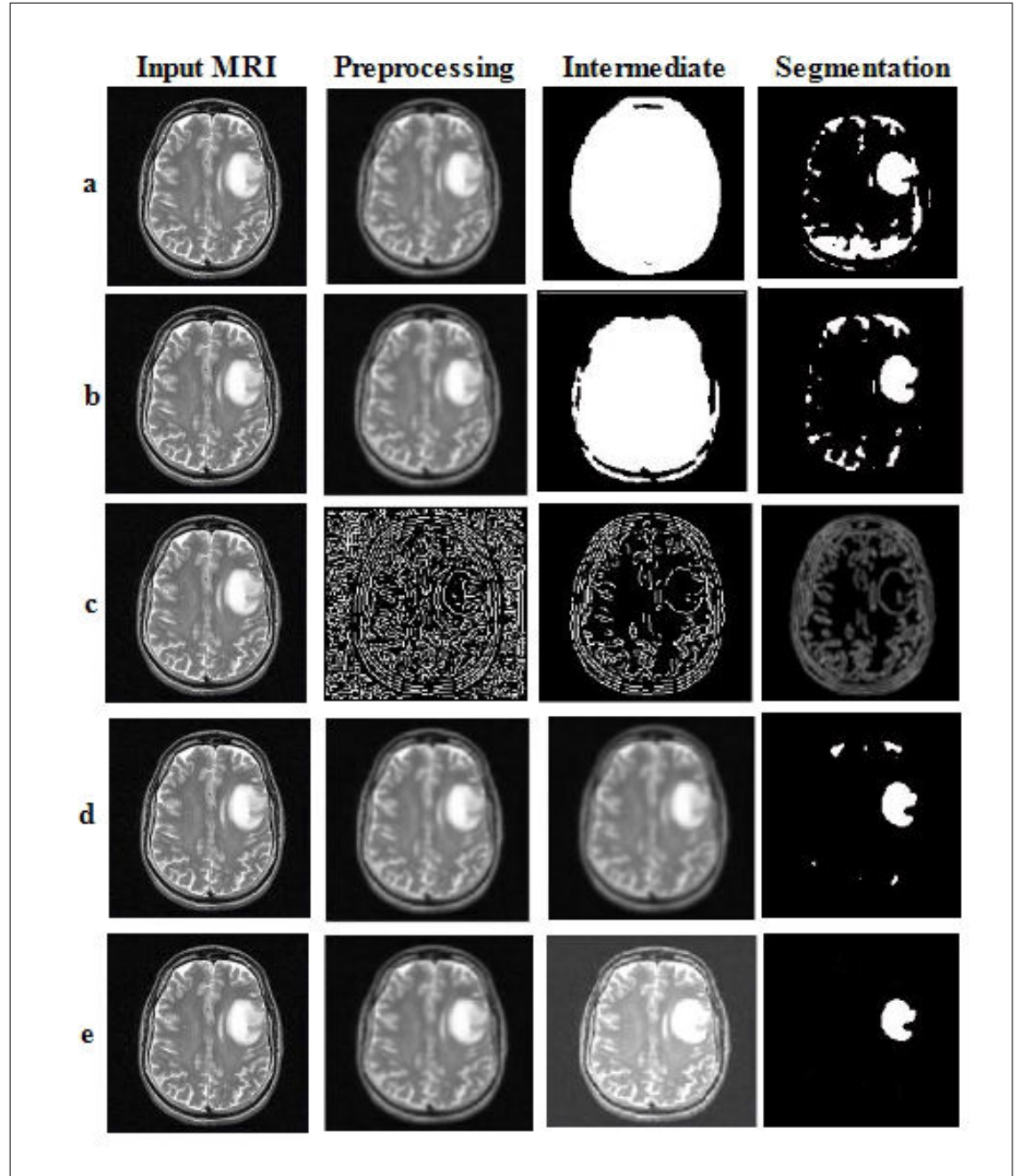


Figure 5.5: Visual Segmentation results of Low grade gliomas (LGG) using MR Images. a. k-means b. FCM, c. Canny edge detection d. FKM, and e. Proposed hybrid WFKM

results is depicted in Fig 5.6 and examination of these results clearly states that the proposed WFKM based on clustering and fuzzification process excels the existing approaches.

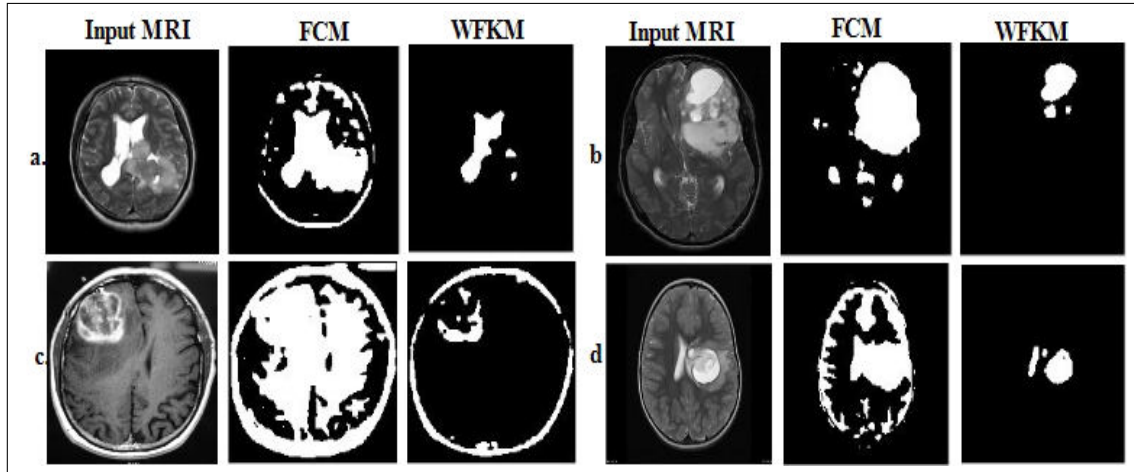


Figure 5.6: The segmentation of MR brain tumors using WFKM a) Meningioma, b) Astrocytoma, c) Glioblastoma Multiforme and d) Pilocytic Astrocytoma

5.1.2 Classification Results

Classification accuracy shows how effectively an approach identifies the type of tumor from MR image scans. Evaluation parameter used to assess efficacy of the classification approach in this research is accuracy. In addition, performance is evaluated in terms of sensitivity, specificity, and dice coefficient index. Confusion matrix parameters, namely true positive rate (TP), true negative rate (TN), false positive rate (FP), and false negative rate (FN) are also computed. All these parameters are defined in Chapter 3 from Eq. 3.17 to Eq. 3.24 respectively. SVM classifiers with four different kernels, namely linear, polynomial, Gaussian (RBF), and quadratic are used to compare the proposed system with the existing systems. Accuracy graphs plotted in Fig. 5.7, Fig. 5.8, Fig 5.10, and Fig. 5.9, compare the performances of all the approaches with respect to each of the four kernels. Higher the classification accuracy better is the approach. Transition is specified in the graph to get a clear picture of accuracy values corresponding to each image. The linear classification accuracy shown in Fig. 5.7, the percentage value of the proposed hybrid WFKM lies between 97% to 98%. However, in case of quadratic, poly, and RBF the average accuracy is reported as 95.5%, 96%, and 92.5%, respectively. The graphs in Fig. 5.8, Fig. 5.10, and Fig. 5.9 show all the percentage accuracies of existing and proposed hybrid WFKM.

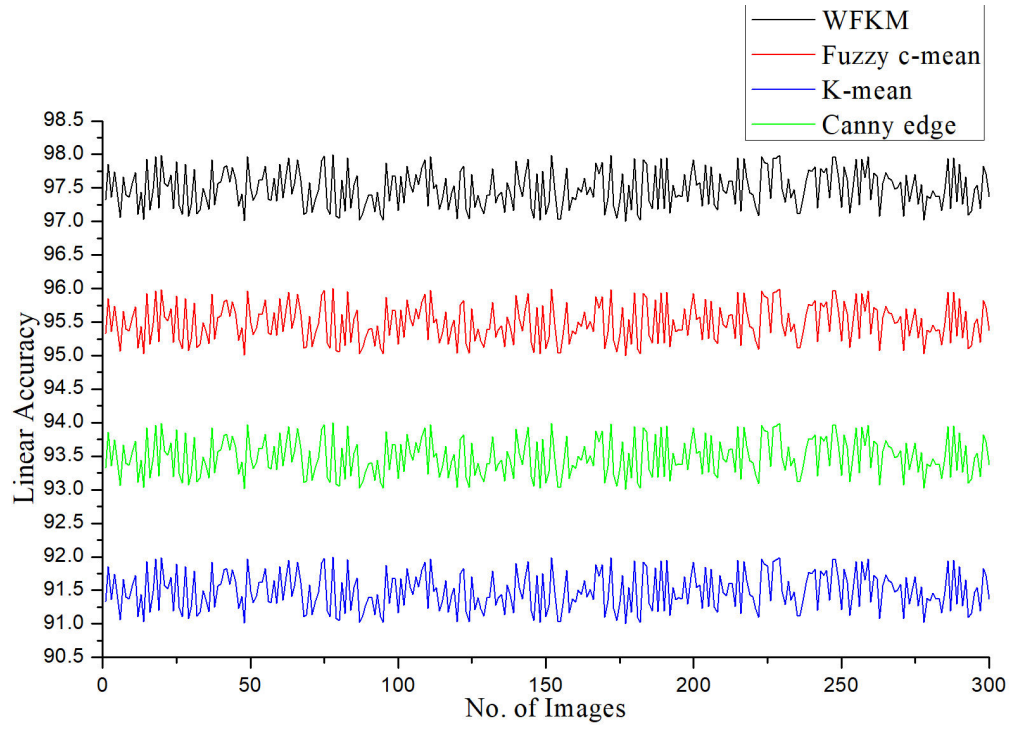


Figure 5.7: Linear classification accuracy of proposed hybrid WFKM compared with existing approaches

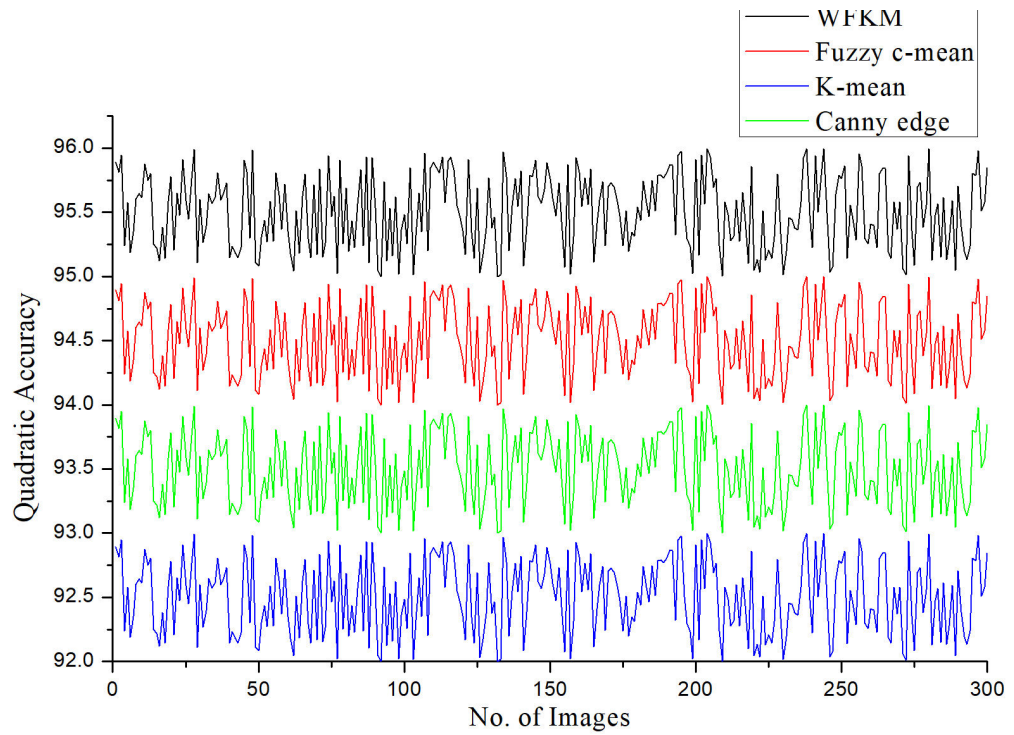


Figure 5.8: Quadratic classification accuracy of proposed hybrid WFKM compared with existing approaches

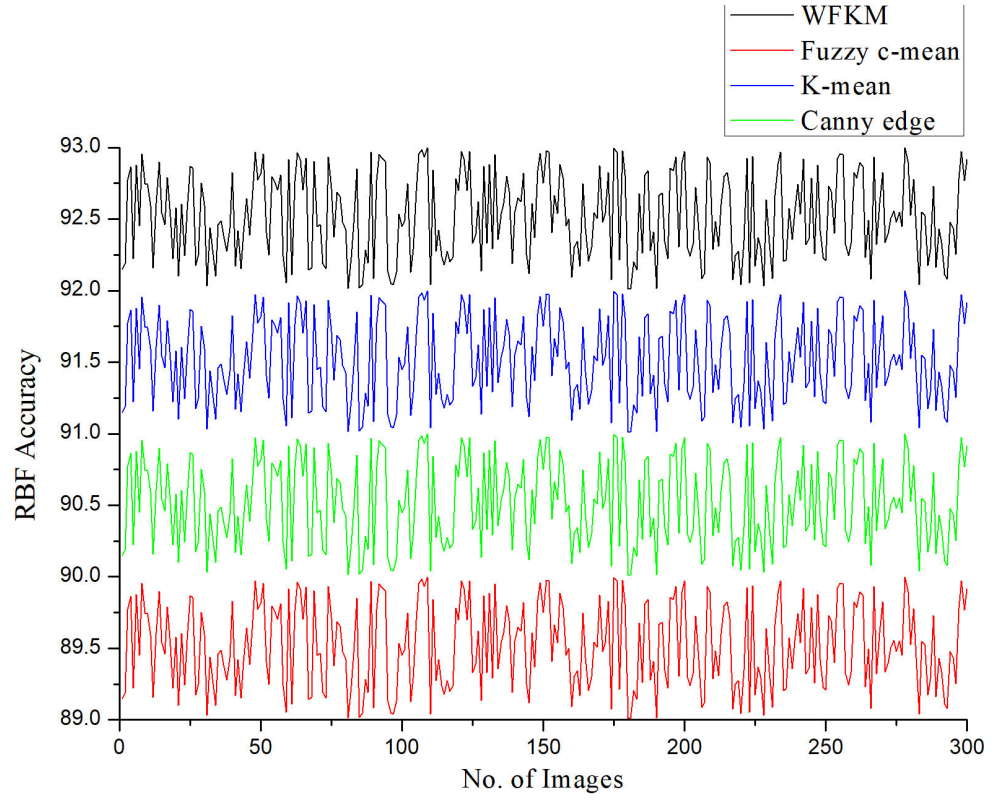


Figure 5.9: RBF classification accuracy of proposed hybrid WFKM compared with existing approaches

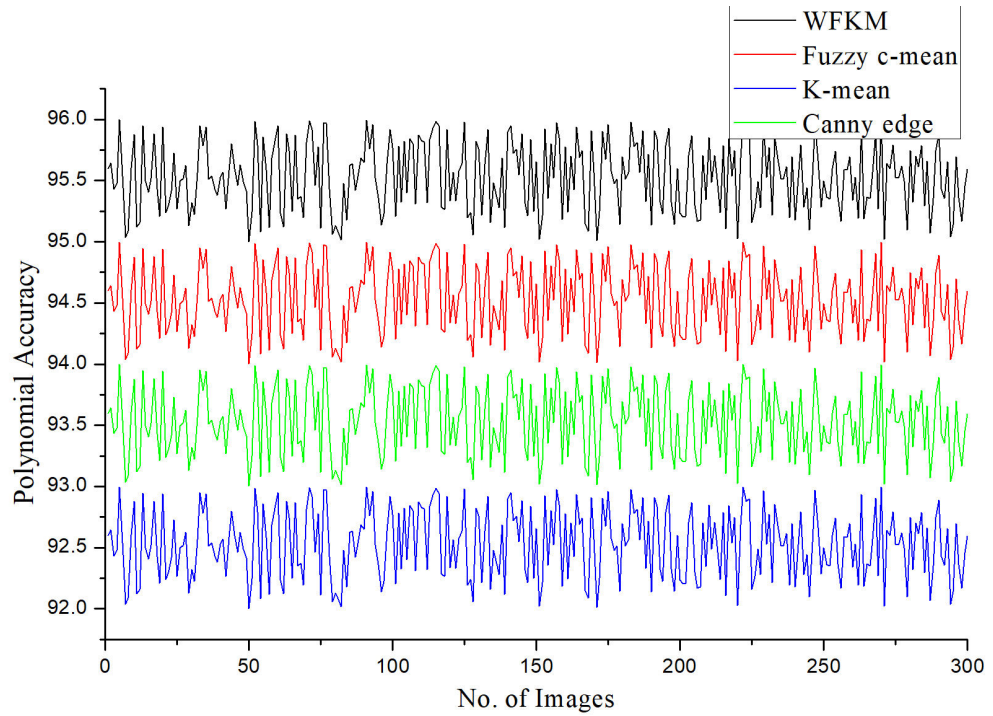


Figure 5.10: Poly classification accuracy of proposed hybrid WFKM compared with existing approaches

Along with this, exact accuracy values are presented in Table 5.1 for better comparison with the existing systems. The reported performance accuracies also prove efficiency of WFKM approach. Comparison results are prepared using the same dataset and then average accuracy is retrieved with the specific number of images which is used in the existing research. Based on the results obtained, it's clear that no matter what the data size is, the proposed WFKM system is always able to accurately detect the tumor in comparison to the existing works. The performance of the proposed scheme is evaluated on different number of images to check the authenticity of the scheme.

Other than accuracy parameter, the proposed work has been evaluated and compared on the basis of sensitivity, specificity and dice similarity index. These parameters demonstrate effectiveness of the proposed approach in spotting abnormal and normal brain tissues within MR brain images, except the dice similarity index which statistically validates the spatial overlap effectiveness of the proposal. The expected outcome and ground truth results for the calculation with reference to these parameters, i.e. sensitivity, specificity, and average dice similarity coefficient are summarized in Table 5.2. Results on 67 normal and 134 abnormal MR images are retrieved and compared with existing approaches, namely adaptive neuro fuzzy inference system (ANFIS) and genetic algorithm (GA) [224]. WFKM has reported specificity, sensitivity, and average dice similarity coefficient as 94.2%, 94.6%, and 95.7%, respectively. Higher values of these specified parameters further supports better overall performance of the proposed detection system.

Table 5.1: Performance comparison with the existing works based on accuracy using same dataset.

Reference	Approach	Size of Dataset (DICOM)	Performance (%)	Performance (%) (Proposed work)
[220]	FSVM	80 images	ACC: 94	ACC: 97.5
[185]	KIFCM	22 images	ACC: 90.5	ACC: 97.71
[221]	EM + adaptive threshold + FFT + MRMR + SVM	100 images	ACC: 98.9	ACC: 97.4
[222]	BWT+ SVM	135 images	ACC: 96.51	ACC: 97.23
[223]	DWT + GLCM + PNN	25 images	ACC: 95	ACC: 97.43
[224]	Watershed, FCM, DCT & BWT + GA	22 images	ACC: 92.03	ACC: 97.52
[225]	MFCN- MOVGA	24 images	ACC: 92	ACC: 97.41
[226]	FKM + (wndchrm classifier & VGG 19 DNN)	168 images & 228 images	ACC: 92.86 (wndchrm classifier) ACC: 98.25 (VGG-19 DNN classifier)	ACC: 97.33
[227]	FKM & FSVm	200 images	ACC: 96	ACC: 97.62

ACC: Accuracy, BWT: Berkeley wavelet Transformation, DCT: discrete cosine transform, DNN: deep convolutional neural network, DWT: discrete wavelet transform, EM: Expectation maximization, FCM: Fuzzy c-means, FFT: Fast Fourier Transform, FKM: Fuzzy c-means with k-means, FSVm: Fuzzy support vector machine, GA: Genetic algorithm, GLCM: gray-level co-occurrence matrix, KIFCM: K-means clustering with Fuzzy C-means algorithm, MFCM-MOVGA: FCM clustering algorithms and multiobjective variable string length based on Genetic Algorithm, MRMR: Minimal-Redundancy-Maximal-Relevance criterion, PNN: Probabilistic neural network, SVM: support vector machine, wndchrm classifier: Weighted Neighbor Distance using Compound Hierarchy of Algorithms Representing Morphology, VGG-19.

Table 5.2: Comparison of specificity, sensitivity, and dice coefficient index of the proposed scheme

Normal images=67, Abnormal images=134			
Parameter	ANFIS [224]	GA [224]	WFKM
True Negative	60	64	66
False Positive	9	6	4
True Positive	118	121	124
False Negative	14	10	7
Specificity (%)	86.95	91.42	94.2
Sensitivity (%)	89.39	92.36	94.6
Accuracy (%)	88.55	92.03	94.5
Average dice coefficient index (%)	91.11	93.79	95.7

5.2 Comparative Results of WFKM Integrated with Hadoop-MDCS

The performance and throughput of integrating the proposed WFKM segmentation approach on Hadoop with MDCS is examined on variable sized datasets i.e. 215MB, 1.2GB and 7.3GB using single node (standalone) as well as 3-node configurations. With the increase of MR Image data size, the processing time also decreases as the number of nodes gets increased, however, authenticate the objective of proposed scheme. It has been already discussed that MDCS supports parallel processing and hence the proposed scheme, developed using it, also follows the same. The behavioral compatibility of WFKM integrated with Hadoop's HDFS environment and MDCS enhances the performance in terms of computation time. The initial step is to upload the dataset in the distributed environment on Hadoop's HDFS. The read time of the uploaded images considering single node and 3-node for each of the three dataset sizes is shown in Fig. 5.11(a). Clearly, read operation is faster on single

node configuration in comparison to 3-node. In case of standalone, i.e. single node configuration, there is no participation of the master node, which results in swift read time. However, master node gets involved and takes more time to distribute data in 3-node configuration, thus reports an increased read time. A single node takes 5.2 seconds, 25.6 seconds, and 133.5 seconds to read 215MB, 1.2GB, and 7.3GB data respectively. Similarly, the 3-node cluster takes 27.5 secs, 127.2 secs, and 652.3 secs to read 215MB, 1.2GB, and 7.3GB data from HDFS platform respectively.

Read operation is followed by the processing of MR Images and the corresponding timings for different sizes of datasets are shown in Fig. 5.11(b) for single node and 3-node configurations. For quick processing on HDFS, MJS allocates task to the connected workers which are created as per the generated workload. A huge margin can be noticed in both the configurations, i.e. the single node and 3-node. A single node configuration consumes 35 minutes, 440.2 minutes, and 2085.2 minutes to process 215MB, 1.2GB, and 7.3GB dataset respectively. On the other hand 3-node configuration takes 3.4 minutes, 48.4 minutes, and 235.4 minutes to process 215MB, 1.2GB, and 7.3GB dataset respectively. 90% reduction in processing-time is observed for 3-node configuration as compared to standalone on a dataset of 215MB. Similarly, 89% and 88.7% decrease in the processing time is noted with 1.2GB and 7.3GB of data, respectively. Last one is the write operation and the write-time for both the configurations is shown in Fig. 5.11(c). Observations similar to read operation are noticed for the write operation. A single node configuration takes lesser time, specifically 6.5 seconds, 28.3 seconds, and 146.6 seconds for 215MB, 1.2GB, and 7.3GB dataset respectively. On contrary, 3-node configuration reported 31.2 seconds, 138.2 seconds, and 672.8 seconds for 215MB, 1.2GB, and 7.3GB dataset respectively.

The performance comparison of the processing time is also compared using variable data size in terms of images as well ranging from 100 to 1000 images. The same dataset and range has been used in Chapter 3 for comparing the performance in terms of classification (accuracy). Fig. 5.12 shows the graphically representation of

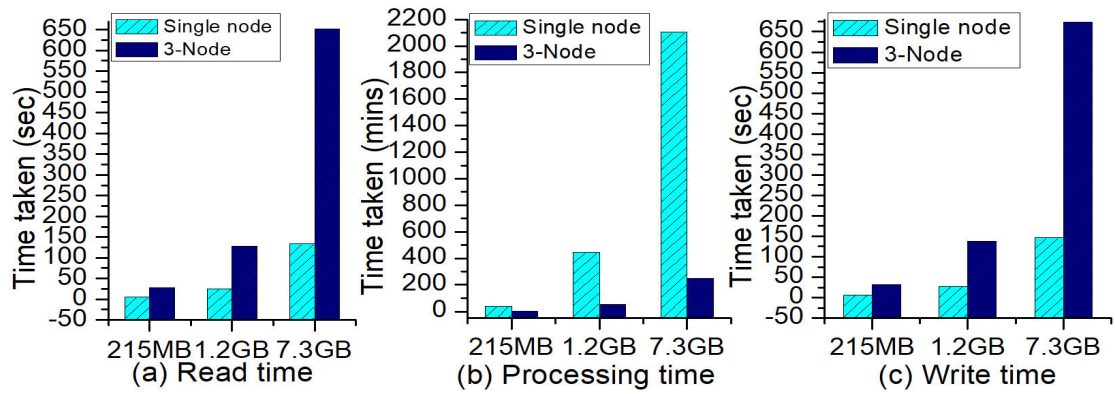


Figure 5.11: Performance Evaluation of standalone and multi-node system with respect to (a) Read time (b) Processing time and (c) Write time

the number of image versus processing time. Processing time of node 1, 2, and 3 is nearly same initially for 100 images. The variation slowly seems to be more viable with an increase in the number of images and nodes.

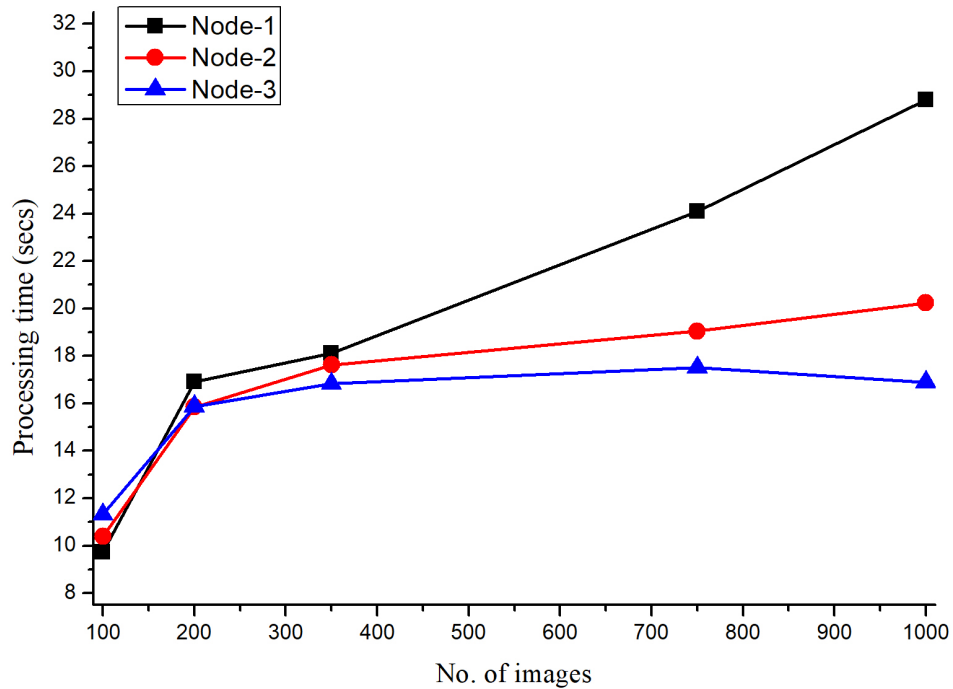


Figure 5.12: Performance comparison of processing time on different nodes

Chapter 6

Conclusions and Future Scope

Development of automatic or semi-automatic brain tumor detection system providing precise results is one of the significant research areas in the domain of medical imaging. Among all modules, the critical as well as the most explored one is MR image segmentation. However, the growing size of medical datasets along with technological developments demand for systems with improved efficiency, accuracy, and computation speed. The work presented in this thesis is an attempt in the same direction and this chapter concludes the same. In addition, a few of the possible future research directions are listed. Section 6.1 presents the main conclusions drawn from the thesis followed by upcoming scope in Section 6.2.

6.1 Conclusions

The existing benefits of MR imaging over other diagnostic imaging techniques drive the focus of this research work on MR image based brain tumor detection techniques. Based on the study and analysis of existing brain tumor detection techniques it is clear that efficient tumor identification, extraction, segmentation, and classification are some of the challenging tasks for physicians and radiologists. Automation of these modules thus occupies a major proportion of research in the domain of medical imaging. Several existing detection techniques are shown to achieve good

performance on different tumor datasets. Irrespective of the accuracy percentages reported by any automatic tumor detection system using the best performing segmentation approach, a second opinion is still required for better diagnosis in any of the cases. The motive of medical image processing handling tumor research is to precisely find and concentrate the tumor region to ease the operation medically.

MR image contrast is a significant factor as it highly influences the process of brain tumor detection. Segmentation approaches have been applied according to the behavior of MR image modality and tumor type which helps to separate abnormal tissue (edema, tumors etc.) from normal tissues (WM, GM and CSF). However, segmentation using MR brain images has to be as efficient as possible and hence there still exists scope for improved accuracy and/or computation speed of existing systems.

The development of a framework for brain tumor segmentation using MR image is achieved by proposing a novel hybrid weighted fuzzy based segmentation algorithm termed WFKM. The algorithm focuses on improved pixel clustering by assigning quantization weights to pixel values and is found to segment well in comparison to standard approaches. It was observed that the concept of fuzzified weights with illumination penalizing factors eases the performance in spatial context. The penalization assists in effectively dealing the issues with pixel's multiple memberships as well as exponential increase in number of iterations to attain more meaningful clusters. WFKM is found to acceptably segment MR image sequences based on anatomical planes (coronal and axial) and is capable in analyzing the tumor type clearly.

Features extracted after effective WFKM segmentation also helps in better training and thus quantitatively, i.e. in terms of classification accuracy WFKM is observed to perform better as compared to the existing works. The proposed approach WFKM with SVM reports an average accuracy of 97% on DICOM dataset with total number of images varying from 100 to 1000. WFKM supersedes the existing approaches, namely k-means, FCM, Canny edge detection, and FKM. Values obtained

using other considered performance parameters (Sensitivity, Specificity, Dice Coefficient Index and several rates computed from Confusion Matrix) further validate the proposal. Experiments on online datasets too reported improved segmentation results with visually enhanced intensity levels.

The approach, WFKM with SVM, is also scaled to handle segmentation and classification of massive MR brain image data aiming high accuracy but with low execution time. A scalable, distributed, and parallel processing environment is incorporated utilizing Hadoop-MapReduce and MDCS platforms. The combination is evaluated on varying sized datasets (215MB, 1.2GB, and 7.3GB) for 1-, 2-, and 3-node configurations. A decrease in the processing-time of the system is observed with configurations having more nodes. The collaboration is found to achieve both the performance parameters, i.e. accuracy as well as execution time. Overall the improved system can easily be used in medical diagnosis as a Decision Support System that will help the practitioners in reducing the turnaround time.

6.2 Future Scope

The real-world problems require proper addressing in the domain of medical image analysis that have been external to the domain of computer vision. Despite various remarkable achievements done in imaging modalities and approaches, still many challenges underline the detection techniques. The computational complexity in handling multi-modal MR image modalities at a single instance is the major challenge in imaging techniques. During MR image scan for diagnosis, correct tuning of parameters and motion applied to images, are the areas that can be explored in future. Exclusive difficulties emerge particularly in each imaging modality, data source, environment, pathological settings, and searching experts (for second opinion) needs to be focused in the coming future to make them more cost effective. Here are some points from research prospective which may be explored to enhance the current state-of-the-art in near future.

Tumor classification is important and critical too. In this research SVM classifier has been explored for classification. The proposal may also be combined with other classification approaches to verify its applicability on a wider range of tumor types or other domains as well. Therefore, one can think of developing a system that can correctly categorize tumor grades along with brain diseases.

In the era of industrialization, a hinge is required to process MR brain tumor images in big data health-care systems. A huge range of machines as well as deep learning techniques are well-accepted in this domain. Integration of such techniques with parallel and distributed platforms can be focused to develop a fast as well as accurate brain tumor detection framework. The same is explored in the research work presented in this thesis but efficiency in terms of read- and write-time is lacking which needs further analysis. Parallelism is achieved via a cloud based environment here. In the future network based environments can also be explored and compared to get a clearer view of these environments in the domain of brain tumor detection using MR images.

Appendix A

A Feature vectors

The 13 features namely, contrast, correlation, energy, homogeneity, mean, standard deviation, variance, entropy, root mean square, smoothness, kurtosis, skewness, and inverse difference moment (IDM) are utilized in this work. Table presents summarizes all the features.³

³where n and m are rows and columns in an image, and $f(n, m)$ and $f^R(n, m)$ are original and reconstructed images

Table A.1: Summarization of feature vectors

Metrics	Formulae
Mean (M)	$M = \left(\frac{1}{n \times m} \right) \sum_0^{n-1} \sum_0^{m-1} f(n, m)$ Measures the average value of intensities at pixel.
Standard deviation (SD)	$SD = \sqrt{\left(\frac{1}{n \times m} \sum_0^{m-1} (f(n, m) - M)^2 \right)}$ Measures the gray level spread in the region of the mean.
Mean square error (MSE)	$MSE = \frac{1}{m \times n} \sum \sum (f(n, m) - f^R(n, m))$ Measures the square of differences in the pixel values between the corresponding pixels of the two images.
Entropy (E)	$E = - \sum_{m=0}^{m-1} \sum_{n=0}^{n-1} f(n, m) \log_2 f(n, m)$ Measures the uncertainty/randomness contained in the image
Skewness (S_k)	$S_k = \left(\frac{1}{m \times n} \frac{\sum (f(n, m) - M)^3 }{SD^3} \right)$ Measures symmetry and non-symmetry of an image.
Kurtosis (K_{urt})	$K_{urt} = \left(\frac{1}{m \times n} \frac{\sum (f(n, m) - M)^4}{SD^4} \right)$ Verifies the shape of a random variable's probability distribution.
Energy (En)	$En = \sqrt{\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} f^2(x, y)}$ Measures similarity of an image.
Contrast (C_{on})	$C_{on} = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (x - y)^2 f(x, y)$ Measures intensity of a pixel and its neighbor over the image.
Homogeneity (H)	$H = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \frac{1}{1 + (x - y)^2} f(x, y)$ Measures single or a range of values to identify whether the image is textured or non-textured.
Coarseness (C_{ness})	$C_{ness} = \frac{1}{2^{m+n}} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} f(x, y)$ Measures roughness for textural analysis of an image.

Appendix B

The clustering approaches applied to process MR images lacks scalable solution in case of massive data. However, FCM always excels any clustering approach based on accuracy parameter whereas effecting the computation time. This time parameter is managed and controlled by the iterative k-means approach. Practically, individual approaches fails to fulfill the desired outcomes of the system. Moreover, even the integration of FCM and k-means (FKM) limits the meaningful cluster formation due to increased number of iterations, in return hampering the execution time. Though, as the size of data accelerates the difficulty to manage the main memory persists. Nevertheless, to withstand such problem a platform to store, process, and distribute bulk data becomes mandatory. The proposed model of hybrid approach integrated with Hadoop-MapReduce Platform is depicted in Fig.B.1.

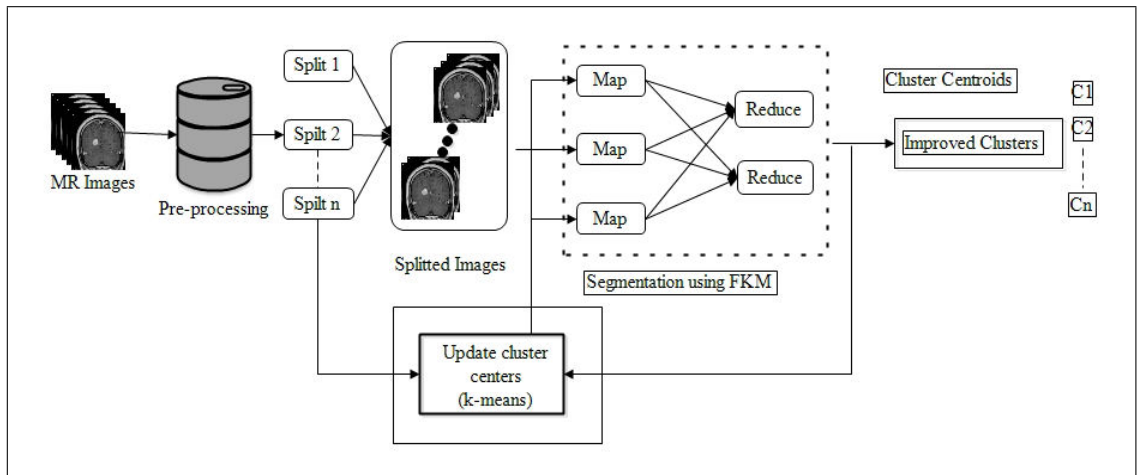


Figure B.1: Proposed fusion based approach for brain tumor detection using Hadoop-MapReduce mechanism

In the first phase, MR brain tumor dataset is collected and preprocessed for

further processing. The preprocessing module executes all the noise removal and filtering mechanism for improved image quality. These preprocessed images are stored in HDFS, which spilt the image data into chunks of 128MB for next level processing. The task is divided here and forwarded to various mapper function. The proposed segmentation approach, FKM is implemented in MapReduce phase to segment tumor from MR brain images parallelly. At the same time through multiple nodes this process of segmentation is carried to save the processing time of the system.

A Segmentation using FKM

The Fuzzy K-mean (FKM) scheme works by fuzzifying all the extracted pixel values of the image based on the membership function. The spatial context of FCM utilizes the information such as illumination and intensity to formulate the clusters. The membership function of the proposed approach is the modification of the standard FCM method and is defined in Eq. B.1. FKM resolves issues occurring due to FCM in handling scattered or multiple membership clusters leading to unclear segments.

$$J(U, V) = \sum_{i=1}^n \sum_{j=1}^k (u_{ij})^m \|x_i - \nu_j\|^2 \quad (\text{B.1})$$

The membership function is denoted by $U = u_{i,j} \in [0, 1], i = 1, 2, \dots, n, j = 1, \dots, c$. The value associated with each pixel $u_{i,j}$ tells the degree to which pixel x_i belongs to cluster j . Membership function ($u_{i,j}$) is calculated using Eq. B.2.

$$u_{i,j} = \frac{1}{\sum_{k=1}^c (d_{ij} - d_{ik})^{\frac{2}{m-1}}} \quad (\text{B.2})$$

$$\sum_{j=1}^c u_j(x_i) = 1 \quad (\text{B.3})$$

$$v_j = \frac{\sum_{i=1}^n (u_{ij})^m x_i}{\sum_{i=1}^n (u_{ij})^m} \quad (\text{B.4})$$

According to Eq. B.3 the summation of each membership for each data point should be equal to 1. The Euclidean distance between the i^{th} data and j^{th} cluster centre is defined in Eq. B.4. The proposed fuzzy c-means algorithm works towards minimization of the objective function $J(U, V)$ defined in Eq. B.1.

B FKM with Hadoop-MapReduce

The FKM approach for MR Image segmentation is discussed in Alg. 3. Experiments on variable sized DICOM dataset, namely 2GB, 6GB, 8GB, reported an overall accuracy of 96% with processing time of In addition, FKM is integrated with on Hadoop-MapReduce and MDCS architecture presented in Chapter 4. Execution on datasets with sizes 215MB, 1.2GB, 7.3GB reported an overall accuracy of 97% and 90% reduction in processing-time is observed for 3-node configuration as compared to standalone on a dataset of 215MB. Similarly, 89% and 88.7% decrease in the processing time is noted with 1.2GB and 7.3GB of data, respectively.

Algorithm 3 Standalone FKM based algorithm

Input: MR Images I

Output: Segmented image I'

- 1: **for** $i \neq null$ **do**
 - 2: **for** $j \leq n$ **do**
 - 3: Randomly select the cluster centers
 - 4: Calculate fuzzy membership
 - 5: $u_{ij} = \frac{1}{\sum_{k=1}^c (d_{ij} - d_{ik})^{\frac{2}{m-1}}}$ Calculate fuzzy center
 - 6: $v_j = \frac{\sum_{i=1}^n (u_{ij})^m x_i}{\sum_{i=1}^n (u_{ij})^m}$ Update the clusters using Euclidean distance
 - 7: $(d_{ij} - d_{ik})$ Update the centers and repeat the steps 3 to 6 till u_{ij} is achieved.
 - 8: Image gradient $\rightarrow I'$
 - 9: **end for**
 - 10: **end for**
-

Bibliography

- [1] A. F. Gaillard. Brain tumors. [Online]. Available: <https://radiopaedia.org/articles/brain-tumours>
- [2] B. C. Lovell and A. J. Maeder, “Proceedings of the 2003 aprs workshop on digital image computing,” 2003.
- [3] Y. Zhu, G. S. Young, Z. Xue, R. Y. Huang, H. You, K. Setayesh, H. Hatabu, F. Cao, and S. T. Wong, “Semi-automatic segmentation software for quantitative clinical brain glioblastoma evaluation,” *Academic radiology*, vol. 19, no. 8, pp. 977–985, 2012.
- [4] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, “Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation,” *Medical image analysis*, vol. 36, pp. 61–78, 2017.
- [5] A. Namburu, S. kumar Samay, and S. R. Edara, “Soft fuzzy rough set-based mr brain image segmentation,” *Applied Soft Computing*, vol. 54, pp. 456–466, 2017.
- [6] T. Logeswari and M. Karnan, “An improved implementation of brain tumor detection using segmentation based on hierarchical self organizing map,” *International Journal of Computer Theory and Engineering*, vol. 2, no. 4, p. 591, 2010.

- [7] A. Basit and S. Egerton, “Bio-medical imaging: Localization of main structures in retinal fundus images,” in *IOP Conference Series: Materials Science and Engineering*, vol. 51, no. 1. IOP Publishing, 2013, p. 012009.
- [8] S. M. Bhandarkar, J. Koh, and M. Suk, “Multiscale image segmentation using a hierarchical self-organizing map,” *Neurocomputing*, vol. 14, no. 3, pp. 241–272, 1997.
- [9] C. S. Anand and J. Sahambi, “Mri denoising using bilateral filter in redundant wavelet domain,” in *TENCON 2008-2008 IEEE Region 10 Conference*. IEEE, 2008, pp. 1–6.
- [10] P. K. Sran, S. Gupta, and S. Singh, “Content based medical image coding with fuzzy level set segmentation algorithm,” in *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012)*. Springer, 2013, pp. 161–171.
- [11] M. P. Gupta, M. M. Shringirishi *et al.*, “Implementation of brain tumor segmentation in brain mr images using k-means clustering and fuzzy c-means algorithm,” *International Journal of Computers & Technology*, vol. 5, no. 1, pp. 54–59, 2013.
- [12] M. Angulakshmi and G. Lakshmi Priya, “Automated brain tumour segmentation techniques a review,” *International Journal of Imaging Systems and Technology*, vol. 27, no. 1, pp. 66–77, 2017.
- [13] D. N. Louis, H. Ohgaki, O. D. Wiestler, W. K. Cavenee, P. C. Burger, A. Jouvet, B. W. Scheithauer, and P. Kleihues, “The 2007 who classification of tumours of the central nervous system,” *Acta neuropathologica*, vol. 114, no. 2, pp. 97–109, 2007.

- [14] L. Ulivi, M. Cosottini, G. Migaleddu, G. Orlandi, N. Giannini, G. Siciliano, and M. Mancuso, “Brain mri in monogenic cerebral small vessel diseases: A practical handbook,” *Current Molecular Medicine*, 2021.
- [15] J. Dolz, N. Betrouni, M. Quidet, D. Kharroubi, H. A. Leroy, N. Reyns, L. Mas-soptier, and M. Vermandel, “Stacking denoising auto-encoders in a deep network to segment the brainstem on mri in brain cancer patients: a clinical study,” *Computerized Medical Imaging and Graphics*, vol. 52, pp. 8–18, 2016.
- [16] D. Chow, J. Qi, X. Guo, V. Miloushev, F. Iwamoto, J. Bruce, A. Lassman, L. Schwartz, A. Lignelli, B. Zhao *et al.*, “Semiautomated volumetric measurement on postcontrast mr imaging for analysis of recurrent and residual disease in glioblastoma multiforme,” *American Journal of Neuroradiology*, vol. 35, no. 3, pp. 498–503, 2014.
- [17] S. Bakas, K. Zeng, A. Sotiras, S. Rathore, H. Akbari, B. Gaonkar, M. Rozycki, S. Pati, and C. Davatzikos, “Glistrboost: combining multimodal mri segmentation, registration, and biophysical tumor growth modeling with gradient boosting machines for glioma segmentation,” in *BrainLes 2015*. Springer, 2015, pp. 144–155.
- [18] X. Xue, Z. Xue, F. Cao, Y. Zhu, G. S. Young, Y. Li, J. Yang, and S. T. Wong, “Pice: prior information constrained evolution for 3-d and 4-d brain tumor segmentation,” in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2010, pp. 840–843.
- [19] N. Gordillo, E. Montseny, and P. Sobrevilla, “State of the art survey on mri brain tumor segmentation,” *Magnetic resonance imaging*, vol. 31, no. 8, pp. 1426–1438, 2013.
- [20] P. Georgiadis, D. Cavouras, I. Kalatzis, A. Daskalakis, G. C. Kagadis, K. Sifaki, M. Malamas, G. Nikiforidis, and E. Solomou, “Improving brain

- tumor characterization on mri by probabilistic neural networks and non-linear transformation of textural features,” *Computer methods and programs in biomedicine*, vol. 89, no. 1, pp. 24–32, 2008.
- [21] G. Vishnuvarthanan, M. P. Rajasekaran, P. Subbaraj, and A. Vishnuvarthanan, “An unsupervised learning method with a clustering approach for tumor identification and tissue segmentation in magnetic resonance brain images,” *Applied Soft Computing*, vol. 38, pp. 190–212, 2016.
- [22] K. Vaidhya, S. Thirunavukkarasu, V. Alex, and G. Krishnamurthi, “Multi-modal brain tumor segmentation using stacked denoising autoencoders,” in *BrainLes*. Springer, 2015, pp. 181–194.
- [23] M. Drozdal, G. Chartrand, E. Vorontsov, M. Shakeri, L. Di Jorio, A. Tang, A. Romero, Y. Bengio, C. Pal, and S. Kadoury, “Learning normalized inputs for iterative estimation in medical image segmentation,” *Medical image analysis*, vol. 44, pp. 1–13, 2018.
- [24] J. L. Foo, “A survey of user interaction and automation in medical image segmentation methods,” *Tech rep ISUHCI 20062, Human Computer Interaction Department, Iowa State University*, 2006.
- [25] J. Yao, “Image processing in tumor imaging,” *New techniques in oncologic imaging*, pp. 79–102, 2006.
- [26] S. D. Olabarriaga and A. W. Smeulders, “Interaction in the segmentation of medical images: A survey,” *Medical image analysis*, vol. 5, no. 2, pp. 127–142, 2001.
- [27] G. Mohan and M. M. Subashini, “Mri based medical image analysis: Survey on brain tumor grade classification,” *Biomedical Signal Processing and Control*, vol. 39, pp. 139–161, 2018.

- [28] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, “Deep learning for brain mri segmentation: state of the art and future directions,” *Journal of digital imaging*, vol. 30, no. 4, pp. 449–459, 2017.
- [29] A. Kamboj, R. Rani, and J. Chaudhary, “Deep learning approaches for brain tumor segmentation: A review,” in *First International Conference on Secure Cyber Computing and Communication (ICSCCC)*. IEEE, 2018, pp. 599–603.
- [30] A. Işın, C. Direkoğlu, and M. Şah, “Review of mri-based brain tumor image segmentation using deep learning methods,” *Procedia Computer Science*, vol. 102, pp. 317–324, 2016.
- [31] J. Amin, M. Sharif, N. Gul, M. Raza, M. A. Anjum, M. W. Nisar, and S. A. C. Bukhari, “Brain tumor detection by using stacked autoencoders in deep learning,” *Journal of Medical Systems*, vol. 44, no. 2, p. 32, 2020.
- [32] T. Sadad, A. Rehman, A. Munir, T. Saba, U. Tariq, N. Ayesha, and R. Abbasi, “Brain tumor detection and multi-classification using advanced deep learning techniques,” *Microscopy Research and Technique*.
- [33] F. Bashir-Gonbadi and H. Khotanlou, “Brain tumor classification using deep convolutional autoencoder-based neural network: multi-task approach,” *Multimedia Tools and Applications*, pp. 1–21, 2021.
- [34] E.-S. A. El-Dahshan, H. M. Mohsen, K. Revett, and A.-B. M. Salem, “Computer-aided diagnosis of human brain tumor through mri: A survey and a new algorithm,” *Expert systems with Applications*, vol. 41, no. 11, pp. 5526–5545, 2014.
- [35] N. J. Tustison, K. Shrinidhi, M. Wintermark, C. R. Durst, B. M. Kandel, J. C. Gee, M. C. Grossman, and B. B. Avants, “Optimal symmetric multi-modal templates and concatenated random forests for supervised brain tumor

- segmentation (simplified) with antsr,” *Neuroinformatics*, vol. 13, no. 2, pp. 209–225, 2015.
- [36] S. Basheera and M. S. S. Ram, “A novel cnn based alzheimer’s disease classification using hybrid enhanced ica segmented gray matter of mri,” *Computerized Medical Imaging and Graphics*, p. 101713, 2020.
- [37] M. Liu, F. Li, H. Yan, K. Wang, Y. Ma, L. Shen, M. Xu, A. D. N. Initiative *et al.*, “A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in alzheimer’s disease,” *NeuroImage*, vol. 208, p. 116459, 2020.
- [38] L. Lazli, M. Boukadoum, and O. A. Mohamed, “Computer-aided diagnosis system for alzheimer’s disease using fuzzy-possibilistic tissue segmentation and svm classification,” in *2018 IEEE Life Sciences Conference (LSC)*. IEEE, 2018, pp. 33–36.
- [39] T. Brosch, L. Y. Tang, Y. Yoo, D. K. Li, A. Traboulsee, and R. Tam, “Deep 3d convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation,” *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1229–1239, 2016.
- [40] S. Roy, J. A. Butman, D. S. Reich, P. A. Calabresi, and D. L. Pham, “Multiple sclerosis lesion segmentation from brain mri via fully convolutional neural networks,” *arXiv preprint arXiv:1803.09172*, 2018.
- [41] R. Kathirvel and K. Batri, “A computer-aided approach for meningioma brain tumor detection using c anfis classifier,” *International Journal of Imaging Systems and Technology*, vol. 27, no. 3, pp. 193–200, 2017.
- [42] Y. Bhanothu, A. Kamalakannan, and G. Rajamanickam, “Detection and classification of brain tumor in mri images using deep convolutional network,” in

2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE, 2020, pp. 248–252.

- [43] K. R. Laukamp, L. Pennig, F. Thiele, R. Reimer, L. Görtz, G. Shakirin, D. Zopfs, M. Timmer, M. Perkuhn, and J. Borggreffe, “Automated meningioma segmentation in multiparametric mri,” *Clinical neuroradiology*, pp. 1–10, 2020.
- [44] Y. Jiang, L. Lv, S. Yin, P. Zhou, and S. Jiang, “Primary spinal pilocytic astrocytoma: Clinical study with long-term follow-up in 16 patients and a literature review,” *Neurosurgical review*, vol. 43, no. 2, pp. 719–727, 2020.
- [45] S. Suja, N. George, and A. George, “Classification of grades of astrocytoma images from mri using deep neural network,” in *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2018, pp. 1257–1262.
- [46] R. R. Monika and A. Kamboj, “Brain tumor classification for mr imaging using support vector machine,” *Progress in Advanced Computing and Intelligent Engineering: Proceedings of ICACIE 2017, Volume 2*, vol. 714, p. 165, 2018.
- [47] L. M. Fletcher-Heath, L. O. Hall, D. B. Goldgof, and F. R. Murtagh, “Automatic segmentation of non-enhancing brain tumors in magnetic resonance images,” *Artificial intelligence in medicine*, vol. 21, no. 1-3, pp. 43–63, 2001.
- [48] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, “Brain tumor segmentation using convolutional neural networks in mri images,” *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.
- [49] C.-H. Lee, M. Schmidt, A. Murtha, A. Bistritz, J. Sander, and R. Greiner, “Segmenting brain tumors with conditional random fields and support vector machines,” in *International Workshop on Computer Vision for Biomedical Image Applications*. Springer, 2005, pp. 469–478.

- [50] R. Rahmat, F. Brochu, C. Li, R. Sinha, S. J. Price, and R. Jena, “Semi-automated construction of patient individualised clinical target volumes for radiotherapy treatment of glioblastoma utilising diffusion tensor decomposition maps,” *The British journal of radiology*, vol. 93, no. 1108, p. 20190441, 2020.
- [51] N. Gordillo, E. Montseny, and P. Sobrevilla, “A new fuzzy approach to brain tumor segmentation,” in *International Conference on Fuzzy Systems*. IEEE, 2010, pp. 1–8.
- [52] J. J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, and A. Yuille, “Efficient multilevel brain tumor segmentation with integrated bayesian model classification,” *IEEE transactions on medical imaging*, vol. 27, no. 5, pp. 629–640, 2008.
- [53] L. Lindner, D. Narnhofer, M. Weber, C. Gsaxner, M. Kolodziej, and J. Egger, “Using synthetic training data for deep learning-based gbm segmentation,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 6724–6729.
- [54] D. P. Augustine and P. Raj, “Performance evaluation of parallel genetic algorithm for brain mri segmentation in hadoop and spark,” *Indian J. Sci. Technol*, 2016.
- [55] G. S. Aujla, R. Chaudhary, N. Kumar, J. J. Rodrigues, and A. Vinel, “Data offloading in 5g-enabled software-defined vehicular networks: A stackelberg-game-based approach,” *IEEE Communications Magazine*, vol. 55, no. 8, pp. 100–108, 2017.
- [56] R. Choy and A. Edelman, “Parallel matlab: Doing it right,” *Proceedings of the IEEE*, vol. 93, no. 2, pp. 331–341, 2005.

- [57] J. Dean and S. Ghemawat, “Mapreduce: simplified data processing on large clusters,” *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.
- [58] R. Chaudhary, G. S. Aujla, N. Kumar, and J. J. Rodrigues, “Optimized big data management across multi-cloud data centers: Software-defined-network-based analysis,” *IEEE Communications Magazine*, vol. 56, no. 2, pp. 118–126, 2018.
- [59] S. Wang and H. Wang, “Big data for small and medium-sized enterprises (sme): a knowledge management model,” *Journal of Knowledge Management*, 2020.
- [60] H. Vo, J. Kong, D. Teng, Y. Liang, A. Aji, G. Teodoro, and F. Wang, “Mareia: a cloud mapreduce based high performance whole slide image analysis framework,” *Distributed and Parallel Databases*, vol. 37, no. 2, pp. 251–272, 2019.
- [61] S. Bao, F. D. Weitendorf, A. J. Plassard, Y. Huo, A. Gokhale, and B. A. Landman, “Theoretical and empirical comparison of big data image processing with apache hadoop and sun grid engine,” in *Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications*, vol. 10138. International Society for Optics and Photonics, 2017, p. 101380B.
- [62] C. Vijayakumar, G. Damayanti, R. Pant, and C. Sreedhar, “Segmentation and grading of brain tumors on apparent diffusion coefficient images using self-organizing maps,” *Computerized Medical Imaging and Graphics*, vol. 31, no. 7, pp. 473–484, 2007.
- [63] K. Shanthi and M. S. Kumar, “Skull stripping and automatic segmentation of brain mri using seed growth and threshold techniques,” in *International conference on intelligent and advanced systems*. IEEE, 2007, pp. 422–426.
- [64] N. M. Saad, S. A. R. Abu-Bakar, S. Muda, and M. Mokji, “Segmentation of brain lesions in diffusion-weighted mri using thresholding technique,” in

- IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE, 2011, pp. 249–254.
- [65] Y.-C. Sung, K.-S. Han, C.-J. Song, S.-M. Noh, and J.-W. Park, “Threshold estimation for region segmentation on mr image of brain having the partial volume artifact,” in *WCC-ICSP. 5th International Conference on Signal Processing Proceedings. 16th World Computer Congress*, vol. 2. IEEE, 2000, pp. 1000–1009.
- [66] P. Gibbs, D. L. Buckley, S. J. Blackband, and A. Horsman, “Tumour volume determination from mr images by morphological segmentation,” *Physics in Medicine & Biology*, vol. 41, no. 11, p. 2437, 1996.
- [67] P. Dvorak, K. Bartusek, W. Kropatsch, and Z. Smékal, “Automated multi-contrast brain pathological area extraction from 2d mr images,” *Journal of applied research and technology*, vol. 13, no. 1, pp. 58–69, 2015.
- [68] S. Koley, A. K. Sadhu, P. Mitra, B. Chakraborty, and C. Chakraborty, “Delineation and diagnosis of brain tumors from post contrast t1-weighted mr images using rough granular computing and random forest,” *Applied Soft Computing*, vol. 41, pp. 453–465, 2016.
- [69] U. Ilhan and A. Ilhan, “Brain tumor segmentation based on a new threshold approach,” *Procedia computer science*, vol. 120, pp. 580–587, 2017.
- [70] M. Sujana, N. Alam, S. A. Noman, and M. J. Islam, “A segmentation based automated system for brain tumor detection,” *International Journal of Computer Applications*, vol. 153, no. 10, pp. 0975–8887, 2016.
- [71] K.-P. Wong, “Medical image segmentation: methods and applications in functional imaging,” in *Handbook of biomedical image analysis*. Springer, 2005, pp. 111–182.

- [72] M. R. Kaus, S. K. Warfield, A. Nabavi, P. M. Black, F. A. Jolesz, and R. Kikinis, "Automated segmentation of mr images of brain tumors," *Radiology*, vol. 218, no. 2, pp. 586–591, 2001.
- [73] S. Lakare and A. Kaufman, "3d segmentation techniques for medical volumes," *Center for Visual Computing, Department of Computer Science, State University of New York*, pp. 59–68, 2000.
- [74] Y. M. Salman, "Modified technique for volumetric brain tumor measurements," *Journal of Biomedical Science and Engineering*, vol. 2, no. 01, p. 16, 2009.
- [75] Y. Salman, M. Assal, A. Badawi, S. Alian, and M. E.-M. El-Bayome, "Validation techniques for quantitative brain tumors measurements," in *IEEE Engineering in Medicine and Biology 27th Annual Conference*. IEEE, 2006, pp. 7048–7051.
- [76] R. Ratan, S. Sharma, and S. Sharma, "Multiparameter segmentation and quantization of brain tumor from mri images," *Indian Journal of Science and Technology*, vol. 2, no. 2, pp. 11–15, 2009.
- [77] I. Epifanio and G. Ayala, "A random set view of texture classification," *IEEE Transactions on Image Processing*, vol. 11, no. 8, pp. 859–867, 2002.
- [78] S. Pandav, "Brain tumor extraction using marker controlled watershed segmentation," *Int J Eng Res Technol*, pp. 2278–0181, 2014.
- [79] C. Benson, V. Lajish, and K. Rajamani, "Brain tumor extraction from mri brain images using marker based watershed algorithm," in *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2015, pp. 318–323.

- [80] S. K. Hasan, Y. Sarkar, and M. Ahmad, “Watershed-matching algorithm: A new pathway for brain tumor segmentation,” in *Advances in Electronics, Communication and Computing*. Springer, 2018, pp. 45–52.
- [81] S. Bauer, C. Seiler, T. Bardyn, P. Buechler, and M. Reyes, “Atlas-based segmentation of brain tumor images using a markov random field-based tumor growth model and non-rigid registration,” in *Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 2010, pp. 4080–4083.
- [82] C. Seiler, P. Büchler, L.-P. Nolte, M. Reyes, and R. Paulsen, “Hierarchical markov random fields applied to model soft tissue deformations on graphics hardware,” in *Recent Advances in the 3D Physiological Human*. Springer, 2009, pp. 133–148.
- [83] K. Held, E. R. Kops, B. J. Krause, W. M. Wells, R. Kikinis, and H.-W. Muller-Gartner, “Markov random field segmentation of brain mr images,” *IEEE transactions on medical imaging*, vol. 16, no. 6, pp. 878–886, 1997.
- [84] R. Rulaningtyas and K. Ain, “Edge detection for brain tumor pattern recognition,” in *International Conference on Instrumentation, Communication, Information Technology, and Biomedical Engineering*. IEEE, 2009, pp. 1–3.
- [85] M. Diwakar, P. K. Patel, and K. Gupta, “Cellular automata based edge-detection for brain tumor,” in *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2013, pp. 53–59.
- [86] A. Aslam, E. Khan, and M. S. Beg, “Improved edge detection algorithm for brain tumor segmentation,” *Procedia Computer Science*, vol. 58, pp. 430–437, 2015.
- [87] N. Mathur, S. Mathur, and D. Mathur, “A novel approach to improve sobel edge detector,” *Procedia Computer Science*, vol. 93, pp. 431–438, 2016.

- [88] J. Sachdeva, V. Kumar, I. Gupta, N. Khandelwal, and C. K. Ahuja, "A novel content-based active contour model for brain tumor segmentation," *Magnetic resonance imaging*, vol. 30, no. 5, pp. 694–715, 2012.
- [89] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Medical image analysis*, vol. 1, no. 2, pp. 91–108, 1996.
- [90] J. Liu, M. Li, J. Wang, F. Wu, T. Liu, and Y. Pan, "A survey of mri-based brain tumor segmentation methods," *Tsinghua Science and Technology*, vol. 19, no. 6, pp. 578–595, 2014.
- [91] T. McInerney and D. Terzopoulos, "Deformable models. handbook of medical imaging processing and analysis," *Bankman, Ed. New York: Academic*, 2000.
- [92] A. K. Law, F. Lam, and F. H. Chan, "A fast deformable region model for brain tumor boundary extraction," in *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society*][*Engineering in Medicine and Biology*, vol. 2. IEEE, 2002, pp. 1055–1056.
- [93] H. Khotanlou, O. Colliot, J. Atif, and I. Bloch, "3d brain tumor segmentation in mri using fuzzy classification, symmetry analysis and spatially constrained deformable models," *Fuzzy sets and systems*, vol. 160, no. 10, pp. 1457–1473, 2009.
- [94] A. Rajendran and R. Dhanasekaran, "Fuzzy clustering and deformable model for tumor segmentation on mri brain image: a combined approach," *Procedia Engineering*, vol. 30, pp. 327–333, 2012.
- [95] C. Li, C. Xu, C. Gui, and M. D. Fox, "Level set evolution without re-initialization: a new variational formulation," in *IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 430–436.

- [96] Y. Chen, J. Zhang, and J. Macione, “An improved level set method for brain mr images segmentation and bias correction,” *Computerized Medical Imaging and Graphics*, vol. 33, no. 7, pp. 510–519, 2009.
- [97] K. Popuri, D. Cobzas, M. Jagersand, S. L. Shah, and A. Murtha, “3d variational brain tumor segmentation on a clustered feature set,” in *Medical Imaging 2009: Image Processing*, vol. 7259. International Society for Optics and Photonics, 2009, p. 72591N.
- [98] D. Cobzas, N. Birkbeck, M. Schmidt, M. Jagersand, and A. Murtha, “3d variational brain tumor segmentation using a high dimensional feature set,” 2007.
- [99] K. Xie, J. Yang, Z. Zhang, and Y. Zhu, “Semi-automated brain tumor and edema segmentation using mri,” *European journal of radiology*, vol. 56, no. 1, pp. 12–19, 2005.
- [100] A. E. Lefohn, J. E. Cates, and R. T. Whitaker, “Interactive, gpu-based level sets for 3d segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2003, pp. 564–572.
- [101] T. Rohlfing, R. Brandt, R. Menzel, D. B. Russakoff, and C. R. Maurer, “Quo vadis, atlas-based segmentation,” in *Handbook of biomedical image analysis*. Springer, 2005, pp. 435–486.
- [102] M. Prastawa, E. Bullitt, N. Moon, K. Van Leemput, and G. Gerig, “Automatic brain tumor segmentation by subject specific modification of atlas priors1,” *Academic radiology*, vol. 10, no. 12, pp. 1341–1348, 2003.
- [103] N. Subbanna and T. Arbel, “Probabilistic gabor and markov random fields segmentation of brain tumours in mri volumes,” *Proc MICCAI Brain Tumor Segmentation Challenge (BRATS)*, pp. 28–31, 2012.

- [104] N. Subbanna, D. Precup, and T. Arbel, “Iterative multilevel mrf leveraging context and voxel information for brain tumour segmentation in mri,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 400–405.
- [105] S. Bauer, L.-P. Nolte, and M. Reyes, “Segmentation of brain tumor images based on atlas-registration combined with a markov-random-field lesion growth model,” in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2011, pp. 2018–2021.
- [106] Z. Ji, Y. Xia, Q. Sun, Q. Chen, and D. Feng, “Adaptive scale fuzzy local gaussian mixture model for brain mr image segmentation,” *Neurocomputing*, vol. 134, pp. 60–69, 2014.
- [107] Z. Ji, Y. Xia, Q. Sun, Q. Chen, D. Xia, and D. D. Feng, “Fuzzy local gaussian mixture model for brain mr image segmentation,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 3, pp. 339–347, May 2012.
- [108] H. Greenspan, A. Ruf, and J. Goldberger, “Constrained gaussian mixture model framework for automatic segmentation of mr brain images,” *IEEE transactions on medical imaging*, vol. 25, no. 9, pp. 1233–1245, 2006.
- [109] Z. Ji, Y. Huang, Q. Sun, and G. Cao, “A spatially constrained generative asymmetric gaussian mixture model for image segmentation,” *Journal of Visual Communication and Image Representation*, vol. 40, pp. 611–626, 2016.
- [110] J. Nie, Z. Xue, T. Liu, G. S. Young, K. Setayesh, L. Guo, and S. T. Wong, “Automated brain tumor segmentation using spatial accuracy-weighted hidden markov random field,” *Computerized Medical Imaging and Graphics*, vol. 33, no. 6, pp. 431–441, 2009.

- [111] R. Donoso, A. Veloz, and H. Allende, “Modified expectation maximization algorithm for mri segmentation,” in *Iberoamerican Congress on Pattern Recognition*. Springer, 2010, pp. 63–70.
- [112] C. Ambroise and G. Govaert, “Spatial clustering and the em algorithm,” 1995.
- [113] T. Hu and S. Y. Sung, “Clustering spatial data with a hybrid em approach,” *Pattern analysis and applications*, vol. 8, no. 1-2, pp. 139–148, 2005.
- [114] Z. Yang, F.-L. Chung, and W. Shitong, “Robust fuzzy clustering-based image segmentation,” *Applied soft computing*, vol. 9, no. 1, pp. 80–84, 2009.
- [115] S. A. Shah and N. C. Chauhan, “An automated approach for segmentation of brain mr images using gaussian mixture model based hidden markov random field with expectation maximization,” *Journal of Biomedical Engineering and Medical Imaging*, vol. 2, no. 4, p. 57, 2015.
- [116] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.
- [117] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.
- [118] T. M. Mitchell, *The discipline of machine learning*. Carnegie Mellon University, School of Computer Science, Machine Learning, 2006, vol. 9.
- [119] W. Cai, S. Chen, and D. Zhang, “Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation,” *Pattern recognition*, vol. 40, no. 3, pp. 825–838, 2007.
- [120] L. Szilagyi, S. M. Szilagyi, and Z. Benyo, “A modified fuzzy c-means algorithm for mr brain image segmentation,” in *International Conference Image Analysis and Recognition*. Springer, 2007, pp. 866–877.

- [121] P. A. Mei, C. de Carvalho Carneiro, S. J. Fraser, L. L. Min, and F. Reis, "Analysis of neoplastic lesions in magnetic resonance imaging using self-organizing maps," *Journal of the neurological sciences*, vol. 359, no. 1-2, pp. 78–83, 2015.
- [122] A. De and C. Guo, "An adaptive vector quantization approach for image segmentation based on som network," *Neurocomputing*, vol. 149, pp. 48–58, 2015.
- [123] K. Nimeesha and R. M. Gowda, "Brain tumour segmentation using k-means and fuzzy c-means clustering algorithm," *Int J Comput Sci Inf Technol Res Excell*, vol. 3, pp. 60–65, 2013.
- [124] L.-H. Juang and M.-N. Wu, "Mri brain lesion image detection based on color-converted k-means clustering segmentation," *Measurement*, vol. 43, no. 7, pp. 941–949, 2010.
- [125] S. Ruan, S. Lebonvallet, A. Merabet, and J.-M. Constans, "Tumor segmentation from a multispectral mri images by using support vector machine classification," in *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2007, pp. 1236–1239.
- [126] B. Tanoori, Z. Azimifar, A. Shakibafar, and S. Katebi, "Brain volumetry: an active contour model-based segmentation followed by svm-based classification," *Computers in biology and medicine*, vol. 41, no. 8, pp. 619–632, 2011.
- [127] S. Ruan, N. Zhang, Q. Liao, and Y. Zhu, "Image fusion for following-up brain tumor evolution," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2011, pp. 281–284.
- [128] A. Demirhan and İ. Güler, "Combining stationary wavelet transform and self-organizing maps for brain mr image segmentation," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 2, pp. 358–367, 2011.

- [129] M. Balafar, A. R. Ramli, M. I. Saripan, R. Mahmud, and S. Mashohor, “Medical image segmentation using fuzzy c-mean (fcm), learning vector quantization (lvq) and user interaction,” in *International Conference on Intelligent Computing*. Springer, 2008, pp. 177–184.
- [130] L. Clarke, “Mr image segmentation using mlm and artificial neural nets,” *Medical physics*, vol. 18, no. 3, p. 673, 1991.
- [131] W. E. Reddick, J. O. Glass, E. N. Cook, T. D. Elkin, and R. J. Deaton, “Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks,” *IEEE Transactions on medical imaging*, vol. 16, no. 6, pp. 911–918, 1997.
- [132] J. C. Bezdek, L. Hall, and L. Clarke, “Review of mr image segmentation techniques using pattern recognition,” *Medical physics*, vol. 20, no. 4, pp. 1033–1048, 1993.
- [133] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [134] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.
- [135] J. Zhou, K. Chan, V. Chong, and S. M. Krishnan, “Extraction of brain tumor from mr images using one-class support vector machine,” in *IEEE Engineering in Medicine and Biology 27th Annual Conference*. IEEE, 2006, pp. 6411–6414.
- [136] M. Maitra and A. Chatterjee, “Hybrid multiresolution slantlet transform and fuzzy c-means clustering approach for normal-pathological brain mr image segregation,” *Medical engineering & physics*, vol. 30, no. 5, pp. 615–623, 2008.
- [137] P. Shanthakumar and P. Ganeshkumar, “Performance analysis of classifier for brain tumor detection and diagnosis,” *Computers & Electrical Engineering*, vol. 45, pp. 302–311, 2015.

- [138] F. G. Zöllner, K. E. Emblem, and L. R. Schad, "Svm-based glioma grading: optimization by feature reduction analysis," *Zeitschrift für medizinische Physik*, vol. 22, no. 3, pp. 205–214, 2012.
- [139] D. A. Dahab, S. S. Ghoniemy, G. M. Selim *et al.*, "Automated brain tumor detection and identification using image processing and probabilistic neural network techniques," *International journal of image processing and visual communication*, vol. 1, no. 2, pp. 1–8, 2012.
- [140] E.-S. A. El-Dahshan, T. Hosny, and A.-B. M. Salem, "Hybrid intelligent techniques for mri brain images classification," *Digital Signal Processing*, vol. 20, no. 2, pp. 433–441, 2010.
- [141] H. Mohsen, E.-S. A. El-Dahshan, and A.-B. M. Salem, "A machine learning technique for mri brain images," in *8th International Conference on Informatics and Systems (INFOS)*. IEEE, 2012, pp. BIO–161.
- [142] A. Demirhan, M. Törü, and I. Güler, "Segmentation of tumor and edema along with healthy tissues of brain using wavelets and neural networks," *IEEE journal of biomedical and health informatics*, vol. 19, no. 4, pp. 1451–1458, 2015.
- [143] N. Gupta and P. Khanna, "A non-invasive and adaptive cad system to detect brain tumor from t2-weighted mris using customized otsu's thresholding with prominent features and supervised learning," *Signal Processing: Image Communication*, vol. 59, pp. 18–26, 2017.
- [144] M. Maitra and A. Chatterjee, "A slantlet transform based intelligent system for magnetic resonance brain image classification," *Biomedical Signal Processing and Control*, vol. 1, no. 4, pp. 299–306, 2006.
- [145] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A modified fuzzy c-means algorithm for bias field estimation and segmentation of

- mri data,” *IEEE transactions on medical imaging*, vol. 21, no. 3, pp. 193–199, 2002.
- [146] A. V. Vasilakos, Y. Tang, Y. Yao *et al.*, “Neural networks for computer-aided diagnosis in medicine: A review,” *Neurocomputing*, vol. 216, pp. 700–708, 2016.
- [147] J. Panic, A. Defeudis, S. Mazzetti, S. Rosati, G. Giannetto, L. Vassallo, D. Regge, G. Balestra, and V. Giannini, “A convolutional neural network based system for colorectal cancer segmentation on mri images,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 1675–1678.
- [148] T. Kooi, G. Litjens, B. Van Ginneken, A. Gubern-Mérida, C. I. Sánchez, R. Mann, A. den Heeten, and N. Karssemeijer, “Large scale deep learning for computer aided detection of mammographic lesions,” *Medical image analysis*, vol. 35, pp. 303–312, 2017.
- [149] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen, “Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans,” *Scientific reports*, vol. 6, p. 24454, 2016.
- [150] A. M. Abdel-Zaher and A. M. Eldeib, “Breast cancer classification using deep belief networks,” *Expert Systems with Applications*, vol. 46, pp. 139–144, 2016.
- [151] W. Sun, B. Zheng, and W. Qian, “Computer aided lung cancer diagnosis with deep learning algorithms,” in *Medical imaging : computer-aided diagnosis*, vol. 9785. International Society for Optics and Photonics, 2016, p. 97850Z.
- [152] P. Lakhani and B. Sundaram, “Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks,” *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.

- [153] Z. Xiao, R. Huang, Y. Ding, T. Lan, R. Dong, Z. Qin, X. Zhang, and W. Wang, "A deep learning-based segmentation method for brain tumor in mr images," in *IEEE 6th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*. IEEE, 2016, pp. 1–6.
- [154] Z. Akkus, I. Ali, J. Sedlar, T. L. Kline, J. P. Agrawal, I. F. Parney, C. Giannini, and B. J. Erickson, "Predicting 1p19q chromosomal deletion of low-grade gliomas from mr images using deep learning," *arXiv preprint arXiv:1611.06939*, 2016.
- [155] H. Ali, M. Elmogy, E. El-Daydamony, and A. Atwan, "Multi-resolution mri brain image segmentation based on morphological pyramid and fuzzy c-mean clustering," *Arabian Journal for Science and Engineering*, vol. 40, no. 11, pp. 3173–3185, 2015.
- [156] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, and A.-B. M. Salem, "Classification using deep learning neural networks for brain tumors," *Future Computing and Informatics Journal*, vol. 3, no. 1, pp. 68–71, 2018.
- [157] M. Soltaninejad, G. Yang, T. Lambrou, N. Allinson, T. L. Jones, T. R. Barrick, F. A. Howe, and X. Ye, "Supervised learning based multimodal mri brain tumour segmentation using texture features from supervoxels," *Computer methods and programs in biomedicine*, vol. 157, pp. 69–84, 2018.
- [158] R. Saouli, M. Akil, R. Kachouri *et al.*, "Fully automatic brain tumor segmentation using end-to-end incremental deep neural networks in mri images," *Computer methods and programs in biomedicine*, vol. 166, pp. 39–49, 2018.
- [159] S. Hussain, S. M. Anwar, and M. Majid, "Brain tumor segmentation using cascaded deep convolutional neural network," in *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017, pp. 1998–2001.

- [160] S. Iqbal, M. U. Ghani, T. Saba, and A. Rehman, “Brain tumor segmentation in multi-spectral mri using convolutional neural networks (cnn),” *Microscopy research and technique*, vol. 81, no. 4, pp. 419–427, 2018.
- [161] M. M. Thaha, K. P. M. Kumar, B. Murugan, S. Dhanasekeran, P. Vijayakarthish, and A. S. Selvi, “Brain tumor segmentation using convolutional neural networks in mri images,” *Journal of medical systems*, vol. 43, no. 9, p. 294, 2019.
- [162] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, “A deep learning model integrating fcnn and crfs for brain tumor segmentation,” *Medical image analysis*, vol. 43, pp. 98–111, 2018.
- [163] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, “Brain tumor segmentation with deep neural networks,” *Medical image analysis*, vol. 35, pp. 18–31, 2017.
- [164] Y. Pan, W. Huang, Z. Lin, W. Zhu, J. Zhou, J. Wong, and Z. Ding, “Brain tumor grading based on neural networks and convolutional neural networks,” in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2015, pp. 699–702.
- [165] Y. Pan, W. Huang, Z. Lin, W. Zhu, J. Zhou, J. Wong, and Z. Ding, “Brain tumor grading based on neural networks and convolutional neural networks,” in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 699–702.
- [166] J. Amin, M. Sharif, M. Yasmin, and S. L. Fernandes, “Big data analysis for brain tumor detection: Deep convolutional neural networks,” *Future Generation Computer Systems*, vol. 87, pp. 290–297, 2018.

- [167] S. Hussain, S. M. Anwar, and M. Majid, “Segmentation of glioma tumors in brain using deep convolutional neural network,” *Neurocomputing*, vol. 282, pp. 248–261, 2018.
- [168] A. Myronenko, “3d mri brain tumor segmentation using autoencoder regularization,” in *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 311–320.
- [169] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, “A reproducible evaluation of ants similarity metric performance in brain image registration,” *Neuroimage*, vol. 54, no. 3, pp. 2033–2044, 2011.
- [170] Z. Akkus, J. Sedlar, L. Coufalova, P. Korfiatis, T. L. Kline, J. D. Warner, J. Agrawal, and B. J. Erickson, “Semi-automated segmentation of pre-operative low grade gliomas in magnetic resonance imaging,” *Cancer Imaging*, vol. 15, no. 1, p. 12, 2015.
- [171] L. Singh, G. Chetty, and D. Sharma, “A novel machine learning approach for detecting the brain abnormalities from mri structural images,” in *IAPR International Conference on Pattern Recognition in Bioinformatics*. Springer, 2012, pp. 94–105.
- [172] K. B. Ahmed, L. O. Hall, D. B. Goldgof, R. Liu, and R. A. Gatenby, “Fine-tuning convolutional deep features for mri based brain tumor classification,” in *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134. International Society for Optics and Photonics, 2017, p. 101342E.
- [173] V. Alex, K. Vaidhya, S. Thirunavukkarasu, C. Kesavadas, and G. Krishnamurthi, “Semisupervised learning using denoising autoencoders for brain lesion detection and segmentation,” *Journal of Medical Imaging*, vol. 4, no. 4, p. 041311, 2017.

- [174] C. Baur, B. Wiestler, S. Albarqouni, and N. Navab, “Deep autoencoding models for unsupervised anomaly segmentation in brain mr images,” in *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 161–169.
- [175] M. Pesteie, P. Abolmaesumi, and R. N. Rohling, “Adaptive augmentation of medical data using independently conditional variational auto-encoders,” *IEEE transactions on medical imaging*, 2019.
- [176] X. Chen and E. Konukoglu, “Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders,” *arXiv preprint arXiv:1806.04972*, 2018.
- [177] H. Chen, Z. Qin, Y. Ding, L. Tian, and Qin, “Brain tumor segmentation with generative adversarial nets,” in *2nd International Conference on Artificial Intelligence and Big Data*. IEEE, 2019.
- [178] K. Sohn, H. Lee, and X. Yan, “Learning structured output representation using deep conditional generative models,” in *Advances in neural information processing systems*, 2015, pp. 3483–3491.
- [179] P. Afshar, A. Shahroudnejad, A. Mohammadi, and K. N. Plataniotis, “Carisi: Convolutional autoencoder-based inter-slice interpolation of brain tumor volumetric images,” in *25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 1458–1462.
- [180] Y. Liang, J. Wang, S. Zhang, and Y. Gong, “Learning visual co-occurrence with auto-encoder for image super-resolution,” in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*. IEEE, 2014, pp. 1–4.
- [181] K.-L. Tseng, Y.-L. Lin, W. Hsu, and C.-Y. Huang, “Joint sequence learning and cross-modality convolution for 3d biomedical segmentation,” in *Proceed-*

ings of the *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6393–6400.

- [182] S. Chaplot, L. Patnaik, and N. Jagannathan, “Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network,” *Biomedical signal processing and control*, vol. 1, no. 1, pp. 86–92, 2006.
- [183] A. Kharrat, K. Gasmi, M. B. Messaoud, N. Benamrane, and M. Abid, “A hybrid approach for automatic classification of brain mri using genetic algorithm and support vector machine,” *Leonardo journal of sciences*, vol. 17, no. 1, pp. 71–82, 2010.
- [184] A. 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)*. IEEE, 2015, pp. 98–102.
- [185] E. Abdel-Maksoud, M. Elmogy, and R. Al-Awadi, “Brain tumor segmentation based on a hybrid clustering technique,” *Egyptian Informatics Journal*, vol. 16, no. 1, pp. 71–81, 2015.
- [186] H. Al-Dmour and A. Al-Ani, “A clustering fusion technique for mr brain tissue segmentation,” *Neurocomputing*, vol. 275, pp. 546–559, 2018.
- [187] Y. Zhang, S. Wang, and L. Wu, “A novel method for magnetic resonance brain image classification based on adaptive chaotic pso,” *Progress In Electromagnetics Research*, vol. 109, pp. 325–343, 2010.
- [188] S. Saha, A. K. Alok, and A. Ekbal, “Brain image segmentation using semi-supervised clustering,” *Expert Systems with Applications*, vol. 52, pp. 50–63, 2016.

- [189] B. Amarapur *et al.*, “Computer-aided diagnosis applied to mri images of brain tumor using cognition based modified level set and optimized ann classifier,” *Multimedia Tools and Applications*, pp. 1–29, 2018.
- [190] A. Deepa and W. S. Emmanuel, “An efficient detection of brain tumor using fused feature adaptive firefly backpropagation neural network,” *Multimedia Tools and Applications*, pp. 1–16, 2018.
- [191] A. Vishnuvarthanan, M. P. Rajasekaran, V. Govindaraj, Y. Zhang, and A. Thiagarajan, “An automated hybrid approach using clustering and nature inspired optimization technique for improved tumor and tissue segmentation in magnetic resonance brain images,” *Applied Soft Computing*, vol. 57, pp. 399–426, 2017.
- [192] W. Li, C. Feng, C. Jin, Q. Chen, H. Liu, and D. Zhao, “A two-level cloud storage system based on asynchronous message for medical image big data,” in *2019 5th International Conference on Big Data Computing and Communications (BIGCOM)*. IEEE, 2019, pp. 54–58.
- [193] K. Meena and J. Sujatha, “Reduced time compression in big data using mapreduce approach and hadoop,” *Journal of Medical Systems*, vol. 43, no. 8, p. 239, 2019.
- [194] J. Patil and D. G. Predeepini, “Two dimensional medical images diagnosis using mapreduce,” *Indian Journal of science and Technology*, vol. 9, no. 17, 2016.
- [195] J. S. Patil and G. Pradeepini, “Three-dimensional mri brain image analysis on hadoop platform,” in *International Conference on Intelligent Computing and Applications*. Springer, 2018, pp. 131–142.

- [196] T. Taylor, N. John, P. Buendia, and M. Ryan, “Map-reduce enabled hidden markov models for high throughput multimodal brain tumor segmentation,” *Multimodal Brain Tumor Segmentation*, vol. 43, 2013.
- [197] A. Palle and R. Kulkarni, “Classification of medical mri brain images based on hadoop,” in *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*, 2016, pp. 1–4.
- [198] R. K. Grace, R. Manimegalai, and S. S. Kumar, “Medical image retrieval system in grid using hadoop framework,” in *2014 International Conference on Computational Science and Computational Intelligence*, vol. 1. IEEE, 2014, pp. 144–148.
- [199] P. Szwarc, J. Kawa, M. Rudzki, and E. Pietka, “Automatic brain tumour detection and neovasculature assessment with multiseriess mri analysis,” *Computerized Medical Imaging and Graphics*, vol. 46, pp. 178–190, 2015.
- [200] S. Gupta, P. Walia, C. Singla, S. Dhankar, T. Mishra, A. Khandelwal, and M. Bhardwaj, “Segmentation, feature extraction and classification of astrocytoma in mr images,” *Indian Journal of Science and Technology*, vol. 9, no. 36, pp. 1–8, 2016.
- [201] K. M. Priya, S. Kavitha, and B. Bharathi, “Brain tumor types and grades classification based on statistical feature set using support vector machine,” in *10th International Conference on Intelligent Systems and Control (ISCO)*. IEEE, 2016, pp. 1–8.
- [202] K. R. Laukamp, F. Thiele, G. Shakirin, D. Zopfs, A. Faymonville, M. Timmer, D. Maintz, M. Perkuhn, and J. Borggrefe, “Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric mri,” *European radiology*, vol. 29, no. 1, pp. 124–132, 2019.

- [203] T. Priya and P. Kalavathi, "Hsv based histogram thresholding technique for mri brain tissue segmentation," in *International Symposium on Signal Processing and Intelligent Recognition Systems*. Springer, 2018, pp. 322–333.
- [204] F. E.-Z. A. El-Gamal, M. M. Elmogy, A. Atwan, M. Ghazal, G. N. Barnes, H. Hajjdiab, R. Keynton, and A. S. El-Baz, "Significant region-based framework for early diagnosis of alzheimer's disease using 11 c pib-pet scans," in *24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 2989–2994.
- [205] S. J. Mohammed, "A proposed alzheimer's disease diagnosing system based on clustering and segmentation techniques," *Engineering and Technology Journal*, vol. 36, no. 2 Part (B) Engineering, pp. 160–165, 2018.
- [206] S. Spasov, L. Passamonti, A. Duggento, P. Liò, N. Toschi, A. D. N. Initiative *et al.*, "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to alzheimer's disease," *Neuroimage*, vol. 189, pp. 276–287, 2019.
- [207] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984.
- [208] J. Vijay and J. Subhashini, "An efficient brain tumor detection methodology using k-means clustering algorithm," in *2013 International Conference on Communication and Signal Processing*. IEEE, 2013, pp. 653–657.
- [209] P. Mohan, D. B. AL V, and B. Kavitha, "Intelligent based brain tumor detection using aco," *Int J Innovat Res Comput Commun Eng*, vol. 1, no. 9, pp. 2143–50, 2013.
- [210] R. Rani, A. Kamboj *et al.*, "Brain Tumor Classification for MR Imaging Using Support Vector Machine," in *Progress in Advanced Computing and Intelligent Engineering*. Springer, 2019, pp. 165–176.

- [211] C. Benson, V. Lajish, and K. Rajamani, "A novel skull stripping and enhancement algorithm for the improved brain tumor segmentation using mathematical morphology," *Int J Image Graph Signal Process*, vol. 8, no. 7, pp. 59–66, 2016.
- [212] N. Gupta and P. Khanna, "A non-invasive and adaptive cad system to detect brain tumor from t2-weighted mris using customized otsu's thresholding with prominent features and supervised learning," *Signal Processing: Image Communication*, vol. 59, pp. 18–26, 2017.
- [213] N. Nabizadeh and M. Kubat, "Brain tumors detection and segmentation in mr images: Gabor wavelet vs. statistical features," *Computers & Electrical Engineering*, vol. 45, pp. 286–301, 2015.
- [214] B. Vendt. (2016; url: <https://wiki.cancerimagingarchive.net/display/Public/Collections>, May) Neuroimaging tools and resources collaboratory. NITRC.
- [215] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in *2010 IEEE 26th symposium on mass storage systems and technologies (MSST)*. IEEE, 2010, pp. 1–10.
- [216] S. Ghemawat, "Mapreduce: Simplified data processing on large clusters," *OSDI, 2004*, 2004.
- [217] M. M. Sonawane and S. S. Kawathekar, "Hadoop and map reduce biomedical images using clustering," in *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*. IEEE, 2017, pp. 945–947.
- [218] J. Shelton, C. Rice, J. Singh, J. Jenkins, R. Dave, K. Roy, and S. Chakraborty, "Palm print authentication on a cloud platform," in *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)*. IEEE, 2018, pp. 1–6.

- [219] R. Dutta and B. Annappa, “A scalable cloud platform using matlab distributed computing server integrated with hdfs,” in *2012 International Symposium on Cloud and Services Computing*. IEEE, 2012, pp. 141–145.
- [220] A. Jayachandran and R. Dhanasekaran, “Brain tumor detection using fuzzy support vector machine classification based on a texton co-occurrence matrix,” *Journal of imaging Science and Technology*, vol. 57, no. 1, pp. 10 507–1, 2013.
- [221] M. Alfonse and A.-B. M. Salem, “An automatic classification of brain tumors through mri using support vector machine,” *Egypt. Comput. Sci. J*, vol. 40, no. 03, 2016.
- [222] N. B. Bahadure, A. K. Ray, and H. P. Thethi, “Image analysis for mri based brain tumor detection and feature extraction using biologically inspired bwt and svm,” *International journal of biomedical imaging*, vol. 2017, 2017.
- [223] N. V. Shree and T. Kumar, “Identification and classification of brain tumor mri images with feature extraction using dwt and probabilistic neural network,” *Brain informatics*, vol. 5, no. 1, pp. 23–30, 2018.
- [224] N. B. Bahadure, A. K. Ray, and H. P. Thethi, “Comparative approach of mri-based brain tumor segmentation and classification using genetic algorithm,” *Journal of digital imaging*, vol. 31, no. 4, pp. 477–489, 2018.
- [225] I. A. Rodríguez-Méndez, R. Ureña, and E. Herrera-Viedma, “Fuzzy clustering approach for brain tumor tissue segmentation in magnetic resonance images,” *Soft Computing*, vol. 23, no. 20, pp. 10 105–10 117, 2019.
- [226] A. M. KV, V. Rajendran *et al.*, “Glioma tumor grade identification using artificial intelligent techniques,” *Journal of medical systems*, vol. 43, no. 5, p. 113, 2019.

- [227] P. K. Chahal, S. Pandey, and S. Goel, “Hybrid approaches for brain tumor detection in mr images,” in *International Conference on Advances in Computing and Data Sciences*. Springer, 2019, pp. 264–274.

LIST OF PUBLICATIONS

Journal Publications (SCI/SCIE):

1. Chahal, P.K., Pandey, S. and Goel, S., 2020. A survey on brain tumor detection techniques for MR images. *Multimedia Tools and Applications*, 79, 21771-21814. [**Impact Factor: 2.757**]
2. Chahal, P.K., Pandey, S., 2020. An efficient Hadoop-based brain tumor detection framework using big data analytic. *Softw Pract Exper.*, 1–14. [**Impact Factor: 2.028**]
3. Chahal, P.K. and Pandey, S., 2021. A hybrid weighted fuzzy approach for brain tumor segmentation using MR images. *Neural Computing and Applications*, 1-15. [**Impact Factor: 5.606**]

Conference Publications (outside India):

1. Chahal, P.K., Pandey, S. and Goel, S., 2019, April. Hybrid approaches for brain tumor detection in MR images. In *International Conference on Advances in Computing and Data Sciences* (pp. 264-274). Springer, Singapore.
2. Chahal, P.K. and Pandey, S., 2020, November. An Efficient Hybrid Approach for Brain Tumor Detection in MR Images using Hadoop-MapReduce. In *2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)* (pp. 926-931).