Brain Tumor Detection & Classification Using Extreme Machine Learning and Implementation through Embedded System.

Team Members

Mr. Tadesse Hailu Ayane	Principal Investigator
Dr. Satyasis Mishra	Co- Investigator
Dr. Harish Kalla	Co- Investigator
Dr. Davinder Singh	Co- Investigator

Department of Electronics and Communication Engineering

This project report is an authentic record of Department of Electronics and Communication Engineering, School of Electrical Engineering and Computing

Office of Research Affairs
Adama Science and Technology University

June 2021

Adama, Ethiopia

ABSTRACT

The field of medical imaging is gaining importance with an increase in the demand for automated, reliable, fast and efficient diagnosis which can provide insight to the image better than human eyes. Brain tumor is the second leading cause for cancer-related deaths in men in age 20 to 39 and fifth leading cause cancer among women in same age group. Diagnosis of brain tumor is a very important part in its treatment. A prime reason behind an increase in the number of cancer patients worldwide is the ignorance towards treatment of a tumor in its early stages. The manual detection and classification of the tumor becomes a rigorous and hectic task for the radiologists. The segmentation, detection, and extraction of infected tumor area from magnetic resonance (MR) images are a primary concern, but a tedious and time taking task performed by radiologists or clinical experts, and their accuracy depends on their experience only.

The FCM (Fuzzy C Means) based segmentation techniques such as spatial information(FCM_S), FCM_S1 and FCM_S2, enhanced FCM algorithm (EnFCM), FGFCM (Fast generalized FCM algorithm), fuzzy local information c-means clustering algorithm (FLICM) has ability to obtain texture and background information but failed to remove in the case of complex rician noisy images. FLICM, improves the segmentation process, but fails to remove Gaussian noise beyond 30%. To improve the performance and reduce the complexity involved in the medical image segmentation process, a novel fast and robost FCM (FRFCM) image segmentation has been investigated in this research work. Furthermore, to improve the accuracy and quality rate of the neural network-based classifier, relevant features are extracted from each segmented tissue and aligned as input to the classifiers for automatic detection and classification of brain tumors. The simulation results of proposed technique have been evaluated and validated for performance and quality analysis on magnetic resonance brain images, based on feature coefficient.

The novel Fast and robust Fuzzy C Means (FRFCM) segmentation technique has been proposed for detection of brain tumor from MR (Magnetic Resonance) images that can inform the radiologist and doctor about the details of brain tumor. This segmentation technique include noise removal and sharpening of the MR image along with basic morphological functions, erosion and dilation, to obtain the background. The MR (Magnetic Resonance) Images features

have been extracted through a popular Gray Level Co-occurrence Matrix (GLCM) feature extraction technique. The extracted features are applied to the, proposed PSO+ELM (Particle Swarm Optimization-Extreme Learning Machine) for classification of the type of malignant **(**cancerous) and benign (non-cancerous) brain tumors for visual localization. The weights of the proposed novel multi class extreme learning machine classifier model has been updated by the PSO (Particle Swarm Optimization) algorithm to increase the performance of the classifiers. Further the classification results have been compared with the existing support vector machine and relevance vector machine model sparse Bayesian extreme learning machine. The proposed PSO+ELM model obtained an accuracy of 99.41% which is higher than the other said models. To validate the robust ness and rician noise removal capability of the proposed FRFCM segmentation two quality indexes are considered as Structural Similarity (SSIM) index and the Quality Index based on Local Variance (QILV). However, SSIM is more sensitive to the noise level in the image and the QILV to blurring of the edges. In addition to the both, the PSNR (Peak signal to noise ratio) is also calculated. It is also observed that the quality measure PSNR value is 37.32 dB and SSIM as 0.9219 dB. The higher value of PSNR indicate better signal-to noise ratio in the extracted image. Also, the larger value of SSIM indicates the noise reduction in the extracted image.

The automatic segmentation, feature extraction and classification has also been presented through GUI (Graphical User Interface) using MATLAB2018b software. Further, the detection and classification has been implemented through the embedded system platform by employing Raspberry PI3B+ along with python which may be the prototype product outcome of the research work. It will help the medical staff, particularly for the radiologist and doctor to understand the seriousness of the tumor.

ACKNOWLEDGEMENT

We would like to express my greatest appreciation to Electronics and Communication Engineering Department, and School of Electrical Engineering and Computing, Adama Science and Technology University, Adama, Ethiopia for giving opportunity and encouragement to propose the research proposal.

We would like to gratefully acknowledge Dr. Dereje Tekilu, Dean, SoEEC, ASTU, Adama, Ethiopia for his continuous encouragement throughout the research work. As well-wisher, his insight, observations and suggestions helped us to establish the overall direction of the research and contributed immensely to the success of the work.

We acknowledge Dr. Fayissa Debo, Former Dean, Technology Center, and Mr. Girma Debele, Former Associate Dean for research & Technology Transfer, Adama Science and Technology University, Adama, Ethiopia for their support and suggestions during the process of the project.

We must acknowledge the hospitals Adama general hospital and medical college, Adama, St. Paulo Hospital, Addis Ababa, American medical center, Addis Ababa for their support in providing data for research.

We acknowledge the academic resource that we received from Adama Science and Technology University, Adama, Ethiopia giving us a comfortable and active environment for pursuing our research work.

Finally, we would like to thank our staff members of the Electronics and Communication Engineering Department, ASTU, Adama for their support and encouragement throughout the research work.

TABLE OF CONTENTS

List of Figures

List of Acronyms

ANN	:	Artificial Neural Network
RBFN	:	Radial Basis Function Neural Networks
MLP	:	Multilayer Perceptron
LLRBFNN	:	Local Linear Radial Basis Function Neural Networks
LLWNN	:	Local Linear Wavelet Neural Networks
LMS	:	Least Mean Square Algorithm
FCM	:	Fuzzy C Means Algorithm
FFCM	:	Fast Fuzzy C Means Algorithm
PSO	:	Particle Swarm Optimization
SVM	:	Support Vector Machine
PNN	:	Probabilistic Neural Network
MRI	:	Magnetic Resonance Image
KNN	:	K-Nearest Neighbor
GLCM	:	Gray Level Co-occurrence Matrix
EnFCM	:	Enhanced Fuzzy c means algorithm
FGFCM	:	Fast generalized FCM algorithm
KWFLICM	:	Kernel metric local information C Means
NDFCM	:	Noise Detection FCM
FRFCM	:	Fast and Robust FCM
GA	:	Genetic Algorithm
DE	:	Differential Evolution
SSIM	:	Structural Similarity index
PSNR	:	Peak signal to noise ratio
QILV	:	Quality Index based on Local Variance
DM	:	Directional Moment
FLICM	:	Fuzzy Local Information C Means

List of Tables

CHAPTER ONE

1. INTRODUCTION

1.1 Background

In recent years, the death rate grows unanimously due the brain tumors at every age groups. According to the report published by the American Brain Tumor Association (ABTA) Feb 24, 2016[1], that the children and young adults between 15-39 age group are affected due to malignant brain tumors. Between the age group 15-19 year olds, the cancer related deaths are growing rapidly as per the report. When the data is analyzed in 5-year age increments, the researchers found that the mentioned age groups are affected by different types of tumor which is not known at its early stage. According to the reports of World Health Organization (WHO), it is estimated that 9.6 million people worldwide died of cancer in 2018 (https://www.who.int/cancer/en/). 30- 50% of these were preventable with early diagnosis. Between types of cancer, brain tumor is one of the deadliest ones. According to statistics, it is estimated that 17,760 adults will die from brain tumors in 2019 [2]. Basically, brain tumors are categorized as malignant and benign tumors, which grows abnormally in the brain. Malignant tumors contain cancerous cells which grows to all the parts of the brain due to non-uniform structure. Benign tumors are of uniform structure and contains non-cancerous cells. When the brain tumor tissues grow slowly, it causes vision problem, vomiting tendency etc. to the patients. The complex structure of tumor diagnosis becomes a challenging task for the clinical doctors. In the study of biomedical imaging, magnetic resonance image plays a vital role of acquisition of brain tumors. Magnetic resonance imaging (MRI) with high resolution becomes most popular imaging techniques in the hospitals. The manual diagnosis of the tumors by visualizing the magnetic resonance images in the clinic becomes a tedious and time-consuming task for doctors due to complex structure of the tumor and noise involvement in magnetic resonance (MR) imaging data. It is impossible to transform the image to a desired simpler condition by manual detection, which motivates us to propose a platform for automatic detection and classification of brain tumor from magnetic resonance image. So, detection of tumor location and identification at earlier stage is essential to reduce the tumor related deaths. At the same time the classification of brain tumor is also essential to know the types of tumor present in the brain. By utilizing the

segmentation and classification techniques, doctors can track and predict the uncontrollable growth of cancer affected areas at different levels to provide suitable diagnosis at early stage.

Segmentation of image from the magnetic resonance images is a consequential and arduous task for detection of brain tumor tissues. It becomes a challenging task due to the involute structure and variations in images. The different segmentation techniques such as FCM based genetic algorithm [3,4], semi supervised learning with graph cuts [5], Berkele Wavelet transform [6,7], Graphcut algorithm with co-segmentation for identification of exact cut point between edema and tumor[8], region growth segmentation[9], K-means clustering[10], spectral clustering[11,12], wavelet transform image segmentation[13,14], hidden markov random field models [15] etc has been proposed for brain tumor segmentation.

1.2 Statement of the Problem

Detection of solid tumors, whether it is benign or malignant, is often difficult in brain MR images. Identifying the exact size and coverage of tumors is also challenging in brain MRI because the original medical image has the problem like noise, low contrast, and bad resolution and so on. So, it takes minutes to get MR images of a subject and more time to view the images generated on a screen and carry out visual assessment.

Visual assessment of the MR images is subjective, often time consuming and hardly repetitive which might give rise to inaccurate diagnosis. Once MRI shows that there is a tumor in the brain the most common way to determine the type of brain tumor is to look at the result from a sample of tissue after biopsy or surgery. There is no mechanism that detects tumors and classifies the tumors as malignant or benign in MR. This could be avoided if there exist a tool that could be used for accurate detection of abnormalities in the brain automatically by clever analysis of the MR images non-invasively. This calls for the development of methodology which could be used for effective analysis of the MR images that could robustly detect and classify brain tumors.

1.3 Objectives

1.3.1 General Objective

The objective of the research work is to detect brain tumor tissues from MR (Magnetic Resonance) images using proposed FRFCM (Fast and Robust Fuzzy C Means algorithm) segmentation technique and classification through novel extreme learning machine model.

1.3.2 Specific Objectives

- Developing a novel FRFCM (Fast and Robust Fuzzy C Means algorithm) segmentation technique to remove noise and detect tumor from MR Images.
- Classification of the normal and abnormal brain tumor tissues using proposed PSO based Extreme learning machine model.
- GUI for automatic visualization of detection and classification through Using Raspberry Pi3.

1.4 Significance of the study

The growing incidence of brain tumors increases the number of images that need to be reviewed by oncologists/radiologists. In addition, particularly in a low resource setting, the high cost of examinations and lack of trained experts hinder patients from receiving the service. In this regard, automating brain tumor detection could deliver many potential benefits. In a screening setting, it allows the examination of a large number of images objectively in less time than current observer driven techniques which are subjective. It can also be an important diagnostic aid and can reduce the workload of trained graders, thereby reducing costs inside clinics.so that this study provides help to radiologist, doctors and surgeons in diagnoses of disease in very short time and with the high accuracy.

1.5 Project Report Organization

Chapter 1 presents a brief background of brain tumor research statistical report and contribution, critical literature review on research work reported on different segmentation and classification techniques of brain tumor form magnetic resonance images, objective and significance of the study followed by statement of the problem.

Chapter 2 presents details of brain tumor and case studies of different patients.

Chapter-3 presents details of conventional Image Segmentation techniques and classification models

Chapter 4 presents the process of segmentation for reduction of noise and smoothening of brain tumor magnetic resonance image an improved fast and robust fuzzy c means algorithm (FRFCM) segmentation algorithm has been presented. Further, an automatic classification of tumor tissues by the proposed hybrid PSO-ELM (Particle swarm optimization based Extreme

learning machine) model. The weights of the ELM (Extreme learning machine) model has been optimized by PSO algorithm. Also, the position equation of the PSO algorithm has been improved to get the faster weight optimization of the ELM model which leads to unique classification of brain tumor tissues.

Chapter 5 presents data collection, results and discussion on classification and segmentation techniques. Further the comparison results with different segmentation technique, and classification models are presented.

Chapter 6 presents the conclusion for the research work followed by reference. It also focuses on contributions of the project and future scope.

CHAPTER TWO

2. LITERATURE REVIEW

2.1 Review on segmentation and Classification

In recent days researchers proposed image segmentation techniques based on FCM algorithms. Segmentation method is predicated on a rudimental region growing method and uses membership grades of pixels to relegate pixels into felicitous segments. From the literature survey, it is found that, FCM algorithm has ability to obtain texture and background information from the simple images but failed in the case of complex noisy images where spatial information's are not considered. To overcome this problem, Ahmed et al. [18] proposed with spatial information (FCM_S) by considering intensity inhomogeneity, but the disadvantages is that during computation the term spatial neighbour is considered in each epoch which takes more computational time. Further to reduce the computational burden, the spatial neighbourhood information has been proposed by Chen and Zhang [19] with spatial term. There are two variants FCM_S1 and FCM_S2 and both are not able to reduce Gaussian noise. Szilagyi et al. [20] proposed enhanced FCM algorithm (EnFCM) by using gray levels lesser value than the size of the image to reduce the time complexity burdens due to the spatial term. In EnFCM the parameter α (adjustable) plays a vital role for the improvement of segmentation results. To reduce noise and guarantee the detail-preservation of segmentation, the FGFCM (Fast generalized FCM algorithm) has been proposed by Cai and Chen [21] with a similarity measure factor which improves the robustness of FCM, but FGFCM requires more parameters than the EnFCM. The fuzzy local information c-means clustering algorithm (FLICM), replaces the parameter α by a fuzzy factor in the objective function to delineate the noise and detail image preservation which is proposed by Krinidis and Chatzis [22]. For the development of a novel FCM algorithm, the algorithm should be free from selection of parameter. In this contrast FLICM, improves the segmentation process, but not free from parameter selection and also not perfect for local information of images and fails to remove Gaussian noise beyond 30%. To enhance the performance of FLICM Gong et al. [23] proposed kernel metric with local information for noise reduction which is free from selection of parameter. FCM with local information and kernel metric (KWFLICM) proposed by Gong et al. [24] to reduce noise and improve the robustness of FLICM. Guo et al. [25] proposed adaptive FCM algorithm based on

noise detection (NDFCM), which is faster since the image filtering is done before starting of iterations. In this, the trade-off parameter has been tuned automatically by considering the local variance of image grey levels. The FLICM technique employs more parameter, and fails to reduce Gaussian noise, salt & pepper noise and uniform noise from the synthetic image beyond 30%.

To ameliorate the robustness to noise and avoid parameter, a fast and robust FCM (FRFCM), which is based on membership filtering and morphological reconstruction has been proposed. FRFCM uses weiner filtering to the membership partition matrix for noise reduction and improvement of segmentation accuracy. The local spatial information of images has been added into FRFCM by employing morphological reconstruction operation to guarantee noise level reduction and image detail-preservation. To improve the rician noise removal and detection capability of FLICM based algorithm, an improved fast and robust FCM algorithm has been proposed by improving the membership partition matrix and employing a wiener filter to the membership partition matrix of the fast and robust FCM algorithm. The fuzzy membership value of the pixel has been updated to maintain image precision and smoothening of image. The proposed segmentation algorithm can provide good segmentation results for magnetic resonance images with high segmentation precision. We have taken eight previously proposed FCM based algorithms for the comparison with our proposed FRFCM segmentation technique.

While considering for classification of brain tumors from magnetic resonance images, the classifiers such as SVM (Support vector machine) with weighted kernel width achieves an accuracy of 89.92% for cancerous tumors proposed by Rezaei and Agahi [26]. Torheim et.al [27], uses texture features, SVM and achieved a classification accuracy of 87% for 3DMR images. Ruixuan Lang et al. [28] uses convolutional neural networks (CNNs) for 3D MR images and achieves the dice accuracy of 80% , Deepa et al. [29] proposed Extreme learning machine and genetic algorithm for classification of tumors and achieves 97.55% accuracy etc. The hybrid model PSO (Particle swarm optimization) based LLRBFNN Algorithm for automatic brain tumor detection proposed by Krishna [30] . There are different hybrid models such as LLRBFNN model with teaching learning-based optimization (TLBO) weight optimization [31, 32] has been used for power signal classification, financial forecasting etc. Also, the particle swarm optimization [33,34,35,36] technique which is based on bird flocking used in controller design, computational intelligence etc. Further the hybrid optimization algorithm PSO-

GA(PSO-Genetic Algorithm) proposed by Garg[37] and de Fátima Araújoa and Uturbey [38] uses PSO-DE(PSO-Differential Evolution) for performance assessment to the dispatch of generation and demand , Dipankar Santra et al. [39] uses PSO-ACO(PSO-Ant colony optimization) to solve economic load dispatch problem. It is observed from the literature survey that, there is not enough research work on automatic detection and classification of brain tumor utilizing hybrid biologically inspired hybrid optimized models. In this research work, we are proposing a modified PSO algorithm to optimize the weights of multiclass ELM model. With the modification in PSO algorithm, the performance of the classifier also increases. In PSO hybrid algorithm, it is observed that still the parameters of the PSO algorithm needs proper tuning for maximizing the performance of optimization. In this research work, the learning parameter of PSO algorithm has been modified by a inverse exponential factor which leads to new position and velocity equations which are utilized for optimization of weights of ELM classifier model.

All mentioned segmentation algorithms aim to partitioning of data into clusters by minimizing the objective function and fails drastically remove noise from the magnetic resonance images. To improve the noise reduction capability and smoothening of magnetic resonance images, a novel improved FRFCM segmentation algorithm is proposed. The improved FRFCM technique is utilized to remove rician noise from the magnetic resonance image in this research work.

Further, the classifiers such as support vector machine (SVM), probabilistic neural network (PNN), are some of the popular classifiers has already been used for classification of brain tumors. Due to the complex mathematical calculations and higher computational time requirements in the mentioned classifiers, we are motivated to propose a hybrid modified PSO based multiclass ELM classification model for classification of brain tumors from magnetic resonance images. The proposed research work presents a novel image segmentation and classification technique for automatic detection and classification of brain tumor from magnetic resonance images.

The research work focuses on two contributions based on segmentation and classification. The contributions are summarized as follows:

- In first aspect, an improvement to fast and robust FCM based segmentation algorithm has been proposed. The improvement has been made to the membership partition matrix of the FCM [16] algorithm and a wiener filter is employed to improve the rician noise

reduction capability. Further, the fuzzy membership value of the pixel has been updated to maintain image precision and smoothening of the magnetic resonance brain tumor images. The complete mathematical derivation has been developed for the new IFRFCM segmentation algorithm to maximize the segmentation accuracy.

- In the second aspect, the modifications to the parameters of particle swarm optimization (PSO) [17] algorithm has been proposed to maximize the optimization performance of the hybrid algorithm. The convergence parameter, position and velocity equations in the algorithm has been modified and the mathematical calculations with modification has been developed. With the modified convergence parameter, new velocity and position equations, the proposed PSO algorithm has been employed to optimize the weights of the ELM model for classification of brain tumor.

2.2 Brain tumor Magnetic resonance image Overview

The complex structure of the human brain complicates the diagnosis of the tumor in the brain region. MRI, a useful method for obtaining high quality brain images, is widely utilized for tumor diagnosis. Especially in brain imaging, the MRI method provides a unique appearance in visualization of soft tissues with spatial resolution and contrast resolution. Magnetic resonance imaging (MRI) is a useful method for diagnosis of tumors in human brain. In this work, MRI images have been analysed to detect the regions containing tumor and classify these regions into three different tumor categories: meningioma, glioma, and pituitary.

As per the Myoclonic Gliomas [40] is a type of tumors begin in the brain or spinal cord and include astrocytomas, ependymomas, glioblastomas, oligoastrocytomas and oligodendrogliomas. Meningiomas is a type of tumor that arises from the membranes that surrounds brain and spinal cord (meninges). Most meningiomas are noncancerous. Acoustic neuromas (schwannomas) are benign tumors that develop on the nerves that control balance and hearing leading from your inner ear to your brain. Pituitary adenomas are mostly benign tumors that develop in the pituitary gland at the base of the brain. These tumors can affect the pituitary hormones with effects throughout the body. Medulloblastomas are the most common cancerous brain tumors in children. A medulloblastoma starts in the lower back part of the brain and tends to spread through the spinal fluid. These tumors are less common in adults, but they do occur. Also the Craniopharyngiomas are the rare, noncancerous tumors start near the brain's pituitary gland, which secretes hormones that control many body functions. As the craniopharyngioma slowly

grows, it can affect the pituitary gland and other structures near the brain. Secondary (metastatic) brain tumors are tumors that result from cancer that starts elsewhere in the body and then spreads (metastasizes) to the brain. According to the medical practitioner, if it is not operated in advance, the chances of survival becomes difficult for a tumor affected patient.

Brain tumor is incited because of an ordinary boom of cells [41]. It comes under two categories that are malignant or benign. The benign tumors do not contain cancerous cells while the malignant tumors are comprised of cancerous cells [42, 43, 44]. Brain tumors associations [45] classify the brain tumor in four grades where grade I and II are referred to as benign and the remaining III and IV are labeled as malignant. The development rate of benign is relatively low when contrasted with malignant. When the benign isn't dealt with opportune then it is changed into malignant tumor. Along these lines, early discovery of tumor is desirable [46]. Grade II patients require ordinary treatment and checking through magnetic resonance imaging (MRI) [47, 48]. MRI depicts one of radio imaging types. Some more types include computed tomography (CT) and positron emission tomography (PET) [49] with the help of which brain tumor can be diagnosed in medical imaging [50]. MRI is mostly useful to analyze tumors in general clinical routines. It provides minute information about human cerebrum. One advantage of MRI is its non-intrusive nature and no ionization radiation.

GLIOMAS are the brain tumors with the highest mortality rate and prevalence [51]. These neoplasms can be graded into Low Grade Gliomas (LGG) and High Grade Gliomas (HGG), with the former being less aggressive and infiltrative than the latter [52], [53]. Even under treatment, patients do not survive on average more than 14 months after diagnosis [54]. Current treatments include surgery, chemotherapy, radiotherapy, or a combination of them [55]. MRI is especially useful to assess gliomas in clinical practice, since it is possible to acquire MRI sequences providing complementary information [56]. The accurate segmentation of gliomas and its intra-tumoral structures is important not only for treatment planning, but also for follow-up evaluations. However, manual segmentation is time-consuming and subjected to inter- and intra-rater errors difficult to characterize. Thus, physicians usually use rough measures for evaluation [57]. For these reasons, accurate semi-automatic or automatic methods are required. However, it is a challenging task, since the shape, structure, and location of these abnormalities are highly variable. Additionally, the tumor mass effect changes the arrangement of the surrounding normal tissues [58]. Also, MRI images may present some problems, such as intensity inhomogeneity [59], or different intensity ranges among the same sequences and acquisition scanners [60].

An important role is played by Medical imaging in current medical research and clinical practice. Tumor detection mainly involves MRI, Computed tomography (CT) and ultrasound images. Precise extraction of

tumor is necessary and advantageous. When we have huge number of sequence of images, the manual observation of tumors is time-consuming, and frequently, the depiction quality depends on the operators. So, for clinical analysis, medical image segmentation has gained much attention and different segmentation methods have been proposed. Imaging techniques allow medical practitioners and researchers, even before performing invasive surgery, to assess activities and disorders in the human brain. The technique magnetic resonance imaging (MRI) is of good quality among other techniques as it gives better contrast details about the brain tissues from a variety of image sequences. The neurosurgeons and medical scientists have great opportunity with the increase in brain MR image data. The diagnosis using computer aids and treatment can be administered through the analysis and processing of image data. Neurosurgeons normally require accurate depiction of the tumor before diagnosing the tumor and performing the surgery. Doctors/ specialists perform manual brain tumor segmentation, but its time-consuming. Hence, to design and develop automated brain tumor segmentation methods, various researchers have been working continually. Gliomas are the most usual and vigorous brain tumors, which in their highest grade can lead to a short life expectancy. Thus, the key stage to boost the life quality of patients is planning the treatment. The automatic segmentation of Magnetic Resonance brain images can be enacted through a extreme learning machine (ELM). The medical image analysis field has made ELMs very popular. On comparison to old machine learning methods, there is no requirement of hand-crafted features in ELMs which are a requisite for classification. Instead, the kernel sets are learned that are specially trained for the problem of classification. The old machine learning methods have been using kernels like Gaussian or Haar-like for getting the occurrence detail, but ELM pick up the kernel set based on the training data as provided.

2.3 Case studies of the brain tumor

The case studies of the brain tumor has been taken from the different agencies such as

- American brain tumor association
- http://www.med.harvard.edu/aanlib/cases/case33/case.html
- St. Louis University

(i) Case study-1

A 73 year old right-handed man sought medical attention because of a grand mal seizure and progressive difficulty with speech. Brain biopsy in 3/95 revealed grade II astrocytoma; however, because of his age and speed of progression, the pathological result was thought to be artifactually benign and due to sampling error. He underwent conventional whole brain radiation consisting of 6440 cGys given in 35 treatments. He tolerated his radiation therapy well but attempts to wean his steroids resulted in prompt seizure recurrence. Concern of tumor recurrence

vs. radiation necrosis prompted evaluation with PET. Case contributed by Drs. C. D. Sturm and R. Bucholz, St. Louis University.

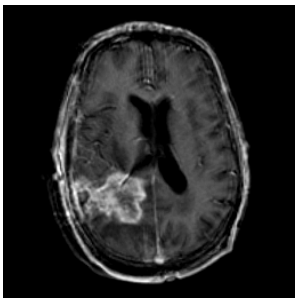

Fig. 2.1 Astrocytoma Tumor

(ii) Case study-2

This 35 year old man began having headaches eight months previously. A CT scan was unremarkable. Two months prior to the current studies, he underwent brain biopsy, after progressive symptoms led to MR imaging. A low grade glioma was found, and he was treated with radiation. The MR demonstrates an area of mixed signal intensity on proton density (PD) and T2-weighted (T2) images in a left occipital region. Contrast enhancement shows the lesion to contain cystic elements. Thallium images show an anterior border of high uptake, consistent with a small region of tumor recurrence.

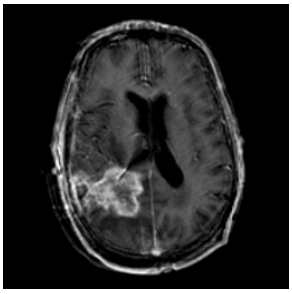

Fig. 2.2 T2 Astrocytoma

(iii) Case study-3

This 62 year old man came to the emergency room after suffering a first seizure, a tonic-clonic convulsion with focal onset, witnessed by his wife. He suddenly became quiet, had eye deviation to the left, and began to twitch in the left face. He became unresponsive to verbal commands, and had generalized jerking movements of arms and legs, lasting for a few minutes. There was a history of carcinoma of the colon, with recent metastasis to the liver and lung. MR images show

a lesion involving the right second frontal convolution and another in the cerebellum, near the fourth ventricle, also visible on the sagittal image map.

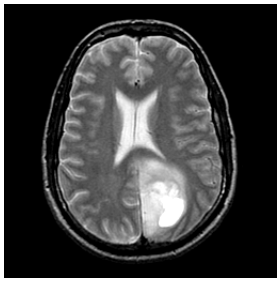

Fig. 2.3 Metastatic Carcinoma of the colonatic

(iv) Case study-4

A 51 year old woman sought medical attention because of gradually increasing right hemiparesis (weakness) and hemianopia (visual loss). At craniotomy (8/90), left parietal anaplastic astrocytoma was found. A right frontal lesion was biopsied in 8/94. Recurrent tumor was suspected on the basis of the imaging and was confirmed pathologically.

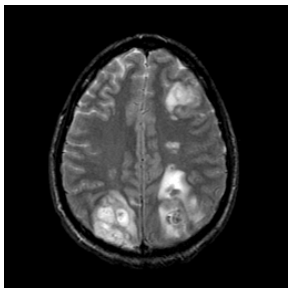

Fig. 2.4 Anaplastic Astrocytoma

(v) Case study-5

The patient was a 22 year old man who was admitted for resection of Ewing's sarcoma (peripheral/primitive neuroepithelialtumor- PNET). Vaguely described visual difficulty was noted retrospectively to have begun approximately one month prior to admission. He underwent thoracotomy for pulmonary metastasis, and was noted after recovery to have trouble with concentration and increasing visual difficulty. On exam, he was inattentive, confused and had a right homonomous hemianopia, a left inferior quadrantanopia, right lower extremity hyperreflexia, and right extensor plantar response.

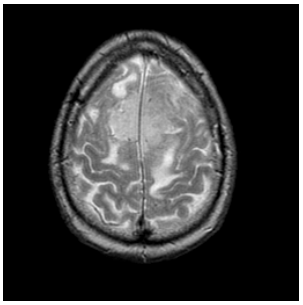

Fig. 2.5 Sarcoma

(vi) Case study-6

The patient was a 75 year old man who had an 8 - 10 month history of progressive difficulty walking. He had noted some left lower extremity weakness and some difficulty with memory and concentration. He was alert and oriented, but had slow and hesitating speech. He could recall only 1 of 3 objects at five minutes. Cranial nerves were normal. Muscle stretch reflexes were active and equal. Babinski signs were negative. Gait was slow, broad based, and tandem walking was ataxic. He underwent craniotomy and tumor resection.

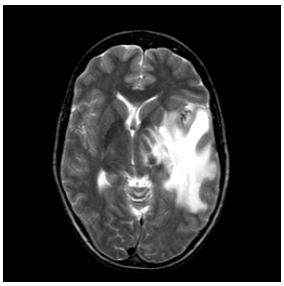

Fig. 2.6 Meningioma

(vii) Case study-7

This 42 year old woman with a long history of tobacco use began having headaches one month before these images were obtained. The headaches intensified during the previous week, and she noted trouble finding her words.

Brain images show a large mass with surrounding edema, and compression of adjacent midbrain structures. The MR demonstrates the tumor as an area of high signal intensity on proton density (PD) and T2-weighted (T2) images in a large left temporal region. Contrast enhancement shows the lesion to contain a cystic component here.

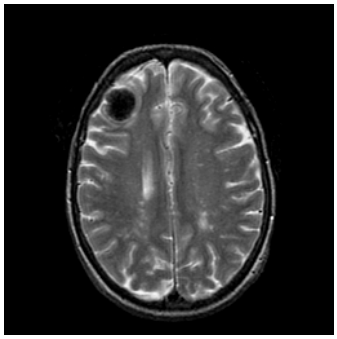

Fig. 2.7 Metastatic bronchogenic carcinoma

Some brain tumor data MR Images are collected from ADNI data set are Benign type as follows.

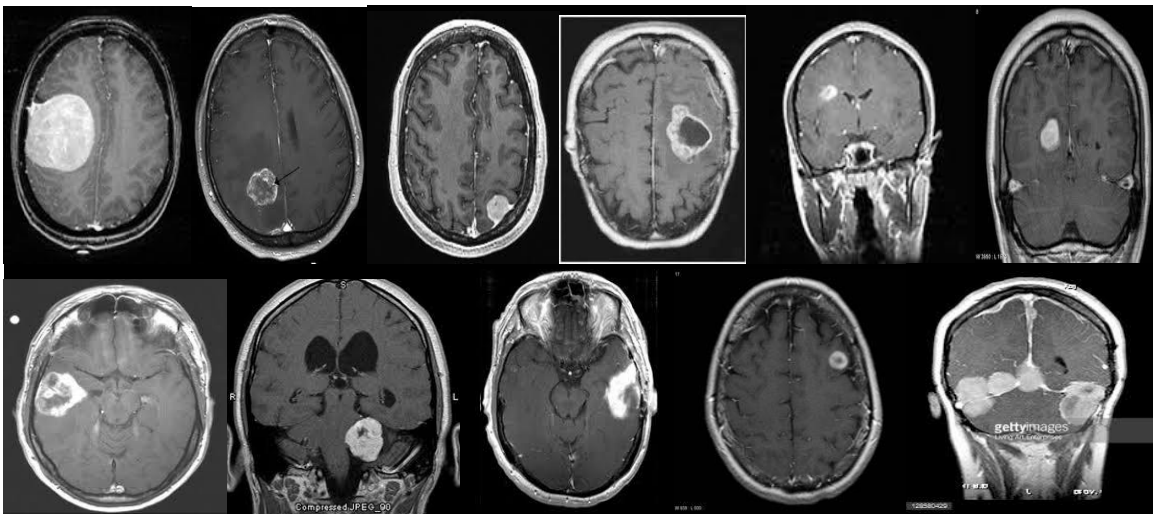

Fig. 2.8. Brain tumor MR images Benign type

The brain tumor dataset belongs to Malignant type and are collected from ADNI data base.

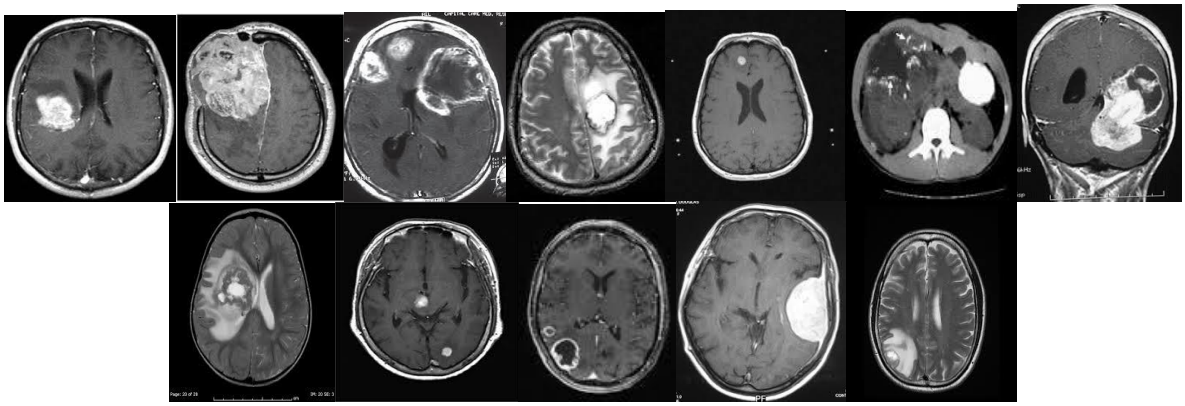

Fig. 2.9 Brain tumor MR images Malignant type

2.4 Segmentation Techniques

There are several FCM based segmentation techniques proposed by the researchers. The details of the segmentation techniques are mentioned in the following section.

2.4.1 Fuzzy C Means Algorithm

In FCM, it is possible for a data sample to belong to multiple clusters at the same time. The similarity is indicated by the membership value. In FCM a data sample is assigned with a membership value based on its similarity with the cluster center. The membership values are between 0 to 1 and more the similarity, higher the membership value.

Defuzzification is applied at the end of the clustering process to decide the clustering. FCM is a repetitive algorithm and the solution is achieved by repetitively updating the cluster center and membership value. These updating equations are obtained by solving the cost function. Let $X = \{x_1, x_2, x_3 \ldots\ldots x_N\}$ denotes the data with N data samples. It has to be partitioned into c-clusters by minimizing the subsequent cost function.

In Fuzzy C means clustering we determine the cluster center v_i and the membership matrix U and we thus determine distinct clusters. Fuzzy C Means [62]-[66] method is based on minimization of the following objective function:

$$J_m = \sum_{k=1}^{N} \sum_{i=1}^{C} u_{ik}^{m} \left\| x_k - v_i \right\|^2 \tag{1}$$

Where m=2, fuzziness coefficient, u_{ik} is the degree of membership of x_k in cluster i, x_k is the i_{th} of n-dimensional measured data, v_i is the n-dimensional center of the cluster. $\|.\|$ is a norm metric and 'm' is a constant. The parameter 'm' decides the fuzziness of the consequential partition. By taking the derivative of the equation and make it equal to zero by using Lagrange method, the following equations are achieved

$$v_i = \frac{\sum_{k=1}^{N} u_{ik}^{m} . x_i}{\sum_{k=1}^{N} u_{ik}^{m}}, \quad u_{ik} = \sum_{k=1}^{C} \left[\frac{\left\| x_k - v_i \right\|}{\left\| x_k - v_k \right\|} \right]^{-2/m-1} \tag{2}$$

Many variations to the FCM have been proposed because conventional FCM couldn't perform well in the presence of noise and intensity inhomogeneity.

2.4.2 Fast Fuzzy C Means Algorithm

According to Fast fuzzy c means algorithm[67], Let $X = [x_1, x_2 \ldots\ldots x_n]$ be a N sample data set and assume that each sample x_k is represented by a set of p features and U is the hard partition matrices whose general term is given by $u_{ik} = 1$ if $x_k \in X_i$, and 0 otherwise. To get partition matrix, the HCM (Hard c Means) algorithm is chosen which minimizes the objective function

$$j = \sum_{k=1}^{N} \sum_{i=1}^{L} u_{ik}^{m} \| x_k - C_i \|^2 \tag{3}$$

Where "L" is the number of clusters and C_i is the cluster center and "m" is the Fuzzifier exponent and $u_{ik} \in [0,1]$.

In many real situations, overlapping clusters reduce the effectiveness of crisp clustering methods. Ruspini first proposed the notion of fuzzy partition [68], where samples may partially belong to several clusters through the idea of partial membership degrees.

Minimization of equation (3) is obtained by an optimization technique that successively updates the cluster centers C_i and partition matrix U by using the formula

$$C_i = \frac{\sum_{k=1}^{N} u_{ik}^{m} x_k}{\sum_{k=1}^{N} u_{ik}^{m}} \tag{4}$$

$$\text{And } u_{ik} = \frac{1}{\sum_{j=1}^{L} \frac{\| x_k - C_i \|}{\left(\| x_k - C_j \| \right)^{2/(m-1)}}} \tag{5}$$

The choice to first initialize a random partition matrix or the cluster centers is let to the user, both being used in the literature. The algorithm stops when the centroids stabilize, i.e. the matrix norm between two successive V is below a given threshold. Equivalently, the entire procedure can be shifted one half cycle, so that initialization and termination is done on U. Naturally, in terms of speed and storage, there are some advantages to initialize and terminating with V. Application to image segmentation consists in taking X as the entire set of pixels X_k of an image I, each of them being described by P features.

2.4.2.1 Centroids initialization:

As defined, iterative fuzzy clustering methods do not guarantee a unique final partition because different results are obtained with different initializations of V (or U). In particular, it has been

shown that these algorithms give better results when the initials (U/V) are sufficiently close to the final partition/centers [69]. However, most of the practitioners initialize in a random manner, which heavily affects the results. Another reason to correctly initialize the cluster centers is that it allows to speed up the convergence, resulting in a more usable algorithm for large scale practical problems. Several methods have been proposed for the initialization of V. However, most of the methods that allow to initialize cluster centers are computationally expensive. For instance, in [70], the method requires to run the HCM algorithm p times on n 1-dimensional samples, and then p times the same algorithm on n p-dimensional samples, which is intractable for large scale data sets.

In this research work we propose a new, efficient, yet simple, manner of initializing the 'c' cluster centers that we call Ordering-split. For each p-dimensional sample x_k, we define its relative mean by

$$m_k = \frac{1}{p}\sum_{j=1}^{p} x_{kj} \tag{6}$$

so that we obtain the n-dimensional vector $m = (m_1, m_2, \ldots m_n)$. Note that we are working on features coming from each channel of an image so that the scale of individual features does not differ. If the features do not hold this property, a normalization is required. Let σ be the permutation function σ such that $m_\sigma(k)$ is an ordered and increasing sequence. We propose to split the n relative means as follows.

Assuming that the clusters are equally distributed, we uniformly split the n-dimensional vector m into c groups. In other terms, we set $c+1$ indices, say $l_0, l_1 .. l_c$, such that the c differences $(l_i - l_{i-1})$ are roughly equal. More formally, each index is given by

$$l_i = i * \lfloor n/c \rfloor \tag{7}$$

Where $\lfloor . \rfloor$ is the floor function. We iteratively build c subsets S_i of n as follows

$$S_i = \{l_{i-1} + 1, \ldots . l_i\}$$

We obtain the subset of indices in each cluster by applying the inverse function:

$$C_i = \sigma^{-1}(S_i) \tag{8}$$

Finally, each cluster center is computed using:

$$v_i = \frac{1}{|C_i|}\sum_{j \in C_i} x_j \tag{9}$$

Where $|C_i|$ is the cardinality of C_i.

2.4.3 Fuzzy clustering with spatial constraints (FCM_S) and its variants

Ahmed et al. [18] proposed a modification to the standard FCM, in that the neighbourhood effect acts as a regularizer and the modified objective function of FCM_S is defined as follows:

$$J_m = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^m \|x_k - v_i\|^2 + \frac{\alpha}{N_R}\sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^m \sum_{r \in N_k}\|x_r - v_i\|^2 \tag{10}$$

where x_k is the gray value of the k^{th} pixel, v_i represents the prototype value of the i^{th} cluster, u_{ik} represents the fuzzy membership of the kth pixel with respect to cluster i, NR is its cardinality, x_r represents the neighbour of x_k and N_k stands for the set of neighbours falling into a window around x_k. The parameter 'm' is a weighting exponent on each fuzzy membership that determines the amount of fuzziness of the resulting classification. The parameter α is used to control the effect of the neighbours term. By definition, each

sample point x_k satisfies the constraint that $\sum_{i=1}^{c} u_{ik} = 1$. Two necessary but not sufficient conditions for J_m to be at its local extreme will be obtained as follows:

$$u_{ik} = \frac{\left(\|x_k - v_i\|^2 + \frac{\alpha}{N_R}\sum_{r \in N_k}\|x_k - v_i\|^2\right)^{-\frac{1}{m-1}}}{\sum_{i=1}^{c}\left(\|x_k - v_i\|^2 + \frac{\alpha}{N_R}\sum_{r \in N_k}\|x_k - v_i\|^2\right)^{-\frac{1}{m-1}}} \tag{11}$$

$$v_i = \frac{\sum_{k=1}^{N} u_{ik}^m\left(x_k + \frac{\alpha}{N_R}\sum_{r \in N_k} x_r\right)}{(1+\alpha)\sum_{k=1}^{N} u_{ik}^m} \tag{12}$$

The second term $\sum_{r\in N_k} x_r \Big/ N_R$ in the numerator of (12) is in fact a neighbour average grey value around x_k, the image composed of all the neighbour average values around all the image pixels forms a so-called local neighbour average image or equivalently mean-filtered image. A shortcoming of (11) and (12) is that computing the neighbour term will take much time in each iteration step.

In order to reduce the computation, Chen and Zhang [19,20] proposed a variant of FCM_S, FCM_S1, which simplified the neighbourhood term of FCM_S. And the low-complexity objective function can be written as follows:

$$J_m = \sum_{i=1}^{c}\sum_{k\in N_k} u_{ik}^m \left\|x_k - v_i\right\|^2 + \alpha \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^m \sum_{k\in N_k} \left\|\overline{x_k} - v_i\right\|^2 \tag{13}$$

Where $\overline{x_k}$ is a means of neighbouring pixels lying within a window around x_k. An iterative algorithm for minimizing (18) with respect to u_{ik} and v_i can similarly in FCM_S be derived, as:

$$u_{ik} = \frac{\left(\left\|x_k - v_i\right\|^2 + \alpha\left\|\overline{x_k} - v_i\right\|^2\right)^{-\frac{1}{m-1}}}{\sum_{i=1}^{c}\left(\left\|x_k - v_i\right\|^2 + \alpha\left\|\overline{x_k} - v_i\right\|^2\right)^{-\frac{1}{m-1}}} \tag{14}$$

$$v_i = \frac{\sum_{k=1}^{N} u_{ik}^m \left(x_k + \alpha\overline{x_k}\right)}{(1+\alpha)\sum_{k=1}^{N} u_{ik}^m} \tag{15}$$

When α is set to zero, the algorithm is equivalent to the original FCM, while when it approaches infinite, the algorithm acquires the same effect as the original FCM on the mean- or median-filtered image, respectively.

2.4.4 Enhanced Fuzzy C Means algorithm

L.Szilágyi et al. [21] proposed the EnFCM algorithm to speed up the segmentation process for gray level image. In order to accelerate FCM_S, a linearly weighted sum image ξ is in advance formed from the original image and its local neighbor average image in terms of:

$$\xi_k = \frac{1}{\alpha}\left(x_k + \frac{\alpha}{N_R}\sum_{j\in N_k} x_j\right) \tag{16}$$

Where ξ_k denote the gray value of the k[th] pixel of the image ξ, x_j represents the neighbors of x_k, N_k stands for the set of neighbors falling into a window around x_k. Concretely, the objective function used for fast segmenting the newly-generated image ξ is defined as:

$$J_s = \sum_{i=1}^{c}\sum_{l=1}^{q}\gamma_l u_{il}^m (\xi_l - v_i)^2 \tag{17}$$

where v_i represents the prototype of the i[th] cluster, u_{il} represents the fuzzy membership of gray value l with respect to cluster i. q denote the number of the gray levels of the given image which is generally much smaller than N. γ_l is the number of the pixels having the gray value equal to l, where l=1,.., q. Naturally, we have

$$\sum_{i=1}^{q}\gamma_l = N \tag{18}$$

And under the constraint that $\sum_{i=1}^{c} u_{il} = 1$ for any l, minimize J_s, specifically, taking the first derivatives of J_s with respect to u_{il} and v_i, and zeroing them, respectively, two necessary but not sufficient conditions for J_s to be at its local extrema will be obtained as follows

$$u_{il} = \frac{(\xi_l - v_i)^{-\frac{2}{m-1}}}{\sum_{j=1}^{c}(\xi_l - v_i)^{-\frac{2}{m-1}}} \tag{19}$$

$$v_i = \frac{\sum_{i=1}^{q}\gamma_l u_{il}^m \xi_l}{\sum_{l=1}^{q}\gamma_l u_{il}^m} \tag{20}$$

EnFCM provides comparable segmenting quality in considerably fast manner for the brain image used there to FCM_S, but the segmenting quality depends on the chosen window size, the parameter α and the filtering method. The value of α has to be chosen large enough so that it can eliminate noise, on the other hand, it also has to be chosen small enough so that the after-segmented image does not lose much of its sharpness and details.

2.5 K- Means Algorithm and process of implementation

Currently the clustering method often used for segmenting large-scale images. Clustering is one of the unsupervised learning method in which a set of essentials is separated into uniform groups. There are different types of clustering: hierarchical clustering, Fuzzy C-means clustering, K-Means clustering. The K-Means method is one of the most generally used clustering techniques for various applications [72,73,74]. K-Means clustering is a partition-based cluster analysis method. The K-means clustering technique is a widely used approach that has been applied to solve low-level image segmentation tasks. The choosing of initial cluster centers is very important since this prevents the clustering algorithm to producing incorrect decisions.The *k*-means clustering algorithm is one of the widely used data clustering methods where the datasets having "*n*" data points are partitioned into "*k*" groups or clusters. The *k*-means grouping algorithm was initially proposed by Mac Queen in 1967 [75] and later enhanced by Hartigan and Wong [76]. Bottou and Bengio [77] demonstrated the merging lands of the K-Means calculations methods and algorithms methodologies. It has been indicated to be exceptionally handy for a corpus of commonsense provisions and widely used applications. The definitive *k*-means algorithm works with in-memory information, yet it could be effectively stretched out for out-of-memory occupant datasets.

K-Means is one of the relaxed unsupervised erudition algorithms that illuminate the well-known clustering issue [75]. The methodology trails after a straightforward and simple approach to group a given information set through a certain number of groups (expect *k* groups) that have been established beforehand. The principle idea is to characterize *k* centroids, one for each group. These centroids have to be set in a guile manner resulting in a distinctive area of diverse effects. The subsequent step is to take each point within a given information set and copartner it with the closest centroid until reaching a state where all the points have been associated with a group. Once the first stage is done and an unanticipated aggregating is carried out automatically, we need to reconfigure *k* new centroids as barycenter of each group due to the last step. After producing these *k* new centroids, another binding must be established between the same centroids set and the closest new centroid. A cycle will be produced. As an after effect of this cycle, we may recognize that the *k* centroids will change their regulated areas and at the end of

the day centroids will not change their positions anymore [77]. The objective function is shown below; and thus, the algorithm aims to reduce the squared error in this function:

$$J_m = \sum_{j=1}^{K} \sum_{i \in S_I} \left\| x_i - c_j \right\|^2 \tag{21}$$

The distance indicator of the n data set points is x_i and their represented centroids c_j are $\left\| x_i - c_j \right\|^2$. Methodically speaking, the k-means algorithm is composed of four steps and they are in the following order.

1) Randomly place k elements in a space representing the items coordinates that are being clustered.
2) Allocate each item in the space to a group that is the most similar to it.
3) After the assignment of all the items in the space, recompute the k centroid elements and change their positions, respectively.
4) Repeat steps (2) and (3) until the centroids reach a position where they no longer change with respect to the distances between all the elements of their group.

Ultimately, the technique will dependably end; however, the k-means calculation does not usually find the most optimal design, or at least comparing it to the minimum global objective function that has been stated. The algorithm is delicate and fundamentally dependent on the beginning arbitrarily elected group centers. The algorithm should run at different times to decrease this impact though it will choose a different set of centroid search time, making it very hard to compare its initial indications [76].

The K-Means [77] Clustering Algorithm starts by picking the number K of centres and randomly assigning the data points x_i to S_i subsets containing N_j data points that minimizes the cost function. It then uses a simple re-estimation procedure to end up with a partition of the data points into clusters containing N data points that minimizes the sum squared clustering function. The clustering process terminates when no more data points switch from one cluster to another based on minimization of the following objective function:

$$J_m = \sum_{j=1}^{K} \sum_{i \in S_I} \left\| x_i - c_j \right\|^2 \tag{22}$$

Where $c_j = \frac{1}{N_j} \sum_{i \in s_j} x_i$

Algorithm:

k-means. The *k*- means algorithm for partitioning, where each cluster's center is represented by the mean value of the objects in the cluster.

Input: *k*: the number of clusters, *D*: a dataset containing *n* objects.

Output: A set of *k* clusters.

Method: (1) randomly choose *k* objects from *D* as the initial cluster centers;

(2) Repeat

(3) (Re) assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster. (Re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;(Re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;

(4) Update the cluster means, i.e., calculate the mean value of the objects for each clusters; until no change;

2.6 Fuzzy Local Information C-Means

More recently, Cai et al. [20] proposed the fast generalized FCM algorithm (FGFCM) which incorporates the spatial information, the intensity of the local pixel neighborhood and the number of gray levels in an image. This algorithm forms a nonlinearly-weighted sum image from both original image and its local spatial and gray level neighborhood. The computational time of FGFCM is very small, since clustering is performed on the basis of the gray level histogram.

However, EnFCM as well as FGFCM, share a common crucial parameter α. This parameter is used to control the tradeoff between the original image and its corresponding ean- or median-filtered image. It has a crucial impact on the performance of those methods, but its selection generally difficult because it should keep a balance between robustness to noise and effectiveness of preserving the details.

In other words, the value of α has to be chosen large enough to tolerate the noise, and, on the other hand, it has to be chosen small enough to preserve the image sharpness and details. Moreover, the value of α is fixed for all pixel neighborhoods over the image. The Fuzzy Local Information C-Means clustering algorithm (FLICM) which supersedes the parameter α by a fuzzy factor in the objective function to delineate the noise and detail image preservation proposed by Krinidis and Chatzis [21].FLICM (Fuzzy Local Information C-Means), is a robust fuzzy local information cmeans clustering algorithm, which can handle the defect of the selection of parameter α as well as promoting the image segmentation performance. In FLICM, a novel fuzzy factor is defined to replace the parameter α used in EnFCM and FCM S and its variants, and the parameter α used in FGFCM and its variants.

The new fuzzy local neighborhood factor can automatically determine the spatial and gray level relationship and is fully free of any parameter selection. Thus, FLICM has the following attractive characteristics:

1) It is relatively independent of the types of noise, and as a consequence, it is a better choice for clustering in the absence of prior knowledge of the noise;

2) The fuzzy local constraints incorporate simultaneously both the local spatial and the local gray level relationship in a fuzzy way;

3) The fuzzy local constraints can automatically be determined, so there is no need of any parameter determination;

4) The balance among image details and noise is automatically achieved by the fuzzy local constraints, enhancing concurrently the clustering performance. All these characteristics make FLICM a more general and suitable for image clustering algorithm.

According to the FLICM [21], the fuzzy factor is given by

$$G_{kv} = \sum_{\substack{k \in N_v \\ v \neq k}} \frac{1}{d_{vk}+1} (1-u_{kv})^m \left\| x_v - v_k \right\|^2$$

By using the definition of G_{kv}, the Fuzzy Local Information C-Means (FLICM) clustering algorithm incorporates local spatial and gray level information into its objective function, defined in terms of:

The objective function of the fuzzy c means algorithm with local information is given by

$$J_s = \sum_{v=1}^{N}\sum_{k=1}^{c} u_{kv}^{m}\|x_v - v_k\|^2 + \sum_{v=1}^{N}\sum_{k=1}^{c} G_{kv} \tag{23}$$

Where the fuzzy factor is given by

$$G_{kv} = \sum_{\substack{k\in N_v \\ v\neq k}} \frac{1}{d_{vk}+1}(1-u_{kv})^m \|x_v - v_k\|^2 \tag{24}$$

Where the spatial Euclidean distance between pixels x_v and x_r is denoted by d_{vr}, N_v is the set of neighbours within a window around x_v and x_r represents the neighbours of x_v and u_{kr} is the neighbours of u_{kv}. With respect to cluster k, x_v is the gray value of the k^{th} pixel, u_{kv} represents the fuzzy membership value of the v^{th} pixel and N is the total number of pixels in the gray scale image $f = [x_1, x_2 .... x_N]$, x_v is the gray value of v^{th} pixel, c denotes the cluster centre and m determines the fuzziness of the consequential partition.

With the fuzzy factor G_{kv}, the capability of noise reduction improves. The fuzzy partition matrix is given by

$$u_{kv} = \frac{1}{\sum_{j=1}^{c}\left(\frac{\|x_v - v_k\|^2 + G_{kv}}{\|x_v - v_j\|^2 + G_{jv}}\right)^{\frac{1}{m-1}}} \tag{25}$$

And $v_k = \dfrac{\sum_{k=1}^{N} u_{kv}^{m} x_v}{\sum_{v=1}^{N} u_{kv}}$ (26)

Further the methods described in the algorithms have yielded effective clustering results for images, but still have some disadvantages:

1. Although the introduction of local spatial information enhances their insensitiveness to noise to some extent, they still lack enough robustness to noise and outliers, especially in absence of prior knowledge of the noise;

2. There is a crucial parameter α in their objective functions, used to balance between robustness to noise and effectiveness of preserving the details of the image. Generally its selection has to be made by experience or trial and error experiments;
3. They are all applied on a static image, which has to be computed in advance. Details of the original image could be lost depending on the method used to generate the new image.

2.7 Modified Fuzzy Local Information C Means (MFLICM) Algorithm

In order to overcome the above mentioned disadvantages, the fuzzy factor in objective function is modified to have some special characteristics:

1. To incorporate local spatial and local gray level information in a fuzzy way in order to preserve robustness and noise insensitiveness.
2. To control the influence of the neighborhood pixels depending on their distance from the central pixel.
3. To use the original image avoiding preprocessing steps that could cause detail missing.
4. To be free of any parameter selection.

Let the N sample data given by $X = \{x_1, x_2, x_3 ..... x_N\}$. Fuzzy C Means [14] algorithm the cost function minimized as :

$$J_m = \sum_{k=1}^{N} \sum_{i=1}^{c} u_{ik}^m \|x_k - v_i\|^2 \tag{27}$$

Where v_i is the cluster center, m is fuzziness coefficient, u_{ik} is membership matrix, x_k is the gray value of the k^{th} pixel. Ahmed et al. [18] introduced a parameter α as a neighbor and the modified objective function is given by

$$J_m = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^m \|x_k - v_i\|^2 + \frac{\alpha}{N_R} \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^m \sum_{r \in N_k} \|x_r - v_i\|^2 \tag{28}$$

Where x_r represents the neighbor of x_k and α controls the neighbors term. Chen and Zhang [6] proposed a variant by taking mean to reduce the complexity, and the objective function is given by

$$J_m = \sum_{i=1}^{c} \sum_{k \in N_k} u_{ik}^m \left\| x_k - v_i \right\|^2 + \alpha \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^m \sum_{k \in N_k} \left\| \overline{x_k} - v_i \right\|^2 \quad (29)$$

Where $\overline{x_k}$ is a mean of neighboring pixels lying within a window around x_k. For faster segmentation, L.Szilágyi et al. [19] proposed enhanced fuzzy c means EnFCM algorithm where image ξ is considered from the original image and is given by

$$\xi_k = \frac{1}{\alpha}\left(x_k + \frac{\alpha}{N_k} \sum_{j \in N_k} x_j \right) \quad (30)$$

Where q represents the number of gray level and less than N, the gray value of k^{th} pixel of image ξ is given by ξ_k, x_j is neighbors of x_k, N_k is set of neighbors around x_k. Now the new objective function is given by

$$J_s = \sum_{l=1}^{N} \sum_{k=1}^{c} \gamma_l u_{kv}^m \left\| \xi_l - v_k \right\|^2 \quad (31)$$

Where u_{il} represents the fuzzy membership of gray value *l*. γ_l is the number of the pixels having the gray value equal to l, and $l = 1,2,...N$

According to the FLICM [21], the fuzzy factor is given by

$$G_{kv} = \sum_{\substack{k \in N_v \\ v \neq k}} \frac{1}{d_{vk} + 1} (1 - u_{kv})^m \left\| x_v - v_k \right\|^2 \quad (32)$$

To improve the noise reduction capability the fuzzy factor is modified, and the new cost function is given by

$$J_s = \sum_{l=1}^{N} \sum_{k=1}^{c} \gamma_l u_{kv}^m \left\| \xi_l - v_k \right\|^2 + \sum_{v=1}^{N} \sum_{k=1}^{c} \exp(G_{kv}) \quad (33)$$

With the new objective function, the segmentation accuracy has been improved and presented in the result section.

2.8 Machine Learning Classification Method

2.8.1 Radial Basis Function Neural Network

A RBFNN is an artificial neural network that uses radial basis functions as activation functions. Figure shows the structure of the RBFNN [79,80]. The RBFNN is three layered feed-forward neural network. The first layer is linear and only distributes the input signal, while the next layer is nonlinear and uses Gaussian functions. The third layer linearly combines the Gaussian outputs. Only the tap weights between the hidden layer and the output layer are modified during training. This algorithm overcomes many issues in traditional gradient algorithms such as stopping criterion, learning rate, number of epochs and local minima.

Due to its shorter training time and generalization ability, it is suitable for real-time applications. The radial basis function selected is usually a Gaussian kernel for pattern recognition application.

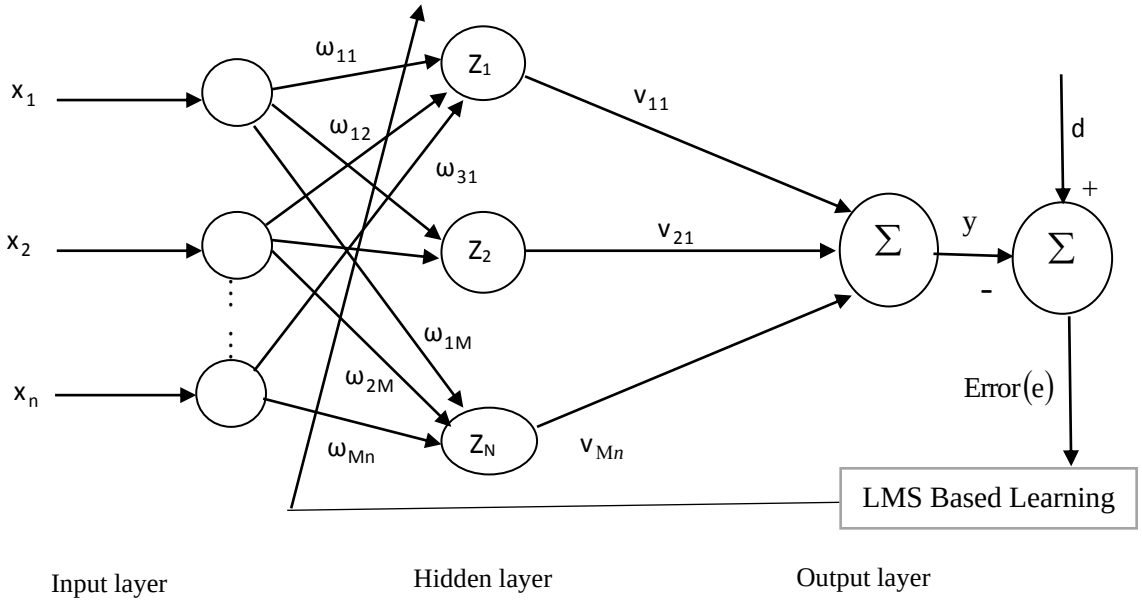

Fig:2.10 LMS Based Radial Basis Function Neural Network

In this model, it is noticed that in RBFNN [80] model the input and number of hidden nodes are equal. In the RBFNN model, a random weight is trained iteratively and weights has been assigned to the computational hidden node. This reduces the overall nodes requirement and provides better approximation to the pattern classification task.

The activation function of the N^{th} hidden neuron is defined by a Gaussian Kernel as

$$Z_N(x) = e^{\left(\frac{-\|x_i - C_j\|^2}{2\sigma_n^2}\right)} \tag{34}$$

Where σ_n^2 is the parameter for controlling the smoothness of the activation function and C_j is the center of the hidden node and $\|x_i - c_j\|$ indicates the Euclidean distance between the inputs and the function center.

The output at the output layer is given by

$$y = \sum_{n=1}^{N} (v_{11}x_1 + v_{12}x_2 ----- v_{Mn}x_n).e^{\left(\frac{-\|(x_N - C_M)\|^2}{2\sigma_M^2}\right)} \tag{35}$$

The error is calculated by subtracting the actual output from the desired output vector:

$$e = d_n - y_n \tag{36}$$

for n=1, 2, 3,……..M

The network weight is now updated by using Widrow Hoff's Least Mean Square algorithm:

$$w_{mn}(p+1) = w_{mn}(p) + \eta(p) x_{kn}^T e_{km} \tag{37}$$

for k = 1, 2, 3,…, K, n = 1, 2, 3,… m = 1, 2, 3,…,M

Here *'p'* is the number of learning cycle and the learning rate parameter 'η' is a function of *'p'*. The weights are updated during training phase. The number of weight column thus needed completely depends on the number of classes present in the data set.

The objective function is to minimize the error and the mean square error is given by

$$MSE(e) = \frac{1}{N}\sum_{n=1}^{N}(d_n - y_n)^2 \tag{38}$$

Where "d" is the desired vector.

2.9 Support Vector Machine

The development of ANN followed a heuristic path, with the applications and extensive experimentation preceding theory and suffers from multiple local minima while the solution to an SVM [81-83] is global and unique. Consider a soft margin SVM for linearly non separable classes. For binary classification in SVM the input vector x_i for linearly separated data defines a space of labeled data point called input space and the target vector is given by $y_i = \pm 1$, where i=1,2,……m is the input pairs or number of samples.

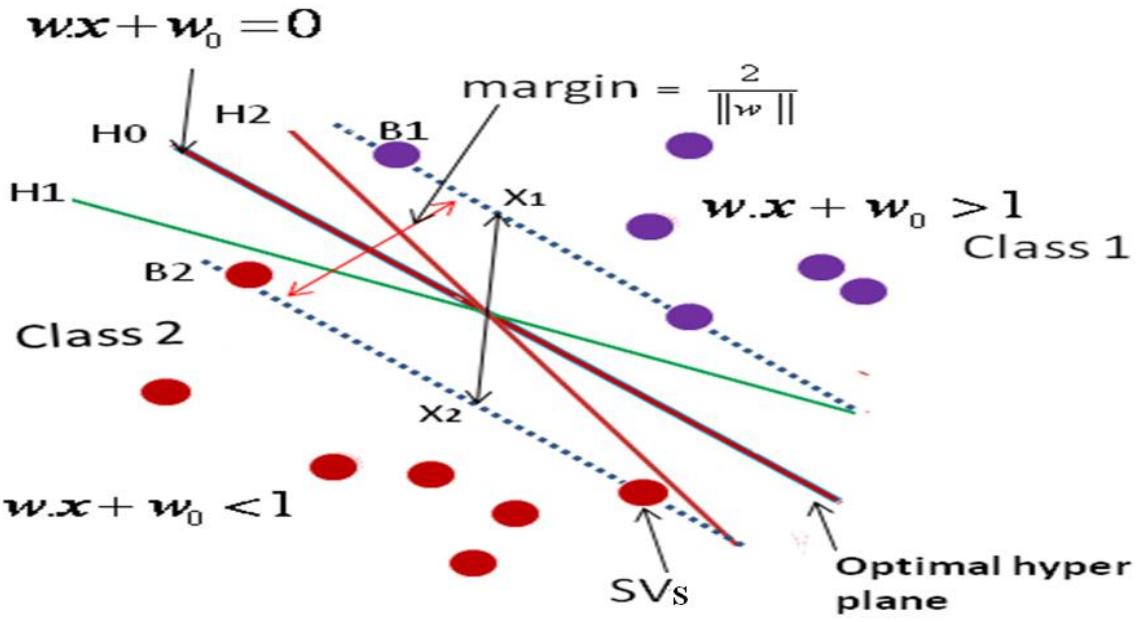

Fig .2.11. Two class problem with hyperplane analysis

The hyper planes H0, H1 and H2 achieve zero empirical risk, we have to maximize the margin which is same as minimizing the distance between the hyperplane and H0 generalizes optimally**.** In an arbitrary dimensional space a decision function separating hyperplane can be written as $w.x + w_0$, where w_0 =bias, which translates the hyperplane away from the origin and w is the weight. Thus, we considered a decision function of the form

$$f(x) = \text{sgn}(w.x + w_0) \tag{39}$$

If $w.x + w_0 \to +ve \qquad f(x) = +1$

And $w.x + w_0 \rightarrow -ve \qquad f(x) = -1$ (40)

Considering two points (x_1, x_2) on the decision boundary B1 and B2, we can write $w(x_1 - x_2) = 2$. So, the margin will be given by the projection of the vector $(x_1 - x_2)$ onto the normal vector to the hyperplane.

Under the circumstance that two supporting planes perfectly separates the two classes of patterns, the margin, measured by the perpendicular distance between the two support planes, is given by,

$$\frac{|+1|}{\|w\|} + \frac{|-1|}{\|w\|} = \frac{|2|}{\|w\|} \tag{41}$$

where $\|w\|$ denotes the Euclidean norm of the weighting vector w

Besides the aim to find a plane separating the patterns, SVMs also try to find the one inducing the largest margin. Without changing the objective, it is possible to minimize the $\|w\|$ instead of maximizing (4.24). Furthermore, the objective to minimize the $\|w\|$ is adapted into the following form:

$$\text{Margin} = \frac{2}{\|w\|} \tag{42}$$

The margin of a classifier is defined as the shortest distance between the separating boundary and an input training vector that can be correctly classified. Optimal solution to an SVM classifier is associated with finding the maximum margin.

Maximizing $\frac{2}{\|w\|}$ is same as minimizing $\|w\|$, so Maximizing the margin is thus equivalent to minimizing the function $\Phi(w) = \frac{1}{2}(w \,.w)$, subject to constraint $y_i[w.x_i + w_0] \geq 1$.

$$\min \quad \left\{\frac{1}{2}\|w\|^2 = \frac{1}{2} w.w\right\} \tag{43}$$

$$w.r.t \qquad y_i(w.x_i + w_0) - 1 \geq 0$$

By introducing Lagrangian multipliers $(\alpha_1, \alpha_2, \ldots\ldots\ldots \alpha_N) \geq 0$, the equation (1.36) can be converted to the Lagrangian primal objective function

$$L(w, w_0, \alpha) = \frac{1}{2}\left(w^T . w\right) - \sum_{i=1}^{N} \alpha_i \left[y_i \left(w.x_i + w_0 \right) - 1 \right] \tag{44}$$

Where $L(w, w_0, \alpha)$ is the Lagrangian function.

The Lagrangian has to be minimized with respect to w, w_0 and maximized with respect to $\alpha \geq 0$.

The minimum with respect to w and w_0 of the Lagrangian , $L(w, w_0, \alpha)$, is given by,

$$\frac{\partial L}{\partial w_0} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0 \tag{45}$$

and $$\frac{\partial L}{\partial w} = 0 \Rightarrow \quad w = \sum_{i=1}^{N} \alpha_i y_i x_i \tag{46}$$

The data points x_i for which $\alpha_i > 0$ are those points that are lie on the margins as well as inside the margin is called support vectors. From the solution given in equation (46), most of the α_i are zero.

Support vectors are the patterns which is most informative for the classification task. At the end of the training period, such a pattern will be one of the support vectors.

Solution for w can be written in a quadratic equation

$$W(\alpha) = \left\{ \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \left\langle x_i, x_j \right\rangle \right\} \tag{47}$$

Subject to $\sum_{i=1}^{N} \alpha_i y_i = 0$ where α_i is the hyper parameter or Lagrangian multiplier.

In the context of classification, SVM classifier first maps the input data vector x into a higher dimensional space through an underlying nonlinear mapping $\Phi(x)$, and then applies linear classification in this mapped space.

For non separable data, we will have alternative of mapping of data into higher dimensional space, as

$$x_i * x_j = \Phi(x_i).\Phi(x_j) \tag{48}$$

This higher dimensional space is called a feature space and then the kernel function is given by

$$\Phi(x_i).\Phi(x_j) \rightarrow K(x_i, x_j) \tag{49}$$

So, the kernel is therefore the inner product between mapped pairs in the feature space.

Therefore, the RBF kernel is given by

$$K(x_i, x_j) = e^{-\|x_i - x_j\|^2 / 2\sigma^2} \tag{50}$$

The point of estimates of the weights $W = \{w_j\}_{j=0}^{N}$ needed by SVM, such that the linear combination of the base functions $\Phi = \{K(x_i, x_j)\}_{i,j=1}^{N}$ fits the training targets closely. At the same time, it reduces the complexity of computation by forcing the majority of the weights to zero.

The SVM predictions is given by

$$y_i(x, w) = f_{svm}(x, w) = \sum_{i=1}^{N} w_i K(x, x_i) + w_0 \tag{51}$$

$$y(x, w) = \Phi(x) w \tag{52}$$

And $w = [w_0, w_1, w_2, ...w_N]^T$

Where $K(x, x_i)$ is a RBF kernel function and it is the key factor in SVM to satisfy the "Mercer Condition". And Φ is an augmented kernel matrix and is given by

$$\Phi = \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & \cdots & K(x_1, x_N) \\ K(x_2, x_1) & K(x_2, x_2) & \cdots & K(x_1, x_1) \\ \vdots & \vdots & \ddots & \vdots \\ K(x_N, x_1) & K(x_N, x_1) & \cdots & K(x_N, x_N) \end{bmatrix}$$

This matrix is formed by all the basis functions evaluated at all the training points that is with the RBF kernel functions.

For pattern classification problem, SVM we have to maximize

$$W(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i y_i \alpha_j y_j K(x_i, x_j) \tag{53}$$

Subject to $\sum_{i=1}^{N} \alpha_i y_i = 0$ where α_i, α_j are the hyper parameters and $i = 1, 2, ...N$

The decision function is given by

$$f(x) = \text{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x_j) + w_0 \right) \tag{54}$$

Some advantages of SVM are (i) it smoothly handles the non linear problems (ii) good prediction accuracy and involved simple mathematical calculations.

CHAPTER THREE

3. METHODOLOGY

3.1 Overview

Basically, brain tumors are categorized as malignant and benign tumors, which grows abnormally in the brain. Malignant tumors contain cancerous cells which grows to all the parts of the brain due to non-uniform structure. Benign tumors are of uniform structure and contains non-cancerous cells. By utilizing the segmentation and classification techniques, doctors can track and predict the uncontrollable growth of cancer affected areas at different levels to provide suitable diagnosis at early stage. Segmentation of image from the magnetic resonance images is a consequential and arduous task for detection of brain tumor tissues. It becomes a challenging task due to the involute structure and variations in images. Further, the classifiers such as support vector machine (SVM), probabilistic neural network (PNN), are some of the popular classifiers has already been used for classification of brain tumors. Due to the complex mathematical calculations and higher computational time requirements in the mentioned classifiers, we are motivated to propose a hybrid modified PSO based ELM classification model for classification of brain tumors from magnetic resonance images. This chapter presents a novel optimization algorithm based on Particle Swarm Optimization (PSO) hybridized with ELM (Extreme Learning Machine) model for automatic brain tumor detection and classification. The image segmentation techniques based on FCM algorithms are predicated on a rudimental region growing method and uses membership grades of pixels to relegate pixels into felicitous segments and it is found that, FCM algorithm has ability to obtain texture and background information from the simple images, but failed in the case of complex noisy images where spatial information's are not considered. So, for reduction of noise and smoothening of brain tumor magnetic resonance image an improved fast and robust fuzzy c means algorithm (FRFCM) segmentation algorithm has been proposed in this research work. The gray level co-occurrence matrix (GLCM) technique has been employed to extract features from brain tumor magnetic resonance images and the extracted features are fed as input to the proposed modified PSO based ELM model for classification of benign and malignant tumors. In this research work the ELM model's weights are optimized by using PSO algorithm which provides a unique solution to get rid of the hectic task of radiologist from

manual detection. The classification accuracy results from different conventional models are compared with the proposed PSO based ELM model and presented.

3.2 Contribution of the Research Work

Due to the complex mathematical calculations and higher computational time requirements in the mentioned classifiers, motivates to propose a hybrid PSO based ELM classification model for classification of brain tumors from magnetic resonance images. The proposed research work presents a novel image segmentation and classification technique for automatic detection and classification of brain tumor from magnetic resonance images.

The research work focuses on three contributions based on segmentation and classification. The contributions are summarized as follows:

- In first aspect, an improvement to fast and robust FCM based segmentation algorithm has been proposed. The improvement has been made to the membership partition matrix of the FCM algorithm and a wiener filter is employed to improve the rician noise reduction capability and purpose of segmentation. Further, the fuzzy membership value of the pixel has been updated to maintain image precision and smoothening of the magnetic resonance brain tumor images. The complete mathematical derivation has been developed for the new FRFCM segmentation algorithm to maximize the segmentation accuracy.

- In the second aspect, modifications to the parameters of particle swarm optimization (PSO) algorithm has been proposed to maximize the optimization performance of the hybrid algorithm. The convergence parameter, position and velocity equations in the algorithm has been modified and the mathematical calculations with modification has been developed. With the modified convergence parameter, new velocity and position equations the proposed PSO algorithm has been employed to optimize the weights of the ELM model for classification of brain tumor.

- In the third aspect the results of automatic segmentation and classification using GUI and implementation through Raspberry PI3 for visual detection and classification has been presented.

3.3 Research workflow diagram

The research work follows the steps such as (i) The magnetic resonance images has been first collected and segmented by the novel FRFCM algorithm and the features has been extracted from the images utilizing GLCM (Gray Level Co-occurrence Matrix) technique. Further (ii) the extracted features has been given as input to the proposed Extreme Learning Model (ELM) for the classification of encephalon tumors. In the third stage (iii) the weights are updated utilizing PSO(Particle Swarm Optimization) algorithm to update the weights of the extreme learning machine model. In the 4th stage (iv) The classification comparison results from the proposed extreme learning model, support vector machine and relevance vector machine model will be presented.

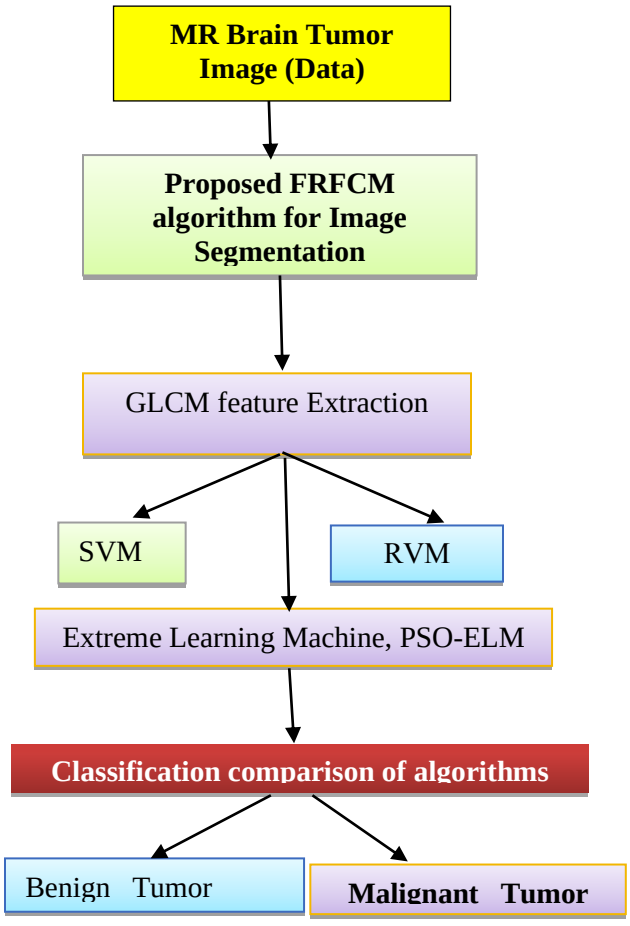

Fig.3.1 Research workflow diagram

3.4 Implementation

The research flow diagram indicates the step by step accomplishment of the research work. Further the block diagram shows the flow of algorithm application for detection and classification of brain tumor.

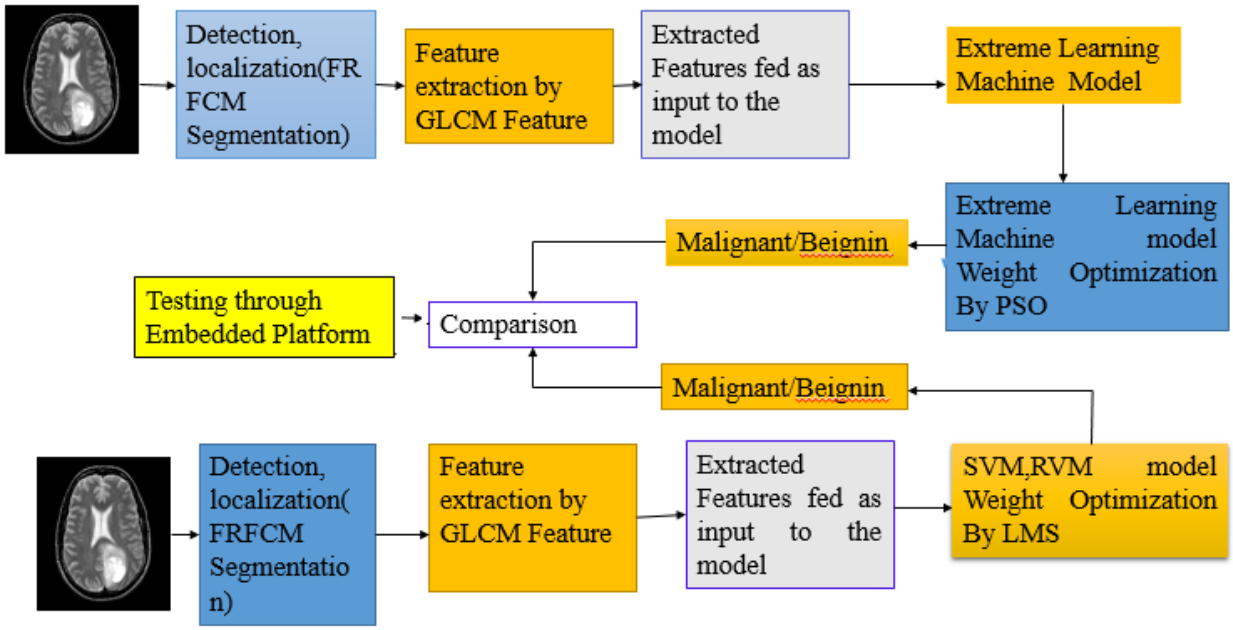

Fig.3.2 Block Diagram Representation of Implementation

3.4.1 Data Collection

According to the World Health Organization (WHO) classification system to identify brain tumors, there are more than 120 types of brain tumors which differ in origin, location, size, characteristics of the tumor tissues [84,85]. In this research work, there are three types of malignant tumors considered which are: Glioblastoma: primary malignant brain tumors that are classified as Grade IV and developed from star-shaped cells, called astrocytes that support nerve cells. It usually starts in the cerebrum. Sarcoma: has different grades that vary from grade I to grade IV and it arises in the connective tissues like blood vessels. Metastatic bronchogenic carcinoma: secondary malignant brain tumors that was spread to the brain from bronchogenic carcinoma brain tumor.

The dataset consists of glioblastoma, sarcoma and metastatic bronchogenic carcinoma tumors collected from Harvard Medical School website (http://med. harvard.edu/AANLIB/) [84]. All the

brain MRIs was in axial plane, T2-weighted and 256 x 256 pixel. A sample of the dataset is illustrated in Fig. 3.3

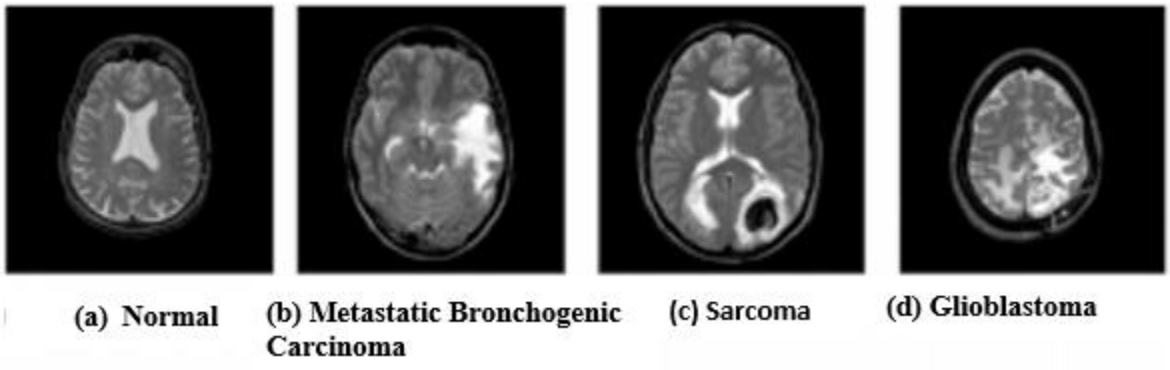

Fig.3.3 Diseased tumor in comparison to Normal

The real time data has been collected from the various renowned hospitals of Ethiopia such as

(i) Tikur Anbessa (Black Lion) Hospital, Addis Ababa
(ii) Armed Forces General Hospital, Addis Ababa
(iii) St.Paulo Hospital, Addis Ababa, American medical center, Addis Ababa
(iv) Adama general hospital and medical college, Adama

The data has been collected from the above hospitals to analyze real time tumor tissue occurrence from the magnetic resonance image. The above hospitals have provided 5000 different types of magnetic resonance scan images for patients. The type of data to be collected involves

(i) The magnetic resonance image of the patients.
(ii) The detailed information of the patients during the growth of tumor.
(iii)The stages of tumor related to dimension.
(iv) The patient's body part condition such as eye vision, headache etc. During the starting and slow growing of tumor tissues.

The method of data collection involves as

(i) We have consulted two doctors and one radiologist from each hospital as a group and take information about the patients.
(ii) Doctors provided information about the quality of image and type of diagnosis they are considered for patients.

(iii) Doctors discussed method of their treatment of the tumor tissues and identification of tumor tissues from images.

To test the model with the real data collected, we have collected nearly 5000 images from the web www.diacom.com /Harvard medical school architecture and Alzheimer's disease Neuroimaging Initiative (ADNI) public database (http://adni.loni.usc.edu/), and BRAT data set.

3.4.2 Feature extraction using GLCM (Gray Level Co-occurrence Matrix) technique

The statistical textural features such as correlation, DM (Directional Moment), entropy, coarseness, kurtosis, homogeneity and energy were extracted using gray-level co-occurrence matrix (GLCM) [86,87] technique for classification. These extracted features were used as input vectors for training and testing the performance of PSO based ELM classifier. The statistics feature formula for some of the useful features is listed below.

(1) Mean (M): The mean of an image is calculated by adding all the pixel values of an image divided by the total number of pixels in an image.

$$M = \left(\frac{1}{m \times n}\right) \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} f(x,y) \tag{55}$$

(2) Standard Deviation (SD): The standard deviation is the second central moment describing probability distribution of an observed population and can serve as a measure of inhomogeneity. A higher value indicates better intensity level and high contrast of edges of an image.

$$SD(\sigma) = \sqrt{\left(\frac{1}{m \times n}\right) \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} (f(x,y) - M)^2} \tag{56}$$

(3) Entropy (E): Entropy is calculated to characterize the randomness of the textural image and is defined as

$$E = -\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} f(x,y) \log_2 f(x,y) \tag{57}$$

(4) Skewness (S_k): Skewness is a measure of symmetry or the lack of symmetry. The skewness of a random variable X is denoted as (X) and it is defined as

$$S_k(X) = \left(\frac{1}{m \times n}\right) \frac{\sum (f(x,y) - M)^3}{SD^3} \tag{58}$$

(5) Kurtosis (S_k): The shape of a random variable's probability distribution is described by the parameter called Kurtosis. For the random variable X, the Kurtosis is denoted as $K_{urt}(X)$ and it is defined as

$$K_{urt}(X)=\left(\frac{1}{m\times n}\right)\frac{\sum\left(f(x,y)-M\right)^4}{SD^4} \tag{59}$$

(6) Energy (En): Energy can be defined as the quantifiable amount of the extent of pixel pair repetitions. Energy is a parameter to measure the similarity of an image. If energy is defined by Haralicks GLCM feature, then it is also referred to as angular second moment, and it is defined as $Energy=\sqrt{\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}f^2(x,y)}$ (60)

(7) Contrast (C_{on}): Contrast is a measure of intensity of a pixel and its neighbour over the image, and it is defined as

$$C_{on}=\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(x-y)^2\,f(x,y) \tag{61}$$

(8) Inverse Difference Moment (IDM) or Homogeneity: Inverse Difference Moment is a measure of the local homogeneity of an image. IDM may have a single or arrange of values so as to determine whether the image is textured or non -textured.

$$IDM=\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}\frac{1}{1+(x-y)^2}\,f(x,y) \tag{62}$$

(9) Directional Moment (DM): Directional moment is a textural property of the image calculated by considering the alignment of the image as a measure in terms of the angle and it is defined as

$$DM=\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}f(x,y)|x-y| \tag{63}$$

(10) Correlation (C_{orr}). Correlation feature describes the spatial dependencies between the pixels and it is defined as

$$C_{orr}=\frac{\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(x,y)f(x,y)-M_xM_y}{\sigma_x\sigma_y} \tag{64}$$

where M_x and σ_x are the mean and standard deviation in the horizontal spatial domain and M_y and σ_y are the mean and standard deviation in the vertical spatial domain.

(11) Coarseness ($Cness$). Coarseness is a measure of roughness in the textural analysis of an image. For a fixed window size a texture with a smaller number of texture elements is said to be coarser than the one with a larger number. The rougher texture means higher coarseness value. Fine textures have smaller values of coarseness. It is defined as

$$C_{ness} = \frac{1}{2^{m+n}} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} f(x,y) \tag{65}$$

The normalized features are presented in Table 3.1.

Table 3.1 Normalized feature extraction

Images	Correlation	DM	Coarseness	Skewness	Kurtosis	Energy
Img-1(Healthy)	0.1233	0.1593	0.1322	0.3448	0.9614	0.2677
Img-2 (Benign)	0.1211	0.1416	0.3418	0.4282	0.4641	0.1454
Img-3(Benign)	0.1084	0.0701	0.2503	0.2826	02704	0.1611
Img-4 (Malignant)	0.1061	0.0552	0.2602	0.3703	0.2117	0.1458
Img-5(Malignant)	0.1309	0.0106	0.1928	0.3882	0.3304	0.1066

3.5 RVM (Relevance Vector Machines)

RVM model adopt a Bayesian approach to learning and also use a prior over the model weights governed by a set of hyper parameters. For SVM the support vectors are typically formed by "borderline", difficult to classify samples in the training set, which are located near the decision boundary of the classifier; in contrast, for RVM the relevance vectors are formed by samples appearing to be more representative of two classes. In RVM the kernel function need not satisfy 'Mercer Condition' for arithmetic's validity and does not depend on Structural Risk Minimization Principle and VC dimension. RVM attains good accuracy comparing to the SVM.

Michel E. Tipping [88] proposed RVM to recast the main ideas behind SVM in a Bayesian context. The RVM decision function can be much sparser than SVM classifier. RVMS are based on a Bayesian formulation of linear model with an appropriate prior that results in a sparse representation. From the training set we wish to learn model of the dependency of the targets on the inputs with the objective of making accurate predictions. To overcome the disadvantages, we illustrate our approach with a particular algorithm the `Relevance Vector Machine' (RVM)[89], a model which is identical to the `support vector machine' (SVM). The RVM produces a much sparser approximation than the SVM.

Considering a two class problem with the training points $X=(x_1,x_2,-----x_N)$ and corresponding class labels or target $t=(t_1,t_2,----t_N)$ with $t\in(-1,1)$ .In this case it is desired to predict the posterior probability of the membership of one of the classes for given the input x, the RVM model follow the sigmoid function $\sigma(y)=1/(1+e^{-y})$

A RVM classifier model applying the logistic sigmoid function is given by

$$p(t_i=1,w)=[\sigma(y(x_i;w))]=1/(1+e^{-y(x_i;w)}) \quad (66)$$

The RVM classifier function is given by

$$y(x;w)=f_{RVM}(x)=\sum_{i=1}^{N}w_iK(x,x_i)+w_0=\Phi w \quad (67)$$

Where N is the length of the data and the weight vector $w=[w_0,w_1,w_2,.....w_N]^T$

And Φ is a $N\times(N+1)$design matrix with $\Phi=[\phi(x_i).....\phi(x_N)]^T$ where

$\phi(x_n)=[1,K(x_1,x_1).....K(x_n,x_N)]^T$ and $K(x_n,x_N)$is a kernel function. The kernel function is used to form expansion basis functions for f_{RVM} , and in theory, is not limited by the Mercer's Condition.

Those training vectors associated with the non-zero weights are called Relevance Vectors and predictions are made based on the posterior distribution over the weights. The next section provides the method of constructing a decision tree based on the extracted features of the power signal disturbance waveforms.

3.6 Proposed fast and robust FCM (FRFCM) segmentation technique:

After data collection the segmentation of the images will be done by proposed FRFCM (Fast and Robust Fuzzy C Means algorithm to detect the brain tumor tissues and removal of the ricin noise. From literature survey, some preliminary segmentation technique has been applied with brain tumor web data, but all the algorithm fails to provide required noise removal from the image and detection of tumor. To improve the drawback that of the FCM related algorithms presented above such as FLICM, FGFCM, NDFCM, etc which is sensitive to noise, we are proposing a Fast and Robust FCM (FRFCM) algorithm by incorporating local spatial information to FCM algorithm to get better precision result.

The proposed improved fast and robust FCM segmentation improves noise reduction capability by employing a Weiner filter to the modified membership partition matrix of the objective function of FCM algorithm with local information. The objective function of the fuzzy c means algorithm with local information [22] is given by

$$J_s = \sum_{v=1}^{N}\sum_{k=1}^{c} u_{kv}^{m}\|x_v - v_k\|^2 + \sum_{v=1}^{N}\sum_{k=1}^{c} G_{kv} \tag{68}$$

Where the fuzzy factor is given by

$$G_{kv} = \sum_{\substack{r\in N_v \\ v\neq r}} \frac{1}{d_{vr}+1}(1-u_{kr})^m \|x_r - v_k\|^2 \tag{69}$$

Where the spatial Euclidean distance between pixels x_v and x_r is denoted by d_{vr}, N_v is the set of neighbours within a window around x_v and x_r represents the neighbours of x_v and u_{kr} is the neighbours of u_{kv}. With respect to cluster k, x_v is the gray value of the k^{th} pixel, u_{kv} represents the fuzzy membership value of the v^{th} pixel and N is the total number of pixels in the gray scale image $f = [x_1, x_2 .... x_N]$, x_v is the gray value of v^{th} pixel, c denotes the cluster centre and m determines the fuzziness of the consequential partition.

With the fuzzy factor G_{kv}, the capability of noise reduction improves. The fuzzy partition matrix is given by

$$u_{kv} = \frac{1}{\sum_{j=1}^{c}\left(\frac{\|x_v - v_k\|^2 + G_{kv}}{\|x_v - v_j\|^2 + G_{jv}}\right)^{\frac{1}{m-1}}} \tag{70}$$

And $$v_k = \frac{\sum_{k=1}^{N} u_{kv}^{m} x_v}{\sum_{v=1}^{N} u_{kv}} \tag{71}$$

From equation (3), it is found that the factor G_{kv} is completely free of using any parameter that controls the balance between the image noise and the image details, but computational complexity increases. Clearly, there is a contradiction between improving the robustness and reducing the computational complexity simultaneously for FCM.

To reduce the computational complexity, the membership partition matrix is modified as

$$G_{kv}^{'} = \sum_{\substack{r \in N_v \\ v \neq r}} \frac{\log\left(\xi^{\tau}\right)}{\exp(d_{vr}) + 1} u_{kr}^{m} \left\| x_r - v_k \right\|^2 \tag{72}$$

Where u_{kr} is the neighbours of u_{kv}, ξ is gray value of image and τ is the smoothness parameter between 0 and 1. Further, considering the morphological reconstruction operations such as dialation and erosion, the reconstruction of the image is considered as ξ_p, which is given by

$$\xi_p = R_e^C(f) \tag{73}$$

Where R_e^C represents the morphological closing reconstruction which is efficient for noise removal and f denotes an original image and reconstruction operators considering morphological closing reconstruction is given by

$$R_e^C(f) = R_{R_f^{\beta}(\chi(f))}^{\chi}\left(\beta\left(R_f^{\beta}(\chi(f))\right)\right) \tag{74}$$

Where χ is the erosion operation, β is the dilation operation, c is the closing operation and f represents original image. With image reconstruction operation, the filtering capability increases. Now, with morphological closing reconstruction the objective function is modified as

$$J_s = \sum_{k=1}^{c} \sum_{p=1}^{q} \left[u_{kp}^{m} \left\| \left(\xi_p - v_k\right) \right\|^2 \right] + \sum_{k=1}^{c} \sum_{p=1}^{q} G_{kp} \tag{75}$$

From equation (80), it is evident that u_{kp} represents the degree of membership of gray value p in cluster k, ξ_p is a gray level, $1 \le p \le q$, q represents the gray levels contained in ξ. Then the modified fuzzy factor is given by

$$G_{kp}^{'} = \sum_{\substack{r \in q_v \\ v \neq r}} \frac{\log\left(\xi^{\tau}\right)}{\exp(d_{vr}) + 1} u_{kp}^{m} \left\| \xi_p - v_k \right\|^2 \tag{76}$$

Where $$u_{kp} = \frac{1}{\sum_{j=1}^{c} \left(\frac{\left\| \xi_p - v_k \right\|^2 + G_{kp}^{'}}{\left\| \xi_p - v_j \right\|^2 + G_{jp}^{'}} \right)^{\frac{1}{m-1}}} \tag{77}$$

And $$v_k = \frac{\sum_{k=1}^{c} u_{kp}^{m} \xi_p}{\sum_{p=1}^{q} u_{kp}} \tag{78}$$

Now, we can write the membership partition matrix in the form as $U = \left[u_{kp}\right]^{c\times q}$. Further, considering convergence speed of the algorithms and the performance of the partition matrix U we employ a wiener filter [82]**.** The new membership partition matrix is given by

$$U^{l} = wiener\left[U\right] \tag{79}$$

The steps of implementation of the algorithm is as follows:

Step 1: Choose the cluster value c, filtering window size w, and fuzzification coefficient m, maximum number of iteration at the beginning.

Step 2: Set loop counter $l = 0$ and update the cluster centre.

Step 3: Initialize the membership partition matrix randomly

Step 4: By using the equation $\xi_p = R_\epsilon^C(f)$, compute the new image.

Step 5: Update U^l according to equation (84) until convergence of objective function, else go to step 2.

3.7 Proposed PSO based ELM model

3.7.1 Motivation

The motivations of introducing local liner model as follows

1. ELM provide a frugal interpolation in high dimension spaces when modeling samples are sparse.
2. ELM replaces the connection of weights between the hidden layer and output layer of conventional RBFNN to have faster convergence speed.
3. ELM requires less number of neurons to approximate the nonlinear system and it is more efficient because the ability of approximation in the hidden layer units.
4. The input and number of hidden layer nodes are equal which reduces overall node requirement.

3.7.2 ELM Model

Recently, Bayesian methods are exploited [90, 91] to learn the output weights of ELM to gain higher generalization. Bayesian methods have advantages for machine learning problems that they try to estimate the probability distribution of output values instead of fitting to data, and hence avoiding data over fitting. A non-sparse Bayesian approach for ELM (BELM) for regression was proposed recently, where the output weights are with Gaussian priori distribution

conditioned on a shared hyper prior parameter. Meanwhile, the classification task has not been solved in due to a more complicated Bayesian learning process.

In this research work, we propose a sparse Bayesian learning (SBL) approach [92], [93] for ELM (called SBELM) that learns the output weights of ELM classifier, where the parameters of hidden layer are randomly generated as in the conventional ELM. SBELM finds sparse representatives for the output weights assumed with priori distribution instead of adding/deleting hidden neurons. The SBL is a family of Bayes methodology that aims to find a sparse estimate of output weights to $w_k, k = 0$ to L (L is the number of hidden neurons), by imposing a hierarchical-independent hyperprior α_k on each w_k in which some w_k's are automatically tuned to zeros during learning phase. Derivatives of SBL mainly vary in the distribution of the hyperprior $p(\alpha_k)$ among which the automatic relevance determination (ARD) prior [94], which commonly assumes Gamma distribution as the learning strategy. A classic of such algorithm is relevance vector machine (RVM) [95], which is the Bayesian approach for support vector machine (SVM) [96]. The proposed SBELM assumes the output weights w_k are conditioned on ARD prior to gain sparsity by tuning some w_k to zeros, leading to pruning the corresponding hidden neurons. Hence, SBELM has the advantages of both SBL (high generalization and sparsity) and ELM (universal approximation and efficient learning speed).

Given a set of N training dataset $D = (x_i, d_i), i = 1$ to N with each x_i is a vector and d_i is the expectation output. The output function of ELM with L hidden neurons is represented by

$$y = \sum_{k=0}^{L} \beta_k h_k(w_k; x) \tag{80}$$

where $h(w; x) = [1, h_1(w_1; x), \ldots\ldots, h_L(w_L; x)]$ is the hidden feature mapping with respect to input $x = [x_1 \ldots . x_N]$ are randomly generated parameters of hidden layer and β is the weight vector of all hidden neurons to an output neuron to be analytically analyzed. $h_k(\cdot)$ which is the activation function of hidden layer. Equation (80) can be written as

$$\mathrm{H}\beta = y \tag{81}$$

Where H is the $\mathrm{N} \times (L + 1)$ hidden layer feature-mapping matrix, whose elements are as follows:

$$\mathrm{H}=\begin{bmatrix} 1 & h_1(w_1;x_1) & \cdots & h_L(w_N;x_1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & h_1(w_1;x_N) & \cdots & h_L(w_N;x_N) \end{bmatrix} \tag{82}$$

And $h_L(w_N;x_N)=[w_1x_1+w_1x_1.....w_Nx_N]e^{\left(\frac{-\|(x_N-v_i)\|^2}{2\sigma_n^{\ 2}}\right)}$

Where σ_n^2 is the parameter for controlling the smoothness of the activation function and v_j is the center of the hidden node and $\|x_i - c_j\|$ indicates the Euclidean distance between the inputs and the function center.

The i th row of H is the hidden layer's output vector for an instance x. Equation (81) is a linear system, which is solved by

$$\beta=\mathrm{H}^{\dagger}d, \qquad \mathrm{H}^{\dagger}=\left(\mathrm{H}^T\mathrm{H}\right)^{-1}\mathrm{H}^T \tag{83}$$

Where $\mathrm{H}^{\dagger}$ is the Moore–Penrose generalized inverse of matrix H

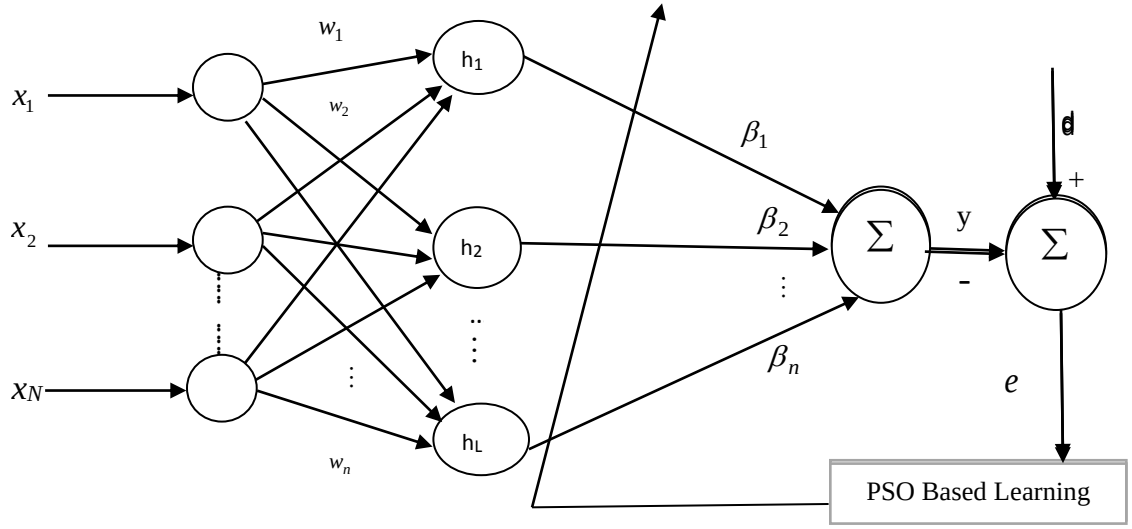

Fig. 3.4 PSO Based ELM Model

And $d=\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$, $\beta=\begin{bmatrix} \beta_1 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$

3.7.3 Proposed learning algorithm steps:

Thus, a simple learning method called extreme learning machine (ELM) can be summarized as follows:

Algorithm ELM: Given a training set $\aleph = \left\{ (x_i, d_i) \middle| x_i \in R^n, t_i \in R^m, i = 1,.....N \right\}$ activation function $h(x)$, and hidden node number,

Step 1: Randomly assign input weight w_i $, i = 1,...,N$.

Step 2: Calculate the hidden layer output matrix H .

Step 3: Calculate the output weight β.

$\beta = \mathrm{H}^{+} d$

Where $d = [d_1,.....d_N]^{\mathrm{T}}$

3.7.4 Choosing center of the ELM model using Enhanced Fuzzy C Means algorithm

This process of choosing center is as follows

Step1: Let $X = \|(x_{j1}, x_{j2},...,x_{jn})\|$, $j = 1,2,...,N$ is the data set. The centers 'C' have been randomly initialized from the data set.

Step2: Initially take random centers and the data points as the input features.

Step 3: For each data point the center having the maximum probability of finding the nearest mean to each data point, and reassigning the data points to the associated centers, and then re-computing the cluster means is chosen as the corresponding center by using the formula

$$v_i = \frac{\sum_{i=1}^{q} \gamma_l u_{il}^m \xi_l}{\sum_{l=1}^{q} \gamma_l u_{il}^m}$$

The center of the ELM model is updated by the formula

$$v_i(n+1) = v_i(n) + \eta(x - v_i(n))$$

Step 4: Repeat step-2 to step-3 for each data point and the optimized center was obtained at the end of iteration.

3.7.5 Sparse Bayesian Approach to ELM

We begin with binary classification, in which every training sample can be treated as an independent Bernoulli event, whose probability $p(t|x)$ is Bernoulli distribution. We write the likelihood as

$$p(t|\beta,h)=\prod_{i=1}^{N}\sigma\{Y(h_i;\beta)\}^{t_i}\left[1-\sigma\{Y(h_i;\beta)\}\right]^{1-t_i} \tag{84}$$

Where $\sigma(\cdot)$is sigmoid function

$$\sigma\{Y(h;\beta)\}=\frac{1}{1+e^{-\sigma\{Y(h;\beta)\}}} \tag{85}$$

and $y(h;\beta)=h\beta$

Where $t=(t_1\ldots t_N)^{\mathrm{T}}\ t_i\in\{0,1\}$ and $\beta=(\beta_0\ldots\beta_L)^{\mathrm{T}}$. A zero-mean Gaussian prior distribution over each parameter w_k conditions on an ARD(automatic relevance determination) hyper parameter α_k is given by

$$(\beta_k|\alpha_k)=N(\beta_k|0,\alpha_k^{-1}) \tag{86}$$

$$p(w/\alpha)=\prod_{k=0}^{L}\frac{\alpha_k}{\sqrt{2\pi}}\exp\left(-\frac{\alpha_k\beta_k^2}{2}\right) \tag{87}$$

Where $\alpha=[\alpha_0\alpha_1\alpha_2\ldots\alpha_L]^{\mathrm{T}}$. Importantly, there always exists an independent α_k associated with each w_k. The ARD prior selects the significant hidden neurons by controlling some values of w_k's to zero. Next step is to marginalize the likelihood over t conditioned on α and H .The values of α are determined by maximizing the marginal likelihood by integrating the weight parameters w

$$p(t|\alpha,\mathrm{H})=\int p(t|\beta,\mathrm{H})p(\beta|\alpha)dw \tag{88}$$

The core learning procedure involves establishing the distribution of $p(t/\alpha,\mathrm{H})$to determine α by maximizing the marginal likelihood. The integral of (93) is, however, intractable; ARD approximates a Gauss for it with Laplace approximation approach, which is achieved by evaluating a quadratic Taylor expansion of log-posterior function. The mean and covariance of the approximated Gauss are the Laplace's mode and negated inversed second derivative (Hessian matrix), respectively. Thus

$$\ln\{p(t|\beta,\mathrm{H})p(\beta|\alpha,)\}$$

$$= \ln\left\{\prod_{i=1}^{N} y_i^{t_i}(1-y_i)^{1-t_i}\right\} + \ln\left\{\prod_{k=0}^{L}\frac{\alpha_k}{\sqrt{2\pi}}\exp\left(-\frac{\alpha_k\beta_k^2}{2}\right)\right\}$$

$$= \sum_{i=1}^{N}\{t_i \ln y_i + (1-t_i)\ln(1-y_i)\} - \frac{1}{2}\beta^{\mathrm{T}}\mathrm{A}\beta + const \tag{89}$$

where $y_i = \sigma\{y(h_i;\beta)\}$and A = diag ($\alpha$).Here, we let $const = \sum_{k=0}^{L}(\ln\alpha_k - 1/2\ln 2\pi)$, which is not associated with w. Generally, it is efficient to find the Laplace's mode $\hat{w}$ using Newton–Raphson method iterative reweighted least squares (IRLS). Given (94), its gradient $\nabla\mathrm{E}$ and Hessian matrix Φ need to be figured out

$$\nabla\mathrm{E} = \nabla_w \ln\{p(t|\beta,\mathrm{H})p(\beta|\alpha,)\} = \sum_{i=1}^{N}(t_i - y_i)h_i - \mathrm{A}\beta$$

$$= \mathrm{H}^{\mathrm{T}}(t-y) - \mathrm{A}\beta \tag{90}$$

$$\Phi = \nabla_w\nabla_w n\{p(t|\beta,\mathrm{H})p(\beta|\alpha,)\} = -(\mathrm{H}^{\mathrm{T}}\mathrm{B}\mathrm{H} + \mathrm{A}) \tag{91}$$

where $y = [y_1 y_2,\ldots,y_N]^{\mathrm{T}}$, B is an $N \times N$ diagonal matrix with element $\gamma_i = y_i(1-y_i)$and $d\sigma/dy = \sigma(1-\sigma)$. Therefore, by using IRLS, $\hat{w}$ is obtained by

$$\beta_{new} = \beta_{old} - \Phi^{-1}\nabla\mathrm{E} = (\mathrm{H}^{\mathrm{T}}\mathrm{B}\mathrm{H} + \mathrm{A})^{-1}\mathrm{H}^{\mathrm{T}}\mathrm{B}\hat{t} \tag{92}$$

Where $\hat{t} = \mathrm{H}w + \mathrm{B}^{-1}(t-y)$. The center $\hat{w}$ and covariance matrix $\sum$of Gauss distribution over w by Laplace approximation are

$$\hat{w} = \sum \mathrm{H}^{\mathrm{T}}\mathrm{B}\hat{t} \tag{93}$$

$$\sum = (\mathrm{H}^{\mathrm{T}}\mathrm{B}\mathrm{H} + \mathrm{A})^{-1} \tag{94}$$

Therefore, we obtain $p(t|\beta,\mathrm{H})p(w|\alpha) \propto \mathrm{N}(\hat{w},\Sigma)$. After gaining Gaussian approximation for w, the integral in (10) becomes tractable. The log marginal likelihood is as follows:

$$\ell(\alpha) = \ln p(t|\alpha,\mathrm{H})$$

$$= -\frac{1}{2}\left[\mathrm{N}\ln(2\pi) + \ln|\mathrm{B} + \mathrm{H}\mathrm{A}\mathrm{H}^{\mathrm{T}}| + (\hat{t})^{\mathrm{T}}(\mathrm{B} + \mathrm{H}\mathrm{A}\mathrm{H}^{\mathrm{T}})^{-1}\hat{t}\right]$$

$$= -\frac{1}{2}\left[\mathrm{N}\ln(2\pi) + \ln|C| + (\hat{t})^{\mathrm{T}}C^{-1}\hat{t}\right] \tag{95}$$

Where $C = \mathrm{B} + \mathrm{HAH}^{\mathrm{T}}$. Set the differential of $\ell(\alpha)$ with respect to α to zero

$$\frac{\partial \ell(\alpha)}{\partial \alpha_k} = \frac{1}{2\alpha_k} - \frac{1}{2}\sum_{kk} - \frac{1}{2}\hat{\beta}_k^2 = 0$$

$$\Rightarrow \alpha_k^{new} = \frac{1-\alpha_k \sum_{kk}}{\hat{\beta}_k^2} \tag{96}$$

By setting β_k and α_k with initial values, $\hat{w}$ and $\sum$ are updated from (95) and (97). Using these two values, α is updated through (18) again, the operation continues to maximize the marginal likelihood function until reaching the convergence criteria (e.g., when the difference between the maximum α_k

(empirically, $\log \alpha_k$ is preferred) in two successive iterations is lower than a predefined accuracy, or the maximum number of iterations). After $\hat{w}$ converges, the probability distribution $p(t_{new}/x_{new}, \hat{w})$ is predicted.

A. Property of Sparsity

The mechanism of ARD prior tunes a part of output weights to zeros resulting in pruning their associated hidden neurons. Decompose C based on the term α_k, $\ell(\alpha)$ is rewritten as

$$\ell(\alpha) = -\frac{1}{2}\left[\mathrm{N}\ln(2\Pi) + \ln|C_{-k}| + \left(\hat{i}\right)C_{-k}^{-1}\hat{t} - \ln\alpha_k + \ln\left(\alpha_k + \hbar_k^{\mathrm{T}} C_{-k}^{-1}\hbar_k\right) - \frac{\left(\hbar_k^{\mathrm{T}} C_{-k}^{-1}\hat{t}\right)}{\alpha_k + \hbar_k^{\mathrm{T}} C_{-k}^{-1}\hbar_k}\right]$$

$$\ell(\alpha - k) + \frac{1}{2}\left[\ln\alpha_k - \ln(\alpha_k + s_k) + \frac{q_k^2}{\alpha_k + s_k}\right] \tag{98}$$

where $s_k = \hbar_k^{\mathrm{T}} C_{-k}^{-1}\hbar_k, q_k = \hbar_k^{\mathrm{T}} C_{-k}^{-1}\hat{t}$, and $\hbar_k$ is the kth column of H. $\ell(\alpha - k)$ is the marginal likelihood with α_k omitted and C_{-k} represents the matrix C with term k and α_k removed, thus $\ell(\alpha - k)$ is irrelevant to α_k but other components. By resetting the derivative of $\ell(\alpha)$ with respect to α_k to zero, the stationary point is obtained

$$\hat{\alpha}_k = \begin{cases} \dfrac{s_k^2}{q_k^2 - s_k}, if & q_k^2 > s_k \\ \infty \quad , if & q_k^2 \le s_k \end{cases}$$

In the process of carrying out the iteration of α to maximize the marginal likelihood, some α_k 's tend to grow to infinity, which influences the mean $\hat{w}$ and covariance $\sum$ as

$$\lim_{\alpha_k \to \infty} \Sigma_{kk} = \lim_{\alpha_k \to \infty} \left\{ \hbar_k{}^{\mathrm{T}} \beta_i \hbar_k + \alpha_k \right\}^{-1} = 0 \tag{99}$$

$$\lim_{\alpha_k \to \infty} \hat{w}_k = \lim_{\alpha_k \to \infty} \left\{ \varepsilon_k \hbar_k{}^{\mathrm{T}} \beta_i \hat{t}_i \right\} = 0 \tag{100}$$

where ε_k is the *k*th row of matrix $\sum$. (103) and (104) indicates the $\hat{w} \sim \mathrm{N}(0,0)$ equivalent to zero. Therefore, the corresponding $h_k(\theta_k; x)$ is pruned, with some that can best contribute to the maximization of likelihood $p(t/\alpha, \mathrm{H})$ are left, which results in sparsity.

B. Multiclassification

For multiclass classification selecting the state-of-the-art approach pairwise coupling in, which combines all the outputs of every pair of classes to the overall reestimate probabilistic densities of all classes for a new instance.

Let r_{mn} be the predicted probability of a binary classifier for a new instance denoted as $r_{mn} = \mathrm{P}\{t = m / t = m \text{ or } n, x\}$, which is the sigmoid output of SBELM and p_m, $m = 1 \ldots K$ be the final probability to be estimated by the pairwise coupling strategy, where *K* is the label of class. Therefore, the problem of estimation of p_m is equivalent to solve an optimization problem

$$\min \sum_{m=1}^{\mathrm{K}} \sum_{n:n \neq m} \left(r_{mn} p_n - r_{mn} p_m \right)^2$$

$$\text{Subject to} \quad \sum_{m=1}^{\mathrm{K}} p_m = 1 \tag{101}$$

Where $r_{mn} \approx p_m / (p_m + p_n)$. Rewrite the objective function of (101) as

$$\min_p 2\mathrm{P}^{\mathrm{T}} Q \mathrm{P} \equiv \min_p \frac{1}{2} \mathrm{P}^{\mathrm{T}} Q \mathrm{P} \tag{102}$$

with

$$Q_{mn} = \begin{cases} \sum_{s:s \neq m} r_{sm}^2 & ,if \quad m = n \\ -r_{nm} r_{mn} & ,if \quad m \neq n \end{cases} \tag{103}$$

Therefore, (106) is a linear equality–constrained convex quadratic programming problem, which can be solved by Lagrange multiplier method. The unique p is a global minimum if it satisfies

the optimal condition. By importing the Lagrange multiplier ξ for the condition of (103), the optimization of (106) is changed to the following form:

$$\begin{bmatrix} Q & e \\ e^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} p \\ \xi \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{104}$$

Where Q_p are a derivative of the right-hand side of (23), ***e*** and **0** are the K×1 vector of all ones and zeros, respectively. The linear system (101) provides the solution for (103).

3.7.6 Particle Swarm Optimization (PSO):

PSO is a population based stochastic optimization technique inspired by social behavior of bird flocking. PSO uses a population of individuals, to search feasible region of the function space. In this context,

(i) Each candidate solution is called particle and represents one individual of a population (features).

(ii) The population is set of vectors and is called SWARM (set of feature data points).

(iii) The particles change their components and move (fly) in a search space.

(iv) They can evaluate their actual position using the function to be optimized. The function is called FITNESS FUNCTION.

(v) Particles also compare themselves to their neighbors and imitate the best of that neighbors.

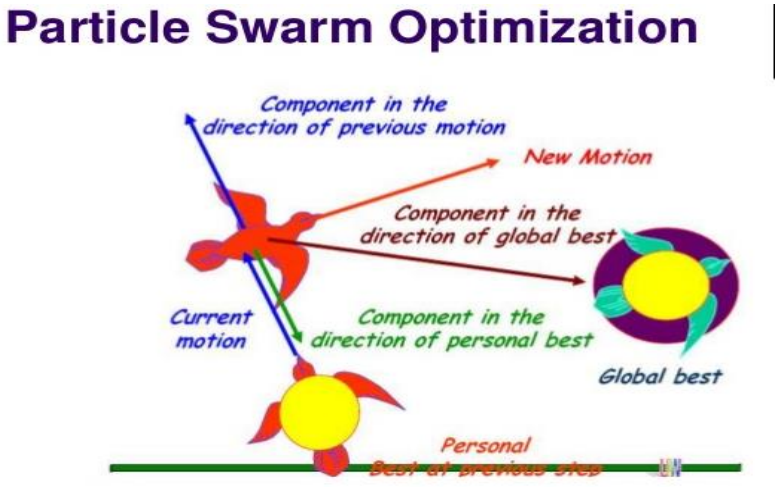

Fig. 3.5 PSO details

3.7.7 Weight optimization by PSO

Particle swarm optimization (PSO) [97] has been applied to virtually every area in optimization, computational perspicacity, and design/scheduling applications.

Let x_i and v_i be the position vector and velocity for particle i[th], respectively. The new velocity vector is determined by the following formula

$$v_i^{n+1} = v_i^n + \alpha \in_1 \left[g^* - x_i^n\right] + \gamma \in_2 \left[x_i^* - x_i^n\right] \tag{105}$$

Where $\in_1$ and $\in_2$ are two random vectors, between 0 and 1. The parameters α and β are the learning parameters or acceleration constants, which can typically be taken as, say, $\alpha \approx \beta \approx 2$.

In standard PSO algorithm, and the most noticeable improvement is probably to use an inertia function $\theta(n)$ so that v_i^n is replaced by $\theta(n)\, v_i^n$

$$v_i^{n+1} = \theta v_i^n + \alpha \in_1 \left[g^* - x_i^n\right] + \gamma \in_2 \left[x_i^* - x_i^n\right] \tag{106}$$

Where $\theta \in (0,1)$.

In the expedited particle swarm optimization (PSO) [98], the velocity vector is engendered by a simpler formula

$$v_i^{n+1} = v_i^n + \alpha \in_n + \beta\left[g^* - x_i^n\right] \tag{107}$$

Where $\in_n$ is drawn from N (0, 1) to replace the second term. The update of the position is simply

$$x_i^{n+1} = x_i^n + v_i^{n+1} \tag{108}$$

In order to increase the convergence even further, we can also write the update of the location in a single step

$$x_i^{n+1} = (1-\gamma)x_i^n + \gamma g^* + \alpha \in_n \tag{109}$$

Typically, $\alpha = 0.1 \sim 0.5$ while $\beta = 0.1 \sim 0.7$ is taken for most of the research work. To reduce the randomness, a further improvement to the accelerated PSO is is done by using a monotonically decreasing function such as

$$\alpha = \alpha_0 e^{-\chi n} \tag{110}$$

Or $\alpha = \alpha_0 \chi^n, \qquad (0 < \gamma < 1)$ (111)

Where $\alpha_0 \approx 0.5 \sim 1$ is the initial value of the randomness parameter. Here t is the number of iterations or time steps. $0 < \chi < 1$ is a control parameter [98].

Where n∈ [0, Nmax] and Nmax is the maximum of iterations.

Further the weights are mentioned as $W = [w_{i0} + w_{i1}x_1 + \ldots\ldots\ldots w_{iN}x_N]$ and the weigthts are mapped and updated using

$$W_i^{n+1} = (1-\gamma)W_i^n + \gamma g^* + \alpha \in_n \quad (112)$$

Traditionally, in order to train ELM, to minimizing the cost function, the minimum square error is given by

$$MSE = \sum_{j=1}^{N}\left(\sum_{i=1}^{N} d_j - \beta_i h(w_i \cdot x_j)\right)^2 \quad (113)$$

When H is unknown gradient-based learning algorithms are generally used to search the minimum of $\|H\beta - d\|$.

3.7.8 Pseudo code of the proposed algorithm

Pseudo code: **PSO Algorithm implementation for weight optimization of ELM Model.**

1. Initializing particles (weights of the model) with random position and velocity vectors.
2. Evaluating fitness function for each particle's position using equation
3.The Population size =50
4.The control parameter α = 0.85n, γ =0.75
5. Initialize the position velocity equation and map with the weights
6.Initialize the weights as W_i^n
7.%starting of loop
If fitness is better than fitness (gbest) then gbest= g^*
update (W_i^n) as

$W_i^{n+1} = (1-\gamma)W_i^n + \gamma g^* + \alpha \in_n$

End

update W_i^{n+1} to obtain minimum weight values

end of for loop

8.Stopping criteria: Continue till optimization gets minimum error values
9. If not converges, repeat until nearly zero error satisfied.

3.8 Set up for visualization of detection and classification of brain tumor using Raspberry Pi 3 B+

- As per the Myoclonic Gliomas is a type of tumors begin in the brain or spinal cord and include astrocytomas, ependymomas, glioblastomas, oligoastrocytomas and oligodendrogliomas.
- Meningioma is a type of tumor that arises from the membranes that surrounds brain and spinal cord (meninges). Most meningiomas are noncancerous.

- Medulloblastomas are the most common cancerous brain tumors. A medulloblastoma starts in the lower back part of the brain and tends to spread through the spinal fluid.

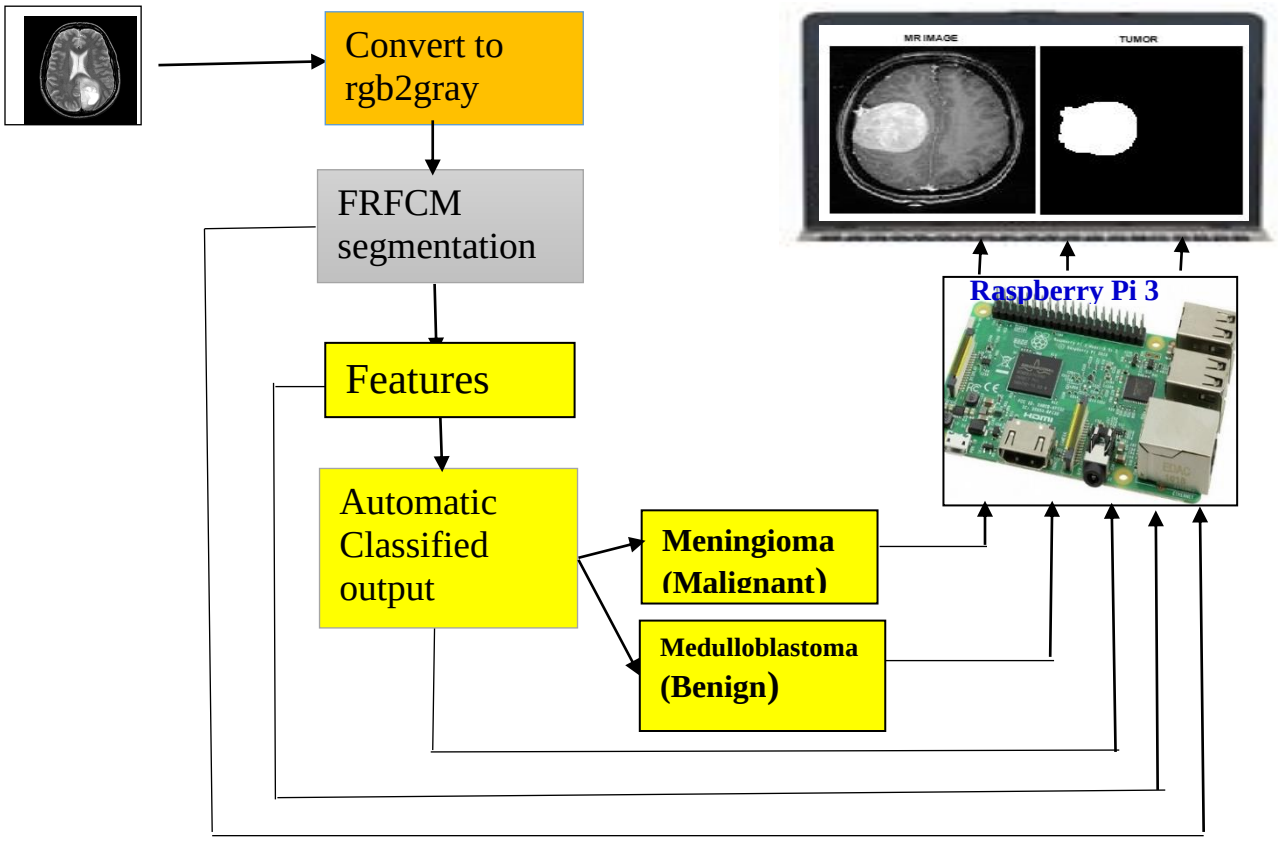

Fig.3.6 GUI for visualization of segmentation, Feature and classification results

Type-1: Meningioma
Type-2: Medulloblastoma

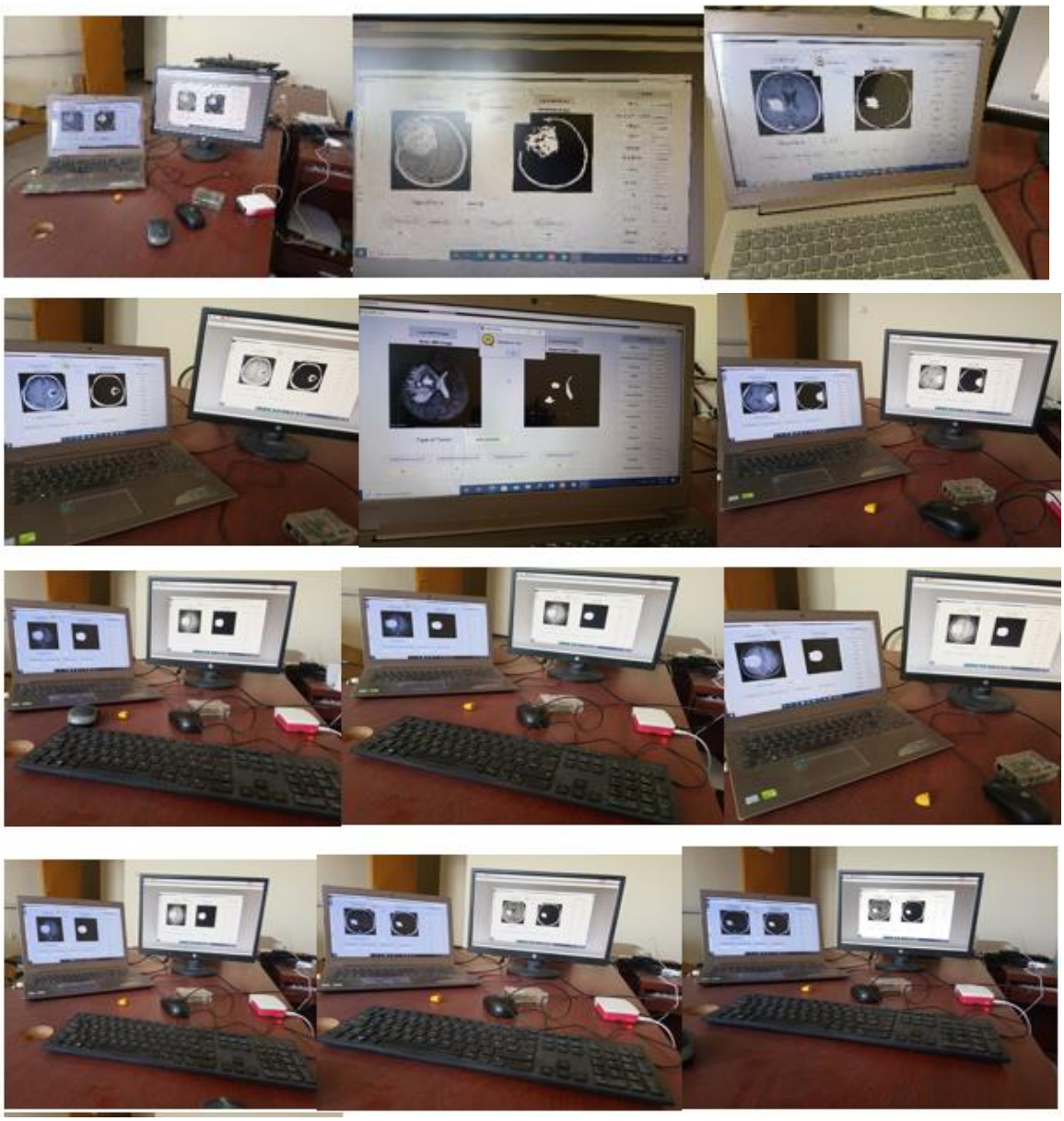

Fig.3.7 GUI for visualization results on system

3.8.1 Raspberry pi and its Components

The Raspberry Pi is a wonderful microcomputer that brims with potential. With a Raspberry Pi you can build robots, learn to code, and create all kinds of weird and wonderful projects. It is a cheap, credit card sized computer running a Linux operating system (Raspbian/Ubuntu) designed for the researchers to learn programming. The **Raspberry Pi 3 Model B+** is the latest product in the Raspberry Pi 3 range, boasting an updated 64-bit quad core processor running at **1.4GHz**

with built-in metal heatsink, dual-band 2.4 GHz and **5 GHz wireless LAN, faster (300 mbps) Ethernet**, and **PoE capability via a separate PoE HAT**.

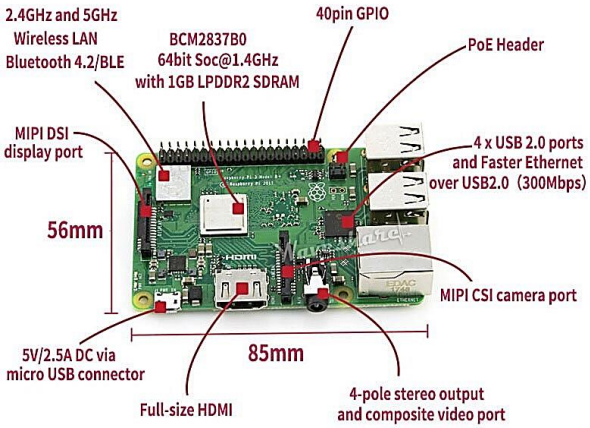

Fig. 3.8 Raspberry Pi peripherals module

CHAPTER FOUR

4. RESULTS AND DISCUSSION

4.1 Pre-processing

The primary task of pre-processing is to improve the quality of the MR images and make it in a form suited for further processing by human or machine vision system. In addition, pre-processing helps to improve certain parameters of MR images such as improving the signal-to noise ratio, enhancing the visual appearance of MR image, removing the irrelevant noise and undesired parts in the background, smoothing the inner part of the region, and preserving its edges. To improve the signal-to-noise ratio, and thus the clarity of the raw MR images, we applied adaptive contrast enhancement based on modified sigmoid function [99].

4.2 Segmentation and Morphological Operation

The segmentation of the infected brain MR regions is achieved through the following steps: In the first step, the pre-processed brain MR image is converted into a binary image. With a threshold value being selected. The pixel values greater than the selected threshold are mapped to white, while others are marked as black; due to this two, different regions are formed around the infected tumor tissues, which is cropped out. In the second step, in order to eliminate white pixel, an erosion operation of morphology is employed. Finally, the eroded region and the original image are both divided into two equal regions and the black pixel region extracted from the erode operation is counted as a brain MR image mask. In this study, FRFCM segmentation is employed for effective segmentation of brain MR image. The proposed segmentation is employed to develop functions, operators, data, or information into components of different frequency, which enables studying each component separately.

4.3 FCM algorithm-based Image segmentation results

In this research work, Fuzzy C Means, Fuzzy C Means Algorithm with Spatial Constraint, Enhanced Fuzzy C Means algorithm has been employed for effective segmentation of brain MR image and comparison segmentation results are presented which are shown from Fig.4.1 to Fig.4.8.

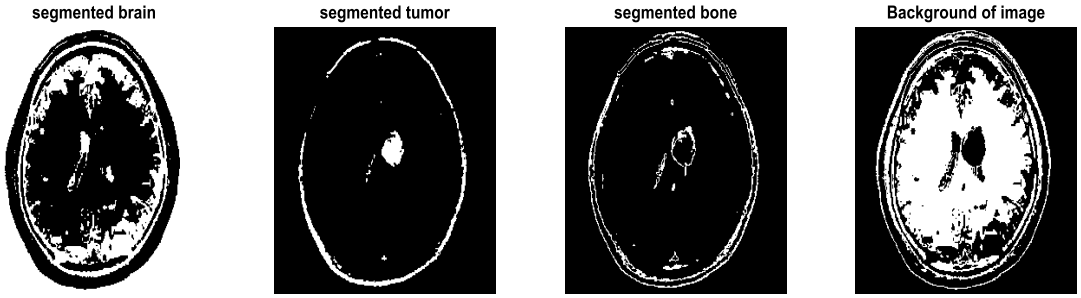

Fig.4.1 Image Segmentation Using Fuzzy C Means Algorithm.

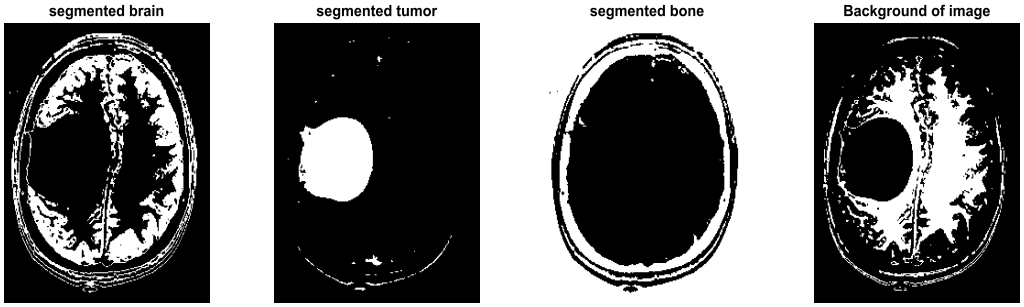

Fig. 4.2 Image Segmentation Using FCM Algorithm with Spatial Constraint (FCM_S1).

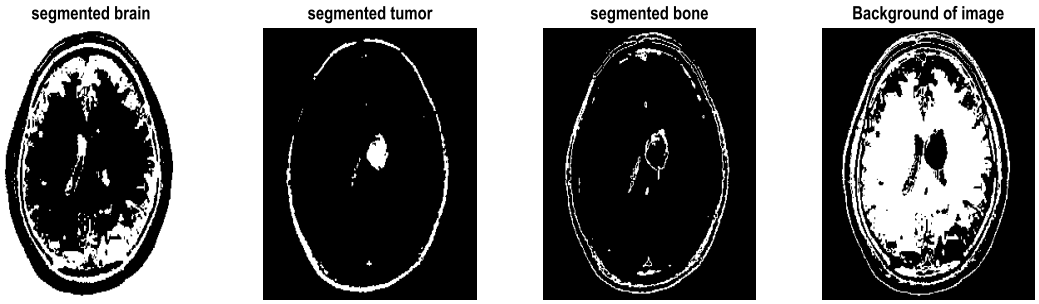

Fig. 4.3 Image Segmentation Using FCM Algorithm with Spatial Constraint (FCM_S2).

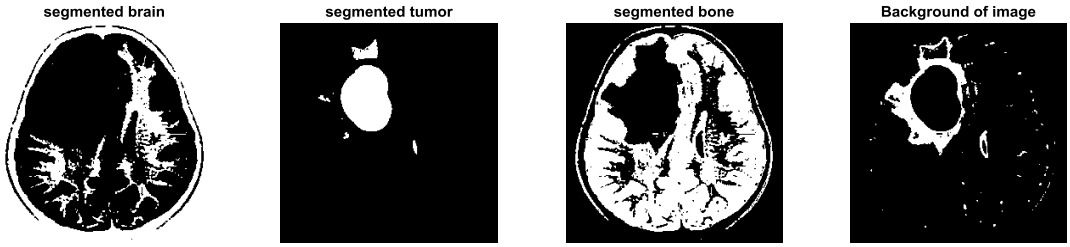

Fig. 4.4 Segmentation Using Enhanced Fuzzy C Means Algorithm.

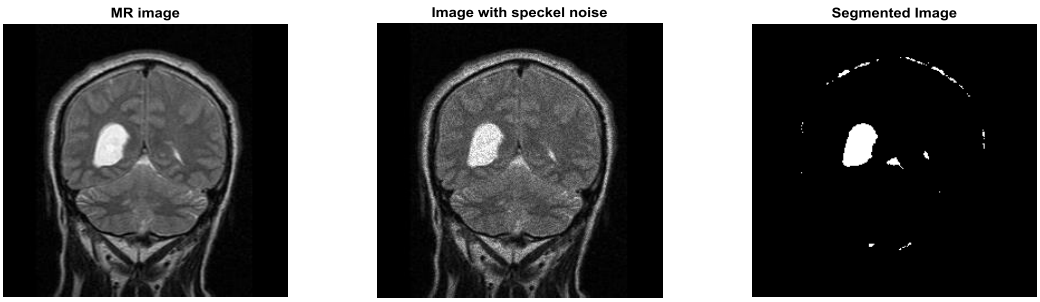

Fig. 4.5 Segmentation Using NDFCM Algorithm.

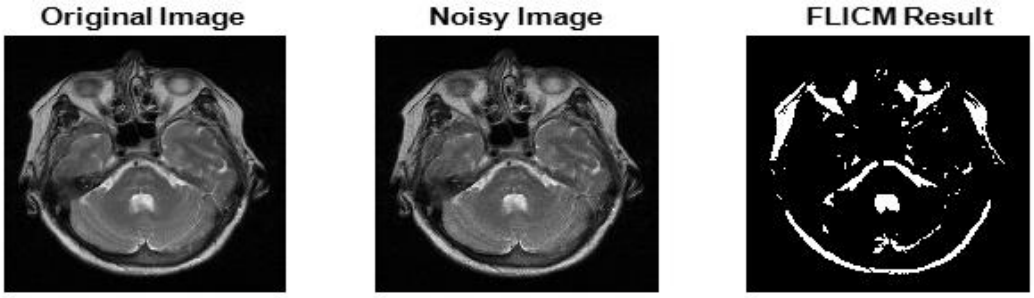

Fig.4.6 Segmentation Using FLICM Algorithm

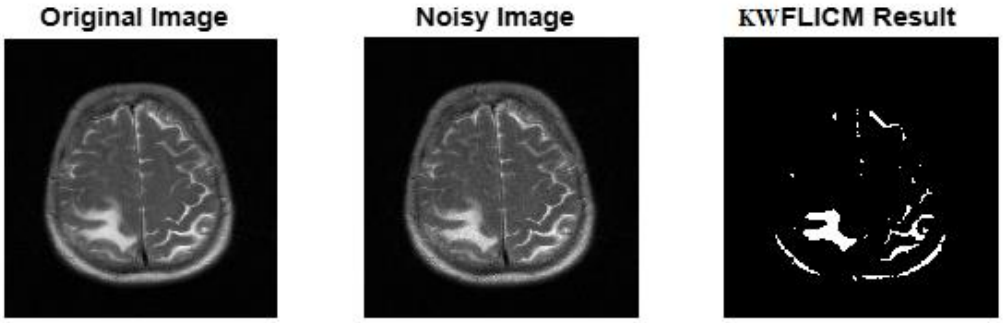

Fig. 4.7 Segmentation Using KWFLICM Algorithm.

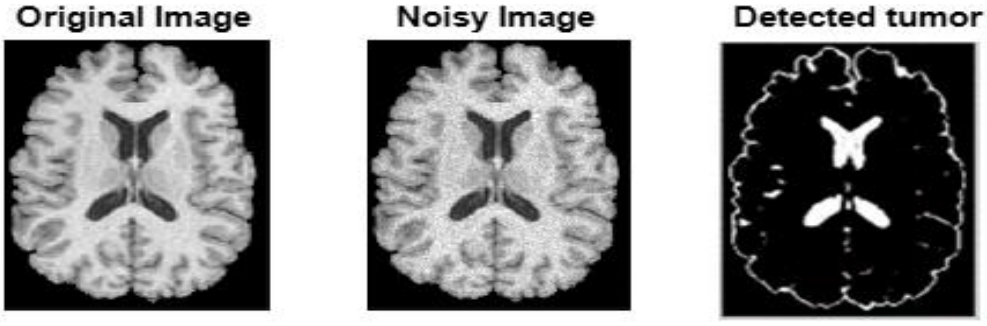

Fig4.8 Segmentation Using FRFCM Algorithm.

The experiments are carried out with the software MATLAB2018a. These results of segmentation using FRFCM is displayed in the Figure 4.8. with rician noise. It is depicted that the FRFCM gives better segmentation results than the other segmentation which is clearly visualized from the figure.

4.4 Segmentation performance evaluation

The performance result of segmentation is evaluated by segmentation accuracy (SA), and a quantitative index score(S) [102]. The Segment accuracy is given by

$$SA = \sum_{k=1}^{c} \frac{A_k \bigcap C_k}{\sum_{j=1}^{c} C_j} \tag{114}$$

And $$S = \sum_{k=1}^{c} \frac{A_k \bigcap C_k}{A_k \bigcup C_k} \tag{115}$$

Where c represents number of the cluster and A_k represents the set of pixels which belongs to the k^{th} class while C_k is the set of pixels in the Ground Truth.

Table 4.1 Segmentation Accuracy

Algorithm	Noise level	
	Rician Noise ($\sigma_n = 10$)	Rician Noise ($\sigma_n = 20$)
FCM	91.28	88.12
FCM S1	96.85	90.34
FCM S2	98.81	95.62
En FCM	98.72	96.84
FLIFCM	98.65	96.27
KWFLICM	98.82	97.01
NDFCM	99.92	98.61
FRFCM	**99.96**	**99.36**

4.4.1 Quality Measures

Apart from the above textural feature extraction, the following quality assessment parameters are also needed to ensure better result analysis on brain MR images.

(1) Structured Similarity Index (SSIM). The Structural Similarity Index (SSIM) is a perceptual metric that signifies that the degradation in image quality may be caused by data compression or losses in data transmission or by any other means of the image processing. It is defined as

$$SSIM = \left(\frac{\sigma_{xy}}{\sigma_x \sigma_y}\right)\left(\frac{2\bar{x}\bar{y}}{(\bar{x}^2)+(\bar{y}^2)+C_1}\right)\left(\frac{2\sigma_x\sigma_y}{(\sigma_x)^2+(\sigma_y)^2+C_2}\right) \quad (116)$$

A Higher value of SSIM indicates better preservation of luminance, contrast, and structural content.

(2) Peak Signal-to-Noise Ratio (PSNR) in dB: Peak signal to-noise ratio is a measure used to assess the quality of reconstruction of processed image and it is defined as

$$PSBR\,in\,dB = 20\log_{10}\frac{(2^n-1)}{MSE} \quad (117)$$

Lower value of MSE and higher value of PSNR indicate better signal-to-noise ratio.

To compare the performance of the different algorithms, two quality indexes are considered as Structural Similarity (SSIM) index and the Quality Index based on Local Variance (QILV). Both quality indexes provide structural similarity between the ground truth and the estimated images. However, SSIM is more sensitive to the noise level in the image and the QILV to blurring of the edges. In addition to the both, the PSNR (Peak signal to noise ratio) is also calculated. Table 4.2 shows the experimental results for two different values of $\sigma_n = 10$ and $\sigma_n = 20$. It is also observed that when $\sigma_n = 10$, the quality measure PSNR value is 37.32 dB in case of proposed FRFCM segmentation technique. The higher value of PSNR in case of IFRFCM indicate better signal-to noise ratio in the extracted image. Also, the larger value of SSIM indicates the noise reduction in the extracted image which is presented in Table 4.2.

Table: 4.2 Quality measures for the MR image with rician noise

	Rician Noise					
	$\sigma_n = 10$			$\sigma_n = 20$		
Algorithm	SSIM	QILV	PSNR(dB)	SSIM	QILV	PSNR(dB)
FCM	0.7632	0.6756	18.14	0.6912	0.6329	16.54
FCM S1	0.7859	0.7358	18.57	0.7146	0.6478	17.31
FCM S2	0.7893	0.7321	19.32	0.7722	0.7154	17.64
En FCM	0.7927	0.8148	20.54	0.7813	0.8023	19.11
FLIFCM	0.8449	0.9125	25.98	0.8182	0.8512	23.27
KWFLICM	0.8562	0.9234	27.32	0.8287	0.8624	24.72
NDFCM	0.8739	0.9482	29.89	0.8542	0.8178	26.83
FRFCM	**0.9219**	0.9781	**37.32**	**0.9027**	0.9127	**35.18**

4.4.2 Dataset Details

The dataset details are presented in Table 4.3. Out of the 5000 images dataset, 75% of the data has been utilized for training, 20 % of data for testing and rest of the images are used for validation.

Table 4.3 Brain Tumor MRI dataset Details

Type of MRI Images	Training	Validation	Testing
Benign and Malignant	3796	255	949
		Total Images: 5000	

4.5 Performance Measure of classifiers

Sensitivity, specificity, accuracy are the measure of system performance in classification of normal and abnormal brain tumor MR images [58]. Sensitivity shows the true positive rate in identifying the brain tumor, which calculates correctly classified number of abnormal images out of total number of abnormal brain MR images. Specificity shows the true negative rate in identifying the condition of normal brain, which calculates the number of normal brain MR images correctly classified out of the total number of normal MR images. The accuracy is the measurement of system effectiveness in conducting the whole classification, which calculates the total number of brain MR images that are correctly classified itself. The terms [105] utilized for performance measure evaluations are as follows:

TP= Number of Abnormal images correctly classified

TN= Number of Normal images correctly classified

FP= Number of Normal images classified as Abnormal

FN= Number of Abnormal images classified as Normal

$$Sensitivity = \frac{TP}{TP + FN}, Specificity = \frac{TN}{TN + FP}, \ Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Table 4.4 Performance measure of different classifiers - 1

Classifier	Sensitivity	Specificity	Accuracy in (%)
RVM	0.96	0.87	94.60
SB-RVM	0.98	0.92	96.86
ELM	0.99	0.94	98.82
SB-ELM	0.98	1.0	**96.85**

Table 4.5 Performance measure of different classifiers 2

Classifier	Sensitivity	Specificity	Accuracy in (%)
SVM	0.98	0.94	96.26
RVM	0.96	0.87	94.60
ELM	0.99	0.94	98.82
ELM+PSO	0.99	1.0	**99.41**

Table 4.6 5×5-fold cross validation procedure for each run

Run	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Total	Accuracy (%)
Run-1	51	51	50	51	50	253	99.2591
Run-2	51	51	51	51	51	255	100
Run-3	51	51	50	51	50	253	99.2591
Run-4	51	51	50	51	50	253	99.2591
Run-5	51	51	50	51	50	253	99.2591
Final result							**99.4078**

Table 4.7 5×5-fold cross validation procedure of Run-1

Fold	Test instances	TP	FN	TN	FP	Accuracy (%)
Fold -1	51	44	0	7	0	100
Fold -2	51	44	0	7	0	100
Fold -3	51	43	1	7	0	98.229
Fold -4	51	44	0	7	0	100
Fold -5	51	43	1	7	0	98.229
Final result						**99.2591**

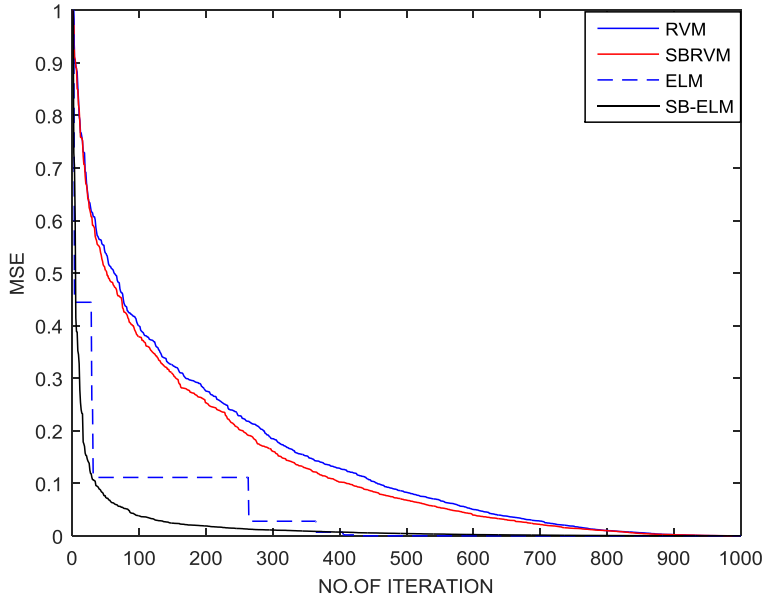

Fig .4.9 Mean Squared Error Convergence -1

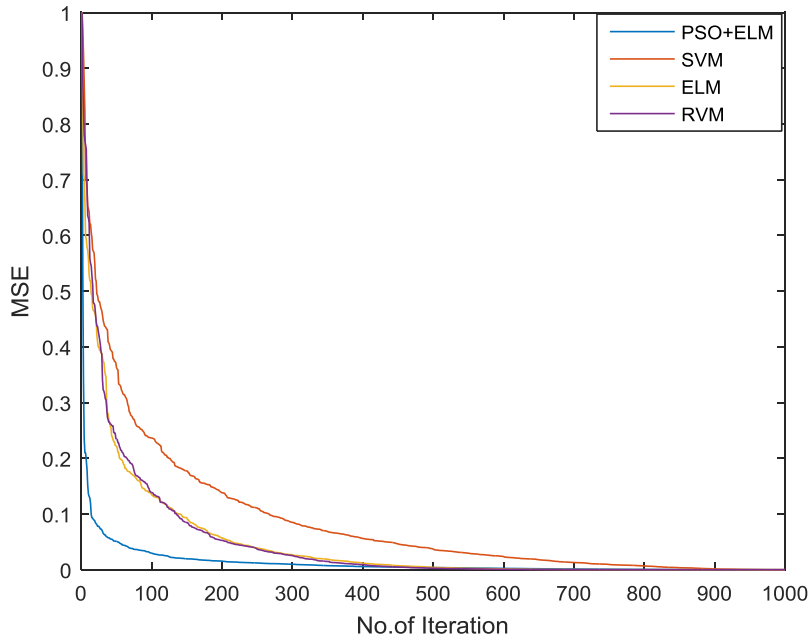

Fig .4.10 Mean Squared Error Convergence -2

Computation time play an important role for evaluation of a classifier. The computation time of the proposed ELM+PSO method is calculated as 8.4127 sec. Further the computation time for SVM, RVM, ELM has been calculated as 16.2564 sec, 12.4125 sec, and 10.3234 sec.

4.6 GUI (Graphical User Interface) for automatic visualization of detection and classification through Using Raspberry Pi3.

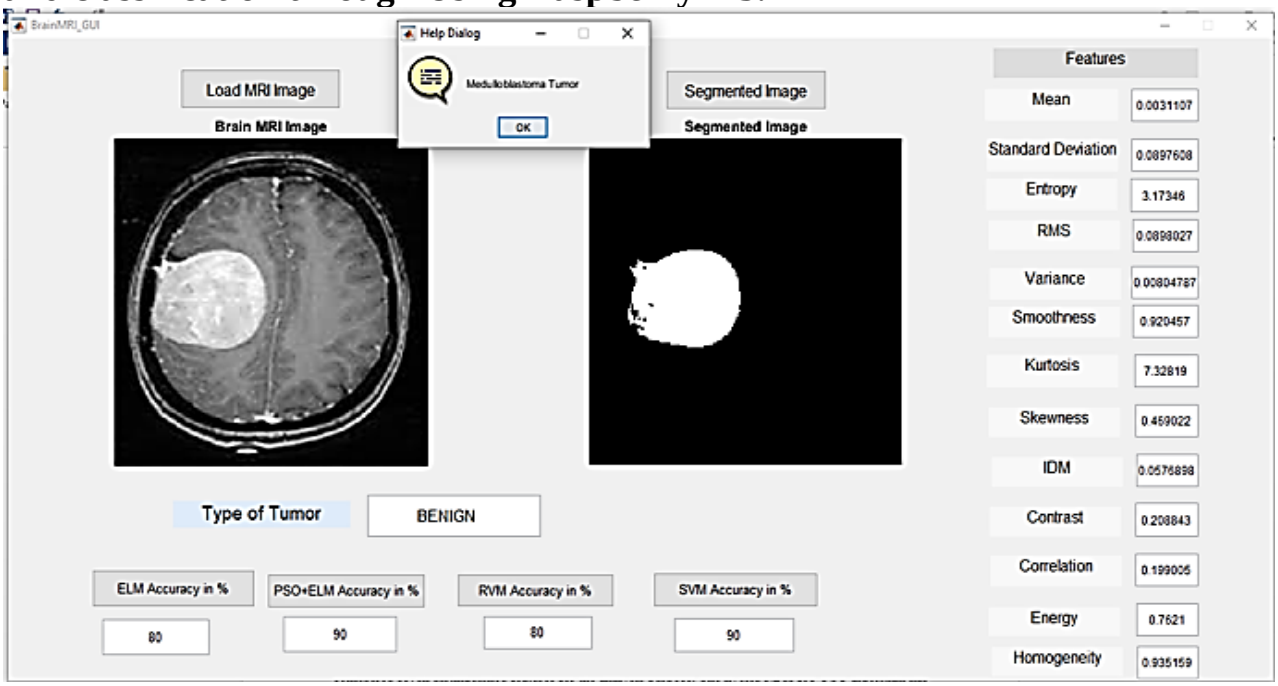

Fig. 4.11 Segmentation and classification result using GUI (Benign-1)

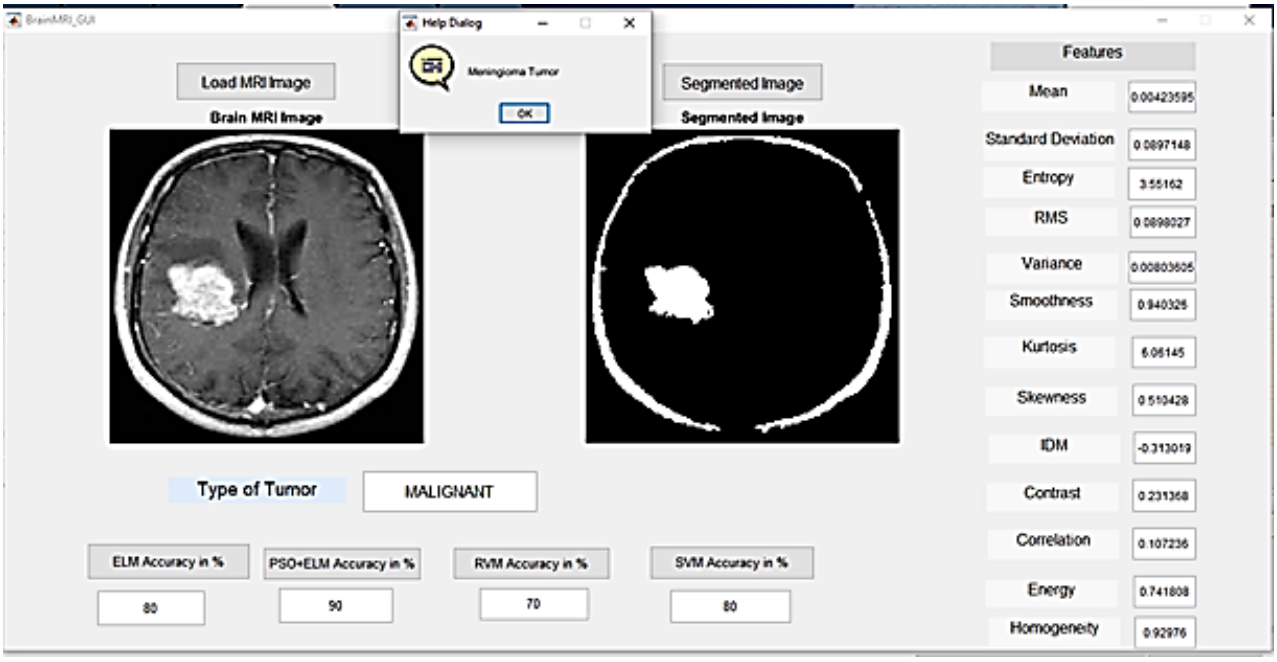

Fig. 4.12 Segmentation and classification result using GUI (Malignant-1)

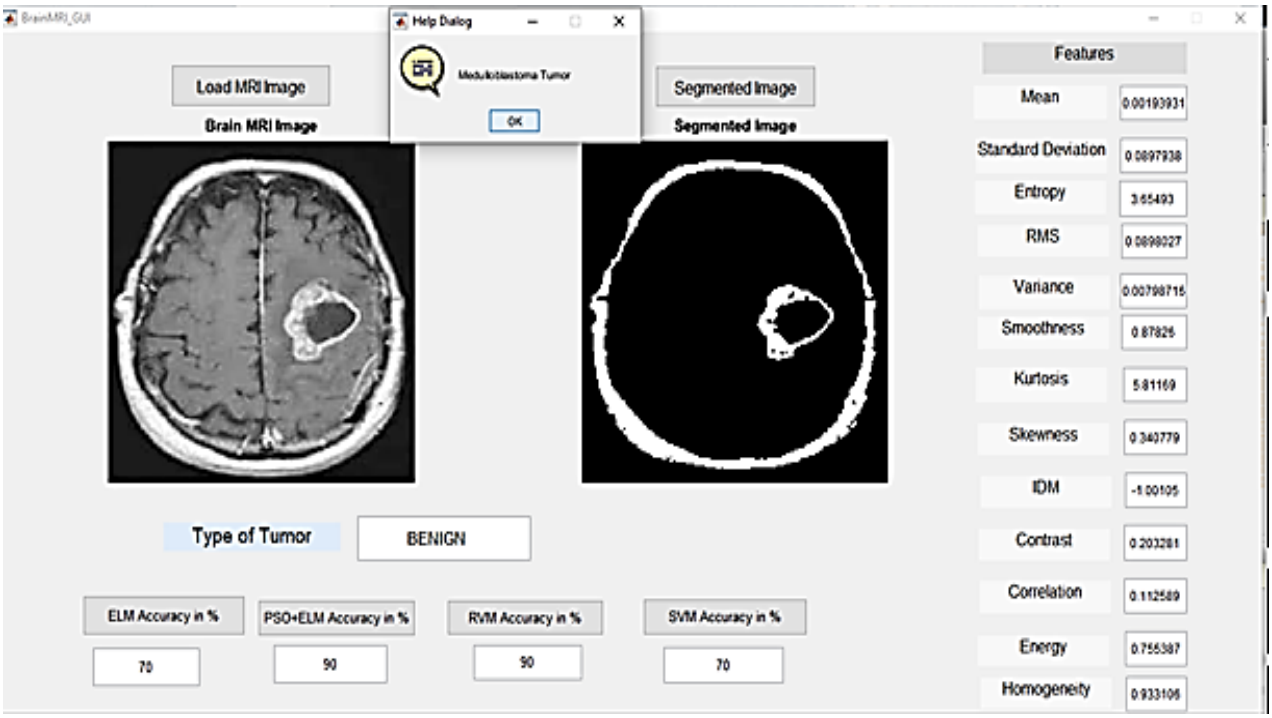

Fig. 4.13 Segmentation and classification result using GUI (Benign-2)

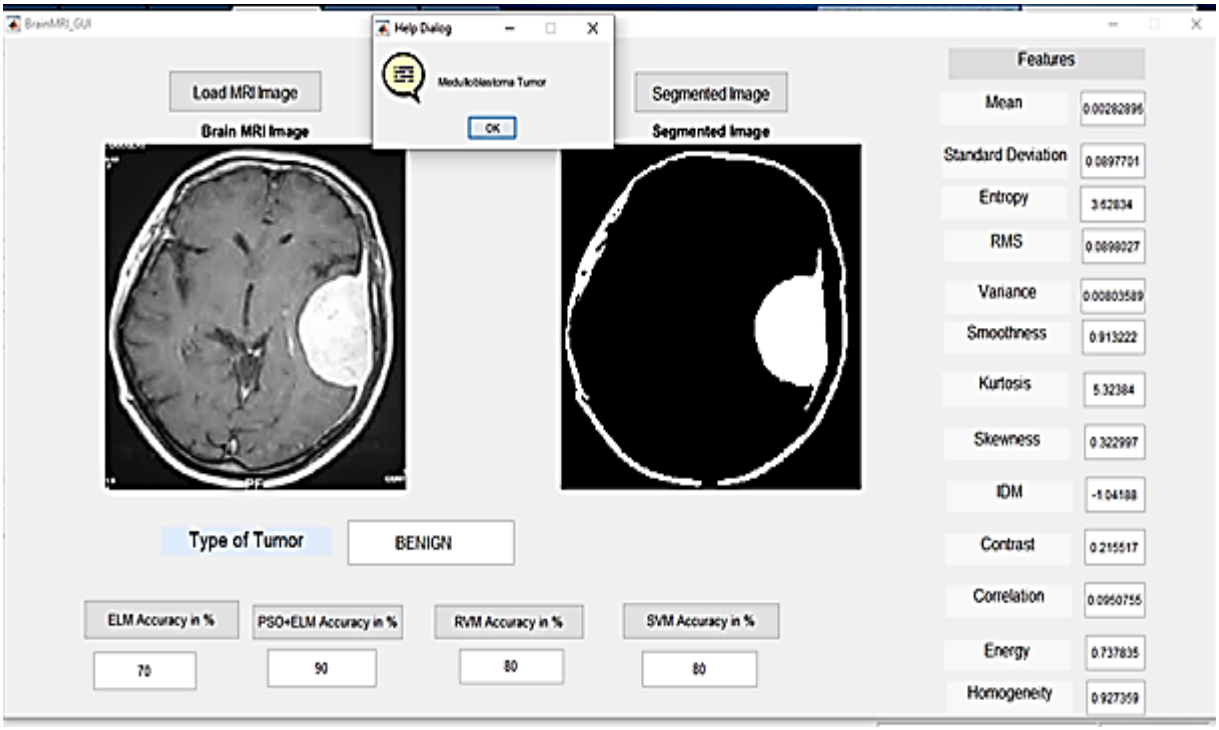

Fig. 4.14 Segmentation and classification result using GUI(Benign-3)

4.7 Set up for classification using raspberry pi3 B+ (Hardware Implementation Step)

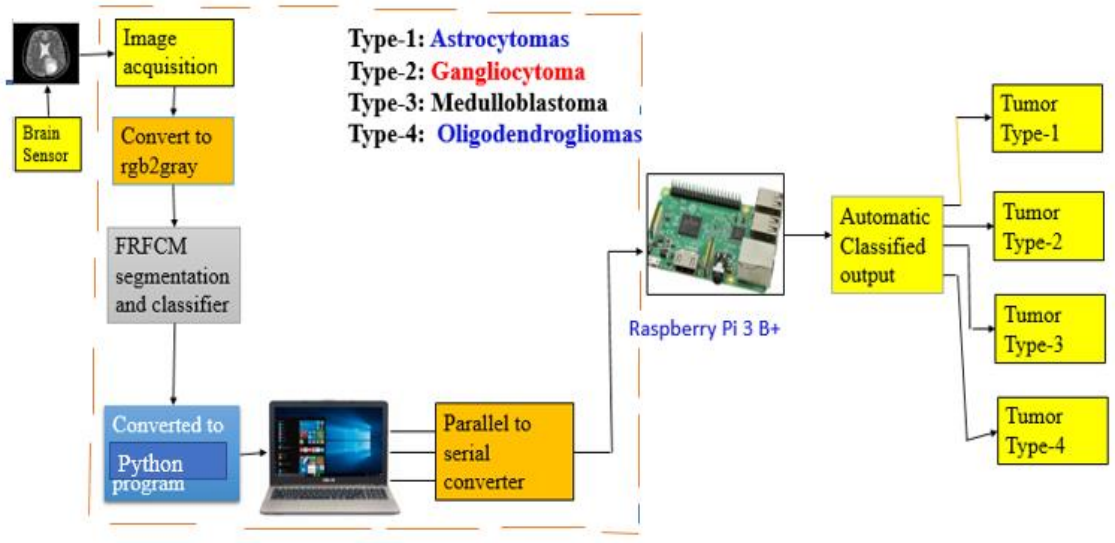

Fig.4.15 Hardware implementation step procedure

Astrocytoma is a type of cancer that can occur in the brain or spinal cord. Medulloblastomas starts in the lower back part of the brain and tends to spread through the spinal fluid.

Gangliocytoma is a rare type of central nervous system (CNS) tumor and may occur in all age groups. Oligodendrogliomas are a type of glioma that are believed to originate from the glial precursor cell.

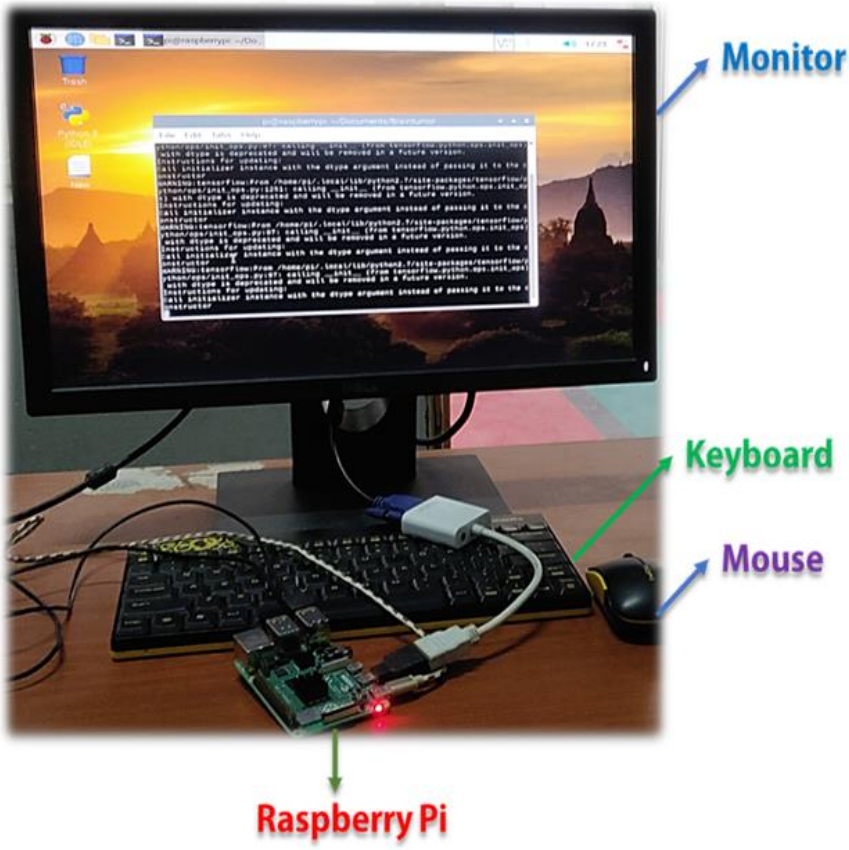

Fig.4.16 Hardware implementation with Raspberry PI3 B+

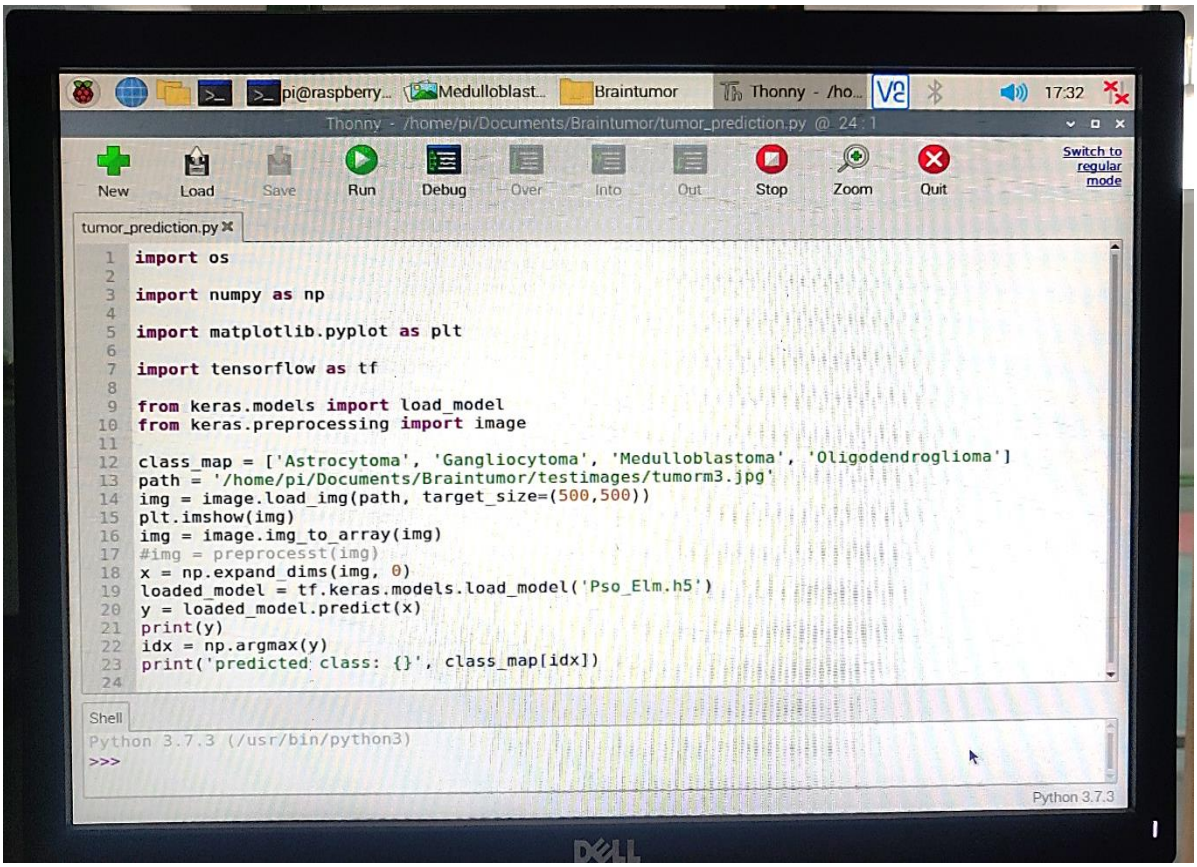

Fig.4.17 Hardware with python code with Raspberry PI3 B+

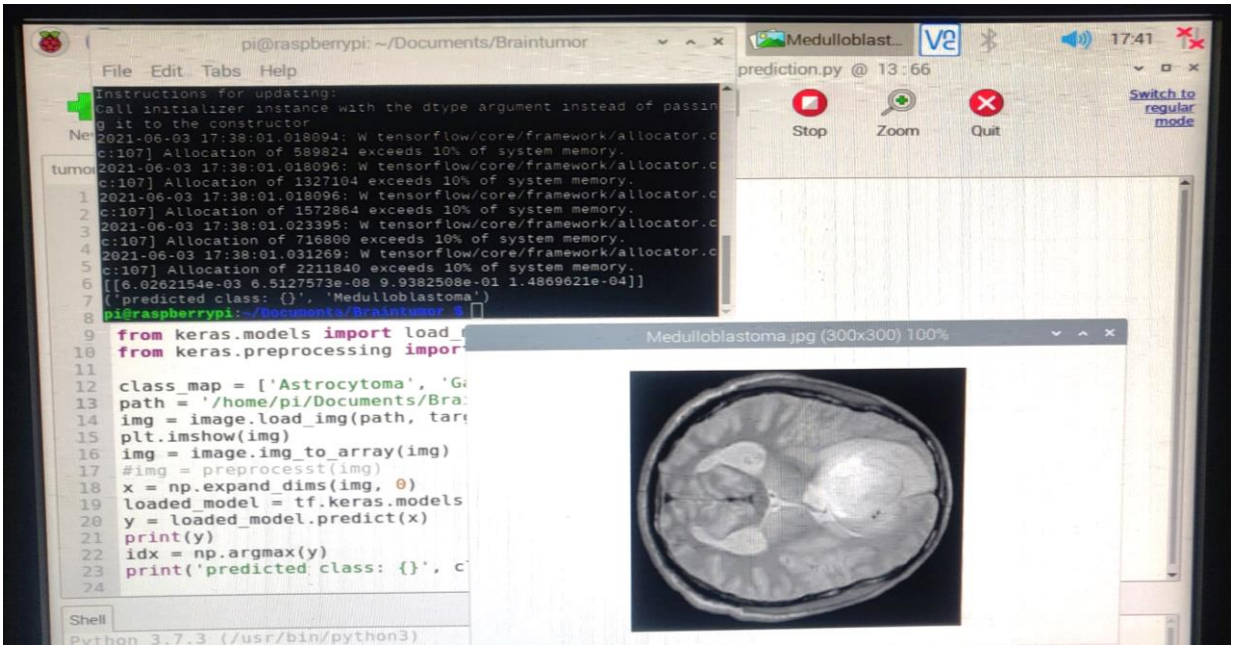

Fig.4.18 Medulloblastomas classification **with Raspberry PI3 B+**

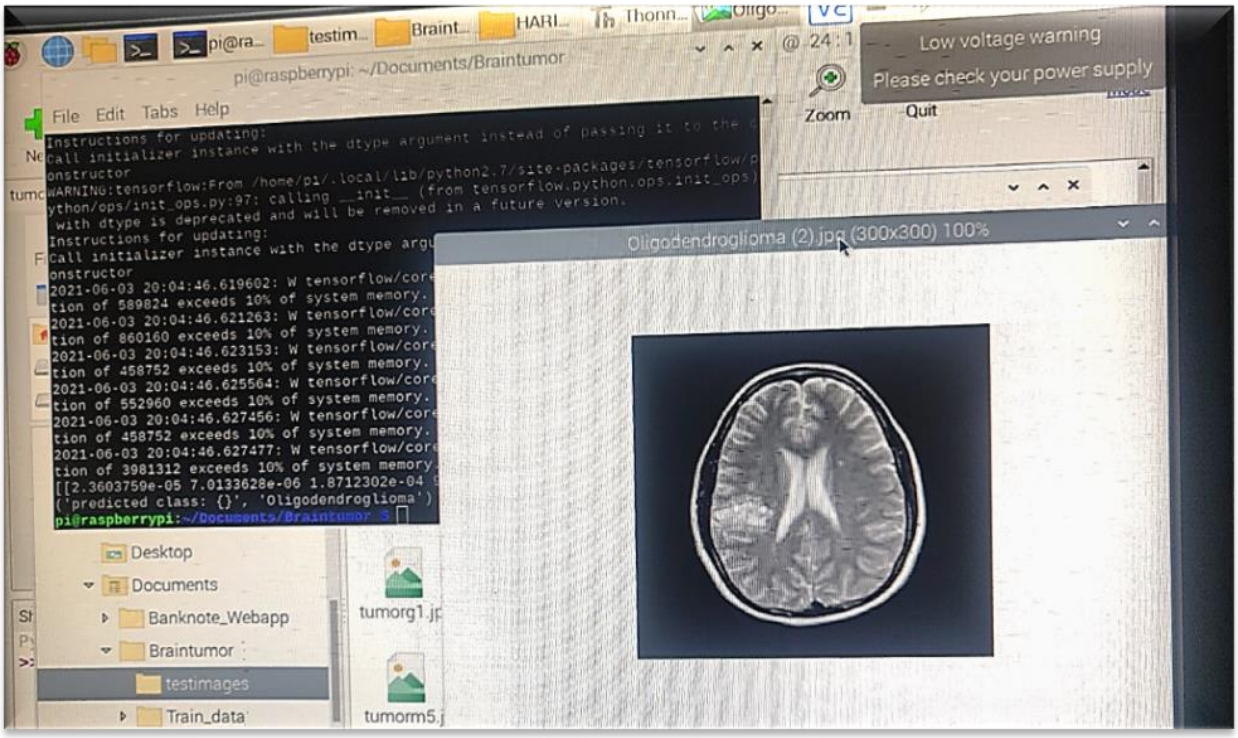

Fig.4.19 Oligodendrogliomas classification implementation with Raspberry PI3 B+

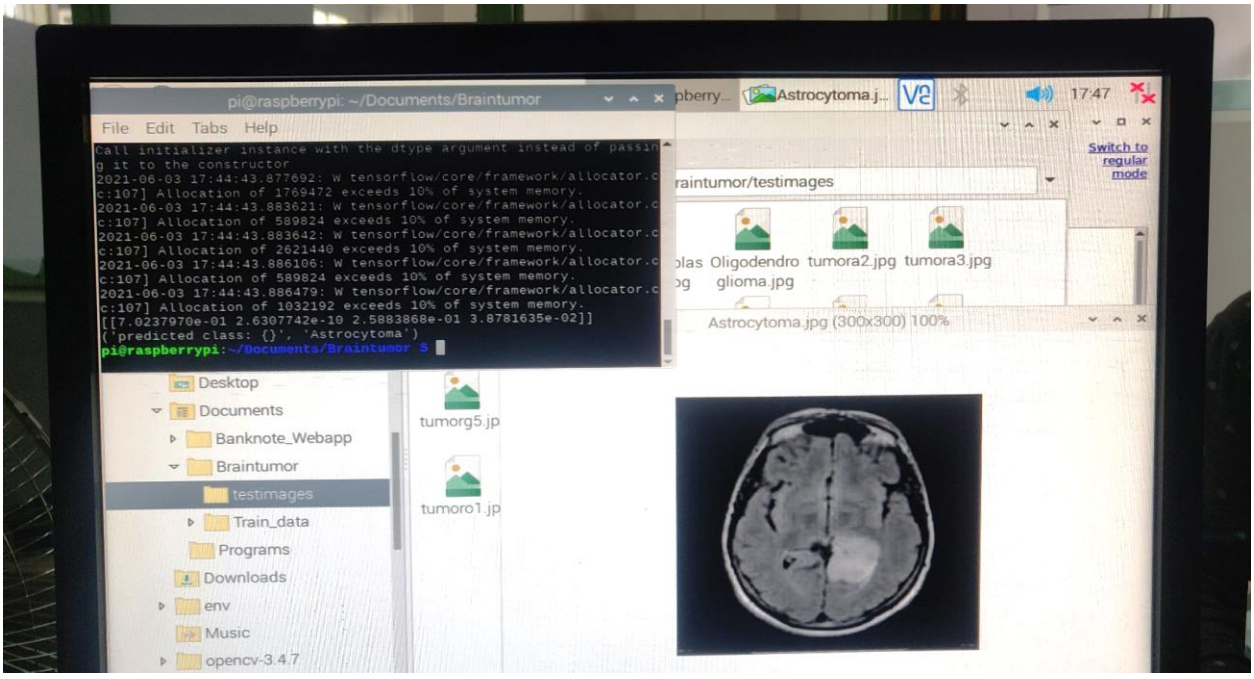

Fig.4.20 Astrocytoma classification implementation with Raspberry PI3 B+

4.7 Discussion

The proposed improved FRFCM segmentation technique provides noise free magnetic resonance images for tumor detection. The segmentation accuracy by the proposed FRFCM is better than the other FCM based algorithm is shown in Table 4.1. In some cases both the segmentation processes NDFCM and FRFCM shown the nearly similar results, still the proposed FRFCM segmentation is preferred due to robust capability of reduction of rician noise. Table 4.2 shows the quality measures containing SSIM and PSNR values. The higher value of PSNR shows the noise reduction from the images. Table 4.3 shows the dataset details of the brain tumor images. In case of classification, it is found that the proposed modified ELM is better than other classifiers in terms of sensitivity, specificity and accuracy. Table 4.4 and Table 4.5 shows the performance measure of all classifiers. The accuracy obtained by utilizing RVM, SB-RVM,ELM and SB-ELM are 94.60%, 96.86%, 98.82%, 96.85% respectively. The SVM, RVM, ELM, ELM+PSO accuracies are 96.26%, 94.60%, 98.82%, 99.41% respectively. The ELM-PSO model is simple in construction and free from complex mathematical calculations in comparison to other methods. Table 4.6 and Table 4.7 shows the 5x5 validation accuracy for all runs and a single run.

Computation time play an important role for evaluation of a classifier. At first, the computation time of each stage of the proposed ELM+PSO method is recorded, and finally the average value is calculated as 8.4127 sec. Further the computation time for SVM, RVM, ELM for each stage has been calculated and the average value is obtained as 16.2564 sec, 12.4125 sec, and 10.3234 sec. It is observed that the proposed ELM+PSO method shows faster convergence than the other mentioned algorithms. Further, the errors of the classifiers are expressed in terms of mean squared error (MSE) value shown in Fig.4.9 and Fig.4.10. The performance rate of proposed ELM+PSO classifier depends on the convergence parameters setting. It is observed from the Fig.4.9 that the proposed SB-ELM model takes near about 340 iterations to converge. Fig.4.10 shows that the SVM model takes near about 790 iterations, whereas the RVM and ELM model takes 430 and 400 iterations to converge. It is found from the mean squared error results that, the accuracy in ELM and ELM+PSO are nearly similar, but the convergence is faster in the case of ELM+PSO classifier.Fig.4.11 to Fig.4.14 shows the GUI model outputs of segmentation along with feature values and classification accuracy for different test brain tumor images. The set up

for hardware implementation has been shown in Fig.4.15. Further, the Raspberry PI B+ along with the screen has been shown in Fig.4.16. Fig.4.17 shows the python program for PSO+ELM algorithm for classification. Fig.4.18 to Fig. 4.20 shows the classification of four types of tumor for visual localization.

From these results and discussions, it is evident that the proposed PSO algorithm provides superior optimized results with respect to accuracy and computational time. With the help of proposed segmentation technique, hybrid PSO algorithm and statistical textural features the brain tumor images were classified into cancerous and non-cancerous tumors. The performance of proposed technique of segmentation and textural features were found to be very useful to justify the performance of the modified PSO based ELM classifier during training and testing.

CHAPTER FIVE

5. CONCLUSION AND RECOMMENDATION

5.1 Conclusion

In this research work, the brain magnetic resonance images are used for the purpose of segmentation and classification. To remove rician noise and smoothen the image, an improved fast and robust FCM based segmentation technique has been employed. After segmentation, the texture features are extracted from magnetic resonance images using GLCM feature extraction technique. We have considered seven distinguished features for the analysis purpose. The PSO algorithm has been proposed for updating of weights of ELM model. The extracted features are given as input to the PSO based ELM classifier for classification of benign and malignant brain tumors from magnetic resonance images and compared with SVM, RVM, ELM classifiers. The accuracy of the classifiers SVM, RVM, ELM and PSO+ELM has been shown in Table 4.5. The proposed classifier model has shown good potentiality in classifying the tumor into cancerous and non-cancerous brain tumors. By employing feature reduction method with our proposed technique may provide better performance accuracy. The results presented in this research work shows uniqueness of the model and comparison results also depict clear classification accuracies. Dataset has been collected from hospitals of Ethiopia, Harvard medical school of architecture for the purpose of segmentation and classification. The results presented by utilizing the proposed PSO based ELM model is suitable automatic classification of cancerous and non-cancerous brain tumors and may help clinical diagnosis process by the radiologists or clinical experts. Further the GUI model and hardware setup simplify the classification, segmentation and feature extraction in platform which is easier for the radiologist to analyze the type of tumor.

5.2 Recommendation

In the future research work, different hybrid algorithm such as BAT optimization algorithm with PSO and the harmony search for weight optimization of classifier, deep learning method, and feature reduction method with PSO-ELM for classification will be used to increase the performance accuracy. Also, the advanced textures features can be considered except the mentioned texture features for the classification. Further, more efficient segmentation techniques with morphological reconstruction and membership filtering can be used for large dataset of magnetic resonance images.

REFERENCE

1. Webpage. https:// consumer.healthday.com /cancer-information5/ brain-cancer-news-93 /brain- cancers-both-common-and-deadlyamong-teens-young-adults-report-708339.html
2. Cancer.Net Editorial Board. (2019). Brain Tumor: Statistics. Retrieved from https://www.cancer.net/cancer-types/braintumor/statistics
3. Ding Y, Fu X (2016) "Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm". Neurocomputing 188:233–238. https://doi.org/10.1016/ j.neucom.2015 .01. 106
4. Pereira DC, Ramos RP, do Nascimento MZ (2014) "Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm". Comput Methods Programs Biomed 114(1):88–101. https://doi.org/10.1016/j.cmpb.2014.01.014
5. Mahapatra D (2017) "Semi-supervised learning and graph cuts for consensus based medical image segmentation". Pattern Recognit 63:700709. https:// doi.org /10.1016 / j.patcog. 2016.09.030
6. Bahadure NB, Ray AK, Thethi HP (2017) "Image analysis for MRI based brain tumor detection and feature extraction using biologically inspired BWT and SVM". Hindawi Int J Biomed Imaging 2017, Article ID 9749108. https://doi.org/10.1155/2017/9749108
7. Satheeskumaran S, Sabrigiriraj M (2014) "A new LMS based noise removal and DWT based R-peak detection in ECG signal for biotelemetry applications". Natl Acad Sci Lett 37(4):341–349. https://doi.org/10.1007/s40009-014-0238-3
8. Shanmuga Priya S, Valarmathi A (2018) "Efficient fuzzy c-means based multilevel image segmentation for brain tumor detection in MR images". In: Design automation for embeded system. Springer, Berlin. https://doi.org/10.1007/s10617-017-9200-1. ISSN: 1572-8080
9. Javed A, Kim YC, Khoo MCK, Ward SLD, Nayak KS (2016) "Dynamic 3-D MR visualization and detection of upper airway obstruction during sleep using region-growing segmentation". IEEE Trans Biomed Eng 63(2):431–437. https://doi.org/ 10.1109/ TBME.2015.2462750.

10. Abd-Ellah MK, Awad AI, Khalaf AM, Hamed FA (2016) “Design and implementation of a computer-aided diagnosis system for brain tumor classification”. In: 28th international conference on microelectronics (ICM), Cairo, pp 73–76
11. Li Z, Chen J (2015) “Super pixel segmentation using linear spectral clustering”. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Boston, pp 1356–1363
12. Nandi AK, Basel AJ, Rui F (2015) “Integrative cluster analysis in bioinformatics”. Wiley, Berlin
13. Demirhan A, Güler I (2011) “Combining stationary wavelet transform and self-organizing maps for brain MR image segmentation”. Eng Appl Artif Intell 24:358–367. https://doi.org/10.1016/j.engappai.2010.09.008
14. Shree NV, Kumar TNR (2018) “Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network”. Brain Inform 5:23–30. https://doi.org/10.1007/s40708-017-0075-5
15. Chatzis SP, Varvarigou TA (2008) “A fuzzy clustering approach toward hidden markov random field models for enhanced spatially constrained image segmentation”. IEEE Trans Fuzzy Syst 16(5):1351– 1361. https://doi.org/10.1109/TFUZZ.2008.2005008
16. Lei T, Jia X, Zhang Y, He L, Meng H, Nandi AK (2018) “Significantly fast and robust fuzzy c-means clustering algorithm based on morphological reconstruction and membership filtering”. IEEE Trans Fuzzy Syst 26(5):3027–3041. https:// doi.org /10.1109/ tfuzz. 2018.2796074
17. Issa M, Hassanien AE, Oliva D, Helmi A, Ziedan I, Alzohairy A (2018) “ASCA–PSO: adaptive sine cosine optimization algorithm integrated with particle swarm for pairwise local sequence alignment”. Expert Syst Appl 99(1):56–70. https://doi.org /10.1016 /j.eswa. 2018.01.019
18. Ahmed MN, Yamany SM, Mohamed N, Farag AA, Moriarty T (2002) “A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data”. IEEE Trans Med Imaging 21(3):193– 199. https://doi.org/10.1109/42.996338
19. Chen S, Zhang D (2004) “Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure”. IEEE Trans Syst Man Cybern B Cybern 34(4):1907–1916. https://doi.org/10.1109/tsmcb.2004.831165

20. Szilagyi L, Benyo Z, Szilagyii SM, Adam HS (2003) “MR brain image segmentation using an enhanced fuzzy c-means algorithm”. In: Proceeding of the 25th annual international conference of the IEEE EMBS, pp 17–21
21. Cai W, Chen S, Zhang D (2007) “Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation”. Pattern Recognit 40(3):825–838. https://doi.org/10.1016/j.patcog.2006.07.011
22. Krinidis S, Chatzis V (2010) “A robust fuzzy local information cmeans clustering algorithm”. IEEE Trans Image Process 19(5):1328– 1337. https://doi.org/ 10.1109/ tip. 2010.2040763 AQ7
23. Gong M, Zhou Z, Ma J (2012) “Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering”. IEEE Trans Image Process 21(4):2141–2151. https://doi.org/10.1109/TIP.2011.2170702
24. Gong M, Liang Y, Shi S, Ma J (2013) “Fuzzy c-means clustering with local information and kernel metric for image segmentation”. IEEE Trans Image Process 22(2):573–584. https://doi.org/10.1109/TIP.2012.2219547
25. Guo F, Wang X, Shen J (2016) “Adaptive fuzzy c-means algorithm based on local noise detecting for image segmentation”. IET Image Process 10(4):272–279. https://doi.org/10.1049/iet-ipr.2015.0236
26. Rezaei K, Agahi H (2017) “Malignant and benign brain tumor segmentation and classification using SVM with weighted kernel width”. Sig Image Proc Int J (SIPIJ). https://doi.org/10.5121/sipij.2017.8203
27. Torheim T, Malinen E, Kvaal K (2014) “Classification of dynamic contrast enhanced MR images of cervical cancers using texture analysis and support vector machines”. IEEE Trans Med Imaging 33(8):1648–1656. https://doi.org/10.1109/TMI.2014.2321024
28. Lang R, Zhao L, Jia K (2016) “Brain tumor image segmentation based on convolution neural network”. In: 2016 9th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI), Datong, pp 1402–1406
29. Deepa SN, Arunadevi B (2013) “Extreme learning machine for classification of brain tumor in 3D MR images”. Informatologia 46(2):111–121. ISSN 1330-0067
30. Krishna TG, Sunitha KVN, Mishra S (2018) “Detection and classification of brain tumor from MRI medical image using wavelet transform and PSO based LLRBFNN

algorithm". Int J Comput Sci Eng 6(1). https://doi.org/10.26438/ijcse/v6i1.1823. E-ISSN: 2347-2693

31. Nayak PK, Mishra S, Dash PK, Bisoi Ranjeeta (2016) "Comparison of modified teaching–learning-based optimization and extreme learning machine for classification of multiple power signal disturbances". Neural Comput Appl 27(7):2107–2122. https://doi.org /10.1007 /s00521-0152010-0
32. Patra A, Das S, Mishra SN, Senapati MR (2017) "An adaptive local linear optimized radial basis functional neural network model for financial time series prediction". Neural Comput Appl 28(1):101–110. https://doi.org/10.1007/s00521-015-2039-0
33. Liu B, Wang L, Jin YH (2007) "An effective PSO-based mimetic algorithm for flow shop scheduling". IEEE Trans Syst Man Cybern B Cybern 37(1):18–27. https://doi.org/10.1109/tsmcb.2006.883272
34. Senapati MR, Vijaya I, Dash PK (2007) "Rule extraction by training radial basis functional neural network with particle swarm optimization". Am J Sci 3(8):592–599. ISSN: 1549-3636
35. X-S Yang, S Deb, S Fong (2011) "Accelerated particle swarm optimization and support vector machine for business optimization and applications". In: International conference on networked digital technologies, NDT 2011. Communications in computer and information science, vol 136, pp 53–66. Springer, Berlin
36. Kaur T, Saini BS, Gupta S (2016) "Optimized multi threshold brain tumor image segmentation using two dimensional minimum cross entropy based on co-occurrence matrix". In: Medical imaging in clinical applications. Part of the studies in computational intelligence, vol 651. Springer, Berlin, pp 461–486. https://doi.org/10.1007/978-3-319-337937_20
37. Garg H (2016) "A hybrid PSO–GA algorithm for constrained optimization problems". Appl Math Comput 274(1):292–305. https://doi.org/10.1016/j.amc.2015.11.001
38. de Fátima Araújoa T, Uturbey W (2013) "Performance assessment of PSO, DE and hybrid PSO–DE algorithms when applied to the dispatch of generation and demand". Int J Electr Power Energy Syst 47:205–217. https://doi.org/10.1016/j.ijepes.2012.11.002
39. Santra D, Mukherjee A, Sarker K, Chatterjee D (Oct 2016) "Hybrid PSO–ACO algorithm to solve economic load dispatch problem with transmission loss for small scale power

system". In: 2016 international conference on intelligent control power and instrumentation (ICICPI), pp 21–23

40. https:// www.mayoclinic.org /diseases-conditions/ brain-tumor/symptoms-causes/syc-20350084
41. Sajjad S., Hanan Abdullah A., Sharif M., and Mohsin S., "Psychotherapy through video game to target illness related problematic behaviors of children with brain tumor," Current Medical Imaging Reviews, vol. 10, pp. 62-72, 2014.
42. Amin J., Sharif M., Yasmin M., and Fernandes S. L., "A distinctive approach in brain tumor detection and classification using MRI," Pattern Recognition Letters, 2017.
43. Rajinikanth V., Satapathy S. C., Fernandes S. L., and Nachiappan S., "Entropy based segmentation of tumor from brain MR images–a study with teaching learning based optimization," Pattern Recognition Letters, vol. 94, pp. 87-95, 2017.
44. Rajinikanth V., Fernandes S. L., Bhushan B., and Sunder N. R., "Segmentation and analysis of brain tumor using Tsallis entropy and regularised level set," in Proceedings of 2nd International Conference on Micro-Electronics, Electromagnetics and Telecommunications, 2018, pp. 313-321.
45. DeAngelis L. M., "Brain tumors," New England Journal of Medicine, vol. 344, pp. 114-123, 2001.
46. Naik J. and Patel S., "Tumor detection and classification using decision tree in brain MRI," International Journal of Computer Science and Network Security (IJCSNS), vol. 14, p. 87, 2014.
47. Acharya U. R., Fernandes S. L., WeiKoh J. E., Ciaccio E. J., Fabell M. K. M., Tanik U. J., Rajinikanth V., and Yeong C. H., "Automated Detection of Alzheimer's Disease Using Brain MRI Images–A Study with Various Feature Extraction Techniques," Journal of Medical Systems, vol. 43, p. 302, 2019.
48. Fernandes S. L., Tanik U. J., Rajinikanth V., and Karthik K. A., "A reliable framework for accurate brain image examination and treatment planning based on early diagnosis support for clinicians," Neural Computing and Applications, pp. 1-12, 2019.
49. Gatenby R. A., Grove O., and. Gillies R. J, "Quantitative imaging in cancer evolution and ecology," Radiology, vol. 269, pp. 8-14, 2013.

50. Amin J., Sharif M., Anjum M. A., Raza M., and Bukhari S. A. C., "Convolutional neural network with batch normalization for glioma and stroke lesion detection using MRI," Cognitive Systems Research, vol. 59, pp. 304-311, 2020.
51. Bauer S. , "A survey of MRI-based medical image analysis for brain tumor studies," Phys. Med. Biol., vol. 58, no. 13, pp. 97–129, 2013.
52. Louis D. N. , "The 2007 who classification of tumors of the central nervous system," Acta Neuropathologica, vol. 114, no. 2, pp. 97–109, 2007.
53. Van Meir E. G. , "Exciting new advances in neuro-oncology: The avenue to a cure for malignant glioma," CA, Cancer J. Clinicians, vol. 60, no. 3, pp. 166–193, 2010.
54. Tabatabai G. , "Molecular diagnostics of gliomas: The clinical perspective," Acta Neuropathologica, vol. 120, no. 5, pp. 585–592, 2010.
55. Menze B. , "The multimodal brain tumor image segmentation benchmark (BRATS)," IEEE Trans. Med. Imag., vol. 34, no. 10, pp. 1993–2024, Oct. 2015.
56. Tustison N. J. , "N4ITK: Improved n3 bias correction," IEEE Trans. Med. Imag., vol. 29, no. 6, pp. 1310–1320, Jun. 2010.
57. Nyúl L. G., Udupa J. K., and Zhang X., "New variants of a method of MRI scale standardization," IEEE Trans. Med. Imag., vol. 19, no. 2, pp. 143–150, Feb. 2000.
58. Silvia Franchini, Antonio Gentile, Filippo Sorbello, Giorgio Vassallo, and Salvatore Vitabile, "Conformal ALU: A Conformal Geometric Algebra Coprocessor for Medical Image Processing", IEEE Transactions on Computers, Vol. 64, No. 4, April 2015.
59. Shang-Ling Jui, Shichen Zhang, Weilun Xiong, Fangxiaoqi Yu, Mingjian Fu, and Dongmei Wang, "Brain MRI Tumor Segmentation with 3D Intracranial Structure Deformation Features", IEEE Intelligent Systems,V.31 n.2, p.66-76, March 2016.
60. Bauer S , " A Survey of MRI-based medical image analysis for brain tumor studies", Phys. Med. Biol., vol. 58, no. 13, pp. 97-129, 2013.
61. Wang Y., "Efficient Stockwell Transform with Applications to Image Processing", University of Waterloo, Waterloo, Ontario, Canada, 2011.
62. Ladan J., "An Analysis of Stockwell Transforms, With Applications to Image Processing", University of Waterloo, Ontario, Canada, 2014.
63. Singh Choudhry M., Kapoor Rajiv "Performance Analysis of Fuzzy C-Means Clustering Methods for MRI Image Segmentation" Twelfth International Multi-Conference on

Information Processing-2016 (IMCIP-2016), Science Direct, Procedia Computer Science 89 (2016) 749 – 758.

64. Chen Songcan, Zhang Daoqiang, "Robust Image Segmentation using FCM with Spatial Constraints Based on New Kernel-Induced Distance Measure", IEEE Transaction on Systems, Man and Cybernetics-Part B, vol. 34(4), August (2004).
65. Kannana S. R., Ramathilagam S., Devia R. and Hines E., Strong "Fuzzy C-Means in Medical Image Data Analysis", The Journal of Systems and Software, pp. 2425–2438, (2012).
66. Zexuan Ji A., Yong Xiab, Chena Qiang, Quansen Suna, Deshen Xiaa and David Dagan Feng, "Fuzzy C-Means Clustering with Weighted Image Patch for Image Segmentation", Applied Soft Computing, vol. 12, pp. 1659–1667, (2012).
67. Hoel Le Capitaine and Carl Frélicot , "A fast fuzzy c-means algorithm for color image segmentation" EUSFLAT-LFA 2011 July 2011 Aix-les-Bains, France.
68. Ruspini E. H.. "A new approach to clustering. Information and control", 15(1):22–32, 1969
69. Jain A., Murty M., and Flynn P.. "Data clustering: A review. ACM Computing Surveys", Vol.31, pp.264–323, 1999.
70. Saravanan Thirumuruganathan "A Detailed Introduction to K-Nearest Neighbor (KNN) Algorithm book" May 17, 2010.
71. Ramzi A.Haraty,Mohamad Dimishkieh,1 and MehediMasud, "An Enhanced k-Means Clustering Algorithm for Pattern Discovery in Healthcare Data", International Journal of Distributed Sensor Networks ,Vol.1, 2015, Article ID 615740, 11 pages.
72. Khan S. K. and Ahmad A.. "Cluster center initialization algorithm for K-Means clustering", Pattern Recognition Letters,25:1293–1302, 2004.
73. Jain A. K. and Dubes R. C., "Algorithms for Clustering Data", Prentice Hall, Upper Saddle River, NJ, USA,1988.
74. Mac Queen J.B, "Some methods for classification and analysis of multivariate observations," in Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, pp.281– 297, University of California Press,1967.
75. .Hartigan J.A and Wong M.A., "Algorithm AS136: a k-means clustering algorithm," Journal of the Royal Statistical Society, SeriesC,vol.28, no.1,pp.100–108,1979.

76. Bottou L. and Bengio Y., "Convergence properties of the k means algorithms," in Advances in Neural Information Processing Systems, pp.585– 592, MIT Press, New York, NY, USA, 1995.
77. Heba Mohsen, "Classification using Deep Learning Neural Networks for Brain Tumors", Future Computing and Informatics, pp 1-4 (2017).
78. Shen W, Guo X, Wu C, and Wu D, Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm, In: Knowl-Based Syst, Vol. 24, No. 3, 2011, pp. 378-385.
79. Sun YF, Liang YC, Zhang WL, Lee HP, and Lin WZ, "Optimal partition algorithm for the RBF neural network for financial time series forecasting", In: Neural Comput. Appl, Vol. 14, No. 1, 2005, pp. 36-44.
80. Kumar, P. & Vijayakumar, B. (2015). "Brain Tumor MR Image Segmentation and Classification Using by PCA and RBF Kernel Based Support Vector Machine". Middle-East Journal of Scientific Research, 23(9), 2106-2116.
81. Mohd Fauzi Bin Othman, Noramalina Bt Abdullah, "MRI Brain Classification using Support Vector Machine," IEEE, 2011
82. Rezaei K, Agahi H (2017) "Malignant and benign brain tumor segmentation and classification using SVM with weighted kernel width". Sig Image Proc Int J (SIPIJ). https://doi.org/10.5121/sipij.2017.8203
83. Suhag Sonu, Saini Lalit Mohan, "Automatic brain tumor detection and classification using svm classifier". In: Proceedings of ISER 2nd international conference, Singapore; July 2015. p. 55-59.
84. http://braintumor.org /brain-tumor- information/understanding-brain-tumors/ tumor - types/.
85. Mohanaiah P, Sathyanarayana P, GuruKumar L (2013) "Image texture feature extraction using GLCM approach". Int J Sci Res Publ 3(5):1 AQ9
86. Das S, Chowdhury M, Kundu MK (2013), "Brain MR image classification using multiscale geometric analysis of ripplet". Prog Electromagn Res 137:1–17. https://doi.org/10.2528/PIER13010105
87. Tipping M.E.. "Sparse Bayesian Learning And The Relevance Vector machine". *J Mach Learn Res*. Vol.1, pp.211–244, 2001.

88. Datcu D. and Rothkrantz L. J. M., "Facial Expression Recognition With Relevance Vector Machines," *IEEE intel. Conf. on Multimedia and Expo*, pp. 193-196, 2005.
89. Gonçalves H, Gonçalves JA, Corte-Real L (2011) HAIRIS: a method for automatic image registration through histogram-based image segmentation. IEEE Trans Image Process 20(3):776–789. https://doi.org/10.1109/TIP.2010.2076298
90. Soria-Olivas E., Gomez-Sanchis J., Martin J. D., Vila-Frances J., Martinez M., Magdalena J. R., and Serrano A. J., "BELM: Bayesian extreme learning machine," *IEEE Trans. Neural Netw.*, vol. 22, no. 3,pp. 505–509, Mar. 2011.
91. Vong C. M., Wong P. K., and Ip W. F., "A new framework of simultaneous-fault diagnosis using pairwise probabilistic multi-label classification for time-dependent patterns," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3372–3385, Aug. 2013.
92. Bishop C. M., *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006, ch. 4.
93. MacKay D. J. C, "Bayesian methods for backpropagation networks," *Models Neural Netw. III*, vol. 6, pp. 211–254, Jan. 1996.
94. Tipping M. E., "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, May 2001.
95. Wu T. F., Lin C. J., and. Weng R. C, "Probability estimates for multiclass classification by pairwise coupling," *J. Mach. Learn. Res.*, vol. 5,pp. 975–1005, Jan. 2004.
96. Senapati M.R, Vijaya I. and ,.Dash P.K "Rule Extraction by Training Radial Basis Functional Neural Network with Particle Swarm Optimization" published in American Journal of Science, Science Publication, ISSN: 1549-3636, vol 3, issue 8, pp.592-599,2007
97. Xin-She Yang, Suash Deb and Simon Fong, "Accelerated Particle Swarm Optimization and Support Vector Machine for Business Optimization and Applications", Communications in Computer and Information Science, Vol. 136, Springer, pp. 53-66 (2011).
98. Lal S. and Chandra M.,"Efficient algorithm for contrast enhancement of natural images," International Arab Journal of Information Technology, vol.11 ,no.1, pp.95–102, 2014.

99. Pal C, Das P, Chakrabarti A, Ghosh R (2017) Rician noise removal in magnitude MRI images using efficient anisotropic diffusion filtering. Int J Imaging Syst Technol 27(3):248–264. https://doi.org/10.1002/ima.22230
100. Aja-Fernandez S, Alberola-Lopez C, Westin C-F (2008) Noise and signal estimation in magnitude MRI and Rician distributed images: A LMMSE approach. IEEE Trans Image Process 17(8):1383–1398. https://doi.org/10.1109/tip.2008.925382
101. Szilagyi L, Benyo Z, Szilagyii SM, Adam HS (2003) MR brain image segmentation using an enhanced fuzzy c-means algorithm. In: Proceeding of the 25th annual international conference of the IEEE EMBS, pp 17–21
102. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612. https://doi.org/10.1109/TIP.2003.819861
103. Aja-Fernandez S, San-José-Estépar R, Alberola-Lopez C, Westin C (Sept 2006) Image quality assessment based on local variance. In: Proceeding of the 28th IEEE EMBS, New York, pp 4815–4818. https://doi.org/10.1109/TIP.2003.819861
104. Nayak DR, Dash R, Majhi B, (2016), Brain MR image classification using two-dimensional discrete wavelet transform and AdaBoost with random forests. Neurocomputing 117:188–197. https://doi.org/10.1016/j.neucom.2015.11.034i