REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

A MEDICAL DECISION MAKING SYSTEM FOR BRAIN TUMOR IDENTIFICATION FROM MAGNETIC RESONANCE IMAGES USING MACHINE LEARNING TECHNIQUES

Zahraa Abd Al Rahman Mohammed AL-SAFFAR

DOCTOR OF PHILOSOPHY THESIS
Department of Electronics and Communications Engineering
Electronics Program

Advisor
Prof. Dr. Tülay YILDIRIM

February, 2021

REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

A MEDICAL DECISION MAKING SYSTEM FOR BRAIN TUMOR IDENTIFICATION FROM MAGNETIC RESONANCE IMAGES USING MACHINE LEARNING TECHNIQUES

A thesis submitted by Zahraa Abd Al Rahman Mohammed AL-SAFFAR in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY** is approved by the committee on 16.02.2021 in Department of Electronics and Communications Engineering, Electronics Program.

Prof. Dr. Tülay YILDIRIM
Yildiz Technical University
Advisor

Approved By the Examining Committee

Prof. Dr. Tülay YILDIRIM, Advisor
Yildiz Technical University _____

Prof. Dr. Nurhan TÜRKER TOKAN, Member
Yildiz Technical University _____

Prof. Dr. Ali Yılmaz ÇAMURCU, Member
Fatih Sultan Mehmet Vakıf University _____

Prof. Dr. Hayriye KORKMAZ , Member
Marmara University _____

Assoc. Prof. Dr. Nihan KAHRAMAN, Member
Yildiz Technical University _____

I hereby declare that I have obtained the required legal permissions during data collection and exploitation procedures, that I have made the in-text citations and cited the references properly, that I haven't falsified and/or fabricated research data and results of the study and that I have abided by the principles of the scientific research and ethics during my Thesis Study under the title of A MEDICAL DECISION MAKING SYSTEM FOR BRAIN TUMOR IDENTIFICATION FROM MAGNETIC RESONANCE IMAGES USING MACHINE LEARNING TECHNIQUES supervised by my supervisor, Prof. Dr. Tülay YILDIRIM. In the case of a discovery of false statement, I am to acknowledge any legal consequence.

Zahraa Abd Al Rahman
Mohammed AL-SAFFAR

Signature

Dedicated to my family
and my best friend

ACKNOWLEDGEMENTS

Yildiz Technical University is one of the seven government universities located in Istanbul and the 3^{rd} oldest university in Turkey with a history dating back to 1911. It is also considered to be one of the best universities in the world. As a student of such a prestigious university, I am very happy, and proud of that.

I would like to express my sincere gratitude to my supervisor Prof. Dr. Tülay YILDIRIM for all her enthusiasm and encouragement over the past few years. She has always been a supporting and motivating professor and sister during my study period and in life. It was really a great honor for me to work under her guidance.

I would like to thank the committee members; Prof. Dr. Nurhan TÜRKER TOKAN and Prof. Dr. Ali Yılmaz ÇAMURCU, who have followed the progress of this thesis from the time it was proposed to the present moment for all their valuable advice, suggestions and time.

I am so grateful to the Iraqi Ministry of Higher Education and Scientific Research and the university of Baghdad in Iraq to give me this invaluable opportunity for completing my PhD study here, in Turkey.

Last but not least, many thanks to my mother for her prayers and unconditional love, my husband for his support and patience, my sisters for their assistance and motivation, and my children; Zain, Sedin, and Emin, for being my primary source of inspiration. I feel incredibly lucky to have such an exceptional family and I love you all very deeply.

Zahraa Abd Al Rahman Mohammed AL-SAFFAR

TABLE OF CONTENTS

LIST OF SYMBOLS

S	A diagonal of eigenvalues matrix in MATLAB
$\varnothing_1, \varnothing_2$	Coordinate system
V	Left singular vector
P	Probability value
U	Right singular vector
Σ	The diagonal matrix

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
BWT	Berkeley Wavelet Transformation
BRaTS	International Brain Tumor Segmentation
CAD	Computer Aided Diagnosis
CDF	Cumulative Distribution Function
CT	Computed Tomography
D-based MI	Mutual Information based on Distance metric
DC	Dice coefficient
DICOM	Digital Imaging and Communications in Medicine
D-SEG	Diffusion-based Segmentation
DTI	Diffusion Tensor Imaging
FKM	Fuzzy K-means
FLAIR	Fluid Attenuated Inversion Recovery
GLCM	Grey Level Co-occurrence Matrix
GVF	Gradient Vector Flow
HGG	High Grade Glioma
KNN	k-Nearest Neighbor
LDI-Means	Local Difference in Intensity-Means
LGG	Low Grade Glioma
MATLAB	Matrix Laboratory (programming language)
MGLCM	Modified Grey Level Co-occurrence Matrix
MI	Mutual Information
MIFS	Mutual Information Feature Selection

MLP	Multi-layer Perceptron
MRI or MRG	Magnetic resonance imaging or manyetik rezonans görüntüleme
MRIs	Magnetic Resonance Images
mRMR	Minimum Redundancy Maximum Relevance
P-based MI	Mutual Information based on Probability
PCA	Principal Component Analysis
PCA-ANN	Principal Component Analysis - Artificial Neural Network
PET	Positron Emission Tomography
PFS	Potential Field Segmentation
SOM	Self-Organizing Map
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SWA	Weighted Aggregation Algorithm
TCIA	The Cancer Imaging Archive
T1-w	T1-weighted image
T2-w	T2-weighted image
RBF	Radial Basis Function
RFE	Recursive Feature Elimination
RGB	Red Green Blue (colored image)
RNN	Residual Neural Network
ROC	Receiver Operating Characteristics
ROI	Region of Interest
VOI	Volume of Interest
WHO	World Health Organization

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

A MEDICAL DECISION MAKING SYSTEM FOR BRAIN TUMOR IDENTIFICATION FROM MAGNETIC RESONANCE IMAGES USING MACHINE LEARNING TECHNIQUES

Zahraa Abd Al Rahman Mohammed AL-SAFFAR

Department of Electronics and Communications Engineering
Doctor of Philosophy Thesis

Advisor: Prof. Dr. Tülay YILDIRIM

Brain tumor is an abnormal and uncontrolled growth of the cells. Early brain tumor detection is essential to save lives. In fact, brain tumors are difficult to diagnose, requiring specialized equipment and training. A medical decision making system facilitates diagnostic process by visualizing the data produced by a classification system, allowing doctors to make a right diagnosis.

This study proposes an automated system for segmentation and classification the brain tumor grades in MRI into three classes: normal, LGG and HGG. In the proposed system, a new segmentation method named LDI-Means algorithm (Local Difference in Intensity-Means algorithm) is used. It is a clustering technique based on the difference in the intensity level of one pixel than another. Furthermore, a new approach in selecting the sub-significant set of attributes is used, denoted MI+SVD (Mutual Information + Singular Value Decomposition). The robust features are later used as an input to the classifier. The new network structure called simplified RNN (Residual Neural Network) is also offered by this study.

The proposed automated system has six stages; the pre-processing, clustering by LDI-Means, feature extraction, feature selection and dimension reduction by MI+SVD, and classification by simplified RNN.

The experimental findings at the end of the segmentation stage presented an approximate match of 99.02% with the hand-labeled images. In addition, in comparison to the original feature space and two standard dimension reduction

methods, PCA and SVD, the MI+SVD algorithm offered a more efficient result for improving the classification process to achieve a satisfied grading of brain tumors. Furthermore, using a simplified RNN as a classifier provides a high level of effectiveness to the proposed system. In comparison with other published studies, it is found that the proposed system is very sufficient to offer a meaningful real-time estimation for identification the brain tumor grades.

Keywords: Medical decision making system, brain image classification, brain tumor segmentation, clustering, image processing, machine learning, mutual information (MI), principal component analysis (PCA), singular value decomposition (SVD), support vector machine (SVM), artificial neural network (ANN), residual neural network (RNN).

ÖZET

Makine Öğrenimi Tekniklerini Kullanarak Manyetik Rezonans Görüntülerinden Beyin Tümörünün Belirlenmesi için Tıbbi Karar Verme Sistemi

Zahraa Abd Al Rahman Mohammed AL-SAFFAR

Elektronik ve Haberleşme Mühendisliği Anabilim Dalı
Doktora Tezi

Danışman: Prof. Dr. Tülay YILDIRIM

Beyin tümörü, hücrelerin anormal ve kontrolsüz büyümesidir. Erken beyin tümörü tespiti, hayat kurtarmak için çok önemlidir. Aslında, beyin tümörlerinin teşhis edilmesi zordur, özel ekipman ve eğitim gerektirir. Tıbbi bir karar verme sistemi, bir sınıflandırma sistemi tarafından üretilen verileri görselleştirerek teşhis sürecini kolaylaştırır ve doktorların doğru tanı koymasına olanak tanır.

Bu çalışma, MRG'deki beyin tümörü derecelerini 3 sınıfa ayırmak ve sınıflandırmak için otomatik bir sistem önermektedir: normal, LGG ve HGG. Önerilen sistemde LDI-Ortalamalar algoritması (Yoğunluk Ortalamalarında Yerel Fark algoritması) adlı yeni bir segmentasyon yöntemi kullanılmıştır. LDI-Ortalamalar, bir pikselin yoğunluk seviyesinin diğerinden farklı olmasına dayanan bir kümeleme tekniğidir. Ayrıca, MI + SVD (Karşılıklı Bilgi + Tekil Değer Ayrışımı) olarak adlandırılan alt anlamlı öznitelikler kümesinin seçilmesinde yeni bir yaklaşım kullanılır. Sağlam özellikler daha sonra sınıflandırıcıya girdi olarak kullanılır. Basitleştirilmiş RNN adı verilen yeni ağ yapısı da bu çalışmada sunulmaktadır.

Önerilen otomatik sistemin altı aşaması vardır; ön işleme, LDI-Ortalamalar ile kümeleme, özellik çıkarma, özellik seçimi ve MI + SVD ile boyut küçültme ve basitleştirilmiş RNN ile sınıflandırma.

Segmentasyon aşamasının sonundaki deneysel bulgular, elle etiketlenmiş görüntülerle yaklaşık 99,02%'lik bir eşleşme sunarak güvenilir bir beyin tümörü segmentasyon süreci ile sonuçlanmıştır. Ek olarak, orijinal özellik uzayına ve standart PCA ve SVD boyut indirgeme yöntemlerine kıyasla; MI+SVD algoritması, beyin tümörlerinin

tatmin edici bir derecelendirmesini elde etmek için sınıflandırma sürecini iyileştirmede kesin ve daha verimli bir sonuç sunmuştur. Ayrıca, basitleştirilmiş bir RNN'yi sınıflandırıcı olarak kullanmak, önerilen sisteme yüksek düzeyde etkililik sağlar. Yayınlanmış diğer çalışmalarla karşılaştırıldığında, önerilen sistemin beyin tümörü derecelerinin belirlenmesi için anlamlı bir gerçek zamanlı tahmin sunmak için çok yeterli olduğu bulunmuştur.

Anahtar Kelimeler: Tıbbi karar verme sistemi, beyin görüntüsü sınıflandırması, beyin tümörü segmentasyonu, kümeleme, görüntü işleme, makine öğrenimi, karşılıklı bilgi (MI), temel bileşen analizi (PCA), tekil değer ayrıştırma (SVD), destek vektör makinesi (SVM), yapay sinir ağı (YSA), artık sinir ağı (RNN).

1
INTRODUCTION

1.1 Literature Review

In recent years, there have been many studies that have presented different methods for segmentation, detection, and classification of the tumor area using brain magnetic resonance images (MRIs). Here, it is important to remember that there is no such an ideal technique and it is necessary to keep researching for new methods to achieve better outcomes by reducing the limitations of the methods mentioned in this section. Some of the published works in the field of brain tumor segmentation and classification are discussed as follows:

1.1.1 Brain Tumor Segmentation

Dhanachandra et al. [1] presented a technique to compute the initial value of cluster centres using subtractive algorithm for accurate segmentation. Their algorithm involved the use of partial contrast stretching, k-means clustering and median filter. The contrast stretching algorithm was employed to improve the quality of the image. Later, the subtractive clustering algorithm was implemented to find the centroids, based on the image potential value. These centres were used as the initial values in k-means algorithm.

Vishnuvarthanana et al [2] proposed an effective way for identification and segmentation of brain tumor. A new method was used for segmentation, combining a self-organizing map (SOM) and fuzzy k-means (FKM). Their method needs the user to choose three different tissues within the brain. Despite the efficiency of the results, their way was complex and not easy to apply.

A novel density computation of data points is defined in the study of Abubaker et al [3] to overcome some of the limitations of K-means algorithm. Their algorithm depended on two versions of K- nearest neighbor. In the first version, they used the kn (the number of nearest neighbors) and k (the number of clusters) as inputs. Later, a group of points is checked until the number of centroids attain a certain k. The other version uses one input, kn. Then, k and the centroids are obtained. By using this technique,

K-means algorithm can find the accurate number of clusters.
Soltaninejad et al. [4] classified the tumor grades by using SVM. The segmentation process was based on a super-pixel method. In their study, the comparison between a manual segmentation and the proposed method can be found.
An automated detection and segmentation system based on the super-pixel technique using only one modality of MRI was proposed in the study of Soltaninejad et al [5]. By grouping the voxels with similar properties and extracting features from superpixels, the accuracy of feature extraction as well as the computation time has been improved.
Javadpour et al [6] suggested a valuable method to enhance the segmentation process. they used a region growth and genetic algorithm. By thresholding the initial points are selected , each one represents the required segmented area. Later the other points are checked. Then the Genetic algorithm is applied by using Fitness function to find the difference between the segmented area and the image resulted from region growth. Their method may be used for huge dataset of images.
Angulakshmi et al [7] presented an unsupervised algorithm to detect tumor. Their method has 3 steps: firstly, finding the tumor slices by using bilateral asymmetry property of the brain. Secondly, finding the ROI (region of interest) by using quad-tree decomposition. Lastly, using spectral clustering to perform the segmentation.
Bahadure et al [8] investigated a Berkeley wavelet transformation (BWT) for brain tumor segmentation. The relevant features that were extracted from each segmented tissue were used as an input to the SVM for the classification process.
Cabria et al [9] proposed an algorithm named Potential Field Segmentation (PFS). in order to find a fused segmentation, they used a combination between the findings obtained by PFS and other methods. They exploited the potential criterion to perform segmentation. For each pixel in the MRI, the potential field is computed. They used 22 images only and it is not sufficient to estimate the segmentation accuracy.
Corso et al. [10] introduced a technique of combining two approaches to performing automatic segmentation: a generative model-based technique and a graph-based affinities method. Then, the resulting model was integrated into multi-level segmentation by applying a weighted aggregation algorithm (SWA).
Dong et al. [11] presented an automated system for brain tumor detection depending on the U-Net-based deep convolutional neural network. A set of data augmentation methods were applied. Their network architecture consisted of an encoding (down sampling) and decoding (up sampling).

1.1.2 Brain Tumor Classification

R. Battiti [12] presented a method to select features based on the mutual information theory called MIFS. This method calculates the relationships between each feature

with all other features and between each feature with all classes. Later, a new subset of attributes is created by picking up a feature which gives a high rate of information about the class label.

Also, based on the MI theory, H. Peng et al. [13] proposed a technique called mRMR. They applied mRMR to increase the relevancy and decrease the redundancy among features.

Vinod Kumar et al. [14] suggested a Principal Component Analysis - Artificial Neural Network (PCA-ANN) system for multiclass brain tumor classification. It has 4 stages: Gradient Vector Flow (GVF), feature extraction, feature reduction by using PCA and classification by using ANN.

Also, a multiclass brain tumor classification using PCA-ANN is proposed in the study of J. Sachdeva et al. [15]. The 856 of ROIs are found by using Content-based active contour (CBAC). Then more than one experiment are performed in this study in order to obtain the accuracy by using ANN with and without and PCA.

The study of L. Fang et al. [16] calculated the value of MI by using the Kozachenko Leonenko information entropy estimation algorithm for obtain a significant attributes. later, it factorized the feature matrices. Their method can successfully decrease the multidimensional time series dimensions of clinical data.

N. Hoque et al. [17] described a new method for feature selection depend on Fuzzy MI with a non-dominated solution. The features are selected according to the fuzzy mutual information between each feature and classes as well as between each feature and the others. Also, they presented a k-nearest neighbor (KNN) classifier modification to classify inputs according to the distance.

A diffusion tensor imaging (DTI) algorithm is used by T. L. Jones et al. [18] to find the tumor VOIs according to the isotropic and anisotropic properties of the diffusion tensor. Later, diffusion-based segmentation (D-SEG) spectra are considered within each VOI. By using SVM, the classification using D-SEG spectra is applied.

More than one test was carried out in the work of Zacharaki et al. [19]. Their proposed system consists of ROI definition, feature extraction, feature selection and classification. For feature subset selection, SVM with recursive feature elimination (RFE) was applied, whereas for classification three methods were investigated: LDA with Fisher's Discriminant Rule, k-nearest neighbour (kNN), and nonlinear SVM. A multiclass classification was performed using a one versus all voting scheme.

Zollner et al. [20] offered the comparative investigation to find an optimal method of feature reduction for improving SVM-based classification in order to achieve a brain glioma grading.

In the research of Khawaldeh et al.[21], a convolution neural network was presented to perform brain tumor classification by using data set from The Cancer Imaging Archive (TCIA). In their results the MR brain images were classified into 3 classes.

A CAD system was proposed by Hsieh et al [22]. Their system was developed to identify the malignancy of diffuse brain gliomas. The classification performances was calculated by using local features, global features, and both groups. The results were achieved 83%, 76%, and 88%, respectively.

The essential step for MR brain tumor images classification is the extraction of significant features. Many researches have presented various methods for feature extraction in order to classify the tumor in the brain MRI scans such as the research of Hasan et al. [23] where an interesting deep learning feature extraction algorithm was proposed to extract the relevant features. A modified grey level co-occurrence matrix (MGLCM) method combined with deep features (DF) learning were used to improve the classification performance of MR brain scans. Later, for binary classification, SVM was implemented.

In the study [24], Bakas et al. offered an overview to the various machine learning techniques used for MR brain tumor images processing of the International Brain Tumor Segmentation (BRaTS) challenge from 2012 to 2018.

In a survey by Litjens et al. [25], the use of deep learning for medical image classification, abnormal tissue detection, segmentation, registration, and other tasks was established. Their paper reviewed the major deep learning models used in medical image processing and summarized over 300 contributions in this field, most of which appeared before and during 2016.

A valuable review on the recent segmentation and tumor grade classification techniques of brain MR images was also found in the survey of Mohan et al. [26]. This survey clarified the methodologies used up to and including year 2017 for segmentation and grading of brain tumors. These methods can be included in the standard imaging procedures. In addition, a vital assessment of the state-of-the-art, upcoming developments and trends was offered. In particular areas, feature selection has been successfully used in medical applications. It can diminish the dimensionality and provide a better understanding of the causes of a disease.

The survey of Remeseiro et al. [27] described some basic theories used in many medical applications and offered some background concepts on feature selection. The survey also presented a review of the current and useful feature selection methods utilized in different medical problems.

Finally, the study of Jalali et al. [28] should be included in this section. It has been published in November 2020. Their study illustrated works of different researchers using medical imaging for automatic brain tumor identification. In addition, they analyzed the results of these studies in term of accuracy, specificity, and sensitivity parameters and provided a valuable overall comparison tables.

The literature survey above offers a clear view to the methods that have been invented to obtain region of interest, extract features, and train / test features to perform classification. It is apparent that segmentation methods need to be more accurate and an effective classification combining feature selection and dimensionality reduction has not been achieved. This study is therefore intended to design a medical decision-making system to perform brain tumor segmentation and classification which will be more reliable, or takes less time or costs, or easier to implement than the current methods.

1.2 Objective of the Thesis

The main objective of this study is based on the following general motivation:

1. To take advantage of clinical information and databases of patients for discovering and diagnosing their diseases in order to provide decision support to medical specialists. This study is expected to design some models can help doctors in grouping patients in useful patterns based on various risk factors, and how machine learning algorithms can recognize such patterns. This can have a great role in detecting early onset of the disease and its stage, as well as providing a suitable plan of care.

2. To deal with large number of attributes and features, and finding the importance of some features over others. A large number of features can cause a curse of dimensionality, and can make a machine learning algorithm limited in terms of specificity, sensitivity accuracy and time.

Thus , this study aimes to design and implement a medical decision making system for an automated segmentation and grading able to classify the tumors into normal, low grade glioma and high grade glioma using brain tumor MRIs. The designed system will aid physicians to identify the disease, prevent misdiagnosis, and decrease a patient waiting time.

1.3 Hypotheses and Problem Statement

The quantities and complexities of current patient data make medical decisions more difficult for doctors and other carers than ever. This case involves the use of computational methods to process data and make suggestion in order to help the decision makers. Over the past two decades, it has become very necessary to design, implement, and use systems in the form of computer-aided decision support. For this

study the medical decision making system is designed to recognize the benign and malignant tumors. In order to reduce erroneous diagnostic interpretation of brain tumors in MRI scans and workload, as well as helping the clinicians to ignore the MRI brain scans of the patients who have normal brain quickly and focus on those who have pathological brain, the following research questions need to be addressed:

Question 1: How to design an appropriate brain tumor detection system that classifies the MRI brain images into a normal or abnormal brain more effectively than the available systems?

Question 2: Does using the current algorithms will provide a better outcomes for proper diagnosis and treatment in terms of classification accuracy and tolerance to noise?

Question 3: Which pre-processing methods that should be used to improve the classification accuracy of tumors in brain MRI?

Question 4: How to identify the exact tumor location in brain MRI?

Question 5: How to find a new way to increase the segmentation accuracy which can automatically finds the ROI (tumor) in MRI?

Question 6: How to identify the most effective features which will describe the input data in a best possible way, in order to distinguish the tumor types by using a learning-based classification and data mining techniques?

Question 7: Which methods that should be used to discriminate data set features by finding the robust features to improve prediction in context of the right medical decision?

Question 8: To what degree this research will be able to achieve a satisfactory criteria such as computational cost and speed, in addition to accuracy?

1.4 General View

In clinical routine, clinicians spend an increasing time in diagnosing and interpreting medical images due to the increased utilization of diagnostic imaging. High levels of experience are required to carry out manual and accurate delineation and classification of these medical images [29]. Due to the improvement in scanner resolutions and the decreasing in slices' thickness, a much more number of slices can be produced than before. Therefore, clinicians need more time to manage the

image set of each patient because of these huge number of data. Coupled with the increase in patient numbers, this puts pressure on resources and services resulting in significant delays to both diagnosis and treatment [30]. Thus, an automated tumor segmentation and classification have attracted a considerable attention in the past two decades, and various algorithms have developed for interactive, semi-automated, and full automated segmentation and classification of brain tumors.

1.4.1 Medical Decision Making Systems

Medical decision making systems can play an important role in the medical environment. Physicians are prone to making some mistakes in their medical decisions, because of the complexity of medical problems and due to cognitive limitations [31]. Computer-based aids aim to reduce physician's errors by providing an appropriate support for decision making. When the decisions being made can have a profound impact upon the patient, it is of the ultimate importance that diagnosticians have the relevant information presented to them in the most effective manner possible [32]. A typical decision making process contains the knowledge discovery process. Many researchers consider data mining programs as a way to make decision making tools intelligent. The importance of using computer-based tools to eliminate the problems of medical decision-making is realized half a century ago [24]. Medical decision making systems are computer tools for the integrated Decision Support System, which is intended to aid doctors and other health professionals in making right medical decisions, such as evaluating patient data for better diagnosis [32]. A working definition has been suggested by Dr. Robert Hayward of the Centre for Health Evidence "Clinical Decision Support Systems link health observations with health knowledge to influence health choices by clinicians for improved health care". This description has the benefit of giving the Clinical Decision Support its functional impression [33]. The possible advantages of using support systems for clinical decisions fall into three different areas:

1. Patient safety improvement, by decreasing the medication errors and enhancing the tests ordering.

2. Treatment quality improvement, by increasing the available time of specialists' for direct care, enhancing the implementation of clinical pathways and recommendations, and promoting the use of up-to-date clinical data. Also, by providing better clinical reporting and patient satisfaction.

3. Efficiency of healthcare delivery improvement, by lowering costs through decreasing the number of the required tests and modifying the medication

prescription patterns.

1.4.2 Medical Imaging

Medical imaging can be described as a technique used to generate images for diagnosis, treatment, and clinical studies. One of the medical imaging system is a Magnetic Resonance Imaging (MRI). It exploits the properties of magnetic fields for capturing images and provides very useful tissue measurements, including anatomical, structural, and functional details [33]. According to its excellent contrast of soft tissues and its precise resolution, MRI is common method for imaging a growth of brain tumor and detecting its location [34]. The classification of brain images and tumor identification are still depend mainly on direct human examination of the images. This visual assessment and analysis by clinicians are biased by their point of view, also, it is time-consuming and subject to mistakes or inattention [35]. Hence, a medical decision making system for automated brain tumor identification and grading using images from MRI system has been developed to improve the physicians' diagnostic skills and decrease the time needed.

1.4.3 Segmentation

An MR brain image includes three areas, the gray matter, cerebrospinal fluid, and white matter. Actually, it will be very beneficial for an accurate diagnosis of brain diseases, if it is possible to separate each of these regions than others. The separation process in image analysis called segmentation which is not an easy task because of the similarity in the biometric features of brain [6]. Many studies have therefore been carried out on MRI segmentation and it remains an open area for more. There are several methods perform the segmentation process in MR images. For example cluster-based, neural network-based, edge-based and threshold-based. Clustering is the most effective method. There are several kinds of clustering algorithms. For example mountain, K-means, subtractive, and Fuzzy C-means [36]. This study tries to establish a cluster-based segmentation algorithm which is more effective and easier to implement than the current ones.

1.4.4 Classification

The essential goal of the brain tumor classification is to accurately identify which type of tumor the patient suffer from. Glioma is the brain tumor with the highest death rate and incidence. These neoplasms can be graded into Low-Grade Gliomas (LGG) and High-Grade Gliomas (HGG) in term of being infiltrative and aggressive. [37]. A brain tumor management depends on the size of tumor, its type, and its

developing level. In computational methods, the classification process is known as a supervised learning task which determines a relation between the attributes of the dataset (input) and the targets (output). A large number of inputs almost can cause some challenges such as over-fitting, or high computational complexity [38]. To get better outcomes, a medical decision-making systems industry has begun to use data mining techniques to detect and identify tumors. Therefore, physicians can let a brain tumor detection system to be as a second opinion as well as their view to finding the proper brain tumor diagnosis and treatment [37]. In machine learning, the input data quality has a direct effect on the output quality, e.g. accuracy. The input data for any algorithm used in machine learning approach is almost represented by number of features displaying the properties of the problem. Hence, the quality of the feature space has an important role in solving any problem [4]. This study tries to improve the classification process to get more effectively and high accuracy results with respect to the published methods. The required three classes are: healthy, low-grade glioma (LGG) and high-grade glioma (HGG).

1.5 Research Phases

This study included four phases of research; requirement collecting, system designing, implementation, and evaluation. As shown in Fig.1.1.

1.6 Thesis Organization

This thesis has been structured as:

- Chapter two explains some concepts and terminologies.
- Chapter three describes the material and method.
- Chapter four involves the experimental results as well as the analysis of the proposed feature selection and classification techniques.
- Chapter five addresses the conclusion of this study and future enhancement.

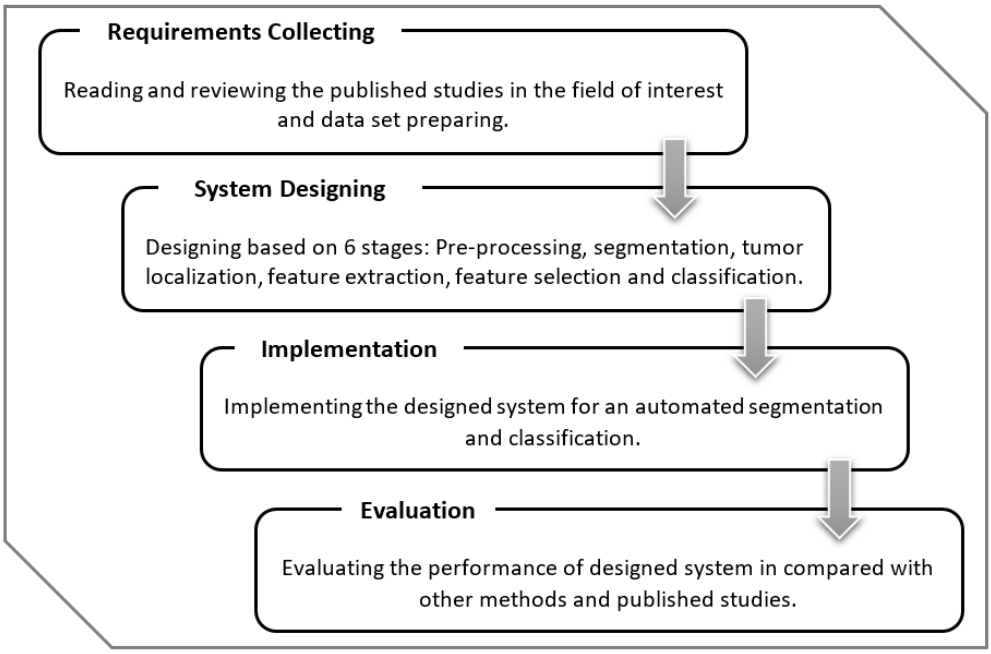

Figure 1.1 Research phases of the proposed system

2
CONCEPTS AND TERMINOLOGIES

2.1 Brain Tumors

Tumor is a mass of abnormal tissues can be solid or fluid filled. It can be classified into primary and secondary. The growth of Primary type is very slow whereas the secondary tumor can be spread quickly. Hence, secondary tumors are caused by cancer cells. Thus it is inferred that all tumors are not cancer, but all cancers are definitely tumors [39]. Brain and spinal cord are both the human Central Nervous System. The command center of human beings is the brain. Brain and spinal cord are made of more than one type of cells such as nerves and glial cells. Glial cells surround and support neuron cells. There are many more glial cells than neurons. Gliomas are cancers of glial cells [40]. Regarding to the WHO grading system, cancers are grouped into 4 grades:

- Grade I implies that the cells of the cancer appear almost normal. Most patients with grade I gliomas stay a live for a long time.
- Grade II implies that the cancer cells appear a little bit abnormal. After treatment, Some patients with grade II return as a higher grade glioma.
- Grade III implies that the cancer cells appear abnormal. Cells of grade III increase in number rapidly.
- Grade IV implies that the cancer cells appear clear abnormal. Cells of grade IV grow and increase in number very rapidly.

Gliomas are almost represented as either high grade cancers or low grade cancers. LGG means grades I (astrocytomas) and II (oligoastrocytoms). HGG means grades III (ependymomas) and IV (Glioblastoma Multiform) [41].

2.2 Medical Imaging Techniques

In radiology, computerized techniques for diagnostic imaging have widely applied to medicine more than any other field for diagnosing injuries, illnesses and other conditions. Computers are important tools for handling data and documents that are downloaded and read out by a central computer bank from various scanning devices. Using such a technology has created computer graphics and anatomical color pictures, has altered diagnostic practices, and has produced a more precise diagnostic images to medical specialists that saves time. Furthermore, assisting doctors to see clear images of the body with no need to surgery is the most significant contribution of imaging technology [42]. Brain imaging techniques can broadly be classified according to the source of energy for the procedure as follows: [43]

- Computed Tomography (CT): uses x-rays to take a several images from different angles to a part of the body.
- Magnetic Resonance Imaging (MRI): uses a magnetic field and radio waves to make images. It provides structural and anatomical information.
- Positron Emission Tomography (PET): in this type of scan a radio-tracer will first be injected into body. Later, by using a special camera the radio-tracer is scanned throughout the procedure. Cancer cells appear brighter than normal cells because they use the radiotracer more quickly.

Brains are often scanned by two-dimensional images (slices). These slices are usually one of the three orthogonal planes: sagittal, coronal and horizontal (axial) as in Fig.2.1. Also, Fig.2.2 shows the three planes by using MRI.

The MRI scanner has two strong magnets which are the major components of the unit. The human body is primarily made up of water molecules or oxygen and hydrogen atoms. At the core of each atom is an even smaller part, it is a proton, that acts as a magnet and it is responsive to any magnetic field. Typically, the water molecules in the body are distributed uniformly, but through MRI scan, the first magnet makes the water molecules to align in a single direction, south or north. Later the second magnetic field is turned on and off to give fast pulses. Hence each hydrogen atom will change its orientation when it is "on" and then very fast turn back to its original relaxing state when it is "off". Also, passing electricity through the gradient coils causes coils vibrating, produces a magnetic field causing a special noisy sound. While the patient can not feel all these changes except the sound, these changes can be identified by the scanner and a cross-section picture for the radiologist can be generated in accordance with the computer [44]. The MRI imaging process can be shown in Fig.2.3.

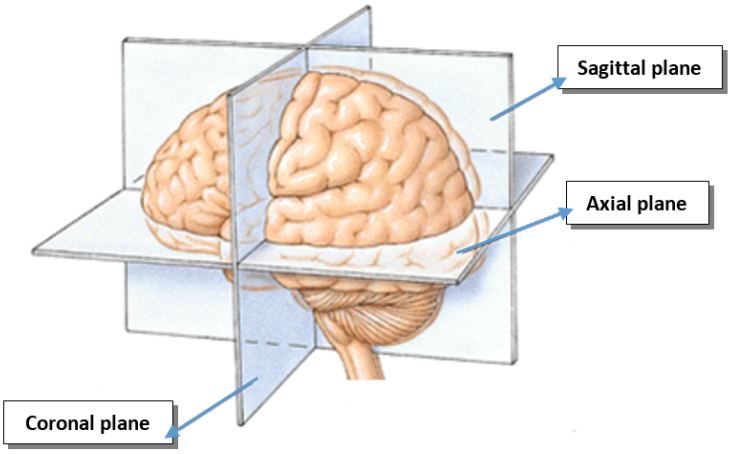

Figure 2.1 The axes of brain imaging [44].

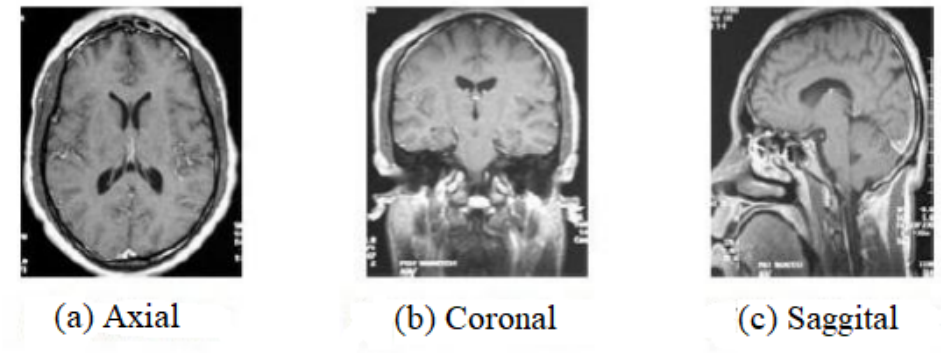

Figure 2.2 The MRI planes [45].

MRI provides a very good soft tissue contrast than CT and that because the intensities of the proton signals are depended on both the water distribution and the nuclear magnetic resonance relaxation properties of the water proton called relaxation times T1 and T2. The tissue molecular composition affects on the value of T1 and T2. The manner in which the image is created can be used to maximize the effect of T1 or T2 for making the intensity of signal more respond to particular aspects of tissue composition. Also, the signal is affected by diffusion. Thus, by changing some acquisition factors the image can be more respond to the diffusion of water molecules [6] . The images of MRI can be as a number of two-dimensional slices or three-dimensional. The thickness of slice is almost much greater than the in-plane resolution; hence, the multi-slice images have less resolution in 1-dimension. Images of three-dimension can be captured with isotropic image resolution but that need more time. In addition, it often have a $1mm$ resolution. Actually these limitations in resolution are not important for brain tumors imaging. The brain MRI is very adequate for diagnostic purposes, radiotherapy targeting, and biopsies management [46]. An acquired image of brain by using MRI system contains three areas. These are cerebrospinal fluid, gray matter, and white

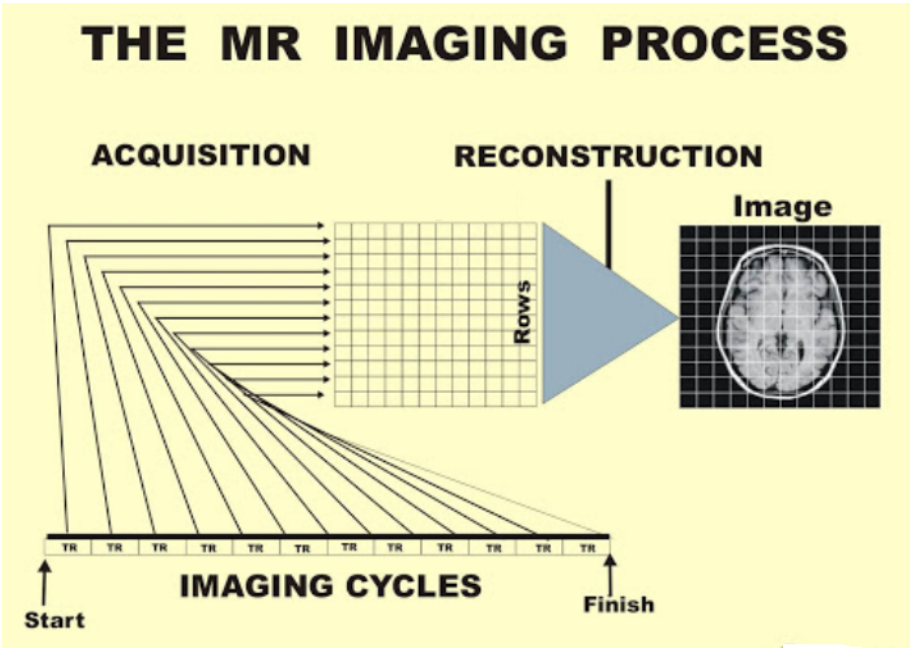

Figure 2.3 Acquisition process by MRI. This image available online at *http* : *//www.sprawls.org/mripmt/MRI*09*/index.html*

matter as shown in Fig.2.4.
Four factors by which the intensity of MRI signal can be determined: Density of proton, T1-weighted, T2-weighted, and Flow. In T1 or T2 scans, the middle of the brain tends to be lighter and have darker colors around it. In FLAIR scans, the middle of brain tends to be darker and have lighter shades around it. In T1-weighted scans, tissues of high amount of fat appear brighter and areas of high amount of water appear darker. This type of scan is used to obtain the useful information about an anatomical properties. While in T2-weighted scans, the opposite will appear and this type of scan is used to obtain the useful information about an pathological properties but of course not all tumors tend to be associated with an increase in water content [47].

These different acquisition image formats can be shown in Fig.2.5 . Each format highlights different tissue. As seen, cerebrospinal fluid and edema are darker in T1-weighted images and brighter in T2-weighted images, while gray matter is not as dark in T1 and not as bright in T2.

2.3 Digital Images

Generally, the medical images are digitally stored in the form of matrix representation. Let is assumed I is a digital image [49]. It can be represented as shown in Fig 2.6. The variables i and j are denoting the row and column starting from zero until m and n, respectively to identify each pixel position within the image whereas m and

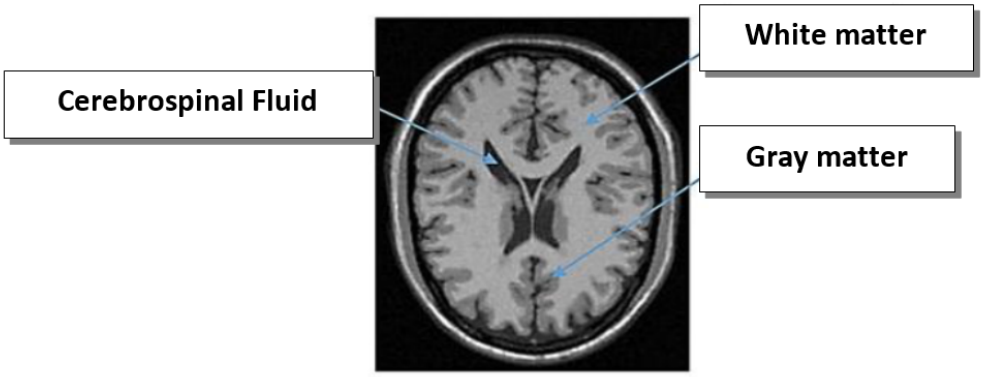

Figure 2.4 The regions of MR brain image [48].

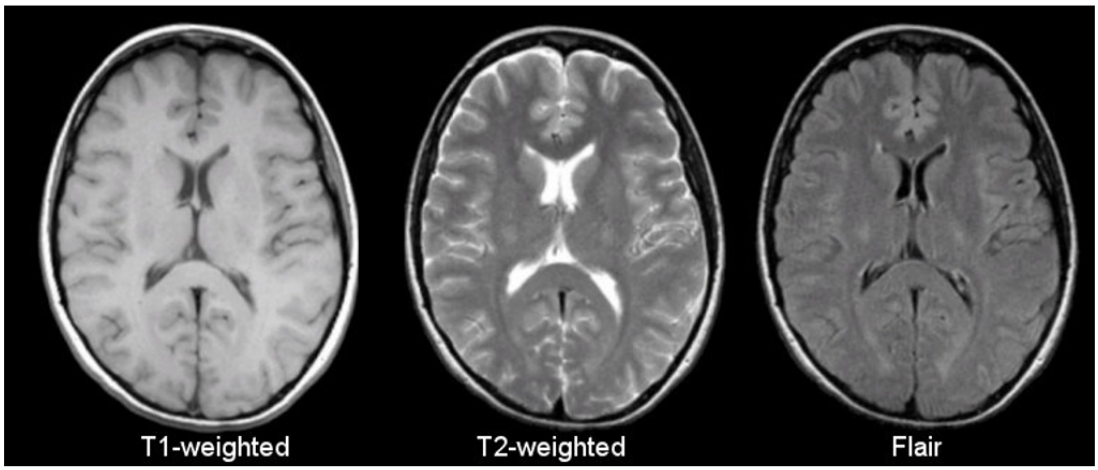

Figure 2.5 The MRI acquisition image formats [45].

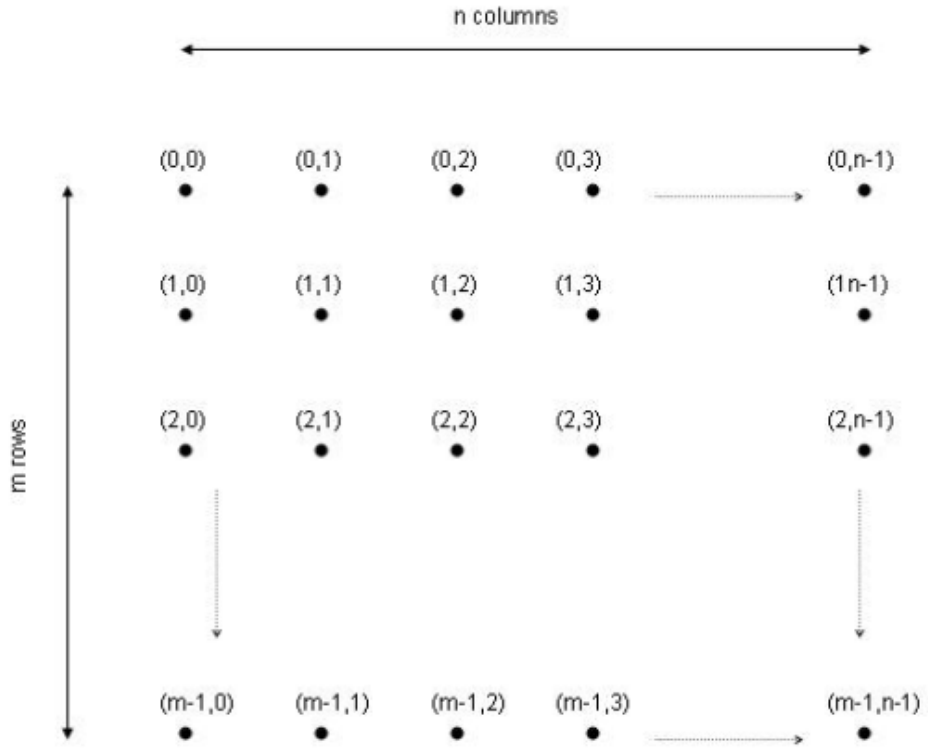

Figure 2.6 The matrix representation of digital image

n are denoting the maximum number of pixels in each row and column. The value of each pixel represents its intensity. For example if *I* is a grey scale image of 8-bit, and I(2,1)= 255. That means the pixel in the third row and second column has an intensity of white.

Digital Imaging and Communications in Medicine (DICOM) services presents an interface for transmitting medical images and information in the DICOM industrial standard. Thus, most medical images files (CT, MRI, PET, and Nuclear Medicine) are received in DICOM file format. In addition to the image data, a single DICOM file contains patient's name, scan type, image dimensions, and other such like information. In MATLAB the dicomread function can read this type of files.

2.4 Image Filtration

The brain images of MRI are almost very noisy due to some acquisition errors. In addition, some errors appear due to image registration. Hence the images need some smoothing techniques before any statistical analysis to get an optimal results. There are many methods of image filtration or noise removing such as Median and Gaussian filter [49]. Median Filters eliminate the high frequency signals from the image and keep on edges. This method involves sorting all pixel by size, later calculating the

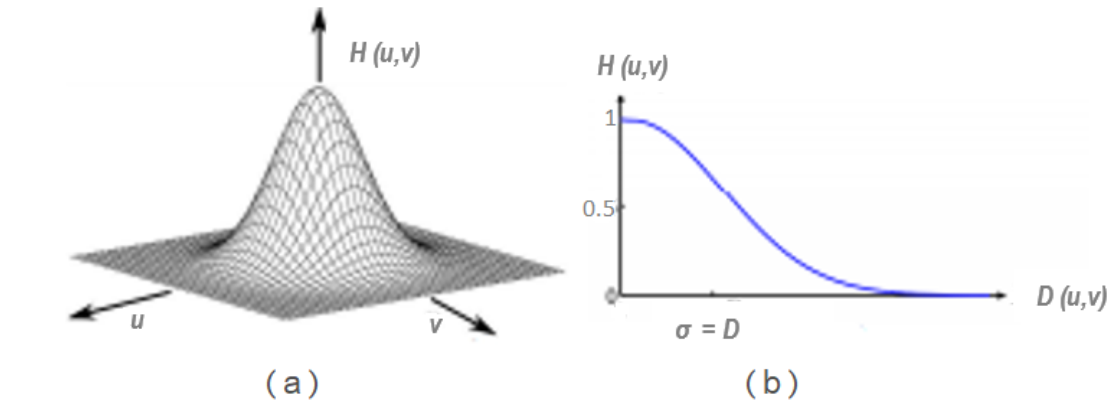

Figure 2.7 Gaussian Transfer function [49], (a) Gaussian Low pass filter, (b) Radial cross section.

median value to be a new pixel value according to following equation:

$$f(x,y) = median(i,j) \in W_{m,n}\{g(x+i, y+j)\} \tag{2.1}$$

where f is the output (filterd image), g is the input (original image), and $W_{m,n}$ is a sliding window of $m \times n$ in size [7].

Gaussian filter is a common smoothing tool in brain imaging [50]. The spread of kernel is usually measured in terms of the full width at the half maximum (FWHM) of Gaussian kernel K_σ. The n-dimensional Gaussian kernel can be described as $n \times 1D$ kernel . Let is assumed $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. Thus the n-dimensional kernel is given by the following equation:

$$K_\sigma(a) = K_\sigma(a_1)K_\sigma(a_2)...K_\sigma(a_n) = \frac{1}{(2\pi)^{n/2}\sigma^n} exp \frac{1}{2\sigma^2} \sum_{i=1}^{n} a_i^2 \tag{2.2}$$

where σ is the standard deviation and by putting σ=D, it is obtained the following expression in terms of the cutoff parameter D_0 as shown in Fig.2.7.

$$H(u,v) = \begin{cases} 0 & D(u,v) > D_0 \\ D(u,v) & D(u,v) \leq D_0 \end{cases} \tag{2.3}$$

2.5 Image Segmentation

A single image can be defined as a number of different pixels. A segmentation process means that pixels of similar features are grouped together.

Cancer is a dangerous and can be a deadly illness. Identifying cancerous cell(s) as early as possible can potentially help in saving people's lives. The sizes of cancer cells can play an important role in deciding the severity of cancer. Here, the image segmentation process have a significant influence. It produces more positive outcomes by segment the abnormal cells [43]. A several techniques can be used for segmentation, such as edge-based, threshold-based, neural network-based and cluster-based methods. The most effective method of these different approaches is clustering. Clustering technique has many algorithm, such as Fuzzy C-means, K-means, mountain and subtractive [1]. In 1967, K-means algorithm is presented by MacQueen. It is one of the simplest unsupervised learning algorithms [51]. The steps of K-means algorithm is shown in Fig.2.8. On local minimal, K-means clustering algorithm often have to converge but a number of measurements must be firstly performed for finding distances and centers of the required clusters. K-means algorithm tries to find the minimum distances between all points to ensure that data points will be separated to make as most variant clusters as possible. In other words, no other iteration could have a lower average distance between the centroids and the data points found within them. To update the centroids, the iterations number increases or decreases according to the initial value of cluster centers [52]. Thus, the K-Means algorithm performance depends highly on the initial value of the cluster centers. This randomization in selection is one of the limitations of using K-means [53]. The other limitation is specifying the required number of clusters (K) and this involves some kind of experience to select a a suitable value which is often not be estimated easily [54]. K-means clustering is an unpredictable algorithm that produces various results each time. Furthermore it works well if the number of clusters increases, but it takes longer. From the above, it is concluded that K-means clustering algorithm is an unstable and provides a different output every time. Additionally its performance will be better when the value of k increases, but that definitely takes time.

There is another common segmentation method, it is a Watershed algorithm which proposed by Beucher 1979 [55]. The Watershed algorithm can be defined by high points and ridge lines that descend into lower elevations and stream valleys. For using the watershed principle in image segmentation, a local minima of the gradient of the image may be chosen as markers, in this case an over-segmentation is produced and a second step involves region merging. Watershed algorithm [36] involves the following steps:

- Calculate the gradient. The gradient used to determine the objects contours in the image as a pre-processing step. It is the lower part of the pixel values in the

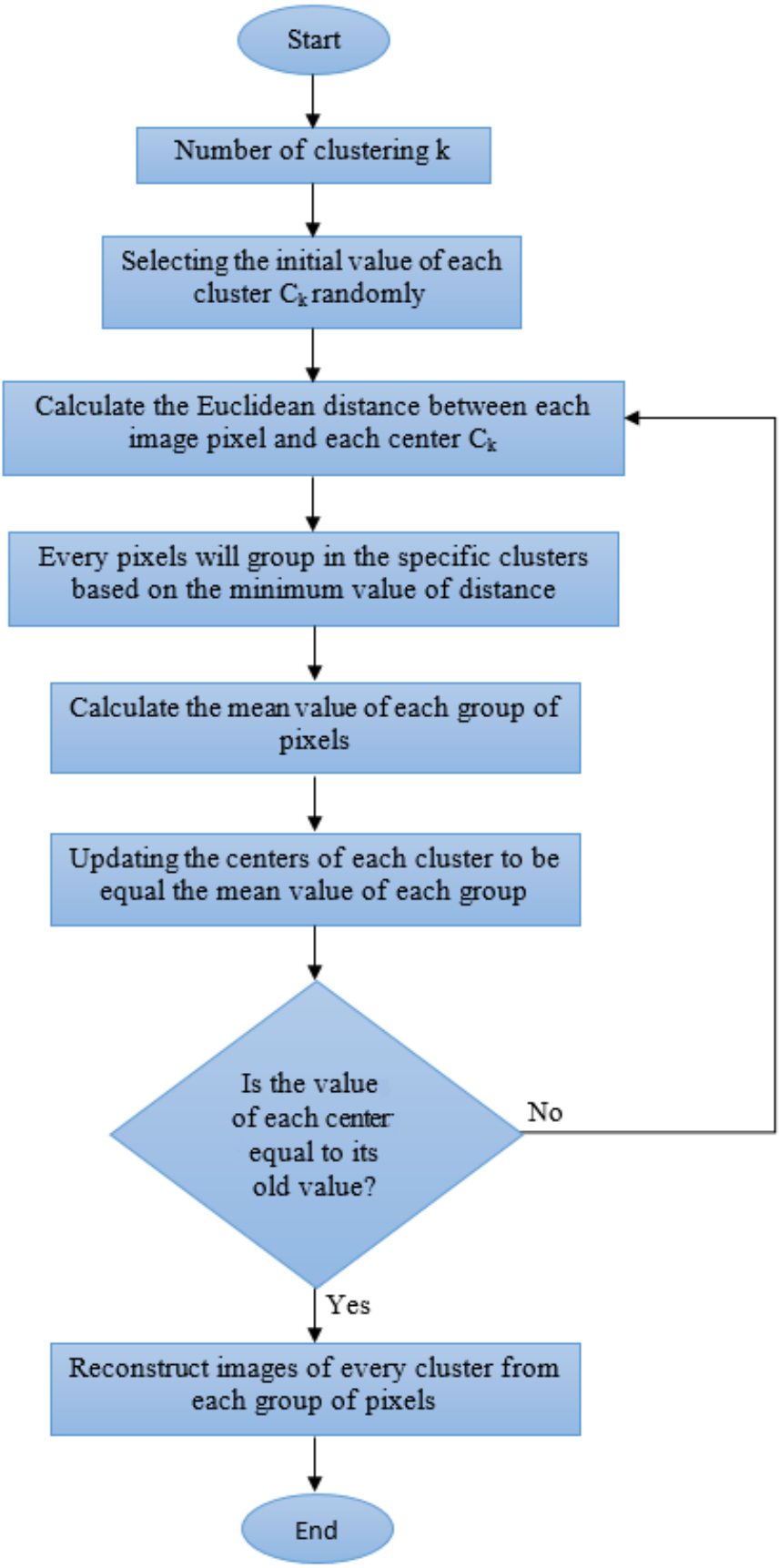

Figure 2.8 K-means algorithm of clustering.

gray scale image. It can be expressed as seen in Eq. 2.4:

$$G(f) = \left[\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2 \right]^{\frac{1}{2}} \tag{2.4}$$

The gradient will be equal to the ratio of the G(f), if f is continuously changing.

- Extract the local minima. The obtained gradient is compared with the pixel next to it and the gradient of the lowest value is selected to be the marker. The marker initialization is described by the following expression:

$$X_{lmin} = \{p \in D \mid f(p) = l_{min}\} = T_{lmin} \tag{2.5}$$

Where $l_{min} = min\{f(r) \mid f(r) < f(p), r \in G(p)\}$ and $G(p)$ is the next pixel. Now by taking pixels one by one and checking them with G_N, the local minima can be found.

- Segment by local minima. The segmentation can be done by the flooding which extends the region of high gradients at lower gradients.

2.6 Feature Selection by using Mutual Information

One of the major goals in machine learning is to discover some relations between output and input. Generally, there is a huge number of features but not all of them are required. Sometimes the output is not defined by using the whole input features space but is decided instead by only a subset of features. This type of decreasing in features number is called feature selection which is used to choose a subset of features to capture the relevant information [56]. Filter feature selection techniques can be categorized into four types (filter, embedded, wrapper, and hybrid) based on their selection mechanisms [17]. Filter feature selection are very interesting methods because they are simple with high computational efficiency. One of the most popular filters is those which uses MI for estimating the relations between each feature and target (mutual relevancy), and between each other (mutual redundancy) [23].
Mutual information can be defined as a statistical method measures the relation between two random variables. In average, it measures how much information does each variable have about the other variable [17]. For instance, let is assumed X and Y are independent, that means X has no information about Y. Hence, mutual information between X and Y is zero. Whereas the mutual information will be as same as the information of X (or Y) if X and Y are same. [57]. It is needed to define entropy, joint entropy and conditional entropy for a better understanding of the terms of mutual information, as shown in Fig.2.9.

The amount of information needed to define any random variable is known as the entropy of that variable which is a measure of its uncertainty. The entropy of the discrete random variable $Z = z_1, z_2, ..., z_N$ is denoted by $H(Z)$ as found in Eq. 2.6 [56]:

$$H(z) = -\sum P(z) log P(z) \tag{2.6}$$

Where $P(z)$ the probability mass function. When another variable c is introduced, hence the conditional entropy is the amount of uncertainty left in z. Therefore, the entropy of both variables more than the conditional entropy [38] which is equal to entropies if, and only if, the two variables are independent. The conditional entropy between z and c is defined in Eq. 2.7.

$$H(z \mid c) = -\sum\sum P(z,c) log P(z,c) \tag{2.7}$$

The relationship between the conditional and the joint entropies is found in Eq. 2.8 [17]:

$$H(z,c) = H(z) + H(c \mid z) = H(c) + H(z \mid c) \tag{2.8}$$

And,

$$H(z,c) \leq H(z) + H(c) \tag{2.9}$$

The MI between any two variables means the amount of information that these variables share as in the equation below:

$$I(z;c) = H(c) - H(c \mid z) \tag{2.10}$$

Symmetry is one of MI properties, therefore,

$$I(z;c) = I(c;z) \tag{2.11}$$

Hence,

$$\begin{aligned} I(z;c) &= H(z) - H(z|c) \\ &= H(z) + H(c) - H(z,c) \end{aligned} \tag{2.12}$$

The theoretical description of the mutual information of two random variables X and Y and their joint distribution is P(X,Y) is given by algorithm (1), as shown in Fig.2.10, that shows how the Mutual Information can be computed between X and Y .

2.7 Dimensionality Reduction

The inputs for any machine learning algorithm is represented by features covering various characteristics of the issue. Consequently, the features quality is the key in

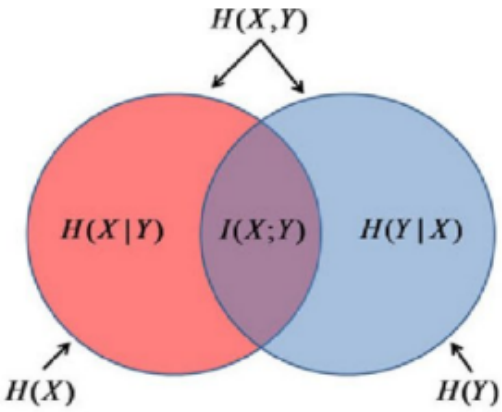

Figure 2.9 Venn diagram to show the relationships between MI and entropies [58].

Mutual Information

Input: Space of all features ***S***
Output: Mutual Scoring ***I***

1. **Repeat**
2. **Calculate the mutual information for *x* and *y* by**
   $$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) log \frac{P(x,y)}{P(x)p(y)}$$
   where; ***X*** is the space of the first group of features, ***Y*** is the space of the second group of features and ***P*** is the probability.
3. **Call** the next two variables.
4. **Until** used up all variables in the whole feature space.
5. **Return** the scoring vector ***I***.
6. **End**

Figure 2.10 Mutual information algorithm.

solving any classification problem [31]. In the field of machine learning, the efficiency of learning data and the analysis of the relationship between features and data are affected by the excessive data dimensions. Sometimes a large feature space leads to over-fitting, high processing sophistication, and poor final model interpretability [38]. However, in some cases it is appropriate to use a special techniques to minimize the dimensions of the data. PCA and SVD are the most common methods for performing dimension reduction, or in the other words for finding matrices with fewer columns.

Fundamentally, there is a difference between the feature selection and the dimension reduction methods as shown in Fig.2.11.

2.7.1 Principal Component Analysis (PCA)

In image analysis, the feature extraction stage is used for specifying acceptable features from data set [11]. The extracted features will be inputs to a next stage. In order to diminish the input dimensions, a principal component analysis method is used. In general, a PCA is a statistical technique helps to define the main directions in which the data are updated. For instant, let is assumed there is a number of variables within a certain data set represented by two original axes X and Y in Fig.2.12 (a). The same variables can be represented by other two axes U and V and it is clear from Fig.2.12 (b), that the main direction in which the variables can modify is U and the second direction which is orthogonal to U is V. Hence, the variables can be represented by one axis U. In other words, by using a PCA procedure it can select a new coordinate system described by the main direction of the variables [59, 60]. The U and V axes in Fig.2.12 (b) are called the principal components. If each variable in $XY-coordinate$ is transformed into its corresponding value in $UV-coordinate$, the whole variables in the data set will be de-correlated or the covariance value between the U and V will be zero.

Fig.2.13 shows a geometric description of the PCA in two dimensional coordinate system [61]. By using all points of the data set, it can find the mean value of the variables $(\mu x_1, \mu x_2)$ and the co-variance matrix Σ which is a 2×2 matrix in this case. If the eigenvectors of the covariance matrix is calculated, the direction vectors $\varnothing_1$ and $\varnothing_2$ is obtained. By putting the two eigenvectors as columns in the matrix $\varnothing = [\varnothing_1, \varnothing_2]$, a transformation matrix is formed. It will take the points from the $[x_1, x_2]$ coordinate system to create the $[\varnothing_1, \varnothing_2]$ coordinate system be using the following equation:

$$P_{\varnothing} = (P_x - \mu_x).\varnothing \tag{2.13}$$

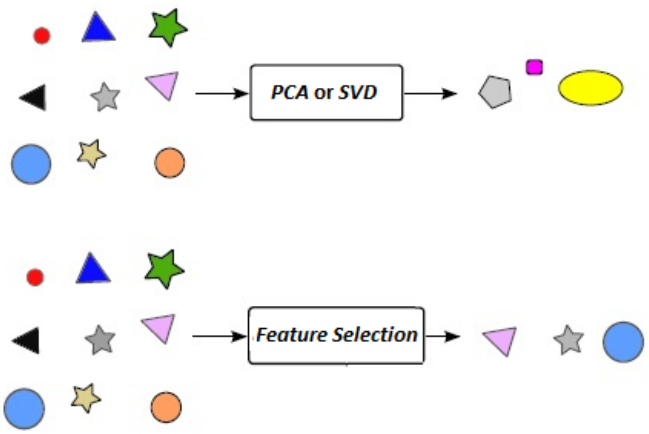

Figure 2.11 The dimension reduction (PCA and SVD) versus feature selection methods [27].

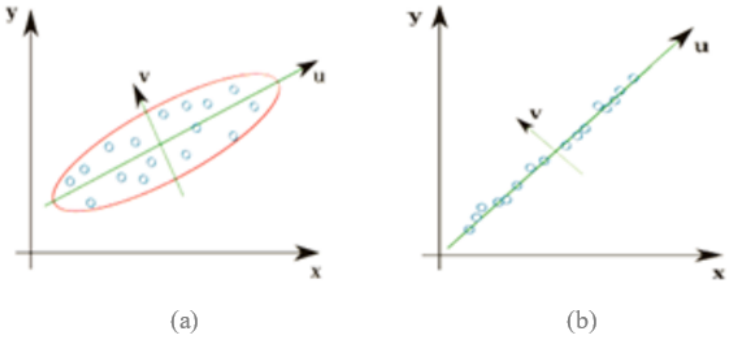

Figure 2.12 Principal Component Analysis in (a) original feature space and (b) reduced dimension space [56].

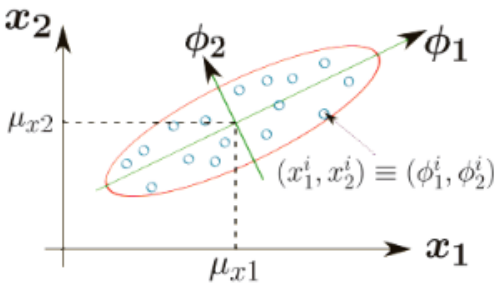

Figure 2.13 PCA data projection [61].

Where P_x is a point in $[x_1, x_2]$ coordinate system, $(\mu x_1, \mu x_2)$ is the mean value, and $P_\varnothing$ is the corresponding point in the $[\varnothing_1, \varnothing_2]$ coordinate system. The steps of PCA algorithm can seen in Fig.2.14.

The parameter K in the PCA represents the number of the components, or the number of dimensions required to reach it. There are two essential factors needed to select K. The average square projection and the total variation in the data set [62]. These two factors can be obtained by using the following equations:

$$\frac{1}{m} = \sum_{i=1}^{m} \left\| x^{(i)} - x^{(i)}_{approx.} \right\|^2 \tag{2.14}$$

$$\frac{1}{m} = \sum_{i=1}^{m} \left\| x^{(i)} \right\|^2 \tag{2.15}$$

By dividing the above two equations on each other, it is obtained:

$$\frac{\frac{1}{m} = \sum_{i=1}^{m} \left\| x^{(i)} - x^{(i)}_{approx.} \right\|^2}{\frac{1}{m} = \sum_{i=1}^{m} \left\| x^{(i)} \right\|^2} \leq 0.01 \tag{2.16}$$

From the above equation, the difference between the original feature and the reduced features, divided by the whole feature space should be less than or equal to 0.01. Usually, K have to be a value within this condition.$\alpha = 0.01$ leads to 99 of the variance can be recovered [63]. In Matlab , a S matrix (a diagonal of eigenvalues) will be found and if $\alpha = 0.01$ that means the summation of the K selected eigenvalues divided by the summation of all eigenvalues have to be greater than or equal to 99 as is shown in the following equation: [59]

$$0.99 \leq \frac{\sum_{i=1}^{k} S_{ii}}{\sum_{i=1}^{m} S_{ii}} \tag{2.17}$$

2.7.2 Singular Value Decomposition (SVD)

In data mining, the Singular value decomposition (SVD) is one of the most common unsupervised algorithm which is mainly used in high dimensions data. It is represented one of the most proper tool for mapping the data of high dimensions to other less dimensions [64]. Of course, the fewer the dimensions that have been chosen, the less accurate will be the approximation. For instance, let is assumed (X) is a $(m \times n)$ matrix of rank (r), where r of any matrix is the maximum number of linearly independent column or row vectors in that matrix [65]. r is equal to the number of

PCA algorithm

Input: Generated Data matrix $\boldsymbol{X}$, Number of principle component $\boldsymbol{d}$
Output: New Dimensions $\boldsymbol{N}$

1. **Repeat**
2. **Calculate** the mean of transactions $\mu \leftarrow \frac{1}{m}\sum_1^m x_i$
3. **Subtract** the mean from each transaction
   $X(t) \leftarrow x_i - \mu$
4. **Compute** eigenvectors $u(t)$ of AA^T from $Co(t)$
5. **Consider** matrix AA^T as a $M \times M$ matrix
6. **Calculate** the eigenvectors $v(t)$ of AA^T such that:
   $$AA^T V_i \rightarrow \mu_i V_i\,, \quad \mu_i V_i \rightarrow AA^T AV_i$$
7. **Compute** the best μ eigenvectors of AA^T: $\mu_i \leftarrow AV_i$
   Keep only K eigenvectors, (K features with their values). $U \leftarrow Top\ eigenvector(C, d)$
8. **Until** All the transactions over the time interval t become as a vector $x(t)_i$
9. **Return** $N \leftarrow U^T X$
10. **End**

Figure 2.14 PCA algorithm.

non zero singular-value (Σ) of X. The matrices U, Σ, and V can be obtained as shown in Fig. 2.15. The mathematical description of the SVD can be defined in the following equation [66, 67]:

$$X = U\Sigma V^T \tag{2.18}$$

where U is $m \times m$ matrix and the columns represents the eigenvectors of XX^T, V is $n \times n$ matrix and the columns of the V represents the eigenvectors of the $X^T X$, and Σ is the diagonal eigenvalues and also called entities or diagonal sigma's values $\Sigma_1, \ldots, \Sigma_2$ which are computed based on the square roots of the non-zero eigenvalues of the XX^T or $X^T X$ matrix. Both of them are the singular values of matrix X and they fill the first r places in the main diagonal of Σ [65, 66]. According to the following two equations, the XX^T and $X^T X$ can be described.

$$XX^T = (U\Sigma V^T)(U\Sigma V^T)^T = (U\Sigma V^T)(V\Sigma U^T) = U\Sigma^2 U^T \tag{2.19}$$

$$X^T X = (U\Sigma V^T)^T(U\Sigma V^T) = (V\Sigma U^T)(U\Sigma V^T) = V\Sigma^2 V^T \tag{2.20}$$

Where U is the eigenvector matrix of XX^T, the Σ is the eigenvalue matrix and the eigenvalues are $\lambda_1 = \sigma_1^2, \ldots, \lambda_r = \sigma_r^2$. At the same way, V is the eigenvector matrix for

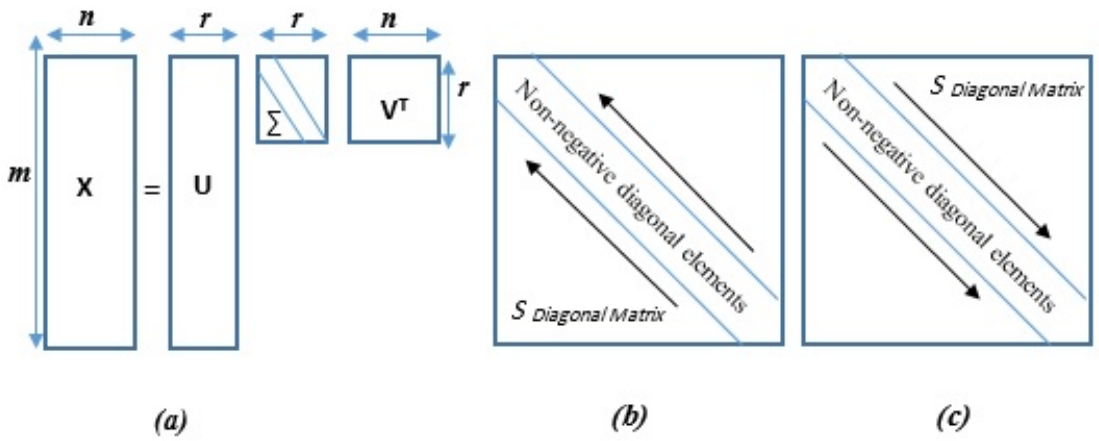

Figure 2.15 (a) The singular-value decomposition matrices, (b) up-ward Σ and (c) down-ward Σ

X^TX. The diagonal matrix Σ has the same property $[\lambda_1 = \sigma_1^2, \lambda_2 = \sigma_2^2, \ldots, \lambda_r = \sigma_r^2]$ [65]. The steps of SVD algorithm are described in Fig.2.16.

There are some important properties of the Singular Value Decomposition [67, 68]:

1. U is a $m \times r$ orthogonal matrix which each of its columns is a unit vector, and for any two columns the result of dot product is zero.
2. V is a $n \times r$ orthonormal matrix. V^T is always used which each of its rows is a unit vector.
3. Σ is a diagonal matrix which means the all elements not on the main diagonal are zero.

SVD can present a low rank approximation by considering the highest singular value that bundles most of the energy included in the image. The approximation of a matrix (X) can be represented as truncated matrix (X_k) of a specific rank r where k is smaller than r [65].

$$X = \sum_{i=1}^{k} U_i S_i V_i^T \simeq u_1 s_1 v_1^T + u_2 s_2 v_2^T + \ldots + u_k s_k v_k^T \tag{2.21}$$

where Eq. 2.21 shows that the partial sum can capture as much energy of X as possible by the truncated matrix X_k [69] i.e., the rank r of X is the non-zero elements of S. For a better understanding, Fig.2.17 shows Eq. 2.21 graphically.

Simply, the value of k is $1 \leq k \leq min(m, n)$ and a proper k can be taken based on the

SVD algorithm

Input: Generated matrix $\boldsymbol{X}$
Output: New Dimensions $\boldsymbol{C}$

1. **Repeat**
2. **Find** the singular values of matrix X by using : $\boldsymbol{X = U\Sigma V^T}$
   where; $\boldsymbol{X}$ is $m \times n$ matrix, $\boldsymbol{m}$ is no. of attributes, $\boldsymbol{n}$ is no. of variables, $\boldsymbol{U}$ is the right singular values matrix, $\boldsymbol{\Sigma}$ is a diagonal matrix, and $\boldsymbol{V}$ is left singular values matrix.
3. **Compute** the covariance matrix by using:
   $$\boldsymbol{XX^T} \leftarrow (\boldsymbol{U\Sigma V})(\boldsymbol{U\Sigma V})^{\boldsymbol{T}} = (\boldsymbol{U\Sigma V})(\boldsymbol{V\Sigma U^T}) = \boldsymbol{U\Sigma^2 U^T}$$
   $$where,\ V(V^TV = I)$$
4. **Calculate** $\sqrt[2]{\text{the eigenvalues of } XX^T}$
5. **Until** all the transactions over the time interval t become as a vector $x_i(t)$
6. **Return** $\boldsymbol{C} \leftarrow U^TX$
7. **End**

Figure 2.16 SVD algorithm.

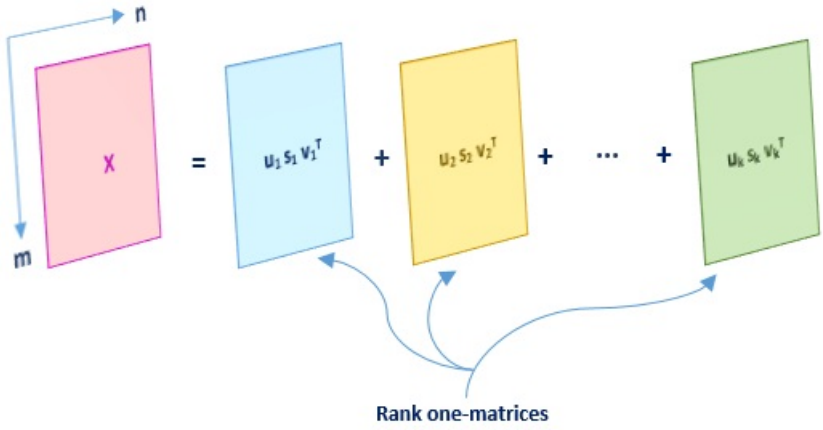

Figure 2.17 Singular value decomposition of matrix X based on the summation of k rank one-matrices.

content of energy measurement E_k as shown in Eq. 2.22 [68].

$$E_k = \frac{\|X_k\|_F}{\|X\|_F} \tag{2.22}$$

Where is the frobenius norm of the truncated matrix which can be calculated for any matrix X as follows:

$$\|X\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} x_{i,j}^2} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} diag(X^T X)} \tag{2.23}$$

In this study, $E_k \geq 0.95$ to ensure the 95% of the variance of the original matrix will be recovered.

2.8 Image Classification

Classification helps to place data into groups. A classifier model is first created based on training samples; then it is used to classify new testing samples, whereas a set of features characterizes each sample. Classification can be defined basically as the process of finding the best boundary between classes. Classification is a machine learning technique used to predict a group membership for data instances. Developing a classifier consists of choosing an analysis method, choosing a set of features, a classifier training, a classifier validating, and evaluating potential classification errors. Each step presents opportunities to introduce bias and error through the process [70]. There are many algorithms can be used for classification purposes such as using artificial neural networks and support vector machine.

2.8.1 Artificial Neural Network (ANN)

In the human being, the human mind is the main motor for making decisions. It consists of billions of nerves that are interconnected in a very complex structure. The artificial neural network (ANN) is an intelligent mathematical algorithm have been found to mimic the function and structure of the human brain neurons network [71]. It consists of three main parts: input layer, hidden layer, and output layer. Basically, the input layer is a non-processing layer in which the information is directly transmitted without any changes. The performance of the neural network model can be greatly affected by the processing operations that takes place inside the hidden layers. The output layer is the last layer and it is responsible to generate the final output after a series of checking with the desired output [72]. There are several kinds of ANNs are developed to manage a wide range of problems in many fields such as signal processing, pattern recognition, object recognition, classification and robotics [45].

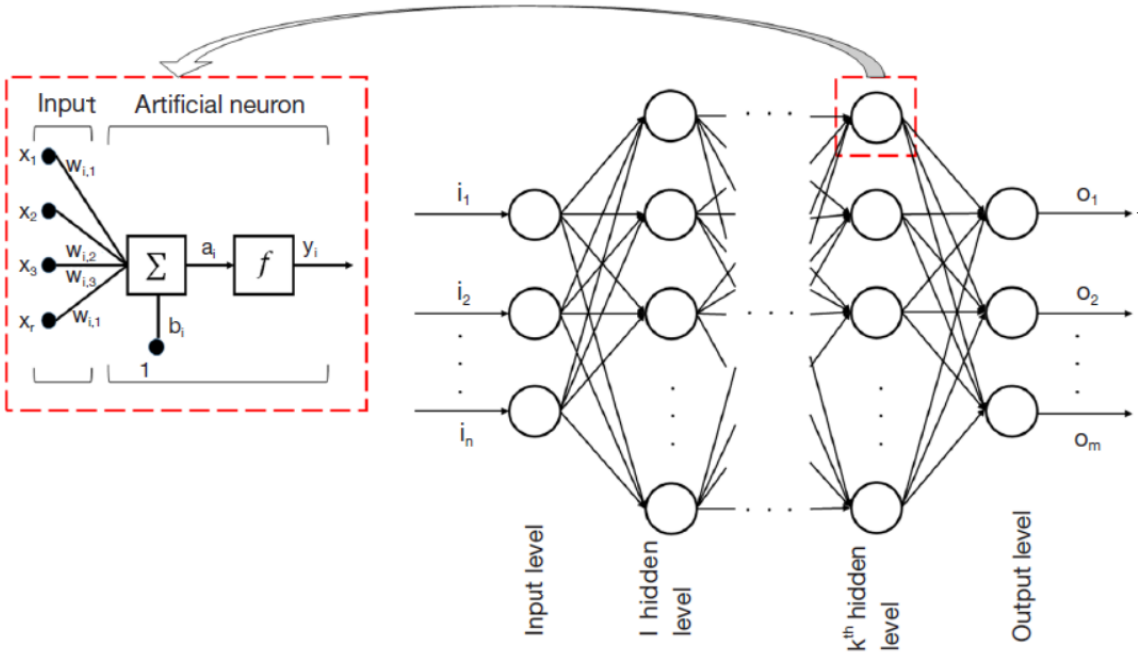

Figure 2.18 A general structure of a feed-forward multi-layer perceptron [75].

ANN applications in CAD systems are the mainstream of computational intelligence. Feed forward networks are particularly proper for medical imaging applications. In ANN, the input / output vectors offer a perfect basis for in a supervised way [73]. A multilayer perceptron (MLP) which is a special type of feed forward network employing more than two layers. It is uses a training algorithm to learn the data set by modifying the weights of neurons according to error rate between the actual and desired output. The typical structure of MLP can be shown in Fig.2.18. In general, MLP uses the back propagation algorithm (supervised learning) as a training algorithm to learn the data sets as shown in Fig.2.19. The most popular one which used an iterative descent method for minimizing mean squared error between the actual and the desired output [74]. There is a range of activation functions used to process weights and bias. The four basic functions that widely found for medical image analysis are shown in Fig.2.20.

2.8.2 Support Vector Machine (SVM)

SVM is a discriminative classifier which requires a training step to find a separating boundary for the feature space. This best decision boundary is called a hyperplane. SVM can be used to solve linear and non-linear problems [76]. SVMs are developed from the theories of statistical learning and structural risk minimization. In linear or non-linear cases, a new decision surface is computed. Then, the input space is mapped by a ϕ function in which samples are separable. In non-linear problem, the mapping process is achieved by applying a non-linear kernel function over each pair of vectors

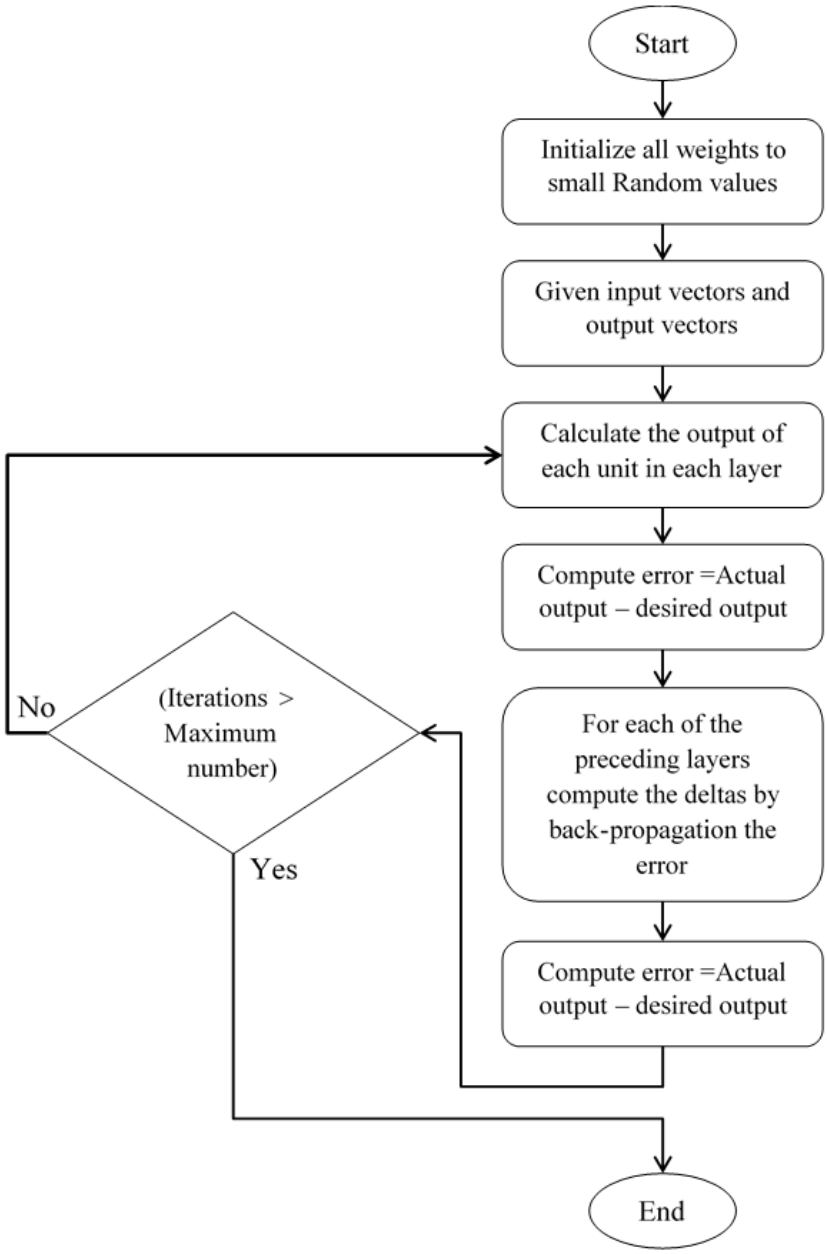

Figure 2.19 The back propagation learning algorithm.

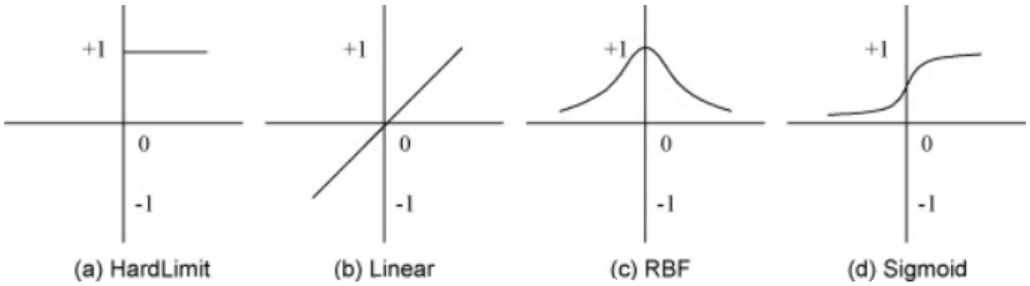

Figure 2.20 Four basic activation functions [73].

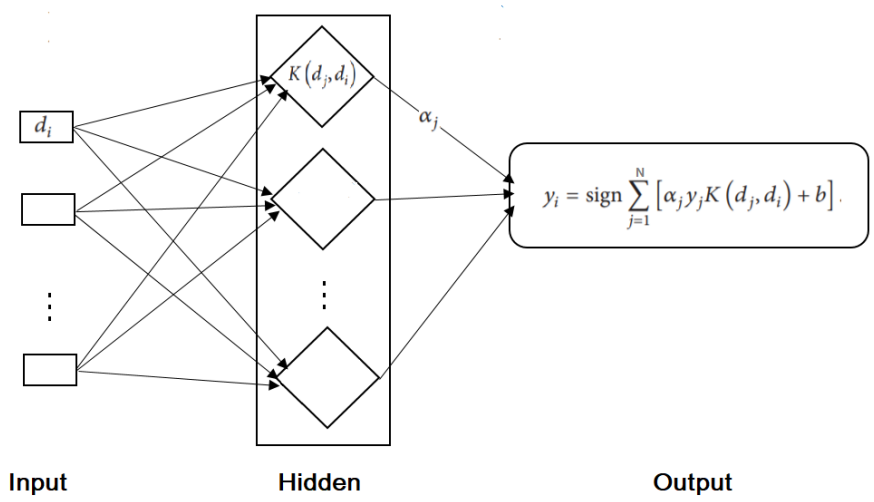

Figure 2.21 The typical structure of SVM.

[77].

The output of SVM can be formulated as follows:

$$y_i = sign \sum_{j=1}^{N} \left[\alpha_j y_j K(d_j, d_i) + b \right] \tag{2.24}$$

where N is the number of samples, d_j is the input, y_j is its class label and $K(d_j, d_i)$ is the kernel. α_j can be obtained by solving the following equations:

$$maximize \qquad \sum_{j=1}^{N} \alpha_j - \frac{1}{2} \sum_{j,i=1}^{N} \alpha_j \alpha_i K(d_j, d_i) \tag{2.25}$$

$$subject\ to \qquad \sum_{j=1}^{N} \alpha_j y_j = 0, \ \ 0 \leq \alpha_j \leq C, \ \ j = 1, 2, ..., N \tag{2.26}$$

The attractive aspect of using SVMs is the possibility of using a kernel function [78]. Polynomial and Gaussian Radial Basis Function (RBF) kernels are two widely used functions for non-linear problems [76]. The mathematical equation of the Gaussian kernel is found in Eq. 2.27.

$$K_{Gaussian}(d_j, d_i) = e^{\frac{\left\| d_j - d_i \right\|^2}{2\sigma}} \tag{2.27}$$

where σ is the Gaussian sigma (kernel width). The typical structure of SVM is shown in Fig.2.21.

For non-linear data, it is impossible to draw a straight line to separate between classes.

Hence, one more dimension has to be added [79]. It can be calculated as:

$$z = x^2 + y^2 \tag{2.28}$$

where z is the new dimension, and (x, y) are the original dimensions.

3
MATERIAL AND METHOD

3.1 Data sets

In this study, two types of data set were used to evaluate the effectiveness of the proposed system. The two data sets are:

- The main data set;
  A standard data set were used. It is one of the most reliable data sets shared by The Cancer Imaging Archive (TCIA) [80, 81]. There are axial plane MR images of 160 patients. All images are FLAIR, RGB, of size 256 × 256 pixels and 8 bit. Samples are shown in Fig.3.1 (a), (b), and (c). Many studies preferred using one acquisition format, FLAIR, to validate their proposed system [5, 21, 82].

  Three classes of MR brain images were classified: normal, HGG, and LGG. For each class, the same number of images was randomly selected as shown in Table 3.1. It is necessary to keep in mind that each subset included patients from various classes to guarantee the validity of the classifier performance.

- The secondary data set;
  A real data set collected from the Iraqi Center for Research and Magnetic Resonance of Al-Kadhimain Medical City in Iraq [23] was used to check the validation of the proposed system. There are 322 (axial plane) MR images of 107 for both normal and abnormal patients. The images are mixed FLAIR and T2-weighted, RGB, of different sizes. Samples are shown in Fig.3.1 (d), and (e).

  The slices were separated into two groups based on the experience of the center specialists. Two classes were classified: normal and abnormal.

The data set was divided into training and testing sets of size 80% and 20%, respectively. It is necessary to keep in mind that each subset included patients from various classes to guarantee the validity of the classifier performance. The training set was divided into five folds. Each time, one of these folds was used as a validation set,

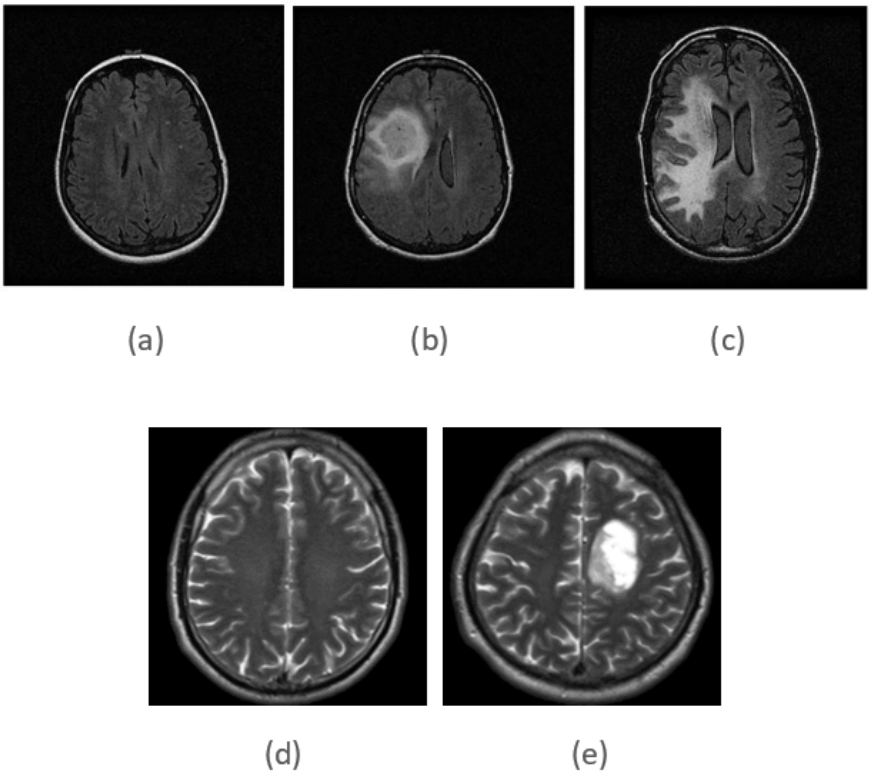

Figure 3.1 Three samples from the standard data set, a) normal, b) LGG, and c) HGG. Also, two samples from the real data set, d) normal, and e) abnormal

and the remaining four folds were used as training set as shown in Fig.3.2. Later, the performance of the trained classifier is evaluated by the votes collected from each fold. Using MATLAB permits the selection of the optimal scaling via a heuristic procedure automatically by sub-sampling [83, 84]. After that, the testing set was fed to the classifier. Basically when a classification method has been validated using a standard cross-validation scheme, an unbiased predictor can be produced. But running a model on an independent and separated data set for testing can provide a more reliable assessment. This method called cross-validation and testing approach [85, 86].

Once again, in this study, the data set was divided into two separate sets (cross-validation set and test set) for only one time . Firstly, different models were trained and validated with five-fold cross-validation. Hence the best set of parameters was decided. The accuracy of prediction and the selected parameters were evaluated on the test set. The division of the selected can be seen in the Table 3.1.

3.2 Software

MATLAB version 15a was used as the platform for programming and experiments in this study since MATLAB demonstrates a high level of performance in integrating

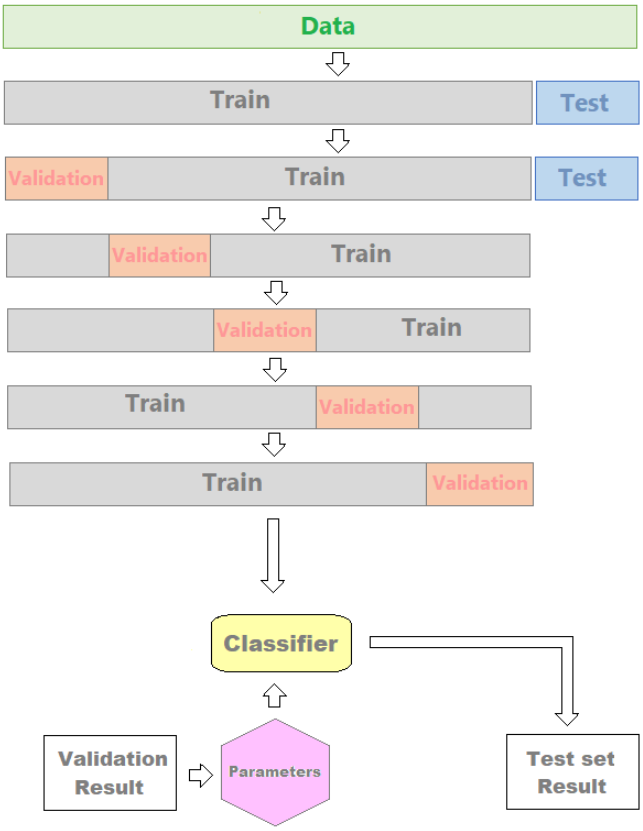

Figure 3.2 Data set division.

Table 3.1 The division of the main standard data set.

Class	**Total**	**Training** Training (80%)	**(80%)** Validation (20%)	**Testing (20%)**
Normal	137	88	22	27
HGG	137	88	22	27
LGG	137	88	22	27
Total	411	264	66	81

computation, visualization and programming. It was installed on Windows 8, Intel R core i7-4500U of CPU 2.40 GHz and RAM 16.0 GB.

3.3 Framework of the Proposed System

This study presented an intelligent model that has six stages. A block diagram which describes these stages can be seen in Fig. 3.3.

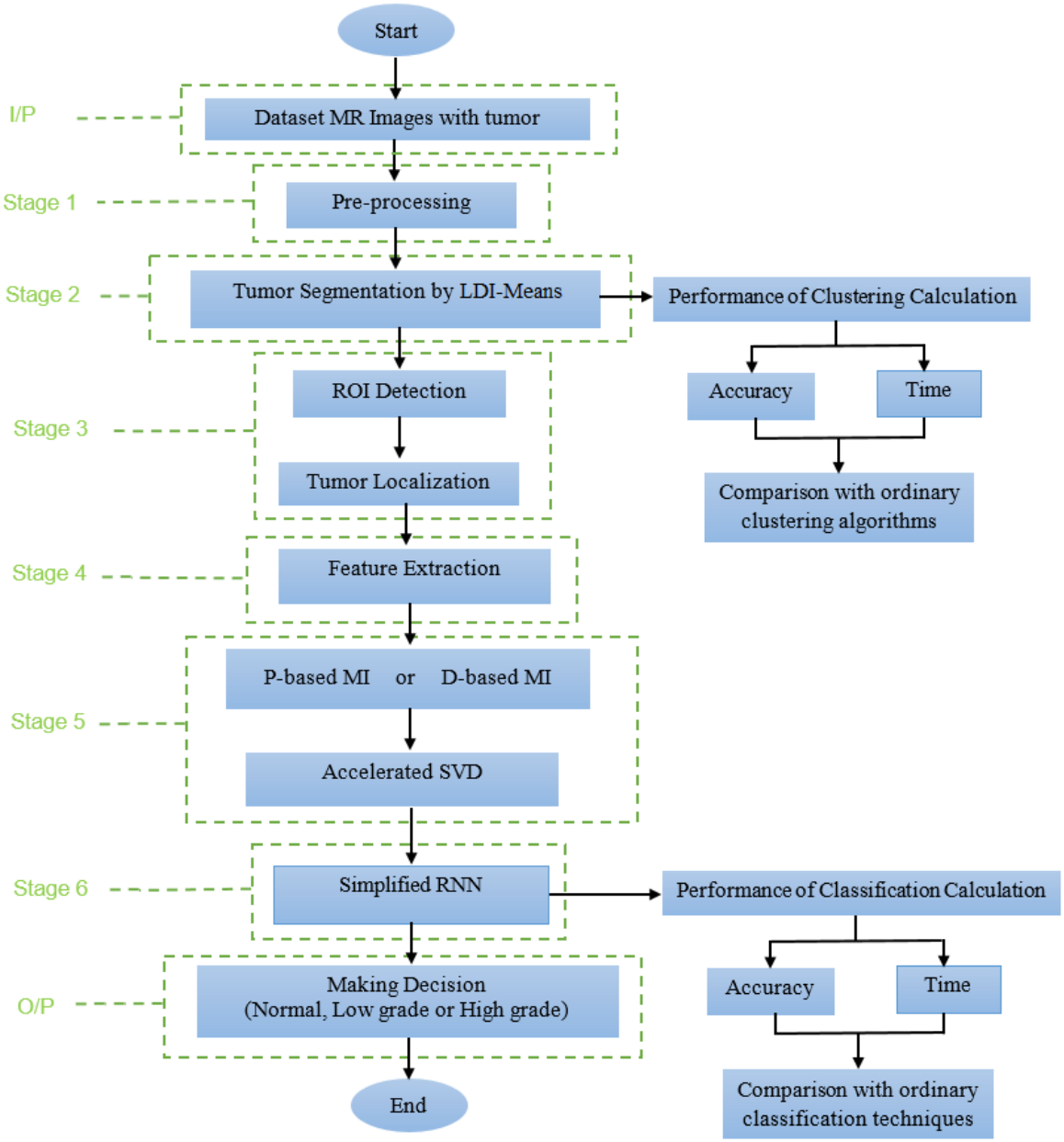

Figure 3.3 Block diagram of the proposed system.

3.3.1 Stage 1: Pre-processing

Several steps were employed on the data set at this stage to make it more proper for next processes. Firstly, the standard data set was converted from $.dicom$ to $.jpg$ format by using MATLAB conversion tool. Furthermore, due to the inhomogeneity, noise, and variety of the intensity ranges and contrast, some pre-processing steps were required to improve the resolution of the images and prepare them for the next stage [49]. Several de-nosing techniques can be used such as median smoothing filter.
In the proposed system, a neighbourhood of 5×5 has been used because if an outsized neighbourhood is selected, a severe smoothing is produced. In addition, skull removal is an important process in brain image analysis and this is required for an efficient examination of brain tissues. The removal of non-brain tissues such as bones, eyes, fat, etc. from the MRI scans helps to increase the accuracy and speeds up the tumor segmentation process [87]. There are various techniques for skull stripping such as by using image contour or histogram analysis [88, 89]. Here, the skull stripping was performed by using threshold value method [8]. Lastly, the intensity level of the image pixels in each channel have to be adjusted which is an essential step for applying LDI-Means clustering algorithm in the next step. The block diagram of pre-processing steps is shown in Fig.3.4.

3.3.2 Stage 2: Clustering by using LDI-Means for Segmentation

Image segmentation is one of the vital tools in medical image analysis. It is important for extracting the ROI from the background. Medical images are segmented using different techniques and the processed outputs are almost used for the further analysis such as classification. A new method named local difference in intensity - Means (LDI-Means) of clustering algorithm has been used for this stage. This algorithm produces a very stable and precise clusters in less processing time compared with the k-means [48]. The flowchart of LDI-Means is shown in Fig.3.5. It involves finding some factors such as the range value, the increment value, the initial value of each cluster centroid, the absolute value of the difference and the mean value by using the following equations:

$$Range = max.Int.value - min.Int.value \tag{3.1}$$

Where, $max.Int.value$ and $min.Int.value$ are the maximum and minimum values of image intensity, respectively.

$$Increment\ \ value = Range/N \tag{3.2}$$

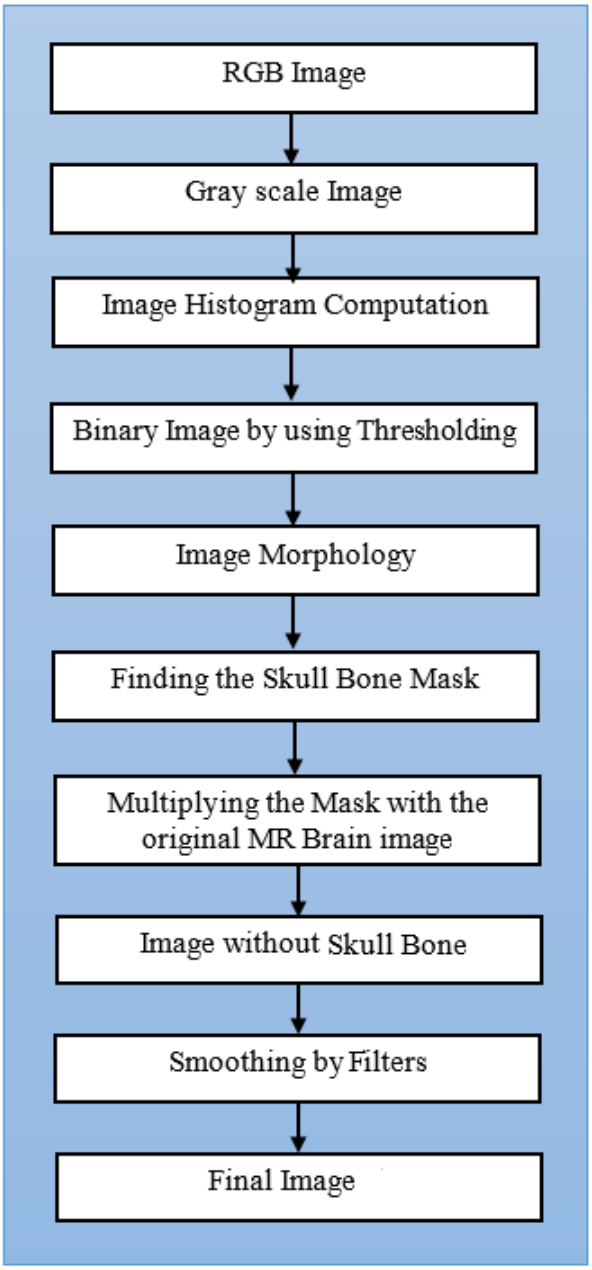

Figure 3.4 Block diagram of pre-processing stage.

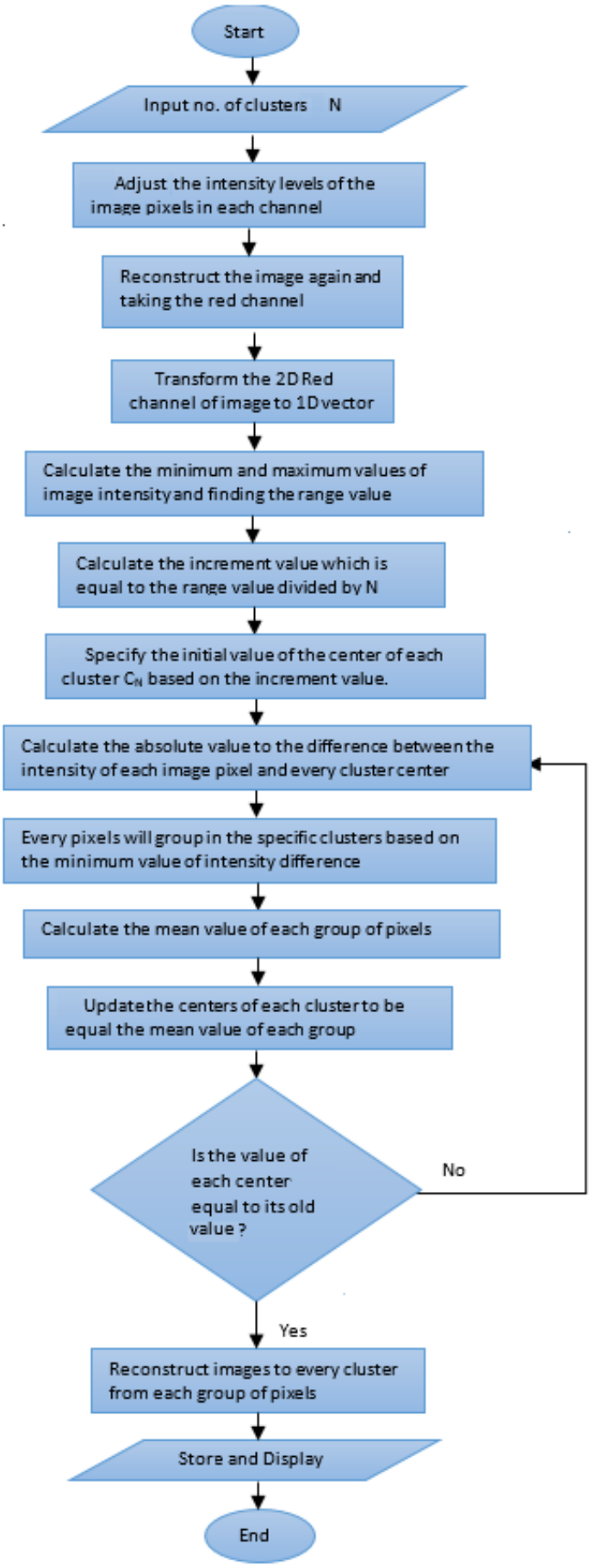

Figure 3.5 The flowchart of the LDI-Means algorithm.

where N is the number of clusters.

$$\begin{aligned} C_1 &= 1 \times Increment\ \ value \\ C_2 &= 2 \times Increment\ \ value \\ &\vdots \\ C_N &= N \times Increment\ \ value \end{aligned} \tag{3.3}$$

where C_1, C_2, ..., C_N are the initial values for the clusters.

$$Difference = |P(i) - C_n| \tag{3.4}$$

where, $P(i)$ is the intensity value of pixel, C_n is the center of the cluster n and $n = 1, 2, \ldots, N$.

$$M = \frac{\sum P(i)}{I} \tag{3.5}$$

where, M is the mean value, $P(i)$ is the intensity value of pixel in each cluster, I is the total number of the pixels in each cluster.

Let is assumed an image I(x×y) needs to be clustered into N, which is the number of clusters. And let p (x, y) be the an input pixel to be clustered and C_n be the cluster centroid of n. where n= 1,2, ..., N. The steps involved in the LDI-Means clustering algorithm are as following:

1. Select N (number of clusters).
2. Find the maximum and minimum values of the image intensity.
3. Use Eq. (3.1) to calculate the range and Eq. (3.2) to find the increment value.
4. Specify the initial centroid value for each cluster ($c_1, c_2, \ldots, c_N$) based on Eq. (3.3).
5. Calculate the difference between the selected centre and each image pixel of by using Eq. (3.4).
6. Assign all the pixels based on the absolute difference value to a cluster which has minimum difference in intensity.
7. After setting all the pixels, recalculate the new centroid value using Eq. (3.5) where the mean value of each cluster will represent the new centroid.
8. Repeat the steps 5, 6, and 7 until it meets the tolerance.

9. Reconstruct the image from each set of cluster pixels.

10. Find the segmented tumor (ROI) in the last cluster for sure.

3.3.3 Stage 3: Tumor Detection and Localization

This stage involves using some mathematical functions to find the location of the tumor within the brain image in terms of (x,y) and finding the tumor to brain tissues ratio as well as the tumor metric size. Firstly, based on the obtained binary tumor image from previous stage (clustering stage) the center of the irregular shape is computed by using some built-in functions in MATLAB. Later, a boundary box is drawn around the lesion area in the original image by using edge detection and shape factor analysis functions in MATLAB. Then, the image of tumor area is cropped to be used in the next stage.

The first three stages of the proposed system can be shown in Fig.3.6 as a block diagram.

3.3.4 Stage 4: Feature Extraction

In general, the image can be transformed into a number of features which describe its main characteristics. Texture is one of the main properties used to identify ROI in an image. Hence, gray level co-occurrence matrix (GLCM) has been presented by Haralick at al. [90] to describe some easily computable textural features in images. GLCM is a statistical method for several properties that are calculated in four directions 0, 45, 90, and 135 [15, 91, 92]. However, the mean and the standard deviation of the feature vector for the four directional features within the distance of one pixel were computed in this study. For feature extraction, the GLCM technique has been employed in two steps: firstly, computing the GLCM, and then, calculating the texture features based on the GLCM [8]. Two types of features, intensity based and GLCM based, were extracted as found in Fig.3.7. The total number of extracted features in this study is 40.

3.3.5 Stage 5: MI+SVD

This stage is the key part of the proposed system. It involves using MI+SVD. It is a novel approach that has been implemented by using a combination of two techniques: feature selection method based on mutual information (MI) theory and dimension reduction method based on singular value decomposition (SVD). Furthermore, this

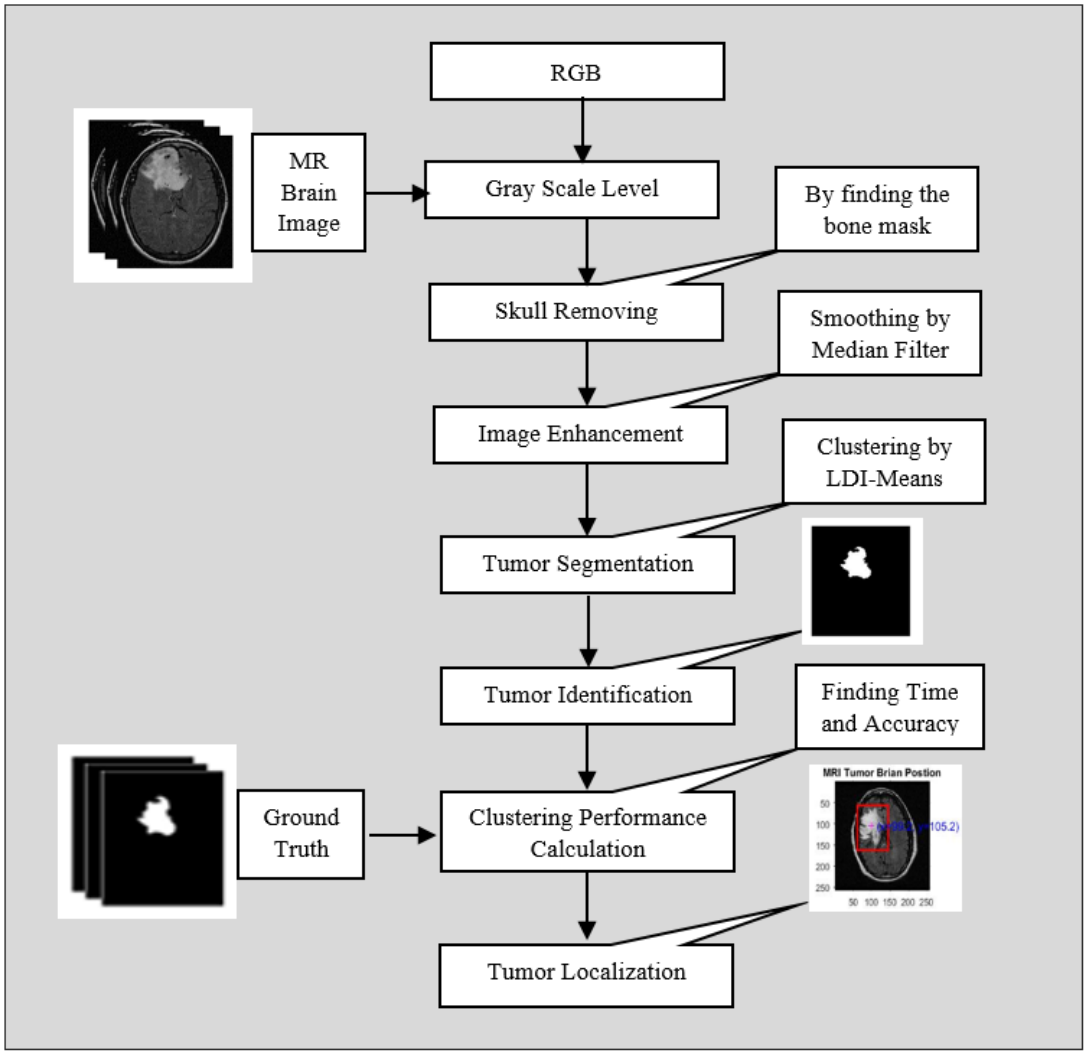

Figure 3.6 Stage 1, 2 and 3 of the proposed model [48].

	Name	Equations
Intensity Based Features	Mean	$f1 = \frac{\sum x}{n}$
	Variance	$f2 = \frac{\sum(x - f_1)^2}{n-1}$
	Standard Deviation	$f3 = \sqrt{\frac{\sum(x - f_1)^2}{n-1}}$
	Skewness	$f4 = \left(\frac{1}{f_2}\right)\sum_{x=1}^{m}\sum_{y=1}^{n}(f(x,y) - f_1)^3$
	Kurtosis	$f5 = \left(\frac{1}{f2}\right)\sum_{x=1}^{m}\sum_{y=1}^{n}(f(x,y) - f_1)^4$
GLCM Texture Features	Entropy	$f6 = -\sum_{i,j} P(i,j) \ \ log \ P(i,j)$
	Correlation	$f7 = \sum_{i,j}\frac{(x-\mu_i)(y-\mu_j)P(i,j)}{\sigma_i\sigma_j}$
	Contrast	$f8 = \sum_{i,j}\lvert i-j\rvert^2 P(i,j)$
	Energy	$f9 = \sqrt{\sum_{i,j} P(i,j)^2}$
	Homogenity	$f10 = \sum_{i,j}\frac{P(i,j)}{1+\lvert i-j\rvert}$
	Dissmilarity	$f11 = \sum_{i,j}\lvert i-j\rvert \ \ P(i,j)$
	Autocorrelation	$f12 = \sum_{i,j}(i,j)P(i,j)$
	Difference entropy	$f13 = -\sum_{i=0}^{2G} P_{x-y}(i)\log\left(P_{x-y}(i)\right)$
	Difference variance	$f14 = \sum_{i=0}^{2G}(i - f_{11})^2 P_{x-y}(i)$
	Max. probability	$f15 = max_{i,j} P(i,j)$
	Sum average	$f16 = \sum_{i=0}^{2G} i \, P_{x+y}$
	Sum entropy	$f17 = -\sum_{i=0}^{2G} P_{x+y}(i)\log\left(P_{x+y}(i)\right)$
	Sum variance	$f18 = \sum_{i=0}^{2G}(i - f_{17})^2 P_{x+y}(i)$
	Inverse difference	$f19 = \sum_{i,j=1}^{G}\frac{1}{1+\lvert i-j\rvert}P(i,j)$
	Inverse difference normalized	$f20 = \sum_{i,j=1}^{G}\frac{1}{1+(\lvert i-j\rvert \ /G)}P(i,j)$

Figure 3.7 Mathematical description of the extracted features [93].

combination produces some development of the ordinary SVD to work in accelerating mode, by exploiting the selection of multi-eigenvalue theory. In the field of machine learning and data mining for brain tumor identification, this novel approach has not been previously described.
The main objective of MI+SVD algorithm is to decrease the space of features and identify a group of meaningful features that allow a valid classification model to be built. MI+SVD is accomplished in two steps:

- Reranking all the extracted features by using MI theory. MI is a model free method with no parameters used for scoring a set of attributes, in which the high MI between a feature and a class label refers to the relevance of that feature. A subset of features is therefore selected based on the previously specified threshold. The obtained relevant features is ready for the next step.
- Applying accelerating SVD. This step is based on a multiple eigenvalues selection to find such a non-biased value for the robust k which the number of required dimension.

A mathematical MI model has been developed by this study to examine the entire feature space. This model specified the threshold value of MI to sort the inputs. It is possible to think of MI as a reduction in uncertainty about one random variable given knowledge of another. In a particular sense, mutual information one of many quantities that measures how much one random variable tells us about another [57]. In general, let is assumed Y_i and Y_j are two variables. Their joint distribution is $H(Y_i|Y_j)$ and their MI is denoted by $I(Y_i;Y_j)$, which can be defined by Eq. 3.6:

$$\begin{aligned} I(Y_i;Y_j) &= H(Y_i) - H(Y_i|Y_j) \\ &= H(Y_j) - H(Y_j|Y_i) \\ &= I(Y_j;Y_i) \end{aligned} \tag{3.6}$$

Fundamentally, the range of MI value can be obtained by using Eq. 3.7:

$$0 \leq I(Y_i;Y_j) \leq min(H(Y_i) - H(Y_j)) \tag{3.7}$$

In this study, MI was performed based on two different calculations; one of them is based on the probability value and the other is based on the distance metric.

1. Probability-Based MI
   The mutual information quantity $I(Y_i, Y_J)$ is non-negative value and can be

obtained based on the Eq. 3.6 as follows:

$$H(Y_i) \geq H(Y_i|Y_j) \tag{3.8}$$

and

$$I(Y_i;Y_j) = H(Y_i) + H(Y_j) - H(Y_i,Y_j) \tag{3.9}$$

The joint MI is defined as in equations below:

$$I(Y_i;Y_j|Y_k) = H(Y_i|Y_j) - H(Y_i|Y_j,Y_k) \tag{3.10}$$

and,

$$I(Y_i,Y_k;Y_j) = I(Y_i;Y_j|Y_k) + I(Y_k;Y_j) \tag{3.11}$$

Interaction information is the amount of information included in all features, but it cannot be found in any subset of features [94]. It can be defined as in Eq.3.12.

$$I(Y_i,Y_k;Y_j) = I(Y_i;Y_k) - I(Y_i;Y_k|Y_j) \tag{3.12}$$

High interaction information means a large amount of information given by the three variables together. In general it can be zero, positive or negative. Furthermore, the positivity for Markov chain can be approved as found in the Eq.3.14.

$$\begin{aligned} I(Y;Y,Y) &= H(Y_i) - H(Y_i|Y_k,Y_j) \\ &= H(Y_i) - H(Y_i|Y_k) \\ &= I(Y_i;Y_k) \end{aligned} \tag{3.13}$$

Thus,

$$\begin{aligned} I(Y_i;Y_k;Y_j) &= I(Y_i;Y_k) - I(Y_i;Y_k|Y_j) \\ &= I(Y_i;Y_k,Y_j) - I(Y_i;Y_k|Y_j) \\ &= I(Y_i;Y_j) \geq 0 \end{aligned} \tag{3.14}$$

In the proposed approach, after the normalization process, the MI scoring became in the range of [0,1]. Also, the cumulative distribution function (CDF) was applied on the whole normalized MI scoring to get the probability distribution of the features instead of population. Later, the positive values were selected, whereas the negative values were neglected based on the uncertainty theory of MI, which is explained in the mathematical equations above.

2. Distance-Based MI

   By definition, for independent variables, MI converges towards 0. MI is not a distance as well as not bounded. But MI can be bounded distance by normalizing its value, then subtracting it from 1. There are two different methods for

normalization. Either by the maximum possible MI of two variables as in the Eq. 3.15:

$$d_{CR}(Y_i, Y_j) = 1 - \frac{I(Y_i, Y_j)}{min(H(Y_i), H(Y_j))} \tag{3.15}$$

Thus,

$$0 \leq d_{CR}(Y_i, Y_j) \leq 1 \tag{3.16}$$

or by the maximum entropy of both variables as shown in Eq. 3.17:

$$d_{CL}(Y_i, Y_j) = 1 - \frac{I(Y_i, Y_j)}{max(H(Y_i), H(Y_j))} \tag{3.17}$$

Thus,

$$1 - \frac{min(H(Y_i), H(Y_j))}{max(H(Y_i), H(Y_j))} \leq d_{CL}(Y_i, Y_j) \leq 1 \tag{3.18}$$

For classification purposes, the distance function d_{CL} can be better choice than d_{CR} because it satisfies the triangle inequality.
d_{CL} can be written as :

$$d_{CL}(Y_i, Y_j) = max\left(\frac{H(Y_i \mid Y_j)}{H(Y_i)}, \frac{H(Y_j \mid Y_i)}{H(Y_j)}\right) \tag{3.19}$$

This is closely related to the similarity metric which proposed by Kolmogorov complexity which has been proven to satisfy triangle inequities up to a constant additive term [95]. By applying the chain rule twice for three variables Y_i, Y_j, and Y_k it will obtain the following equations:

$$H(Y_i \mid Y_j) = H(Y_k, Y_j) + H(Y_i \mid Y_j, Y_k) - H(Y_k \mid Y_i, Y_j) \tag{3.20}$$

Hence,

$$H(Y_i \mid Y_j) \leq H(Y_k, Y_j) + H(Y_i \mid Y_k) \tag{3.21}$$

In order to achieve a small value of distance metric between the selected features, which is called the closed-interval feature score (semi-closed interval). The empirical distribution function is used to project the final MI-based distance score and find the final feature space by setting the threshold value to 0.5 and less. CDF is a cumulative distribution function of a real-valued random variable X as shown in Eq. 3.22.

$$F_X(X) = D(X \leq x \in I_v \leq 0.5) \tag{3.22}$$

where $D(X \leq x)$ represents the distance of the whole feature space X, which

Distance-based MI algorithm

Input: Space of all features $\boldsymbol{S}$
Output: Mutual Scoring $\boldsymbol{I}$

1. **Repeat**
2. **Calculate** the mutual information for $\boldsymbol{x}$ and $\boldsymbol{y}$
$$\boldsymbol{I}(X,Y) = D(X,Y) = \max\left(\frac{H(x|y)}{H(x)}, \frac{H(y|x)}{H(y)}\right)$$
where $\boldsymbol{D}$ is the Euclidean distance function
3. **Find** $H(X|Y) \leq H(Z|Y) + H(X|Z)$
where:
$\boldsymbol{X}$ is the space of the first group of features, $\boldsymbol{Y}$ is the space of the second group of features and $\boldsymbol{Z}$ is the result of applying the chain rules twice on the tested variables.
4. **Call** the next two variables
5. **Until** all the variables of the feature space have been used
6. **Return** vector $\boldsymbol{I}$
7. **End**

Figure 3.8 The MI algorithm based on the distance metric [96].

selects only the values for which the MI score is less than or equal to x. The distance of X in the semi-closed interval is shown in Eq.3.23.

$$Selection_{threshold} = \forall_X F_X(a) F_X(b) = D(a < x \leqslant b) \tag{3.23}$$

In Eq. 3.23, the "less than or equal" sign illustrates the convention of the closest discrete distribution features falling between the lower bound distance score, which is 0 ("very close or the same"), and the upper bound score, which is 0.5 ("fairly close").

After performing MI scoring, the SVD is developed to work in an acceleration mode by estimating a non-biased threshold for the required dimension as shown in Fig.3.9. The following steps explain the MI+SVD algorithm in detail:

1. Load the matrix $X_{m,n}$, where m is number of variables, n is number of attributes.
2. Initialize an empty set R.
3. For all features $f(i) \in F$, compute the first equation for P-based MI and the

second one if the approach is D-based MI:

$$I(f,c) = \sum\sum P(f,c) log \frac{P(f,c)}{P(f)p(c)}$$

or,

$$D(f,c) = max\left(\frac{H(f \mid c)}{H(f)}, \frac{H(c \mid f)}{H(c)}\right)$$

where f and c represent feature and class, respectively.

4. Rank the features according to their scores and store them in the set R.
5. Repeat until using up all variables in the whole features space.
6. Normalize the MI scores.
7. Apply the empirical CDF.
8. Keep the positive values only.
9. Form a new matrix of $X_{m,n}$. Its columns will include the selected group of features, in which the first column contains the features of the highest MI value and so on.
10. Check the following condition:

$$If \quad \frac{variables\ no.}{features\ no.} \geq 1$$

11. Compute the covariance matrix:

$$Y \leftarrow XX^T$$

12. Find the eigenvalues and the left eigenvectors (V) of Y.
13. Sort the eigenvalues in descending order.
14. Order the eigenvectors of V based on their corresponding eigenvalues.
15. Calculate the : $\sqrt{the\ \ eigenvalues\ \ of\ \ XX^T}$
16. Form a diagonal matrix S based on the previous step.
17. Compute matrix D using element-by-element multiplication between each column in V and its corresponding eigenvalues of power -1.
18. Find the left singular decomposition (U).

19. Form a new dataset matrix using the obtained V, Σ, and U.

20. Calculate the frobenius norm and check the E_k .

21. Repeat until convergence.

3.3.6 Stage 6: Classification

In this study, three classifiers were used. These are: MLP, SVM, and simplified RNN.

1. MLP

   A multilayer perceptron (MLP) which is a special type of feed forward network employing three layers was used in this study through the assistance of Neural Network Toolbox for MATLAB. Its structure is shown in Fig.3.10. Output nodes number represent the classes number and they are three in the proposed MLP. Whereas the input nodes number was changeable based on the number of features. Also, based on the inputs, the nodes of hidden layer was selected. In neural networks, the activation function is one of the main components. It is used to take the decision for generating suitable output to a given set of inputs. There are many types of activation functions but the ones in the hidden and the output layers, which were used in this study, can be found in Table 3.2.

   Table 3.2 The activation functions used in the proposed MLP network.

Layer No.	Layer Type	Activation Function	Mathematical Description
Layer 2	Hidden Layer	Sigmoid	$f(x) = \frac{1}{1-e^{-x}}$
Layer 3	Classification Output	SoftMax	$f(x_i) = \frac{e^{x_i}}{\Sigma e^{x_i}}$

   Learning the proposed structure is one of the difficulties in the neural network field. It is the way by which the neural network can learn to do things. Using learning algorithms makes the network able to understand a pattern in different sets of data. This is done based on some mathematical equations that are used to adjust the parameters of the network to their optimum values for a specific task. Gradient Descent Back Propagation is one of the learning algorithms. There are numerous suggested ways to enhance the convergence rate of Gradient Descent

Accelerated SVD algorithm

Input: Data matrix $\boldsymbol{X}$
Output: New Dimensions $\boldsymbol{C}$

1. **Repeat**
2. **Construct** the covariance matrix $\boldsymbol{X}$ from the decomposition according to:
   If $\frac{No.\ of\ Features}{No.\ of\ Samples} \geq 1$ **then** $Data \leftarrow X^T X$
   else $Data \leftarrow XX^T$
   End if
3. **Calculate** the d-dimensional mean vectors for each class from X.
   $XX^T = (USV^T)(USV^T)^T = (USV^T)(VSU^T)$
4. **Calculate** V as an orthogonal matrix
   $(V^T V = I), \quad XX^T = US^2U^T$
5. **Calculate** the scatter matrices (between-class and within-class scatter $\boldsymbol{X}$).
6. **Calculate** $\sqrt[2]{\text{the eigenvalues of } XX^T} =$ singular values of X
7. **Compute** the eigenvectors $(e_1,\ldots e_d)$ and the corresponding eigenvalues $(\lambda_{1,\ldots,}\lambda_d)$.
8. **Order** the eigenvectors by decreasing eigenvalues.
9. **Select** k eigenvectors with the smallest error (square root) from a $\boldsymbol{d} \times \boldsymbol{k}$ dimensional matrix $\boldsymbol{W}$ (where every column represents an eigenvector).
10. **Use** this $\boldsymbol{d} \times \boldsymbol{k}$ eigenvector matrix to transform the samples into the new subspace. $\boldsymbol{Y} \leftarrow \boldsymbol{X} \times \boldsymbol{W}$ where ($\boldsymbol{X}$ is an $\boldsymbol{n} \times \boldsymbol{d}$ dimensional matrix, and $\boldsymbol{Y}$ is the transformed $\boldsymbol{n} \times \boldsymbol{k}$ dimensional samples in the new subspace)
11. **Until** Convergence
12. **End**

Figure 3.9 The accelerated SVD algorithm [96].

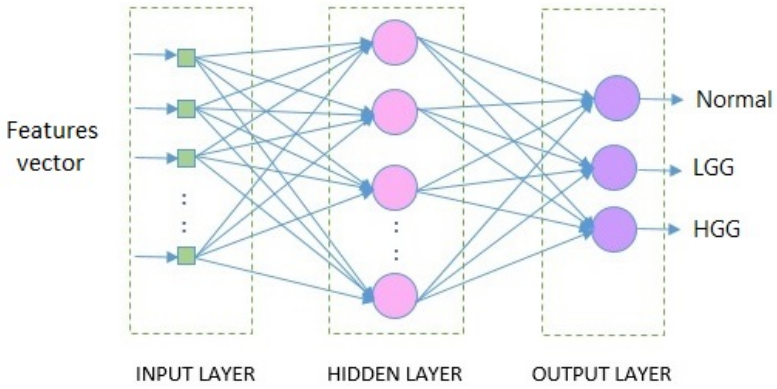

Figure 3.10 Architecture of the proposed MLP network.

Back Propagation algorithm such as choosing of initial weights and biases, network topology, rate of learning, value of momentum, activation function and its gain value [97]. In this study, the Gradient Descent Back Propagation with an adaptive learning rate and momentum method is used to adjust biases and weights [98].

2. RBF-SVM
   For SVMs, the kernel and its parameters control on the complexity of the model. In general, the RBF kernel is a good choice because it can deal with the nonlinear relations between the features and class labels. In addition, the RBF kernel has fewer hyperparameters and fewer numerical difficulties compared with the polynomial kernel [99]. In this study, Gaussian RBF-SVM classification model with ($\sigma = 0.1$) has been used.

3. Simplified RNN
   In an ordinary feed-forward neural network, each layer transfers the parameters to the next layer in one direction (forward direction). In other words, the data passes through input nodes to feed the next layer until eventually it reaches the output nodes [73]. The ANN is known as a universal function approximator because it has the ability to learn weights that map any input to the output and gradually increases the number of layers that are added to the structure [71]. On the other hand, having a limited number of layers remains a critical issue in any ANN design to achieve the desired balance between reducing complexity while improving accuracy. Hence, in some cases, the increasing in the dimension of layers in an ANN harms the ability of the universal learning function [100]. Basically deep learning tends to increase the number of layers with a simple learning function. But deep learning is not easy to implement and these networks represents a black box for users. Researchers decide to add layers when the output of model does not converge to predicted output. The proposed

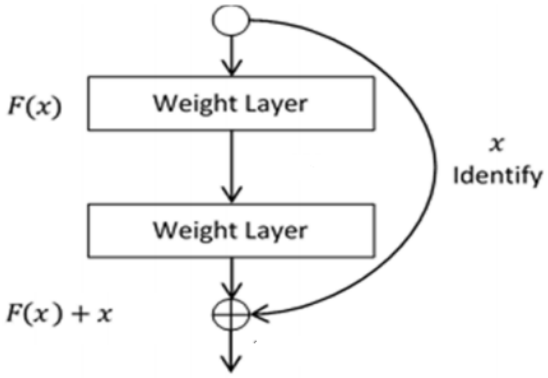

Figure 3.11 A single residual building block which proposed by He et al. [103].

simplified RNN tries to skip a connection or add a shortcut that enables data to flow effectively from one layer to the layer that comes after the next avoiding the full connection and the complex learning functions. Therefore, adding new layers does not decrease the performance of the model but it may increase it slightly due to the residual connection [101, 102].
By adding skip connections to the proposed network, rather than managing the layers number and the significant parameters to tune, the network will be able to skip training for not useful layers and do not add their value to overall accuracy . The skip connections made the proposed network dynamic and it may optimally tune the number of layers during training. In the study of He et al. [103], the residual learning framework was presented to facilitate the network training process. The building block of two weight layers can be shown in Fig 3.11. The difference between the input and the output can be expressed by Eq. 3.24.

$$F(x) = O/P - I/P = H(x) - x \tag{3.24}$$

where the O/P is the new set of weights to be the next layer input, and the I/P is the old set of weights of the previous layer. Hence, the layers in the ordinary ANN tends to learn the output H(x) only by tuning the weights, while the network with residual blocks tends to learn the true output F(x).

The proposed simplified RNN focused on the behavior of the identity shortcut connections in the He et al study [103]. The formulation of one hidden layer network with shortcut connection is given in Fig.3.12 and the structure of the proposed network is given in Fig.3.13. As a classifier, the simplified RNN was implemented by using the gradient descent Back Propagation as a learning algorithm, that has one input layer, three hidden layers, and one output layer. The number of input nodes is dependent on the selected approach (PCA, SVD,

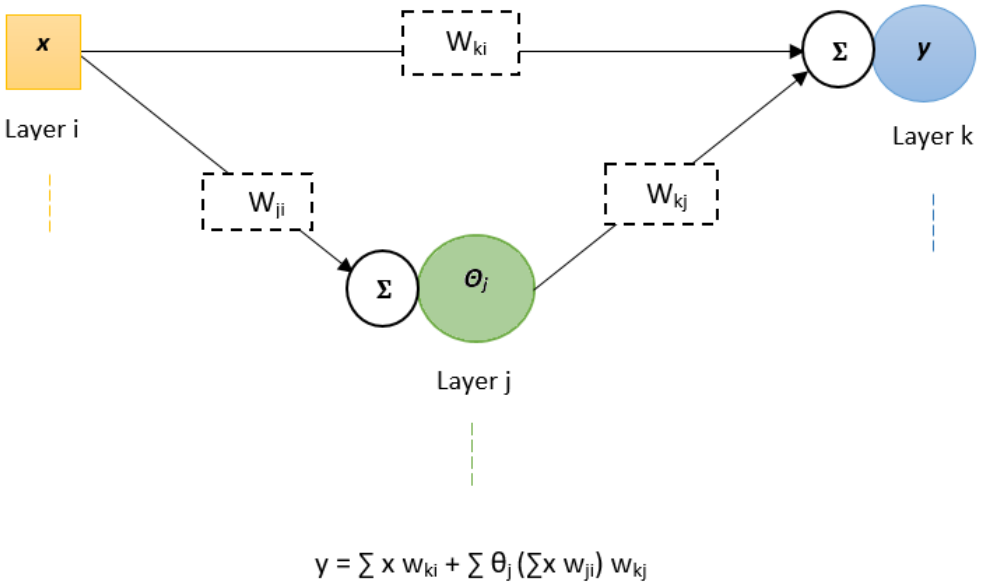

Figure 3.12 One hidden layer network with shortcut connection.

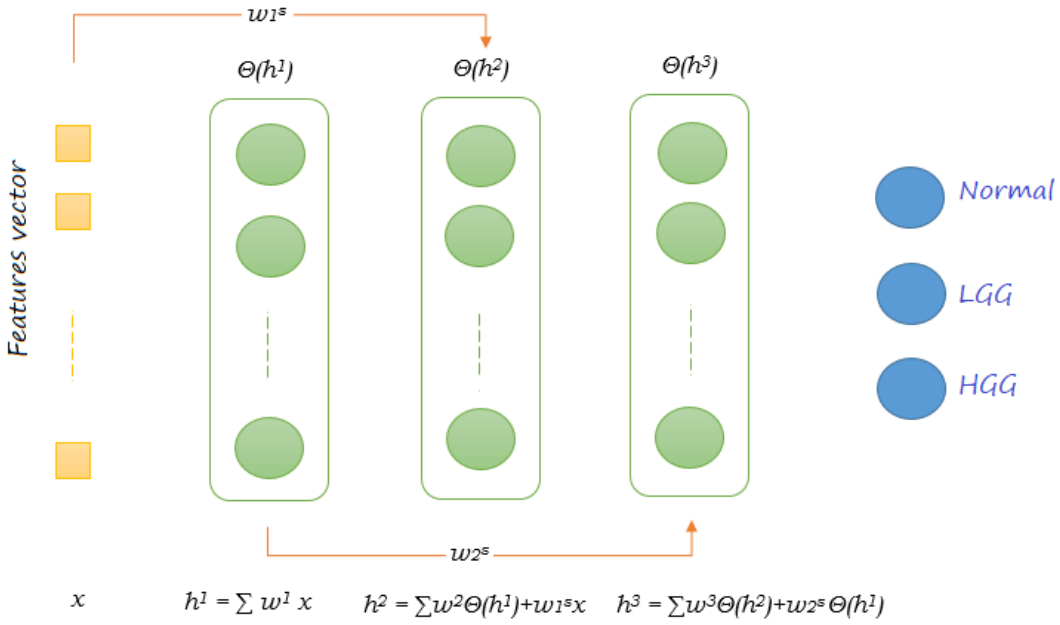

Figure 3.13 The proposed simplified RNN structure.

or accelerated SVD). It depends on the total number of measurements k. The number of neurons was 22, 13, 8 in the three hidden layers, respectively. While the total number of output nodes was the maximum number of class labels, which is three. In the all three hidden layers, every neuron has a $Sigmoid$ as an activation function while $SoftMax$ is the main activation function in the output layer as shown in Table 3.3. In the multi-class classification problems, $SoftMax$ function transforms a vector of numbers into a vector of probabilities. Each probability value is in the range [0-1], and the sum of the probabilities is 1. It can be defined as in the following Eq.:

$$f(S)_i = \frac{e^{S_i}}{\sum_{j=1}^{C} e^{S_j}} \tag{3.25}$$

where $f(S)_i$ is the probability score (predicted) for each class. S is the input vector, and C is the number of classes. In the proposed classifier, the

Table 3.3 The activation functions used in the proposed classifier

Layer No.	Layer Type	Activation Function	Mathmatical Expression
2	Hidden layer 1	Sigmoid	$f(x) = \frac{1}{1-e^{-x}}$
3	Hidden layer 2	Sigmoid	$f(x) = \frac{1}{1-e^{-x}}$
4	Hidden layer 3	Sigmod	$f(x) = \frac{1}{1-e^{-x}}$
5	Output layer Optimization	SoftMax Cross entropy	$f(x_i) = \frac{e^{x_i}}{\sum e^{x_i}}$ $Loss = -log\left(\frac{e_p^S}{\sum_{j=1}^{C} e_j^S}\right)$

cross-entropy loss was the main optimization function. It shows the distance between what the model considers the distribution of output must be, and what the original distribution actually is. In neural network, it is often used alternative of squared error when the output is a probability distribution, i.e. when the activation function is SoftMax in the output layer [104]. It can be expressed as:

$$Loss = -\sum_{i=1}^{C} l_i \; log\,(f(S)_i) \tag{3.26}$$

where l_i is the label of each class.

For multi-class classification labelling, one-hot coding was used which makes it possible to convey categorical data more expressively. In this case, only positive classes (l_p) remain in the loss function (main term) that allows for some extra optimisation. One element in the loss function remains as the target vector (label), is as follows:

$$l_i = l_p \tag{3.27}$$

Based on the target labels, the elements for which the summation is zero are discarded. According to this assumption, the optimization loss function can be written as follows:

$$Loss = -log\left(\frac{e_p^S}{\sum_{j=1}^{C} e_j^S}\right) \tag{3.28}$$

The proposed network was trained by using the following parameters:

- The initial learning rate parameter is 0.0001.
- The momentum factor is used to adjust the step size for the global minimum coverage by setting it to 0.9.
- The learning patch size is 16.
- The epoch size is 20.
- The iteration number for each epoch is 500.
- The data set was augmented by pre-processing step using a Gaussian filter.

3.4 Evaluating Parameters

A confusion matrix can show the performance of a classification process by knowing how many positive or negative cases are predicted truly or falsely [105, 106]. Table 3.4 shows five parameters used to evaluate the proposed system. For better clarification, Fig. 3.14 shows how the five parameters are calculated for class 1 (the normal class). Where TP is a true positive and it refers to the detection of positive events correctly, FP is a false positive and it refers the detection of positive events incorrectly, TN is a true negative and it refers the detection of negative events correctly, and FN is the false negative and it refers the detection of negative cases incorrectly. The two other classes follow the same procedure seen in class 1.

Ultimately, three values of three classes were obtained for each parameter. Later, the overall averages were taken. In addition to the evaluating parameters seen above, the computational time of the proposed method was also calculated because time represents one of the most important factors for model evaluating. Moreover, a receiver operating characteristics (ROC) graph is a technique for visualizing,

Table 3.4 The evaluating parameters.

Parameters	Mathematical Description
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Sensitivity	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$
Precision	$\frac{TP}{TP+FP}$
Error rate	$\frac{FP+FN}{TP+TN+FP+FN}$

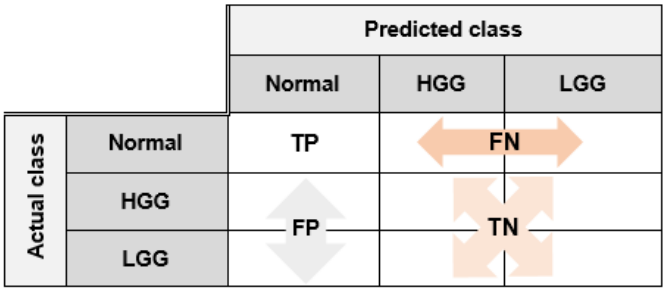

Figure 3.14 The confusion matrix for 3 classes system.

organizing and evaluating the classifiers. The terms associated with ROC curves are True Positive Rate (sensitivity) and False Positive Rate (1-specificity) [107, 108]. It is important to note, there are several important points in ROC space. The lower left point (0,0) indicates that the classifier does not record any false positive errors and at the same time it does not gain any true positives. In contrary, the upper right point (1,1) indicates that the classifier works randomly where as the point (0,1) indicates the perfect classification as shown in Fig.3.15.

Figure 3.15 The ROC space.

4
ANALYSIS AND INTERPRETATION OF THE EXPERIMENTAL RESULTS

4.1 Highlights of the Proposed System

Automatically extracting the brain tumors from magnetic resonance images and classifying the tumor grades using medical decision making system is still a challenge. This study offered a medical decision making system which has six stages as it is mentioned in previous chapter, section 3.3 in details.
Once again, these stages will be listed here in terms of number of experiments that were carried out to test the performance and the efficiency of the proposed system. The stages are:

1. Applying Pre-processing steps. One experiment was performed in this stage.
2. Applying a new clustering method LDI-Mean to obtain a segmented tumor area. The results of using LDI-Means were compared with two common methods (K-Means and watershed). There were two experiments to test the effectiveness of LDI-Means.
3. Applying some mathematical calculations to find the position of brain tumor in terms of x and y. The results of this stage were obtained based on the previous stage results. One experiment was performed.
4. Applying some mathematical expressions for feature extraction stage and no results can be shown in this stage.
5. Applying the novel technique MI+SVD. This method exploited the mutual information theory in two ways:
    - MI based on Probability
    - MI based on Distance metric.

Later, SVD in its accelerated mode was used. The objective of using this method is to decrease the dimensions of extracted feature space and ultimately improve the classification process in the next step. Hence, there were no experiments may be conducted to represent this stage but its robustness was tested in combination with the classification stage.
Logically, using a similar and common algorithms such as PCA and SVD will create a suitable environment for comparison. Therefore, PCA and SVD were performed for the mentioned purpose.

6. Applying a simplified mode of RNN which is a modern approach. In order to examine the performance of the classification with and without MI+SVD, there were three experiments. These are:

   - (a) (P-based MI + SVD) + Simplified RNN
   - (b) (D-based MI + SVD) + Simplified RNN
   - (c) All features space + Simplified RNN

   In addition, there were four experiments to test the quality of using MI+SVD in comparison with PCA and SVD:

   - (d) PCA + Simplified RNN , using k=13 and k=26.
   - (e) SVD + Simplified RNN , using k=13 and k=26.

   Furthermore, two other classifiers, MLP and RBF-SVM, were used to find the best classification performance.There were 14 experiments to cover all the scenarios.

   - (f) (P-based MI + SVD) + MLP
   - (g) (D-based MI + SVD) + MLP
   - (h) All features space + MLP
   - (i) PCA + MLP , using k=13 and k=26.
   - (j) SVD + MLP , using k=13 and k=26.
   - (k) (P-based MI + SVD) + RBF-SVM
   - (l) (D-based MI + SVD) + RBF-SVM
   - (m) All features space + RBF-SVM
   - (n) PCA + RBF-SVM , using k=13 and k=26.
   - (o) SVD + RBF-SVM , using k=13 and k=26.

Totally, 25 experiments were conducted by this study. The obtained results by using the proposed system, from all the experiments mentioned above, were evaluated on the basis of three criteria. These are:

i. A comparison with the entire feature space using neither selection feature nor dimension reduction methods; stage six exp.(c), (h), and (m).

ii. A comparison with the SVD and PCA of two different values of k; stage 6 exp.(d), (e), (i), (j), (n), and (o).

iii. A comparison with some researches in the same field of study, as found in Table 4.10.

4.2 Result of Pre-processing Stage

The result obtained by the proposed pre-processing is shown in Fig.4.1. This result is of two images selected randomly from the data set as an example. The intensity adjustment can be shown in Fig.4.2.

4.3 Result of Segmentation Stage

The result of the segmentation process by using LDI-Means algorithm is shown in Fig. 4.3, this result is of two images selected randomly from the dataset as an example. The result of the segmentation process by using K-means clustering algorithm is shown in Fig.4.4, this result is of same two images. It is necessary to mention that with the help of one of the open annotation tools; named Labelme, 60 images from the data set were hand-labeled by experts, and those images were stored in separate file. In order to find the usefulness of the proposed technique, the obtained hand-labeled images are considered for comparison with the segmented tumor images as ground truth annotation.

From Fig.4.4, the brain tumor was clustered into Three clusters by using LDI-Means and the last cluster represents the tumor binary image in both images. For the same images, Fig.4.4 shows five clusters by using K-means. Definitely, the most important cluster is the tumor cluster. The tumor cluster by using K-means appeared in the second cluster for the first image whereas it appeared in the fifth cluster for the second image.

By using LDI-Means, the tumor image can be easily seen in the third cluster which is the last one. While the tumor image, by using K-means can be seen in any cluster. Here, in the case of using K-means, it is important for users to manually select the brain image cluster to be separately stored which will take a lot of time and effort whereas in the case of LDI-Means, the segmented tumor was obtained in the last cluster for all images in the data set. Therefore, specifying the brain tumor image was performed automatically with no need to make selection by a user.

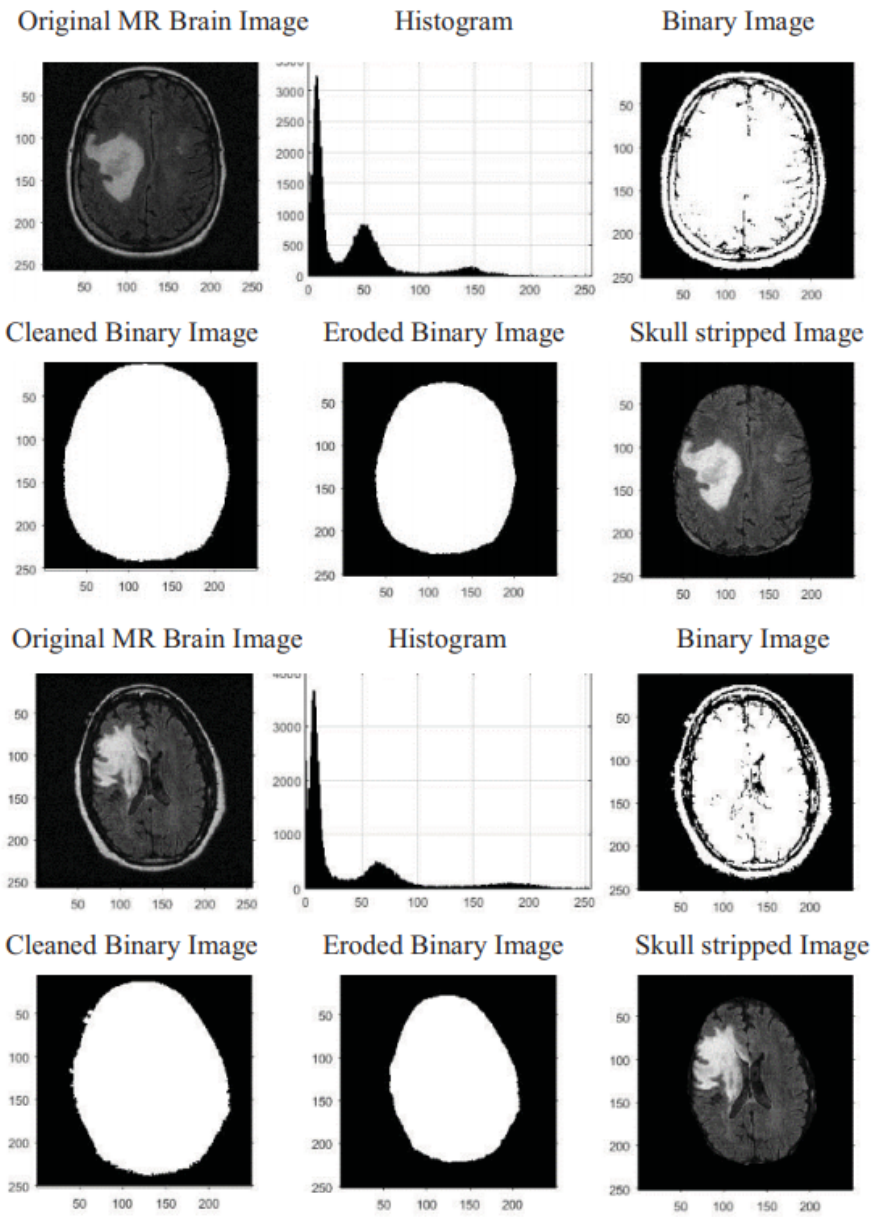

Figure 4.1 The steps of pre-processing.

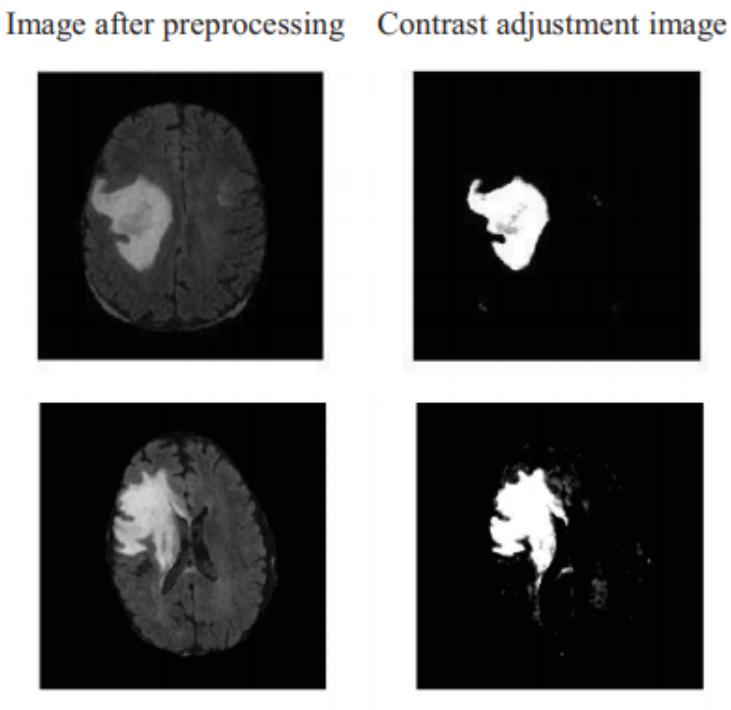

Figure 4.2 Image contrast adjustment.

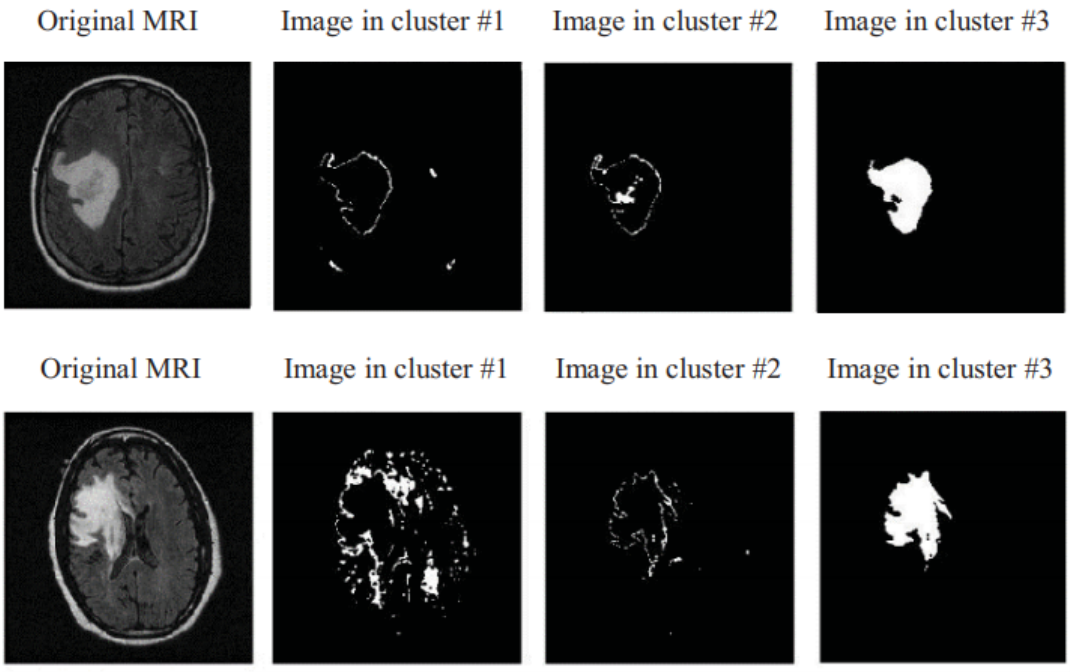

Figure 4.3 Clusters as a result of using LDI-Means algorithm.

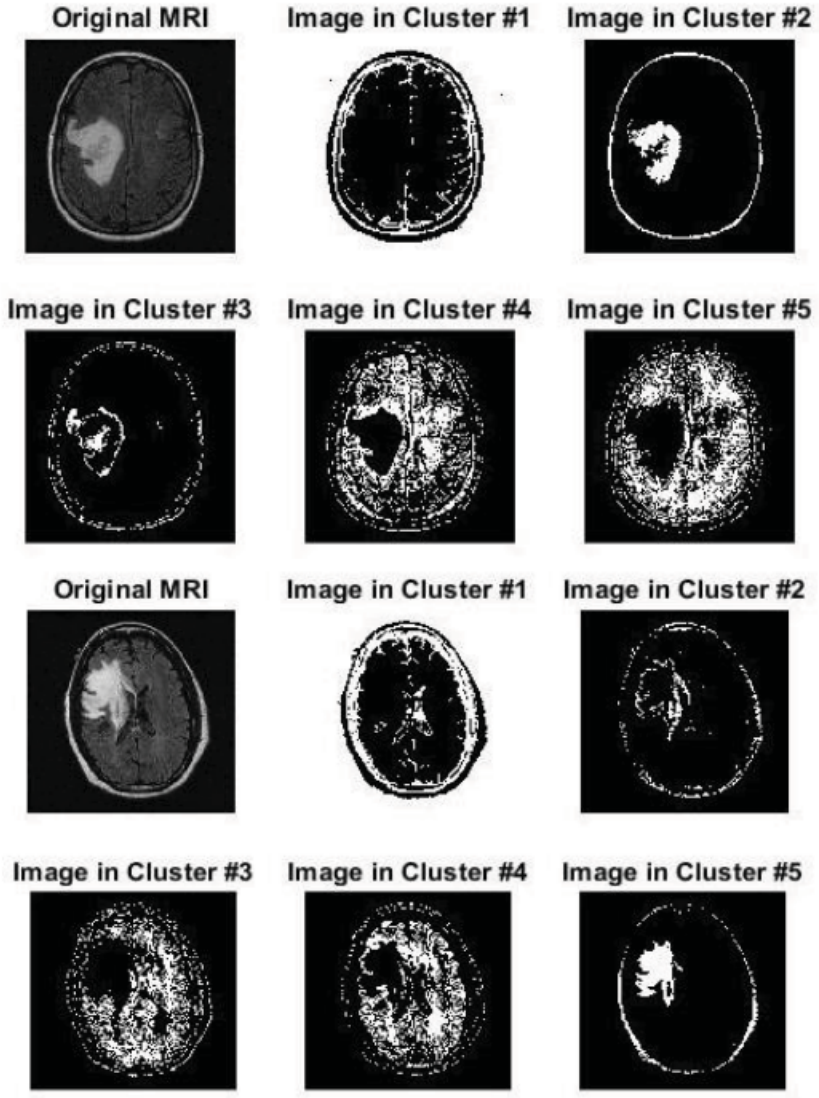

Figure 4.4 Clusters as a result of using standard K-means algorithm.

Furthermore, K-means did not offer a sufficient segmentation result when the number of clusters was set to be less than five clusters. In contrast LDI-Means worked properly even if the number of clusters was less than five. Number of cluster is a significant factor in clustering process. It is directly proportional to computational time, e.g. The processing time increases as the number of cluster increases [53].
The comparative analysis of the effectiveness of both the LDI-Means and the standard K-Means algorithms is shown in Fig.4.5 in terms of similarity with the ground truth, Fig.4.6 in terms of accuracy, specificity and sensitivity, and Fig.4.7 in terms of processing time. Fig.4.5 shows the tumor image which was obtained by using LDI-Means tends to be more similar to its ground truth (hand-labeled image) than the tumor image resulting from K-means algorithm. It must be noted that all steps of the pre-processing stage were applied to images before clustering process, either by LDI-Means or by K-means alike, for the comparison to be fair.
In addition to Fig.4.6 which shows the average value of accuracy, specificity and sensitivity for all the images in data set graphically, there is Table 4.1 which shows the values numerically that were calculated by using the following equations:

$$Average\ Accuracy = \frac{\sum_{n=1}^{N} Accuracy(n)}{N} \tag{4.1}$$

$$Average\ Specificity = \frac{\sum_{n=1}^{N} Specificity(n)}{N} \tag{4.2}$$

$$Average\ Sensitivity = \frac{\sum_{n=1}^{N} Sensitivity(n)}{N} \tag{4.3}$$

where $n = 1, 2, ..., N$ (N is the number of images in the data set to be clustered) and $Accuracy(n)$, $Specificity(n)$, and $Sensitivity(n)$ are the $accuracy$, $specificity$ and $sensitivity$ for image n, respectively. The mathematical equations of accuracy, specificity and sensitivity can be found in section 3.4. But here TP, FP, TN and FN are a little bit different according to the segmentation process. These are their meaning:

- TP (True Positive) indicates the pixel of tumor in the ground truth occurs in the obtained image by clustering as a tumor pixel.
- FP (False Positive) indicates the pixel of non-tumor in the ground truth occurs in the obtained image by clustering as tumor pixel.
- TN (True Negative) indicates the pixel of non-tumor in the ground truth occurs in the obtained image by clustering as non-tumor.

- FN (False Negative) indicates the pixel of tumor in the ground truth occurs in the obtained image by clustering as non-tumor.

Generally, using a similarity measure between the image of segmented tumor and the hand labeled ground truth is widely used in many published studies such as the Dice coefficient (DC) which can be calculated by as follow:

$$DC = \frac{2(X \cap Y)}{X + Y} \tag{4.4}$$

where, X and Y are the obtained tumor image by using the segmentation and the corresponding ground truth image, respectively. The value of DC has to be 1 for a perfect segmentation [34]. It was found that the DC value for all the segmented images by LDI-Means was 0.96 as shown in Table 4.1.

Table 4.1 The performance of clustering process by using K-means and LDI-Means.

Parameters	**K-means**	**LDI-Means**
Accuracy	91.65 %	99.02 %
Specificity	94.71 %	99.39 %
Sensitivity	65.44 %	82.85 %
DC	≥ 0.87	≥ 0.96

By using LDI-Means, It was found that the required segmentation can be obtained in less time than using standard K-means. The computational time can be defined as the processing time that required to obtain the tumor images from all the input data set. It was computed by seconds. Time used for comparison was calculated by using average as below:

$$T = \frac{\sum t(n)}{N} \tag{4.5}$$

where, T is the average time, $t(n)$ is the required time to complete the clustering process of image (n), N is the number of images in data set to be clustered, and $n = 1, 2, ..., N$.

From Fig.4.7, the average time to complete clustering by LDI-Means for one image is about 1.5 seconds whereas by using K-means, it is about 18.9 seconds.

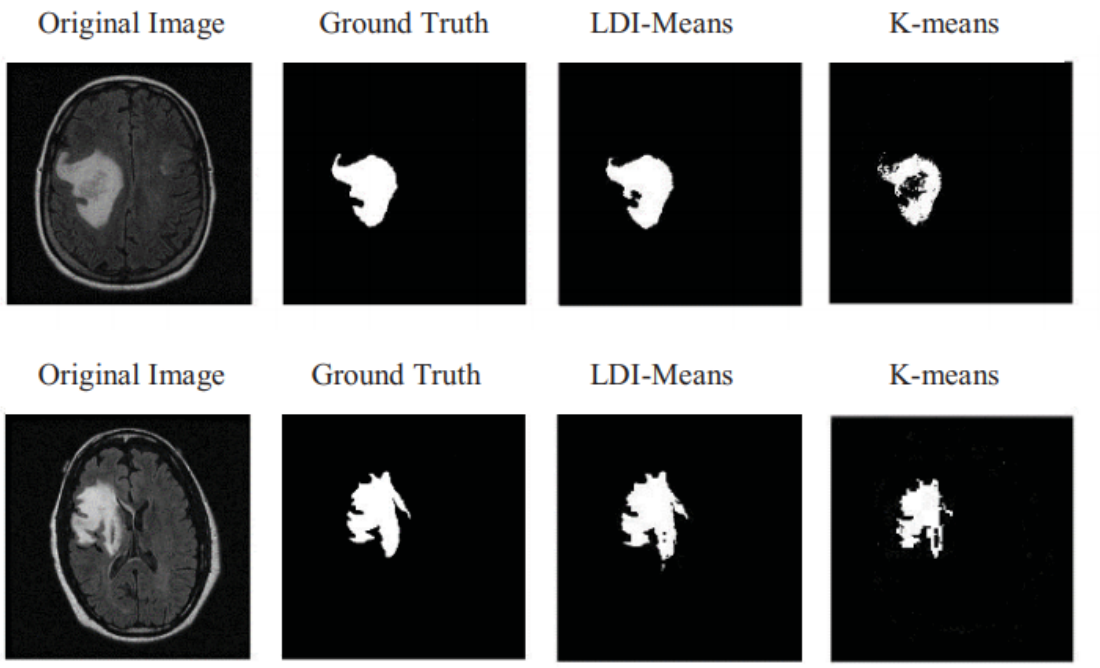

Figure 4.5 The efficiency of tumor segmentation by using LDI-Means and ordinary K-Means algorithm for only two images of the data set as an example.

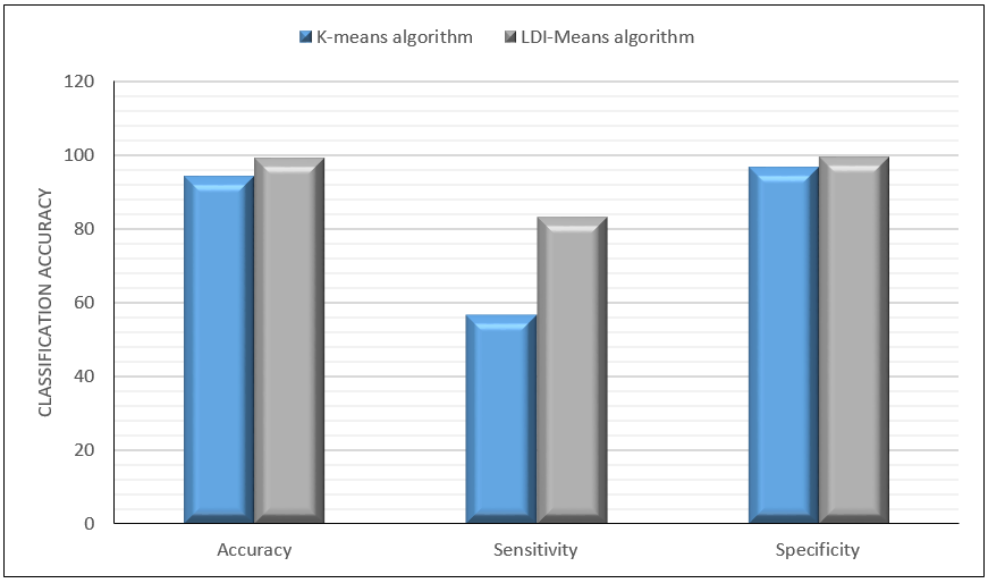

Figure 4.6 The clustering performance by using LDI-Means and ordinary K-Means algorithm.

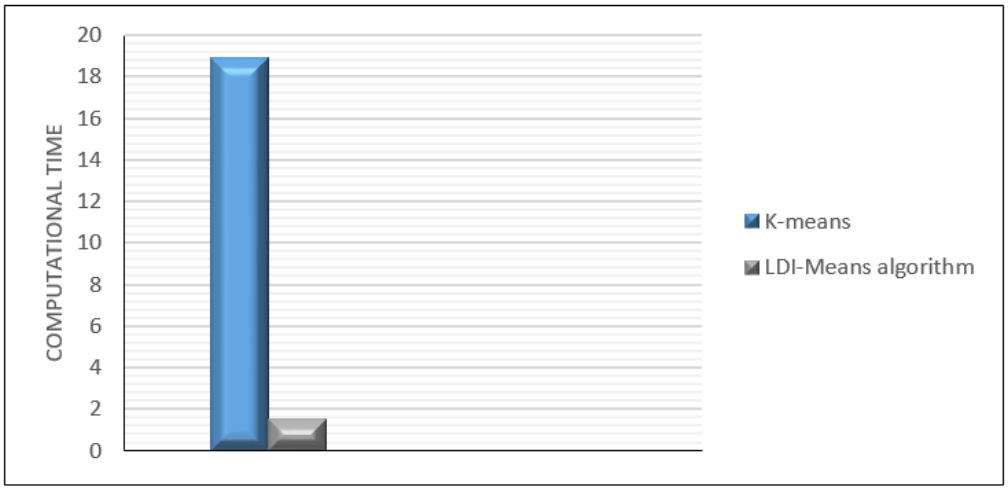

Figure 4.7 The computational time of LDI-Means and the ordinary K-Means algorithm.

In conclusion, K-means algorithm selects K objects randomly from population and sets them as the initial centers and because of this randomization, it mostly will not be able to give a stable and true clustering. On the contrary the LDI-Means algorithm proves its ability to produce very stable and precise clusters. LDI-Means offers an efficient way for assigning pixels to the number of specific clusters in very short time. It can give a better accuracy and shorter computational time than K-means algorithm.

4.4 Result of Tumor Localization Stage

The result of tumor localization is shown in Fig.4.8, this result is for only two images in data set as an example. Also, this stage contained some calculations to find the ratio of tumor size in respect to whole brain and the metric size of the tumor as shown in Fig.4.9 and Fig. 4.10.

Depending on some functions such as edge detection and shape factor analysis in MATLAB, a boundary box is drawn around the tumor in the original image and then to be cropped as shown in Fig.4.9, the tumor image. The cropped tumor images were stored in separated file to be used in the feature extraction stage. From the Fig.4.9, the tumor to all brain tissues ratio can be obtained. This ratio could give indication about how far the tumor has spread within the brain tissues.

It is very appropriate to compare the outcome of the proposed method with the outcome of standard K-means and watershed since they are frequently used for segmentation, for testing the effectiveness of LDI-Means. Fig.4.10 presents the value of tumor metric size by using LDI-Means, K-means, and watershed segmentation

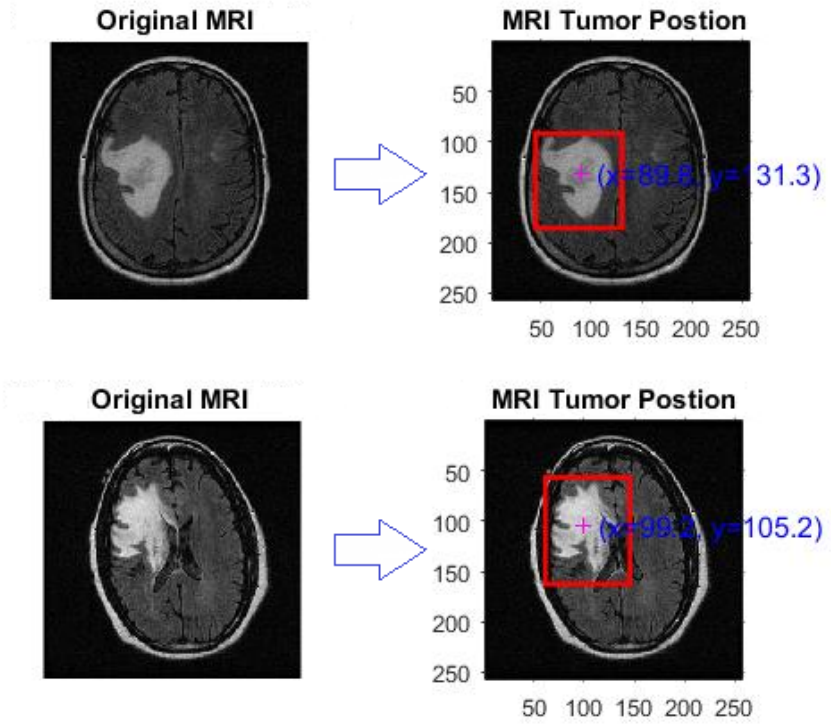

Figure 4.8 The brain tumor position (x, y) for two images in data set.

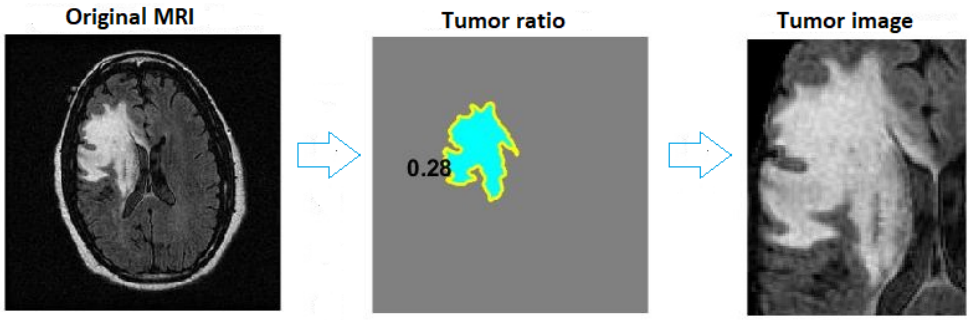

Figure 4.9 The tumor to brain ratio and the tumor image for one image in data set.

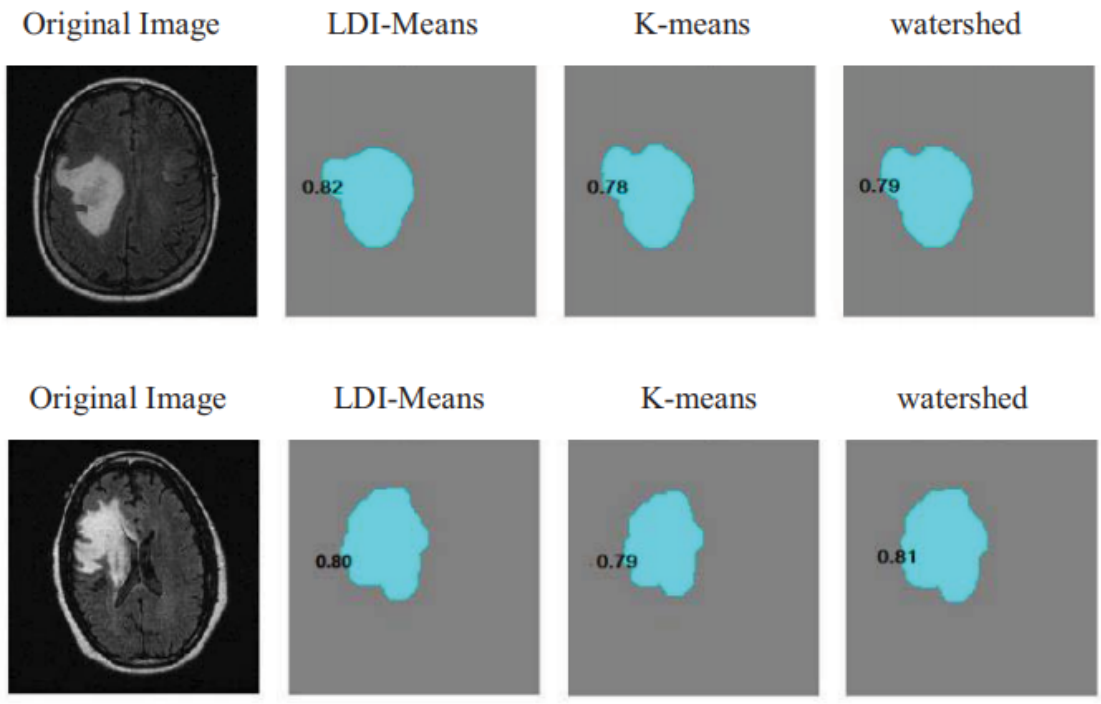

Figure 4.10 Tumor metric size as a result of using LDI-Means, standard K-Means and watershed algorithms.

algorithms. From Fig.4.10, it was found that the value of tumor metric size using the LDI-Means was closer to the value of tumor metric size calculated from the ground truth, than the values of the other two algorithms.

4.5 Result of using MI+SVD

This study proposes using MI+SVD to improve the performance of the classifier by finding the most meaningful features. Here, the MI is used to rerank the extracted features. Later the new sorted features will be the input matrix of SVD and according to the values of MI a series of calculations will take their way and end by allowing SVD to select the robust K (number of the required dimension). As mention before, the mutual information was performed based on two different values; probability [93] and distance metric [96].

1. Using P-based MI+SVD
   The result of new ranking features according to P-based MI+SVD is shown in Table 4.2.

2. Using D-based MI+SVD

Table 4.2 New sorting of features according to P-based MI + SVD

Feature No.	Original Feature	Feature Name	Ranking Score
1	11	Mean of the entropy	0.519892
2	16	Standard deviation of the contrast	0.507263
3	40	Standard deviation of the inverse difference normalized	0.506322
4	14	Standard deviation of the correlation	0.503251
5	29	Mean of the maximum probability	0.492930
6	19	Mean of the homogeneity	0.481151
7	36	Standard deviation of the sum variance	0.474483
8	21	Mean of the dissimilarity	0.430674
9	39	Mean of the inverse difference normalized	0.424282
10	34	Standard deviation of the sum entropy	0.420824
11	20	Standard deviation of the homogeneity	0.417607

The result of new ranking features according to D-based MI+SVD is shown in Table 4.3.

Depending on Table 4.2, P-based MI+SVD algorithm was able to determined 11 features and considered them as the best features that could represent the inputs more accurately than other features. While Table 4.3 shows 13 features was selected by using D-based MI+SVD and considered them as the best features that could represent input more accurately than other features.
In Tables 4.2 and 4.3, it was found some differences in result due to using two values of MI; probability and distance metric. These differences can be interpreted as the superiority of the second algorithm (D-based MI+SVD) over the first (P-based MI+SVD). Fundamentally, this superiority would be because the first algorithm, which is the standard measure of MI, only takes the intensity values of population into its account and that leads to lack of concern on any spatial information which might be occurred in the individual images. The second algorithm pays more attention to the spatial information as it originally depends on the distance.

4.6 Result of Classification

This study presents a novel algorithm to improving the classifier performance and proposes using an new network structure; named simplified RNN. Hence, to achieve a fair comparative analysis, this study used two traditional classifiers (MLP and SVM) in addition to the proposed classifier.

Table 4.3 New sorting of features according to D-based MI + SVD

Feature No.	Original Feature	Feature Name	Ranking Score
1	14	Standard deviation of the correlation	0.155009
2	16	Standard deviation of the contrast	0.155009
3	22	Standard deviation of the dissimilarity	0.152422
4	11	Mean of the entropy	0.150504
5	40	Standard deviation of the inverse difference normalized	0.146310
6	29	Mean of the homogeneity	0.144260
7	38	Mean of the difference entropy	0.143850
8	33	Mean of the sum entropy	0.140517
9	39	Mean of the inverse difference normalized	0.139570
10	36	Standard deviation of the sum variance	0.137944
11	20	Standard deviation of the homogeneity	0.134419
12	21	Mean of the dissimilarity	0.132471
13	34	Standard deviation of the sum entropy	0.130517

1. Using MLP.
   The classification accuracy in both training and testing phases is shown in Table 4.4 and the other evaluating parameters values can be seen in Table 4.5.

2. Using SVM.
   The classification accuracy in both training and testing phases is shown in Table 4.6. The other evaluating parameters values can be seen in Table 4.7.

3. Using simplified RNN.
   The classification accuracy in both training and testing phases is shown in Table 4.8. The other evaluating parameters values can be seen in Table 4.9.

In Table 4.4, The first two rows show the classification accuracy using both versions of MI+SVD; P-based MI+SVD and D-based MI+SVD, with MLP which achieved 88.23% and 89.71% on the training set, and 90.03% and 90.80% on the testing set respectively. While the third row shows the classification accuracy of MLP without using any dimension reduction methods which was 58.72% and 61.60% in both phases respectively. Based on the values in the first three rows in the table, it was found that MI+SVD succeed to improve the classification accuracy of MLP. The last four rows show the accuracy of MLP combined with PCA and SVD in two different values of k for each which was selected manually. According to the 13 features obtained by using MI+SVD, it was decided to select K=13 to be the required reduction for PCA and SVD. PCA achieved an accuracy of 78.99% on the training

Table 4.4 The Classification accuracy by using MLP in both training and testing phases.

Approach	Training (%)	Testing (%)	No. of Features
(P-based MI + SVD) + MLP	88.23	90.03	11
(D-based MI + SVD) + MLP	89.71	90.80	13
All features + MLP	58.72	61.60	40
PCA + MLP	78.99 69.07	79.95 70.82	13 26
SVD + MLP	76.82 67.04	78.20 67.84	13 26

set and 79.95% on the testing set, whereas SVD achieved an accuracy of 76.82% on the training set and 78.20% on the testing set. Moreover, in terms of measuring the performance of PCA and SVD, different numbers of features were used. It was selected K=26; double of 13. For PCA, when using 26 features, the accuracy values were 69.07% and 70.82% on training and testing respectively; whereas, they were 67.04% and 67.84% on training and testing respectively, for SVD. By using PCA and SVD, the classification accuracy has increased in comparison with the value of using the classifier without any dimension reduction method. But The increment in accuracy, when using MI+SVD, was much higher than the increment that done due to using PCA or SVD and that confirmed the efficiency of MI+SVD. It was also found that D-based MI+SVD was better than P-MI+SVD in improving MLP.

Furthermore in Table 4.6, the first two rows show the classification accuracy using both versions of MI+SVD; P-based MI+SVD and D-based MI+SVD, with RBF-SVM which achieved 89.54% and 91.04% on the training set, and 91.02% and 92.21% on the testing set, respectively. While the third row shows the classification accuracy of MLP without using any dimension reduction methods which was 72.87% and 75.66% in both phases, respectively. Once again, based on the values in the first three rows in the table, it was found that MI+SVD succeed to improve the classification accuracy of

Table 4.5 Classification performance of MLP in both training and testing phases.

Approach	Criteria	Training (%)	Testing (%)
(P-based MI + SVD) + MLP	Sensitivity Specificity Precision Error rate	82.35 90.19 82.60 0.110	85.03 93.51 85.56 0.099
(D-based MI + SVD) + MLP	Sensitivity Specificity Precision Error rate	85.13 94.34 86.03 0.110	85.92 95.01 88.32 0.098
All features + MLP	Sensitivity Specificity Precision Error rate	73.22 69.44 69.91 0.134	74.62 71.10 70.05 0.129
PCA + MLP	Sensitivity Specificity Precision Error rate	81.02 80.16 79.96 0.122	81.71 80.92 80.23 0.120
SVD + MLP	Sensitivity Specificity Precision Error rate	80.34 80.00 78.10 0.122	80.82 80.78 79.23 0.121

RBF-SVM. The last four rows show the accuracy of RBF-SVM combined with PCA and SVD in two different values of k for each which was selected manually. According to the 13 features obtained by using MI+SVD, it was decided to select K=13 to be the required reduction for PCA and SVD. PCA achieved an accuracy of 80.78% on the training set and 81.00% on the testing set, whereas SVD achieved an accuracy of 77.24% on the training set and 80.10% on the testing set. Moreover, in terms of measuring the performance of PCA and SVD, different numbers of features were used. It was selected K=26; double of 13. For PCA, when using 26 features, the accuracy values were 73.23% and 74.23% on training and testing, respectively. Whereas, they

Table 4.6 The Classification accuracy by using SVM in both training and testing phases.

Approach	Training (%)	Testing (%)	No. of Features
(P-based MI + SVD) + SVM	89.54	91.02	11
(D-based MI + SVD) + SVM	91.40	92.21	13
All features + SVM	72.87	75.66	40
PCA + SVM	80.78 73.23	81.00 74.23	13 26
SVD + SVM	77.24 70.67	80.10 71.97	13 26

were 70.67% and 71.97% on training and testing, respectively, for SVD. By using PCA and SVD, the classification accuracy has increased in comparison with the value of using the classifier without any dimension reduction method. But the increment in accuracy, when using MI+SVD, was much higher than the increment that done due to using PCA or SVD and that confirmed the efficiency of MI+SVD. It was also found that D-based MI+SVD was better than P-MI+SVD in improving RBF-SVM.

From Table 4.8, it was found, for the third time, the effectiveness of MI+SVD in improving the accuracy of the third classifier. By making a comparison between the first two rows, which represent using MI+SVD, with the third row of the same table, which represents using simplified RNN with all the extracted features. Where MI+SVD achieved 91.03% and 92.69% on the training set, and 93.10% and 94.91% on the testing set, respectively. Whereas the classification accuracy of simplified RNN without using any dimension reduction methods was 77.58% and 79.20% in both phases, respectively. The last four rows show the accuracy of simplified RNN combined with PCA and SVD in two different values of k for each which was selected manually. According to the 13 features obtained by using MI+SVD, it was decided to select K = 13 to be the required reduction for PCA and SVD. PCA achieved an accuracy of 86.21% on the training set and 88.02% on the testing set, whereas SVD achieved an accuracy

Table 4.7 Classification performance of SVM in both training and testing phases.

Approach	Criteria	Training (%)	Testing (%)
(P-based MI + SVD) + SVM	Sensitivity Specificity Precision Error rate	84.31 91.66 84.39 0.101	86.52 94.26 87.07 0.089
(D-based MI + SVD) + SVM	Sensitivity Specificity Precision Error rate	85.61 92.41 86.21 0.101	87.01 94.87 87.89 0.089
All features + SVM	Sensitivity Specificity Precision Error rate	74.31 73.42 69.01 0.121	75.25 75.30 69.92 0.116
PCA + SVM	Sensitivity Specificity Precision Error rate	84.02 83.16 82.33 0.113	84.63 83.64 84.74 0.111
SVD + SVM	Sensitivity Specificity Precision Error rate	82.86 83.21 82.34 0.114	81.14 83.92 83.87 0.111

of 85.89% on the training set and 87.71% on the testing set. Moreover, in terms of measuring the performance of PCA and SVD, different numbers of features were used. It was selected K=26; double of 13. For PCA, when using 26 features, the accuracy values were 75.22% and 76.17% on training and testing, respectively. Whereas, they were 72.90% and 73.85% on training and testing, respectively, for SVD. By using PCA and SVD, the classification accuracy has increased in comparison with the value of using the classifier without any dimension reduction method. But The increment in accuracy, when using MI+SVD, was much higher than the increment that done due to using PCA or SVD and that confirmed the efficiency of MI+SVD. It was also found

Table 4.8 The Classification accuracy by using simplified RNN in both training and testing phases.

Method	Training (%)	Testing (%)	No. of Features
(P-based MI + SVD) + Simplified RNN	91.0	93.1	11
(D-based MI + SVD) + Simplified RNN	**92.69**	**94.91**	**13**
All features + Simplified RNN	77.58	79.20	40
PCA + Simplified RNN	86.21 75.22	88.02 76.17	13 26
SVD + Simplified RNN	85.89 72.90	87.71 73.85	13 26

that D-based MI+SVD was better than P-based MI+SVD in improving simplified RNN. In addition, It should be mentioned that the classification accuracy of the third classifier were the highest among all the classifiers used in this study. Thus, using simplified RNN adds an extra success to the entire proposed system.
In terms of the other evaluating parameters; sensitivity, specificity, precision, and error rate, Tables 4.5, 4.7, and 4.9 show that the values of the three classifiers used in this study with MI+SVD were better than the values without MI+SVD. It was also found that the values of D-based MI+SVD with the simplified RNN were the best among all the mentioned schemes.
Besides that, in Fig.4.11 the D-based MI + SVD + simplified RNN approach (blue loss function line) achieved the lowest score during the training phase within 500 epochs compared with the other dimensionality reduction approaches PCA and SVD, in grey and orange, respectively, as well as in comparison with the all feature space, in purple.

In Fig.4.12 shows the error curve of training set versus the error curve of test set. It is clear from Fig.4.12 the perfect match between training and validation sets during the cross-validation of MI+SVD in comparison with PCA and SVD.

In order to assess the proposed system (D-based MI + SVD + Simplified RNN) ability

Table 4.9 Classification performance of simplified RNN in both training and testing phases.

Method	Criteria	Training (%)	Testing (%)
(P-based MI + SVD) + Simplified RNN	Sensitivity Specificity Precision Error rate	95.61 94.07 87.40 0.081	95.93 94.61 88.35 0.079
(D-based MI + SVD) + Simplified RNN	**Sensitivity** **Specificity** **Precision** **Error rate**	**96.69** **96.70** **89.84** **0.081**	**96.89** **96.37** **91.36** **0.078**
All features + Simplified RNN	Sensitivity Specificity Precision Error rate	76.62 77.12 71.79 0.121	77.70 78.17 72.75 0.117
PCA + Simplified RNN	Sensitivity Specificity Precision Error rate	86.45 86.46 83.56 0.109	86.53 86.54 85.33 0.103
SVD + Simplified RNN	Sensitivity Specificity Precision Error rate	86.44 86.34 83.25 0.109	86.42 86. 62 85.02 0.104

to distinguish among classes, a ROC curve for the three classes was plotted as shown in Fig. 4.13.

In summary, from all the obtained results, it is clear that the proposed system (D-based MI+SVD with the simplified RNN) has the highest classification accuracy compared to using the other reduction algorithms, SVD and PCA, for two values of dimensions. It is expected that for the standard SVD and PCA to have a low classification accuracy because they do not take into consideration the relations between the class labels

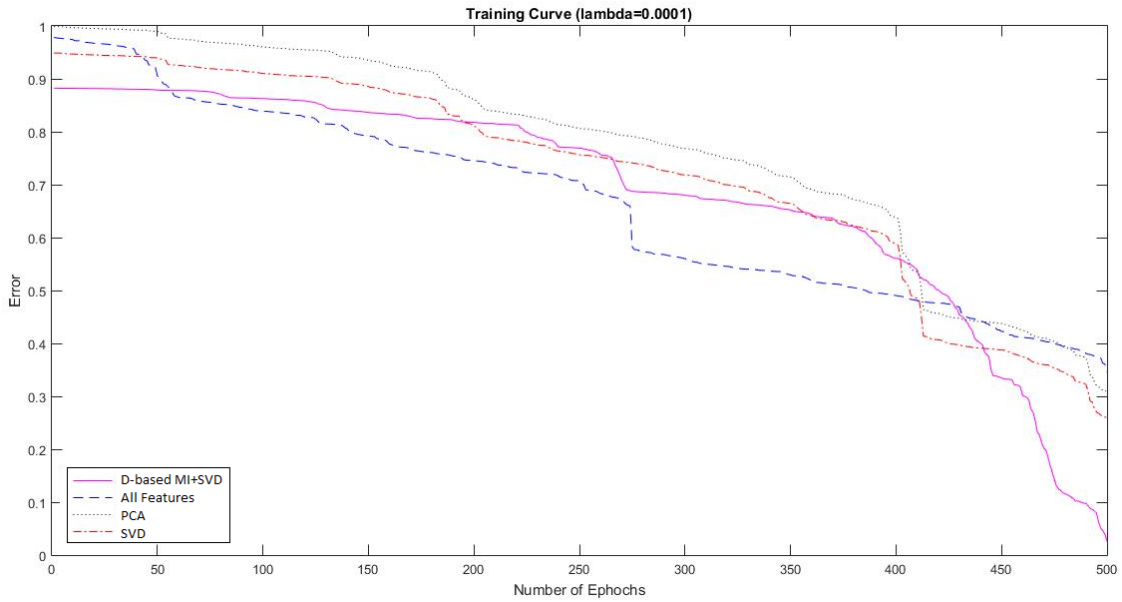

Figure 4.11 Loss function scores of the simplified RNN classification training phase by using all features, D-based MI+SVD, PCA and SVD.

and the features. Actually, this is what inspired this study to consider the MI+SVD algorithm, which includes two steps: ranking the extracted features based on MI and using accelerating SVD. Using MI+SVD reduced the feature space and improved the classifiers performance in relatively less time. Practically, it was found that the decreasing number of inputs of the classifier causes an increase in the speed of training phase compared to using all of the extracted features.

4.7 Comparison to Other State-of-the-Art Techniques

As can be seen from Table 4.10, the comparison with other published studies in same field of study. According to the comparison, it was found that the classification performance of the proposed system is suitable for providing a precise estimation of brain tumor grade.

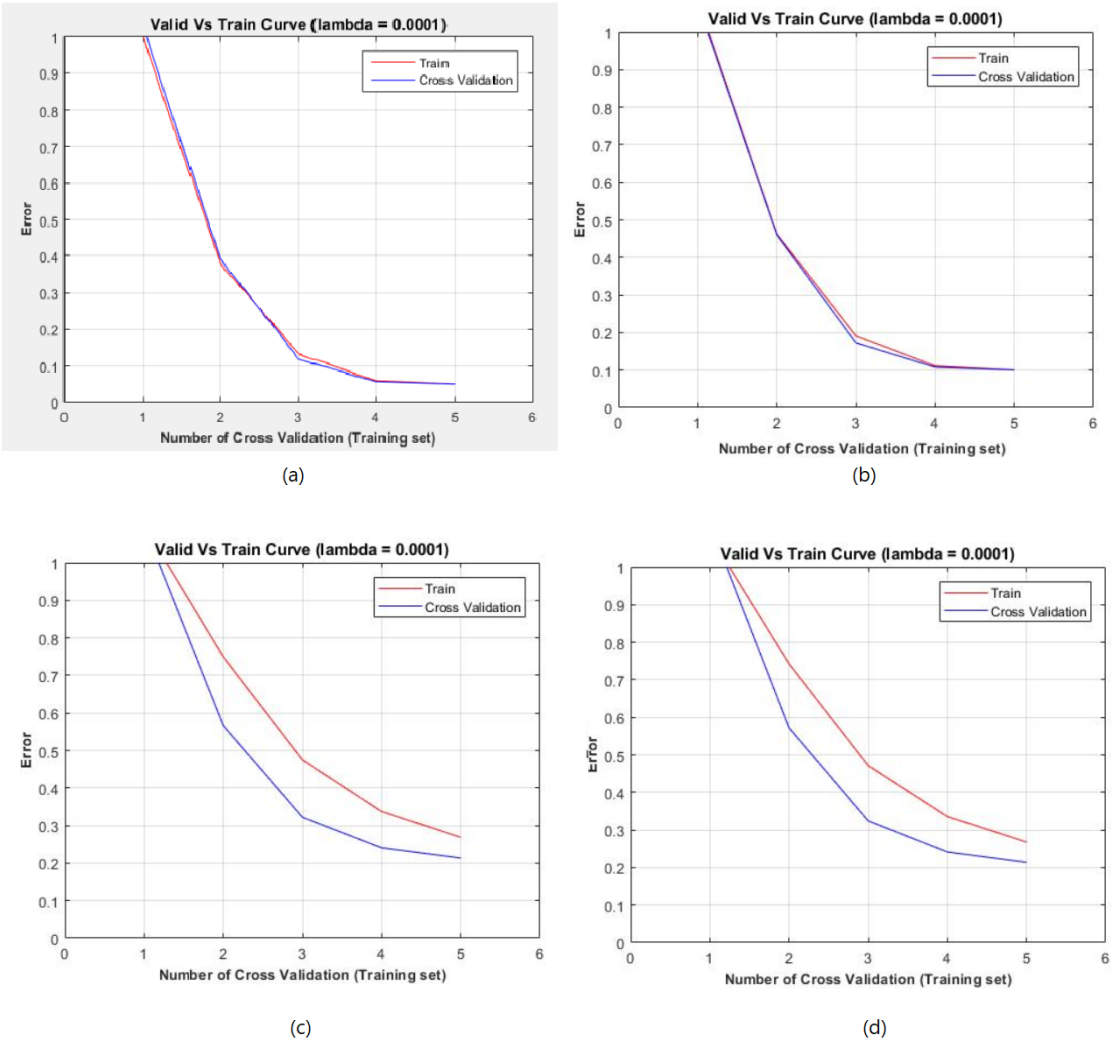

Figure 4.12 The error curve of training vs validation sets of (a)D-based MI+SVD, (b)P-based MI+SVD, (c)PCA, and (d)SVD.

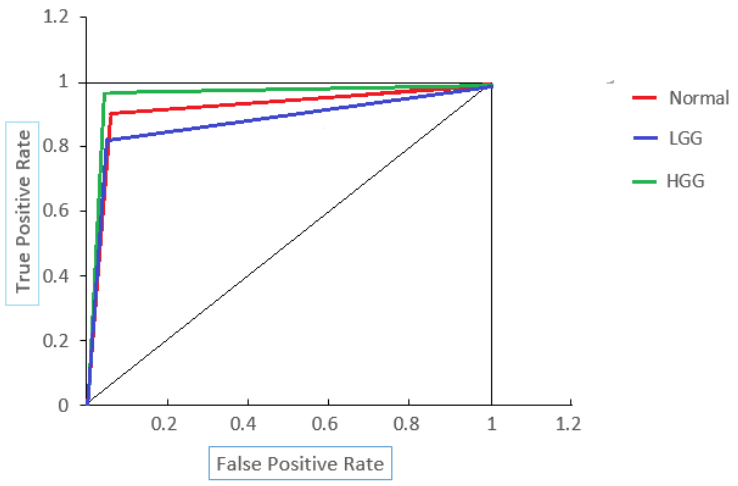

Figure 4.13 The ROC curve of three classes by using simplified RNN.

Table 4.10 Comparison between the results of the proposed system and the results of other published studies.

Study	No. of classes	Method	Data set / Modality	Specificity (%)	Sensitivity (%)	Accuracy (%)
Zollner et al. [20]	2	SVM + PCA	Private data / T1-W and T2-W	84.0	87.0	85.0
E.I. Zacharaki et al. [19]	2	SVM-RFE + SVM	Private data set / Multi-modalities	95.5	84.6	87.8
T. Gupta et al. [82]	2	DWT + PCA + CART DWT + PCA + Random Forest DWT + PCA + KNN DWT + PCA + Linear SVM	FORTIS Memorial Research Institute / FLAIR	88 96 80 96	80 80 80 80	84.0 88.0 80.0 88.0
K.L. Hsieh et al. [22]	2	Logistic Regression	TCIA / T1-W	90	82	88.0
M. Soltaninejad et al. [5]	3	Superpixel mRMR + SVM Superpixel mRMR + ERT	MICCAI BRATS 2012 / FLAIR	83.7 89.0	82.7 88.0	(Dice factor) 0.83 0.88
Khawaldeh et al. [21]	3	ConvNets	TCIA / FLAIR	91.7	92.0	91.1
J. Sachdeva et al. [15]	6	CBAC + PCA + ANN	PGIMER / T1-W	N/A	N/A	91.0
G. Yang et al. [109]	2	D-SEG + SVM	Private data set / Multi-modalities	90	94.4	91.6
T.L. Jones et al. [18]	5	SVM + D-SEG spectra	Private data set / Multi-modalities	> 97	> 90	94.7
Proposed System	**3**	LDI-Means + (P-based MI + SVD) + MLP LDI-Means + (P-based MI + SVD) + RBF-SVM LDI-Means + (P-based MI + SVD)+ Simplified RNN LDI-Means + (D-based MI + SVD) + MLP LDI-Means + (D-based MI + SVD) + RBF-SVM **LDI-Means + (D-based MI + SVD) + Simplified RNN**	**TCIA / FLAIR**	93.5 94.3 94.6 93.3 94.8 **96.3**	85.0 87.0 95.9 86.5 87.4 **96.8**	90.0 91.0 93.1 90.8 92.2 **94.9**

4.8 System Validation by using Real Data Set

The complexity of brain issues in MRIs made automated medical decision making systems a difficult challenge. The existing methods are often limited to research oriented organizations because they are not suitable for clinical use. Most of them are designed to fit specific imaging modalities, or specific tumor types and tested on a relative small set of data or even artificial data set as well as they take a long time to implement.

Any design of automated MRI brain tumor segmentation / grading system should consider the actual problems and be appropriate for regular use by the doctors. The data set which used for training such systems should handle as many tumor grades and imaging protocols as possible. In this field of study, The challenge is still for state-of-the-art to fill the gap between research oriented software and the application of real-world routine.

Proceeding from this principle, this study used a real data set taken from Iraqi MRI testing center to check the validity of the proposed system. Al-Kadhimain Medical city is one of the public hospitals in Baghdad / Iraq. Its capacity is approximately 812 beds. The MRI testing center of the hospital receives every day more than 100 cases for spine test, abdomen test and brain test, see Fig.4.14. In order to perform the necessary examinations it is used 3 scanners of different spacial resolution; Philips *GYROSCAN* of 1.5 Tesla, Siemens *AVANTO* 1.5 Tesla, and Philips *ACHIEVA* 3.0 Tesla.

The collected data are 322 images (Axial plane). 149 images of 60 normal patient and 173 images of 87 abnormal patients. The slices are mixed between FLAIR and T2-weighted, RGB, and of different sizes.

By applying the proposed system using this real data set, the LDI-Means clustering algorithm faced some failure modes. The percentage of failure to successful cases is 8%. That means from 173 images of abnormal brain, there are 14 images didn't give the segmented lesion properly. This percentage is represented a little bit high in medical imaging processing.

In the following, some factors that affected on the obtained results:

1. The effect of the variations in intensity of each image because of the variation in the acquisition formats, FLAIR and T2-W.

2. The effect of noises which may be occurred because of patient movement, such as strong breathing due to fear or nervousness as well as due to some settings of the device itself. The noises or interference produced by raido-frequency signal occupy a predominant position in MRIs. Some MRIs need bias-field

Figure 4.14 The Iraqi center for MRI studies and researches.

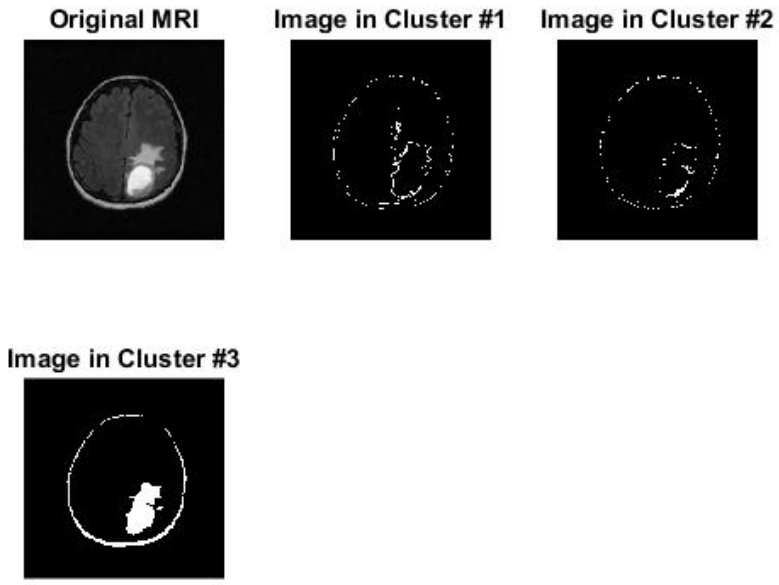

Figure 4.15 Clusters by using LDI-Means

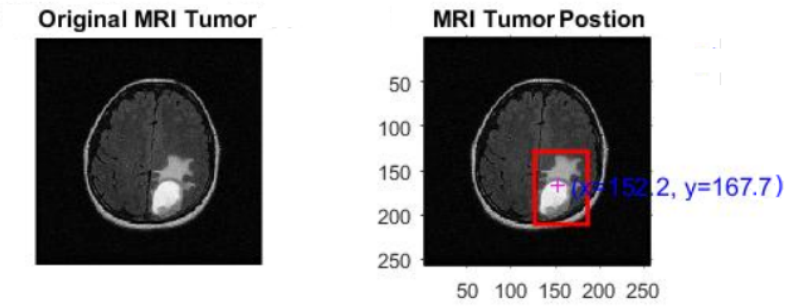

Figure 4.16 Tumor location in term of x and y.

correction. Bias-field is an intensity inhomogeneity caused by physical effects during recording.

3. The effect of the difference in sizes because the images were collected from different MRI scanners. Also, these images were acquired by more than one operator. In a study of Foster et al. [70], the experimental results showed some differences when using images acquired by two different operators.

The example of successful clustering result is shown in Fig.4.15 and the successful tumor position identification is shown in Fig.4.16. The example of failure clustering result is shown in Fig.4.17 and the unsuccessful tumor position identification is shown in Fig.4.18.

It was carefully examined LDI-Means algorithm's failure modes. While most cases are segmented with almost 98.72% accuracy, failure modes will need to be solved before the system is ready for the clinic, which is the objective of this study. To that end,

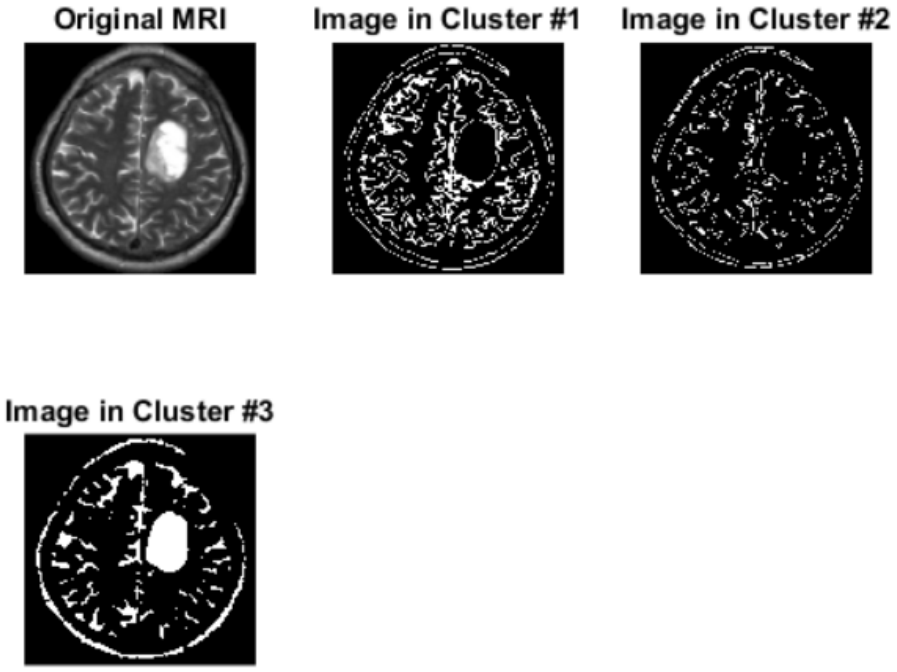

Figure 4.17 Unsatisfied clustering by using LDI-Means.

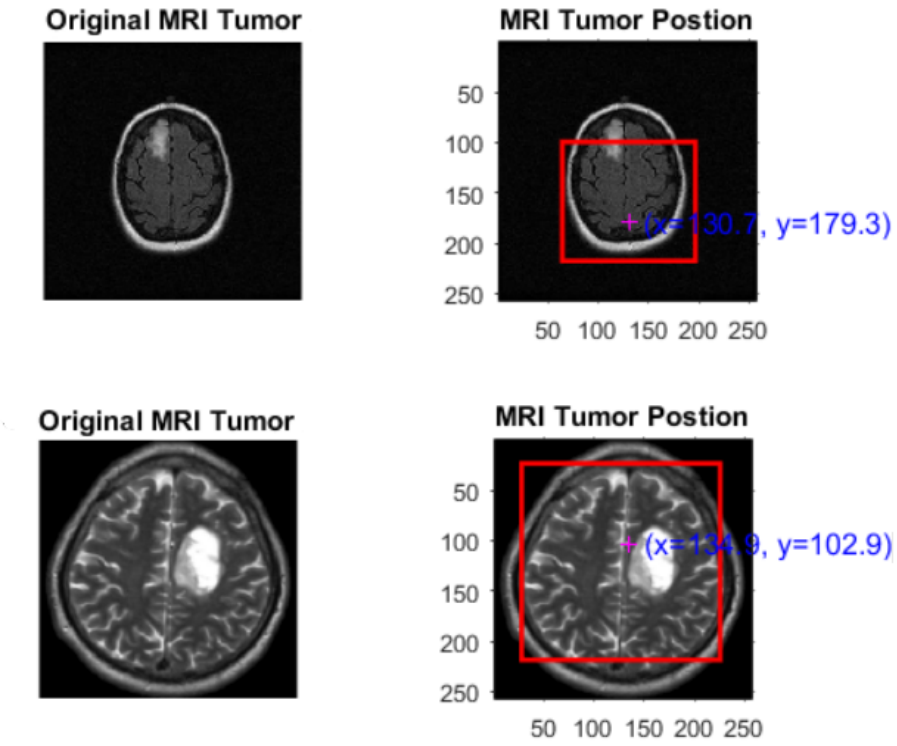

Figure 4.18 The wrong identification for the tumor centre in term of x and y for two samples.

Table 4.11 The classification accuracy of the proposed method by using real data set

Approach	Accuracy (%) Training	Testing	No. of Features
(P-based MI + SVD) + Simplified RNN	93.86	95.20	12
(D-based MI + SVD) + Simplified RNN	94.75	95.83	15
All features + Simplified RNN	82.88	84.33	40

there are some proposed potential ways to repair them. The post-processing step could be added to obtain better result. It is an additional step used to improve clustering result for the failure modes. Here, The post-processing involved some morphological operations to remove the non-circular area which were incorrectly labeled as tumor. The morphological operation enhanced the segmentation by removing distortion and to filter very small non-circular regions. These steps came after detecting the object by using LDI-Means and it will comprises filling the holes, by opening and closing.
After the post-processing step the clustering of the 14 images was corrected as shown in Fig.4.19.

The classification process provided a satisfied results by using the proposed system based on the real data as shown in Table 4.11.

For the second time, the proposed system; the LDI-Means + (MI+SVD) + Simplified RNN has achieved remarkable success when using a real data set. Based on the feature values of real data set, D-based MI+SVD was able to pick up automatically fifteen features from the entire feature space extracted throughout the feature extraction stage of the proposed system. These 15 features was considered the most significant features. Whereas P-based MI+SVD selected only 12 features to be the most meaningful features.
It was found that the (D-based MI+SVD) + Simplified RNN with accuracy of 95.83% achieved more satisfied result than (P-based MI+SVD) + Simplified RNN with accuracy of 95.20%. The D-based MI+SVD followed the same direction as it had previously taken and remained superior to P-based MI+SVD for the classification accuracy improvement purpose as shown in Fig.4.20.

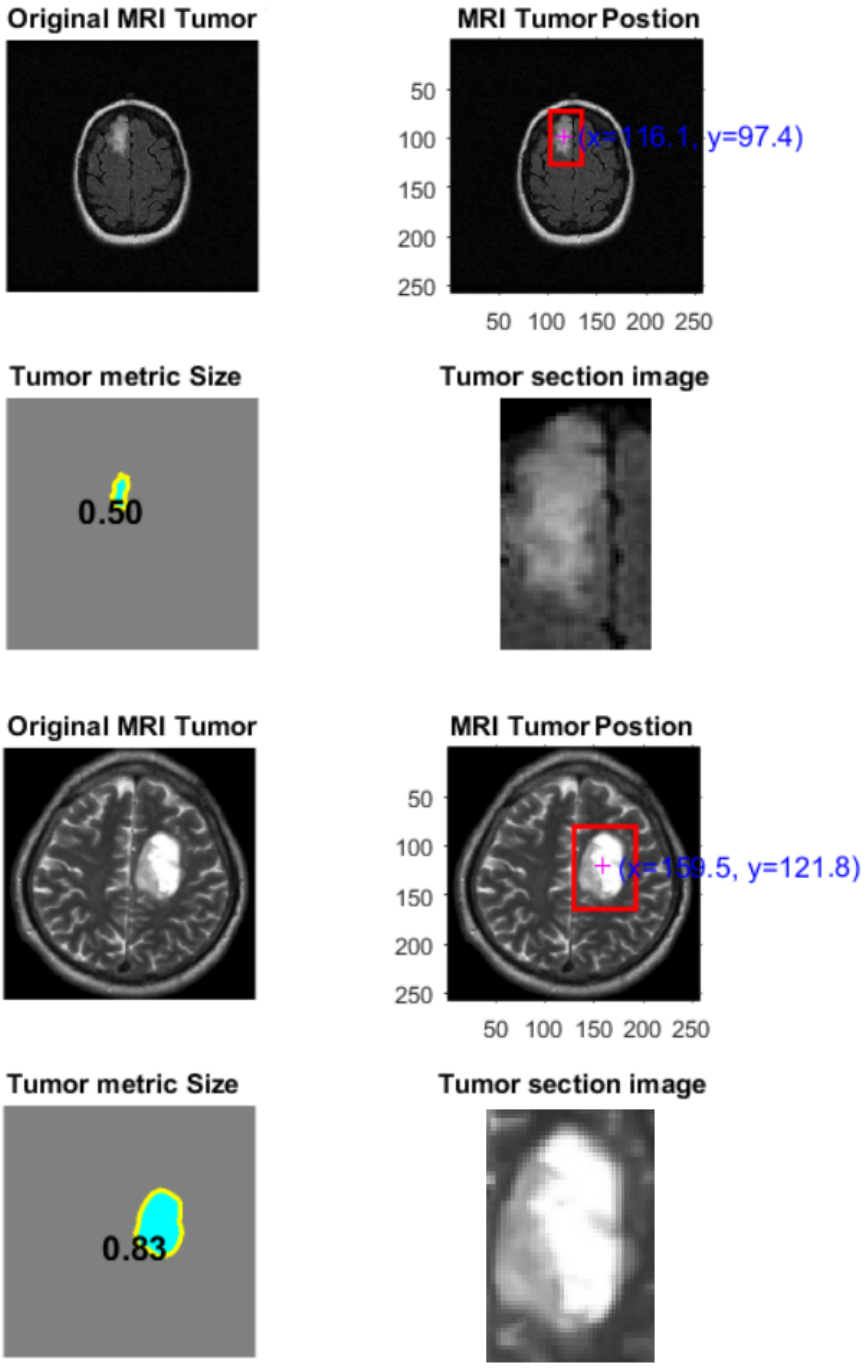

Figure 4.19 The successful clustering for obtain the tumor image

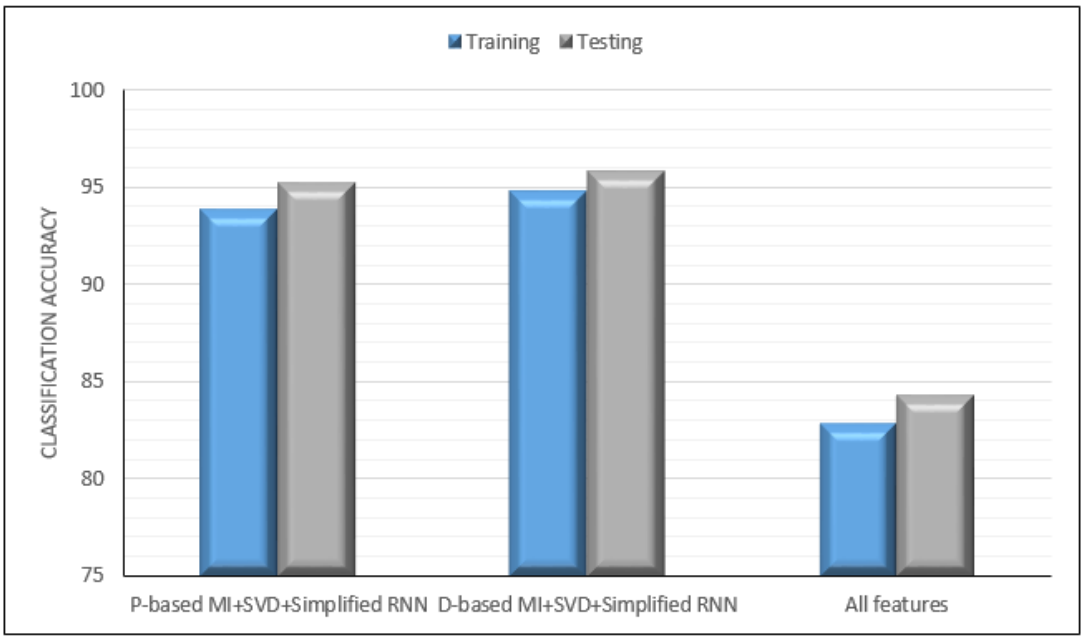

Figure 4.20 The graphical description of the classification accuracy for the proposed system by using real data set.

4.9 Summary

In view of all the experimental results, several findings are summarized:

1. Using LDI-Means to extract the tumor image in the proposed system provided a precise segmented area to be stored in a separated file for to use in the next stage. LDI-Means is very stable algorithm and can automatically perform the segmentation in less number of clusters and less time with higher accuracy. It is simple and can be implemented easily.

2. Using LDI-Means can produce useful information about the tumors such as tumor to brain ratio, and tumor metric size. Tumor size and tumor to brain tissues ratio could give good indication about how far the tumor has spread within the brain tissues. It proves its effectiveness compared to two segmentation methods; K-means and watershed algorithms.

3. Using MI in combination with SVD creates an impressive way to decrease the features space by selecting the robust features that could offer an excellent match to the inputs more than others. It is a novel approach which has not been mentioned before this study.

4. Using MI+SVD to improve the classification performance in the proposed system provided a significant estimation of the brain tumor grades. It is not complex algorithm and easy to implement.

5. Using MI+SVD in its two version; P-based MI+SVD and D-based MI+SVD, offered better result than two common methods used for the same purpose; PCA and SVD. It is expected that for the standard SVD and PCA to have a low classification accuracy because they do not take into consideration the relations between the class labels and the features in contrast to MI+SVD.

6. D-based MI+SVD yielded slightly better result than P-based MI+SVD and that because MI based on probability, depends only on the intensity values of population without any spatial information which might be occurred in the individual images whereas the MI based on distance metric depends on the distance between the elements in population.

7. Using simplified RNN to classify the brain images into three classes in the proposed system provided a reliable model that can make a useful medical decision making system.

8. simplified RNN is very flexible and understandable network in contrary to many deep learning networks used in the filed of interest which full of sophistication and act as a black box.

9. As seen from Table 4.10 the proposed system occupies a superior position compared to current methods in published studies.

5
RESULTS AND DISCUSSION

5.1 Conclusions

In recent days, tumor detection, segmentation and classification methods have been a widely investigated field. While several algorithms have been developed, this field of study remains open to further studies. The essential objective of this study is to build an automated system can identify and classify grades of brain gliomas. This study achieved this goal by proposing an intelligent system which has six stages: image enhancement, segmentation, tumor localization, feature extraction, MI+SVD algorithm employment, and classification. In the first stage, the image noise is removed, non-brain tissues are stripped, and the image is prepared appropriately for the next stage. In both segmentation and tumor localization stages, the new method named LDI-Means is used to segment the ROIs and to obtain the tumor position. In the feature extraction stage, intensity and texture based features are extracted. In the fourth stage, a novel algorithm, MI+SVD, is used to find a small subset of features that have a maximum information about the class label. In the classification stage, MLP, RBF-SVM, and simplified RNN are implemented.

The proposed system combined both unsupervised and supervised learning methods. It can construct a suitable medical decision making system for decreasing the errors in diagnoses and speeding up diagnostic procedures. it may avoid the surgical interventions as well.

K-means algorithm is a common clustering method used for segmentation process. The ordinary K-means algorithm selects randomly K points from the population and assumes these points are the initial centers. Definitely, this selection will not give a stable and true clusters each time. Conversely, the proposed LDI-Means algorithm showed its ability to generate a very stable and accurate clusters. This study offers a successful way for grouping pixels in a short time to the number of clusters. The experimental findings confirmed that LDI-Means algorithm can select specific initial centroids and provide a better accuracy in shorter computational time than the K-means algorithm. Also, can give a better tumor metric size than ones that measured

after using K-means and watershed algorithms.
This study presents a hybrid method of using MI+SVD to enhance the classification performance. Automatically, MI+SVD technique identify the most meaningful features, resulting in excellent recognition of the grade of the tumors and saved time. In addition, adding a shortcut connections to built a new network structure; simplified RNN, and later to used it as a classifier permitted the classification process in the proposed system to be more convenient. This study achieves many benefits, such as:

1. Developing a medical decision making system can rapidly provide a precise suggestions about brain glioma grades to radiologists based on preoperative clinical examinations.

2. Offering a understandable and easy to implement methods instead of using one of deep learning networks which is a complex and ambiguous process. Almost, its complexity is because of the itemized patterns of how information can flow within the model.

3. Providing a an effective automated tumor grading agency that is able to detect and classify brain tumor grades based on as little data as possible. This study used only one MRI acquisition format (FLAIR). Most of the existing algorithms are used depend on information from four different sequences of MRI and frequently, this is not easy to provide.

4. Removing the noisy and unreliable features. Practically, a high amount of training data means more features may significantly slow down the learning process and cause overfitting due to the redundant or irrelevant features that confuse the learning algorithm. Thus using MI+SVD will improve the classification accuracy, reduce the overfitting risk and speed up in training.

5. Lowering the computational costs due to reduced dimensionality in the model training.

6. Improving the generalization ability of the classifiers by reducing the capacity and achieving the early stopping.

7. Saving the time. There is no doubt that the automatically computational segmented lesions could serve as a proper surrogate or even better to manual delineations in term of time and precision.

8. Presenting a comparative experimental study of three segmentation algorithms (K-means, watershed, and LDI-Means), three dimensionality reduction methods (PCA, SVD, and MI+SVD) and 3 classification techniques (MLP, RBF-SVM, and simplified RNN).

9. Addressing an important overview of the existing methods in the task of brain tumor segmentation and classification as seen in section 1.4 which contained more than 25 published studies.

Generally, based on the satisfied results of this study, it is expected to expand it to be able to identify tumors in other organs. If this study can be adapted to fit in different medical fields, it will enable the early diagnosis of many types of diseases and prevents the invasive operation, which can suffer from risks more than its benefits.

5.2 Main Contribution and Novelty

The combination of using LDI-Means, MI+SVD, and simplified RNN is the main novelty which proposed by this study. This study includes multiple contributions, which are summarized as follows:

- A new clustering method named LDI-Means (local difference in intensity - means) clustering method is proposed to overcome the standard k-means clustering limitation.
- A new algorithm named MI+SVD (mutual information + singular value decomposition) is presented to find the robust features in order to improve the classification process.
- A simplified version of RNN is offered to perform classification.
- Several tests are carried out to create a satisfied comparative analysis.
- Finally, a fully automated system is presented for detection, segmentation and grading of the brain tumor.

In summary, a review of previous studies in the last ten years was addressed for comparison purposes. Novel approaches were proposed for brain tumor segmentation and grading. This study can assist the doctors, radiologists and surgeons to make a right decision about diseases diagnosing in very short time and with high accuracy. This study actively contributes to the development of medical decision making systems.

5.3 Limitations

The proposed system may face some limitations regarding to the segmentation and the selected features. In this study, both LDI-Means and MI+SVD were managed to perform harmoniously complementing each other by using the data set of brain MRIs, But definitely the variety of other data sets may present different difficulties and challenges. Additionally each organ within human body can show different tissue intensities through imaging. Therefore, the proposed model in its current version could be unsuitable for all data sets and will need some modification to suit the nature of each data set. It is needed to use various data sets to cover different possible scenarios in order to overcome these limitations in order to make the proposed system comprehensive and wide-ranging and can help in many medical issues.

5.4 Future Perspectives

This study has some potential for future development. These possible future steps includes:

1. Using T1-weighted or T2-weighted MR images or may be mixing more than one acquisition format if available, where FLAIR-weighted is only used in this study could enhance the proposed system accuracy.

2. Using different data sets from various imaging techniques and for different organs could overcome the limitation of the proposed system and make it more general for decision making in more than one medical issue.

3. Using fused images could improve the findings of the proposed system. These images can be created by infusing images from different imaging modalities using some special methods such as wavelet-based fusion image which utilizes the complementary and redundant information from the Computed Tomography (CT) image and Magnetic Resonance Imaging (MRI) images [110].

4. Using 3-D VOIs for evaluation, which could be more convincing.

5. Increasing the number of classes, which could provide more information on the grades of glioma tumors.

6. Using another methods to extract features. There are many techniques used to extract different types features from images such as local binary pattern (LBP), histogram of gradient (HOG) etc. [28]. Furthermore, using global features in addition to local features may increase the quality of the extracted features [22].

7. Using different classifiers to improving the system efficiency. In the future, the researchers can use selective classifier method combined by more than one classifier with methods of feature selection [8].

8. Using deep learning neural networks, because in spite of some successes, the applications of these neural network types remain relatively unexplored effectively in the field of neuroimaging.

REFERENCES

[1] N. Dhanachandra, K. Manglem, Y. J. Chanu, "Image segmentation using k-means clustering algorithm and subtractive clustering algorithm," pp. 764–771, 2015.

[2] G. Vishnuvarthanana, P. M. Rajasekaranb, P. Subbarajc, A. Vishnuvarthana, "An unsupervised learning method with a clustering approach for tumor identification and tissue segmentation in magnetic resonance brain images," *Applied soft computing journal*, vol. 38, pp. 190–212, 2016.

[3] M. Abubaker, W. Ashour, "Efficient data clustering algorithms: Improvements over k-means," *International Journal of Intelligent Systems and Applications*, vol. 5, no. 3, pp. 37–49, 2013.

[4] M. Soltaninejad, X. Ye, G. Yang, N. Allinson, T. Lambrou, "Brain tumour grading in different mri protocols using SVM on statistical features," in *Proc. 18th Conf. Med. Image Understand. Anal. (MIUA)*, Egham,U.K, 2014, pp. 259–264.

[5] M. Soltaninejad, G. Yang, T. Lambrou, N. Allinson, L. Timothy, J. Thomas, R. Barrick, F. Howe, X. Ye, "Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in flair mr," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, pp. 183–203, 2017.

[6] A. Javadpour, A. Mohammadi, "Improving brain magnetic resonance image MRI segmentation via a novel algorithm based on genetic and regional growth," *Journal of Biomedical Physics and Engineering*, vol. 6, no. 2, 2016.

[7] M. Angulakshmi, L. G. G., "Automatic brain tumour segmentation of magnetic resonance images (MRI) based on region of interest (ROI)," *Journal of Engineering Science and Technology*, vol. 12, no. 4, pp. 875–887, 2017.

[8] N. Bahadure, A. Ray, H. Thethi, "Image analysis for mri based brain tumor detection and feature extraction using biologically inspired BWT and SVM," *International journal of biomedical imaging*, vol. 2017, pp. 1–12, 2017.

[9] I. Cabria, I. Gondr, "Mri segmentation fusion for brain tumor detection," *Information Fusion journal*, vol. 36, pp. 1–9, 2017.

[10] J. J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, A. Yuille, "Efficient multilevel brain tumor segmentation with integrated bayesian model classification," *IEEE Trans. Med. Imag.*, vol. 27, no. 5, pp. 629–640, 2008.

[11] H. Dong, G. Yang, F. Liu, Y. Mo, Y. Guo, "Automatic brain tumor detection and segmentation using U-net based fully convolutional networks," *Annu. Conf. Med. Image Understanding Anal.*, pp. 11–13, 2017.

[12] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Trans. Neural. Netw.*, vol. 5, no. 4, pp. 537–550, 1994.

[13] H. Peng, F. Long, C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1266–1238, 2005.

[14] V. Kumar, J. Sachdeva, I. Gupta, N. Khandelwal, C. K. Ahuja, "Classification of brain tumors using PCA-ANN," in *Proceedings of the 24th international conference on Machine learning*, in Proc. World Congr. Inf.Commun. Technol., 2011, pp. 1079–1083.

[15] J. Sachdeva, V. Kumar, I. Gupta, N. Khandelwal, C. K. Ahuja, "Segmentation, feature extraction, and multiclass brain tumor classification," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1141–1150, 2013.

[16] L. Fang, H. Zhao, P. Wang, M. Yu, J. Yan, W. Cheng, P. Chen, "Feature selection method based on mutual information and class separability for dimension reduction in multidimensional time series for clinical data," *Biomed. Signal Process. Control*, vol. 21, pp. 82–89, 2015.

[17] N. Hoque, H. Ahmed, D. Bhattacharyya, J. Kalita, "A fuzzy mutual information-based feature selection method for classification," *Fuzzy Inf. Eng*, vol. 8, no. 3, pp. 355–384, 2016.

[18] T. Jones, T. Byrnes, G. Yang, F. Howe, B. Bell, T. Barrick, "Brain tumor classification using the diffusion tensor image segmentation (d-seg) technique," *Neuro-Oncol.*, vol. 17, no. 3, pp. 466–476, 2015.

[19] E. Zacharaki, S. Wang, S. Chawla, D. Yoo, R. Wolf, R. Methem, C. Davatzikos, "Classification of brain tumor type and grade using mri texture and shape in a machine learning scheme," *Magnetic resonance in medicine*, vol. 62, no. 6, pp. 1609–1618, 2009.

[20] F. Zollner, K. Emblem, L. Schad, "SVM based glioma grading: Optimization by feature reduction analysis," *Zeitschrift fur Medizinische Physik*, vol. 22, no. 3, pp. 205–214, 2012.

[21] S. Khawaldeh, U. Pervaiz et al., "Noninvasive grading of glioma tumor using magnetic resonance imaging with convolutional neural networks," *Appl. Sci*, vol. 8, no. 27, pp. 1–17, 2018.

[22] K. Hsieh, C. Lo, C. Hsiao, "Computer-aided grading of gliomas based on local and global MRI features," *Computer Methods and Programs in Biomedicine*, vol. 139, pp. 31–38, 2017.

[23] M. Hasan, H. Jalab, F. Meziane, H. Kahtan, A. Al-Ahmad, "Combining deep and handcrafted image features for MRI brain scan classification," *IEEE Access*, vol. 7, pp. 79 959–79 967, 2019.

[24] S. Bakas, et al, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," pp. –, 2018.

[25] G. Litjens, T. Kooi, B. Bejnordi, A. Setio, F. Ciompi, M. Ghafoorian, J. V. D. Laak, B. V. Ginneken, C. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.

[26] G. Mohan, M. Subashini, "MRI based medical image analysis: Survey on brain tumor grade classification," *Biomed. Signal Process. Contro*, vol. 39, pp. 139–161-, 2018.

[27] B. Remeseiro, V. Bolon-Canedo, "A review of feature selection methods in medical applications," *Comp. Bio. Med*, vol. 112, 2019.

[28] V. Jalal, D. Kaur, "A study of classification and feature extraction techniques for brain tumor detection," *International Journal of Multimedia Information Retrieval*, vol. 9, pp. 271–290, 2020.

[29] A. M. Hasan, F. Meziane, "Automated screening of MRI brain scanning using gray level statistics," *Comput. Elect. Eng.*, vol. 53, pp. 276–291, 2016.

[30] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[31] E. Tsolaki, "Clinical decision support systems for brain tumor characterization using advanced magnetic resonance imaging techniques," *World J. Radiol.*, vol. 6, no. 4, pp. 72–81, 2014.

[32] P. de Clercq, J. Blom, H. Korsten, A. Hasman, "Approaches for creating computer-interpretable guidelines that facilitate decision support," *Artificial Intelligence in Medicine*, vol. 31, no. 1, pp. 1–27, 2004.

[33] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, D. Firmin, "DAGAN: Deep dealiasing generative adversarial networks for fast compressed sensing MRI," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1310–1321, 2018.

[34] M. Saii, Z. Kraitem, "Automatic brain tumor detection in mri using image processing techniques, biomedical statistics and informatics," *Biomedical Statistics and Informatics*, vol. 2, no. 2, pp. 73–76, 2017.

[35] V. Zeljkovic, C. Druzgalski, Y. Zhang, Z. Zhu, Z. Xu, D. Zhang, P. Mayorga, "Automatic brain tumor detection and segmentation in MR images," in *in Proc. Pan Amer. Health Care Exchanges (PAHCE)*, 2014.

[36] J. Kim, G. Lee, Y. Park, Y. Hong, "Using a method based on a modified k-means clustering and mean shift segmentation to reduce file sizes and detect brain tumors from magnetic resonance MRI images," *Wireless Personal Communications Journal*, vol. 89, pp. 993–1008, 2016.

[37] S. Pereira, A. Pinto, V. Alves, C. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," vol. 35, no. 5, pp. 1240–1251, 2016.

[38] E. Hancer, B. Xue, M. Zhang, "Differential evolution for filter feature selection based on information theory and feature ranking," *Knowl.-Based Syst.*, vol. 140, pp. 103–119, 2018.

[39] F. Ali-Osman, "The gliomas," in *Berger MS, Wilson CW (eds)*, WB Saunders, 1999, pp. 134–141.

[40] J. Smirniotopoulos, "The new WHO classification of brain tumors," in *Neuroimaging Clin North Am.*, vol. 9, 1999, pp. 595–613.

[41] "Brain tumours (primary) and brain metastases in adults," in *guideline, NICE*, https://www.nice.org.uk/guidance/ng99, 2018.

[42] F. R. Nelson, C. T. Blauvelt, "Chapter three: Imaging techniques," in *A Manual of Orthopaedic Terminology (Eighth Edition)*, Elsevier, 2015.

[43] T. A. Dolecek, J. M. Propp, N. E. Stroup, C. Kruchko, "Primary brain and central nervous system tumors diagnosed in the united states in 2005–2009," *Neuro-oncology*, vol. 14, no. 5, pp. 1–49, 2012.

[44] N. Smith, A. Webb, "Introduction to medical imaging: Physics, engineering and clinical applications," in *Cambridge Texts in Biomedical Engineering*, Cambridge University Press, 2010.

[45] R. Birry, "Automated classification in digital images of osteogenic differentiated stem cells," in *PhD*, University of Salford, Manchester, UK., 2013.

[46] R. B. Buxton, "Introduction to functional magnetic resonance imaging: Principles and techniques," Cambridge University Press, 2009.

[47] M. Bynevelt, J. Britton, H. Seymour, E. MacSweeney, N. Thomas, K. Sandhu, "FLAIR imaging in the follow-up of low-grade gliomas: Time to dispense with the dual echo," Neuroradiology, vol. 43, 2001, pp. 129–133.

[48] Z. Al-Saffar, T. Yildirim, "An optimized clustering approach for tumor segmentation using local difference of intensity level in MR brain images," in *Innov. Intell. Syst. Appl. (INISTA)*, Thessaloniki, Greece, 2018, pp. 1–8.

[49] R. C. Gonzalez, R. E. Woods, S. L. Eddins, "Chapter 2: Fundamentals," in *Digital Image Processing Using MATLAB*, Upper Saddle River, NJ, USA: Prentice-Hall, 2003.

[50] M. K. Chung, "Chapter 6: Smoothing on cortical mainfolds," in *Computational Neuroanatomy: The Methods*, World Scientific Publishing Co. Pte. Ltd., 2013.

[51] S. Jigui, L. Jie, Z. Lianyu, "Clustering algorithms research," *Journal of Software*, vol. 19, no. 1, pp. 48–61, 2008.

[52] K. Abdul Nazeer, M. Sebastian, "Improving the accuracy and efficiency of the k-means clustering algorithm," in *Proceeding of the World Congress on Engineering*, London, vol. 1, 2009.

[53] Y. Raykov, A. Boukouvalas, F. Baig, M. Little, "What to do when k-means clustering fails: A simple yet principled alternative algorithm," *PLoS ONE*, vol. 11, no. 9, e0162259, 2016.

[54] A. Chadha, S. Kumar, "An improved k-means clustering algorithm: A step forward for removal of dependency on k," in *International Conference on Reliability, Optimization and Information Technology (ICROIT)*, India, 2014.

[55] B. S., C. Lantuejoul, "Use of watershed in contour detection," in *In Proceedings of international workshop on image processing. CCETT/IRISA*, Rennes, France, 1979.

[56] G. Zeng, "A unified definition of mutual information with applications in machine learning," *Math. Problems Eng.*, vol. 2015, pp. 201–874, 2015.

[57] R. M. Gray, "Entropy and information theory," 1990.

[58] J. Vergara, P. Estévez, "A review of feature selection methods based on mutual information," *Neural Computing and Applications*, vol. 1, no. 24, pp. 175–186, 2014.

[59] L. Smith, "A tutorial on principal component analysis," [Online]. Available: http://www.cs.otago.ac.nz/cosc453/student$_tutorials/principal_components.pdf$, 2002.

[60] A. J. Calder, A. Mike Burton, P. Miller, A. W. Young, S. Akamatsu, "A principal component analysis of facial expressions," *Journal of Vision Research*, vol. 41, pp. 1179–1208, 2001.

[61] Y. Arya, P. Shinde, S. Chandwani, N. Chandwani, M. Roja, "Facial expression recognition," *International Journal of Emerging Technology and Advanced Engineering*, vol. 9001, 2014.

[62] G. James, D. Witten, T. Hastie, R. Tibshirani, "Chapter 10: Unsupervised learning," in *An Introduction to Statistical Learning: with Applications in R,*, Springer, 2013.

[63] E. Barshan, A. Ghodsi, Z. Azimifar, M. Zolghadri Jahromi, "Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds," *Pattern Recognition*, vol. 44, no. 7, pp. 1357–1371, 2011.

[64] O. Chapelle, B. Schölkopf, A. Zien, "Semi-supervised learning," in *Cambridge*, USA, 2006.

[65] R. Sadek, "SVD based image processing applications: State of the art, contributions and research challenges," *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 3, no. 7, pp. 26–34, 2012.

[66] W. A. Shehab, Z. Al-qudah, "Singular value decomposition: Principles and applications in multiple input multiple output communication system," *Int. J. Comput. Netw. Commun.*, vol. 9, no. 1, pp. 13–21, 2017.

[67] E. Biglieri, K. Yao, "Some properties of singular value decomposition and their applications to digital signal processing," *Signal Processing*, vol. 18, no. 3, pp. 277–289, 1989.

[68] W. Ford, "Chapter 15: The singular value decomposition," in *Numerical linear algebra with applications*, Elsevier, 2015.

[69] H. Andrews, C. Patterson, "Singular value decompositions and digital image processing," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP24, pp. 26–53, 1976.

[70] K. Foster, R. Koprowski, J. Skufca, "Machine learning, medical diagnosis, and biomedical engineering research - commentary," *Biomedical engineering online*, vol. 13 94, 2014.

[71] T. Ross, "Artificial neural network," in *Fuzzy logic with engineering applications*, John Wiley Sons, 2009.

[72] A. Dongare, R. Kharde, D. Kachare, "Introduction to artificial neural network," *International Journal of Engineering and Innovative Technology*, vol. 2, no. 1, pp. 2277–3754. 2012.

[73] J. Jiang, P. Trundle, J. Ren, "Medical image analysis with artificial neural networks," *Comput. Med. Imaging Graphics*, vol. 34, pp. 617–631, 2010.

[74] D. Larose, "Discovering knowledge in data an introduction to data mining," in *USA,John*, Wiley Sons, Inc., 2005.

[75] A. Oustimov, V. Vu, "Artificial neural networks in the cancer genomics frontier," *Translational Cancer Research*, vol. 3, no. 3, 2014.

[76] M. Huang, C. Chen, W. Lin, S. Ke, C. Tsai, "Svm and svm ensembles in breast cancer prediction," *PLOS ONE*, vol. 12, no. 1, 2017.

[77] R. Romero, E. Iglesias, L. Borrajo, "A linear-RBF multikernel SVM to classify big text corpora," *BioMed Research International*, 2015.

[78] S. Bauer, L. Nolte, M. Reyes, "Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization," *Medical Image Computing and Computer-Assisted Intervention, MICCAI*, vol. 6893. 2011.

[79] H. Sanz, C. Valim, E. Vegas, E. Vegas, J. Oller, F. Reverter, "SVM-RFE: Selection and visualization of the most relevant features through non-linear kernels," *BMC Bioinformatics*, vol. 19, no. 432, 2018.

[80] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. e. a. Pringle, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, pp. 1045–1057, 2013.

[81] L. Scarpace, A. Flanders, R. Jain, T. Mikkelsen, D. Andrews, "Data from REMBRANDT," in *The Cancer Imaging Archive*, 2015.

[82] T. Gupta, T. K. Gandhi, R. K. Gupta, B. K. Panigrahi, "Classification of patients with tumor using MR FLAIR images," *Pattern Recogn. Lett.*, vol. 139, pp. 112–117, 2020.

[83] J. Rodriguez, A. Perez, J. A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, 2010.

[84] S. Arlot, A. Celisse, "A survey of cross-validation procedures for model selection," *Statist. Surv.*, vol. 4, pp. 40–79, 2018.

[85] M. Bernardini, L. Romeo, P. Misericordia, E. Frontoni, "Discovering the type 2 diabetes in electronic health records using the sparse balanced support vector machine," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 235–246, 2019.

[86] K. Korjus, M. Hebart, R. Vicente, "An efficient data partitioning to improve classification performance while keeping parameters interpretable," *PLoS ONE*, vol. 11, no. e0161788, 2016.

[87] K. Somasundaram, T. Kalaiselvi, "Automatic brain extraction methods for t1 magnetic resonance images using region labeling and morphological operations," *Comput. Biol. Med.*, vol. 41, pp. 716–725, 2011.

[88] C. C. Benson, V. L. Lajish, "Morphology based enhancement and skull stripping of MRI brain images," in *the International Conference on Intelligent Computing Applications, Coimbatore*, 2014, pp. 254–257.

[89] S. Mohsin, S. Sajjad, Z. Malik, A. H. Abdullah, "Efficient way of skull stripping in mri to detect brain tumor by applying morphological operations, after detection of false background," *International Journal of Information and Education Technology*, vol. 2, no. 4, pp. 335–337, 2012.

[90] R. Haralick, K. Shanmugam, I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern*, vol. SMC-3, no. 6, pp. 610–621, 1973.

[91] L. Soh, C. Tsatsoulis, "Texture analysis of sar sea ice imagery using gray level co-occurrence matrices," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 780–795, 1999.

[92] B. Park, W. Jang, S. Yoo, "Texture analysis of supraspinatus ultrasound image for computer aided diagnostic system," *Healthcare Informatics Research*, vol. 22, no. 4, pp. 299–304, 2016.

[93] T. Al-Saffar Z.A.and Yildirim, "A hybrid approach based on multiple eigenvalues selection (MES) for the automated grading of a brain tumor using mri," *Computer Methods and Programs in Biomedicine*, 2021.

[94] M. Bennasar, Y. Hicks, R. Setchi, "Feature selection using joint mutual information maximisation," *Expert Syst. Appl.*, vol. 42, pp. 8520–8532, 2015.

[95] P. Grünwald, P. Vitányi, "Kolmogorov complexity and information theory with an interpretation in terms of questions and answers," *Journal of Logic, Language and Information*, vol. 12, pp. 497–529, 2003.

[96] Z. Al-Saffar, T. Yildirim, "A novel approach to improving brain image classification using mutual information-accelerated singular value decomposition," *IEEE Access*, vol. 8, pp. 52 575–52 587, 2020.

[97] M. Rehman, N. Nawi, "The effect of adaptive momentum in improving the accuracy of gradient descent back propagation algorithm on classification problems," *Software Engineering and Computer Systems, Communications in Computer and Information Science*, vol. 179, 2011.

[98] M. Moreira, E. Fiesler, "Neural networks with adaptive learning rate and momentum term," *IDIAP Technical Report*, no. 95–04, 1995.

[99] S. Song, Z. Zhan, Z. Long, J. Zhang, L. Yao, "Comparative study of SVM methods combined with voxel selection for object category classification on fmri data," *PLOS ONE*, vol. 6, no. 2, 2011.

[100] T. I. Poznyak, I. C. Oria, A. S. Poznyak, "Chapter3: Background on dynamic neural networks," in *Ozonation and Biodegradation in Environmental Engineering*, Elsevier, 2019, pp. 57–74.

[101] Q. Nguyen, M. C. Mukkamala, M. Hein, "On the loss landscape of a class of deep neural networks with no bad local valleys," *Published as a conference paper at ICLR 2019*, 2018.

[102] B. Li, Y. He, "An improved $resnet$ based on the adjustable shortcut connections," *IEEE Access*, vol. 6, pp. 18 967–18 974, 2018.

[103] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," in *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. CVPR*, Las Vegas, NV, USA., 2016, pp. 770–778.

[104] B. Gao, L. Pavel, "On the properties of the softmax function with application in game theory and reinforcement learning," vol. 04, 2017.

[105] C. Beleites, R. Salzer, V. Sergo, "Validation of soft classification models using partial class memberships: An extended concept of sensitivity co. applied to grading of astrocytoma tissues," *Chemom Intell Lab Syst*, vol. 122, pp. 12–22, 2013.

[106] P. Shanthakumar, P. Ganeshkumar, "Performance analysis of classifier for brain tumor detection and diagnosis," *Comput. Electr. Eng.*, vol. 45, pp. 302–311, 2015.

[107] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, pp. 861–874, 2006.

[108] X. He, B. D. Gallas, E. C. Frey, "Three-class ROC analysis—toward a general decision theoretic solution," *IEEE Trans. Med. Imaging*, vol. 29, no. 1, pp. 206–215, 2010.

[109] G. Yang, T. L. Jones, F. A. Howe, T. R. Barrick, "A morphometric model for discrimination between glioblastoma multiforme and solitary metastasis using three-dimensional shape analysis," *Magn. Reson. Med.*, vol. 75, no. 6, pp. 2505–2516, 2016.

[110] V. Angoth, C. Dwith, A. Singh, "A novel wavelet based image fusion for brain tumor detection," *International Journal of Computer Vision and Signal Processing*, vol. 2, no. 1, pp. 1–7, 2013.

PUBLICATIONS FROM THE THESIS

Papers

1. Z. A. Al-Saffar and T. Yildirim, "A Novel Approach to Improving Brain Image Classification Using Mutual Information-Accelerated Singular Value Decomposition," IEEE Access. vol. 8, pp. 52575-52587, 2020.
   doi: 10.1109/ACCESS.2020.2980728.

2. Z. A. Al-Saffar and T. Yildirim, "A Hybrid Approach Based on Multiple Eigenvalues Selection (MES) for the Automated Grading of a Brain Tumor Using MRI," Computer Methods and Programs in Biomedicine.
   (Accepted for publication)

Conference Papers

1. Z. A. Al-Saffar and T. Yildirim, "An Optimized Clustering Approach for Tumor Segmentation using Local Difference of Intensity Level in MR brain Images," in Proc. Innov. Intell. Syst. Appl. (INISTA), Thessaloniki, Greece, July 2018, pp.1–8.
   doi: 10.1109/INISTA.2018.8466316.