

Brain tumor detection in MR image using superpixels, principal component analysis and template based K-means clustering algorithm

Md Khairul Islam ^{a,*}, Md Shahin Ali ^a, Md Sipon Miah ^{b,c}, Md Mahbubur Rahman ^b,
Md Shahariar Alam ^b, Mohammad Amzad Hossain ^{c,d}

^a Department of Biomedical Engineering, Islamic University, Kushtia 7003, Bangladesh

^b Department of Information and Communication Technology, Islamic University, Kushtia 7003, Bangladesh

^c School of Computer Science, National University of Ireland Galway, H91 TK33 Galway, Republic of Ireland

^d Department of Information and Communication Engineering, Noakhali Science and Technology University, Bangladesh

ARTICLE INFO

Keywords:

Magnetic resonance imaging
Segmentation
Feature extraction
Superpixels
Principal component analysis
Template based K-means algorithm

ABSTRACT

In the present era, human brain tumor is the extremist dangerous and devil to the human being that leads to certain death. Furthermore, the brain tumor arises more complexity of patients life with time. As a result, early detection of tumors is most crucial to save and prolong the patient's lifetime. Therefore, enhanced brain tumor detection is required in medical fields. Automatic human brain tumor detection in magnetic resonance imaging (MRI) is playing a vital role in several symptomatic and cures applications. However, the existing schemes (e.g., random forest, Fuzzy C-means, artificial neural network (ANN) and wavelet transform) can detect brain tumors with insufficient accuracy and longer execution time (in minutes). In this paper, we propose an enhanced brain tumor detection scheme based on the template-based K-means (TK) algorithm with superpixels and principal component analysis (PCA) which efficiently detects the human brain tumors in lower execution time. At first, we extract essential features using both superpixels and PCA which helps accurately to detect brain tumors. Then, image enhancement is done using a filter that helps to improve accuracy. Finally, the image segmentation is performed through TK-means clustering algorithm to detect the brain tumor. The experimental results show that the proposed detection scheme achieves a better accuracy and a reduced execution time (in seconds) than other existing schemes for the detection of brain tumor in MR image.

1. Introduction

1.1. Motivation

Brain, the paramount significant and critical structural part of the human body that consists of 50–100 trillion neurons. It is also known as central part of human body. Furthermore, it plays the most significant and vital role in the nervous system that called the processor or kernel of the nervous system (Alam et al., 2019; Islam et al., 2020). To the best of our knowledge, it is too hard and complex to diagnose brain disease due to the existence of the skull around it (Gondal & Khan, 2013).

Generally, the brain is the sensitive part of the human body. Unlike other diseases, the mass growth or abnormal growth of cells in the brain leads to a lot of changes in brain behavior and characteristics. Also, it makes resistance to perform the brain functionality. This dysfunctionality or abnormality is a symptom of a brain tumor. In other

words, a brain tumor is the result of uncontrolled growth of cells within the brain. "It causes cancer that may be the reason for death and accounts for around 13% of all deaths worldwide". The level of threat for a brain tumor depends on some factors such as tumor style, behavior, size, location, and status of growth. Brain and other cancer caused human death more times than another nervous disease. It is 10th leading reason for the passing of men and women. Based on brain tumor estimates, "it is estimated that 23,890 adults (13,590 men and 10,300 women) in the United States will be clinical tested in the current year with primary cancer of the brain and spinal cord". "Moreover, 18,020 adults (10,190 men and 7830 women) are expected to die in 2020 due to primary cancerous brain and CNS tumors". In addition, "the survival rate for people with a cancerous brain or CNS tumor is almost 36% for five years and 31% for ten years" (Bahadure et al., 2017; cancer.net, 2020). Brain tumors are mainly two types such as benign and malignant. The tumors that do not contain cancerous cell and

* Corresponding author.

E-mail addresses: khairul.ice06@gmail.com (M.K. Islam), shahinbme.iu@gmail.com (M.S. Ali), m.miah1@nuigalway.ie (M.S. Miah), mrahman@ict.iu.ac.bd (M.M. Rahman), shahariar.ice12@gmail.com (M.S. Alam), m.hossain3@nuigalway.ie (M.A. Hossain).

URLs: <http://www.iu.ac.bd> (M.K. Islam), <http://iu.ac.bd> (M.S. Ali), <http://iu.ac.bd>, <http://nuigalway.ie> (M.S. Miah), <http://iu.ac.bd> (M.M. Rahman), <http://iu.ac.bd> (M.S. Alam), <http://iu.ac.bd> (M.A. Hossain).

<https://doi.org/10.1016/j.mlwa.2021.100044>

Received 15 January 2021; Received in revised form 12 May 2021; Accepted 19 May 2021

Available online 25 May 2021

2666-8270/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

less harmful to human is called benign. Whereas the tumor that exists cancerous cell that is more harmful to humans is called malignant (Ali et al., 2021; Borole et al., 2015). The radiological analysis and estimation are needed to identify the location, size, behavior, and growth status of the tumor once a brain tumor is clinically suspected. Based on this information, it is easy to take a decision for giving proper treatment to the patient like radiation, chemotherapy, best therapy, and surgery. Moreover, the most significant is the accurate detection of the brain tumor at an early stage can prolong the chance of survival of an infected patient (Coatrieux et al., 2013). There is a radical change in the medical field due to various types of new imaging techniques. The most popular imaging techniques are magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), single-photon emission computed tomography (SPECT), Ultrasound, and X-ray. These medical imaging techniques are used to detect complex diseases like brain tumors, COVID-19, cancerous cells or CNS cancer in humans (Borole et al., 2015). MRI is a more popular non-invasive technique for detecting abnormalities of tissue composition. It is used rather than other medical imaging methods, offering the best contrast images of the brain and cancer tissues. MRI provides rich information, high-quality anatomical image structures for diagnosing clinically suspected disease. In addition, it is performing a very vital role in medical and biomedical research (Liu et al., 2014). The more prominent advantages of MR images are high spatial resolution with cross-section images that is much better for soft-tissue diagnosis (Sharma & Mukharjee, 2013; Sinha & Sinha, 2014). Brain tumor detection means recognizing the infected portion of the brain with the shape, size, position, and boundary of the tumor. MRI scans or CT scans are used more frequently for the detection of brain tumors. However, CT scans emit harmful radiation during the diagnosis of a patient. Whereas, MRI is safe from radiation that is best suited for all classes of a patient like a child, adult, and pregnant women. In addition, MRI produces accurate visualization of an anatomical structure of the human body particularly soft tissues of the brain (Alam et al., 2019). It performs the test based on the radio waves and magnetic field for producing a high-quality cross-section image.

Group pixels identical in color and other low-level properties are known as superpixels. The superpixels function is very useful in feature extracting that uses the simple linear iterative clustering (SLIC) algorithm. It groups pixels into region based on similar values. These regions are used in image processing e.g., segmentation, can reduce complexity. Due to the reduction in image complexity from hundreds of thousands of pixels to only a few hundred superpixels, superpixels are computationally efficient (Wang et al., 2017). Principal Component Analysis (PCA) is a technique that decreases the dimensionality of the dataset, enhances interpretability and minimizes data loss. Sometimes, it is called heart of dimension reduction method. Moreover, It is also used to make data easy to analyze and explore (Uddin et al., 2019). For these reasons, in the proposed scheme both superpixels and PCA are used to reduce complexity of image and execution time. In addition, both method are used to extract features that assist of segmentation and detection.

1.2. Contributions

In this paper, we have achieved the following major contributions:

- We propose a template based K-Means clustering algorithm where we analyzed the human brain tumors detection with better accuracy of detection for different sizes of the tumors.
- We propose superpixels and principal component analysis for extracting features that assist the scheme for segmentation and detection of tumors from complex MR images.
- Experimental results show that the proposed scheme achieves a better detection accuracy and a lower execution time (in seconds) when compared to other conventional schemes.

1.3. Organizations

The rest of this paper is summarized as follows: the related works are presented in Section 2. Section 3 introduce the proposed detection scheme materials and methods. The results and discussion are summarized in Section 4 and the conclusion and future works are discussed in Section 5.

2. Related work

Nowadays, it is getting more challenging to detect brain tumor from MR Images. After a surgery or biopsy from tissue's sample radiologists are able to diagnose of tumor from brain MR images. In recent decades, there are many people died of having inaccurate results with higher variability in detection of brain tumor. After that, a large number of machine learning algorithm like fuzzy C-means (FCM), linear-regression, K-nearest neighbor (KNN), support vector machine (SVM), random forest and so on have been built up and used for the detection of brain tumor. These techniques utilized for decreasing manual interactions in medical field. "However, these techniques had less accuracy and more time-consuming (i.e., non-automated)". As a result, the manual interaction based detection could essential to mitigate these drawbacks. In Chithra and Dheepa (2018), Rai and Chatterjee (2020), the authors proposed a cascaded CNN architecture for the detection of tumors by extracting tumoral characteristics and recognizing tumor images automatically from mind MR images.

However, the above scheme was needed more kernel to convolve feature map and softmax for pixel-wise classification. In addition, this scheme only detected the tumor exists or not using without segment the tumor area MR images.

In Alam et al. (2019), the authors proposed a TKFCM algorithm for detecting brain tumor on MRI image in where firstly, template-based k-means algorithm; secondly, update enrollment to a cluster data point from cluster centroid using Fuzzy C-means and finally, the improved Fuzzy C-means algorithm is used for detecting a tumor in brain image. However, it was more complex because it was extracted six features for prediction. Therefore, the system was needed to decrease complexity.

In Singh et al. (2015), the authors proposed an FCM algorithm for detecting brain MR image. However, this scheme was needed accurate defining the cluster members for automatic interaction. To mitigates this problem, the authors proposed the gray level co-occurrence matrix (GLCM) based features where the features were extracted from brain images (Chavan et al., 2015).

In Bahadure et al. (2017), the authors proposed discrete wavelet transformation (DWT) based techniques where the feature extraction techniques are used there. However, if these feature extraction techniques are used then the significant details of brain images lost. In Sandhya et al. (2017), the authors proposed SVM for tumor identification. The adaptive threshold algorithm was applied for segmentation and features were extracted by using First Fourier Transform (FFT). Finally, SVM was applied for prediction. However, the image size was 512×512 but they could extract feature only 16×16 per epoch. In addition, the system was needed more computational time.

In Sachdeva et al. (2013), the authors proposed a computer-aided diagnostic (CAD) system in which PCA utilized to decrease the dimensionality of feature space and tumor were classified by ANN. It was leading to the development of methods for segmentation, feature extraction, and classification of human brain tumors. Moreover, this scheme was beneficial to radiologists for localization, diagnosis, and interpretation of brain tumors on MR images. However, PCA-ANN enhanced accuracy but that was not enough for all classes. In Shree and Kumar (2018), introduced a probabilistic neural network classifier noise removal techniques, i.e., GLCM, DWT and segmentation to reduce the complexity which improves the accuracy. Due to, it clears the artifacts using the morphological filtering. The identification accuracy of brain tumor location by the probabilistic neural network classifier

where both train and test dataset were used. In [Rathi and Palani \(2015\)](#), the authors studied for detecting tumor images using the feed-forward back-propagation network and feed-forward artificial neural network.

In [Soltaninejad et al. \(2017a\)](#), the researchers proposed superpixel-based method for automated brain tumor identification and segmentation from Fluid-Attenuated Inversion Recovery Magnetic Resonance Imaging (FLAIR-MRI). The experimental results show that the proposed scheme improves the detection and segmentation performance by using Extremely randomized trees (ERT) classifier. For their own clinical dataset, the mean identification sensitivity, balanced error rate and the Dice overlap value for the tumor are 89.48%, 6% and 0.91, respectively; whereas, for the BRATS-2012 dataset, the corresponding assessment outcomes are 88.09%, 6% and 0.88, respectively. Another learning based method was proposed in [Soltaninejad et al. \(2017b\)](#) which used both machine-learned and hand-crafted features for automated segmentation of brain tumor from the multi-modal MRI images.

In [Soltaninejad et al. \(2018\)](#), the authors introduced a 3D supervoxel-based segmentation method for tumor segmentation in multi-modal MRI brain images. They are extracted several feature sets for each supervoxel, including first order intensity statistical features, histograms of texon descriptor to classify each supervoxel into the following three classes tumor core, oedema or healthy brain tissue by using random forest (RF) classifier. The method was evaluated based on two datasets such as (i) local clinical dataset: 11 images of patients and (ii) BRATS 2013 clinical dataset: 30 multimodal images. The experimental results shown that the average detection sensitivity of tumor using local clinical dataset is 86% with balanced error rate (BER) 7% and the Dice score for automatic tumor segmentation is 0.84. The resembling results of the BRATS 2013 dataset are 96%, 2% and 0.89, respectively. In [Zhang et al. \(2021\)](#), the authors proposed a Multi-Encoder Network (ME-Net) architecture with a new loss function named "Categorical Dice" for segmenting the 3D MRI image that reduces the feature extraction difficulty. They also used different weights for different segmented regions of the 3D MRI image to resolving the voxel imbalance issue. However, the segmentation results of this model are not sufficient for the enhancing tumor area of the tumor and multi-class segmentation problem.

3. Materials and methods

An overview of tumor detection using our proposed scheme is presented in this section (see [Fig. 1](#)). The MRI image is either colorful or gray-scale. Preprocessing of MR image plays vital role for getting better result, reducing noise and distortions. Afterward, the segmentation is the most crucial part that is called the heart part of detection ([Byale et al., 2018](#); [Sharma et al., 2014](#)).

After that, it extracts the essential features based on both superpixels and PCA to decide the segment of the brain where the abnormality exists. Furthermore, the processing is required for getting the estimations of mean deviation, variance, co-variance, eigenvalues, intensity, etc. of the images that will be helpful for the detection scheme. Finally, the proposed algorithm is utilized for the segmentation of brain MR images to detect brain tumors.

3.1. Dataset

Dataset is a collection of instances that has a great significance in research field. Many researchers use different types of 2D & 3D image dataset e.g., open access data and BRATS dataset ([Bakas et al., 2018](#)) to train and evaluate their new systems. In this study, we first consider that the MRI scan images of given patients are with a default size of 256×256 , either colorful or grayscale intensity images. If it is a color image, a big matrix with numerical input values between 0 and 255 is used to measure the converted grayscale image, where 0 is black and 255 is white for illustrations. To create our dataset, we have taken some critical images from [kaggle.com \(2018\)](#) open access data, and some images are taken locally ([Alam et al., 2019](#)).

3.2. Preprocessing

In preprocessing, our aim is the enhancement of images that works well for the next step of the detection scheme. After the transformation of color images (RGB) into a grayscale image, additional noises are dispelled using a filter. There are two common types of filters namely mean filter and median filter are used to reduce noises from an image. On one hand, the mean filter uses the mean values of all pixels instead of the central values of an image. On the other hand, the median filter uses the median values of all the pixels instead of the central values of an image. Images are processed successfully after removing unnecessary noise or artifacts ([Hebli & Gupta, 2017](#); [Kapoor & Thakur, 2017](#)). The most typical technique is known as "non-linear" e.g., median filter, filtering technique used for noise elimination ([Rathi & Palani, 2015](#)). The "non-linear" filter is a filter whose output is not linear with the inputs.

Moreover, median filter utilized the median estimation where mean filter depends on the mean estimation of pixels. The most points of interest of the median filter are less multifaceted nature and time utilization when contrasted with the mean filter ([Gopal & Karnan, 2010](#)). In our proposed scheme, we used a median filter for less computation complexity and better smoothing of images. Additionally, it works better in protecting subtlety inside the image than the mean filter. The median filter has two fundamental points of interest over the mean filter ([Srivastava et al., 2009](#)):

- It is a more powerful assessment than the mean. One unrepresentative pixel during a neighborhood will not affect the median significantly.
- Since the median is the value of one of the pixels within the neighborhood, it does not generate new ridiculous pixel values.

3.3. Feature extraction

Feature extraction is the way toward changing the input information into a bunch of features. It is the main advance in the development of any pattern classification, and the objective of extraction to separate relevant data that portrays each class ([Seerha & Kaur, 2013](#)). In this scheme, applicable features are extracted from objects to form features vector. These feature vectors are then utilized by classifiers to perceive the information unit with the objective yield unit. Moreover, GLCM is the factual technique for inspecting the surfaces that think about the spatial relationship of the pixels ([Zulpe & Pawar, 2012](#)). Additionally, large-scale feature selection is more critical for many datasets. Dataset may contains redundant and more irrelevant feature with more local optima. Thus, some existing evolutionary computation (EC) schemes can be used but they are better for small-scale feature selection. Furthermore, they provide a low performance for large-scale feature selection with huge solution size. To overcome this problem, a prominent technique such as the self-adaptive particle swarm optimization (SaPSO) algorithm can be used ([Xue et al., 2019](#)). In addition, the performance of SaPSO is better with smaller solution size. Another approach, self-adaptive parameter and strategy based particle swarm optimization (SPS-PSO) algorithm is used for large-scale feature selection when multiple classifier exists ([Xue et al., 2020](#)). Feature extraction is useful in recognizing a brain tumor where is actually found and helps in foreseeing the class of tumor. In our scheme, we extracted features of image using both superpixels and PCA.

3.3.1. Features extraction using superpixels

Superpixels are the result of the perceptual grouping of pixels by considering the similarity in close proximity and sharing intensity which is different from watersheds and maximally stable extremal regions(MSER). SLIC algorithm is powered by feature extraction with its fast computational time ([Achanta et al., 2012](#); [Tian et al., 2017](#); [Wu et al., 2012](#)). It represents meaningful descriptions with far fewer data

of an image. By reducing the number of primitives and redundancy it reduces the complexity of recognizing the desired area. Generally, this technique is based on measures that look for the same color and measures the region's shape. In addition, it incorporates the edges in intensity to describe regions. It extracts features in two approach like:

- Graph-based methods
- Region growing or clustering methods.

In accordance, the distance measurement between pixel p and the l th superpixel seed consists of three terms like boundary adherence, homogeneous intensity, and compactness (COM). In mathematically,

$$D(p, l) = w_b \times B(p, l) + w_i \times I(p, l) + \alpha w_c \times C(p, l) \quad (1)$$

where, w_b , w_i , and w_c represents the weight functions that are designed for different characteristics during the optimization procedure. Also, $B(p, l)$, $I(p, l)$, $C(p, l)$ denotes the boundary adherence matrix, homo intensity matrix, COM matrix, respectively and α denote a parameter to control flexibly of COM for resulting superpixels. These weight functions are defined as follows:

$$w_b = \frac{I(p, l) + C(p, l)}{A} \quad (2)$$

$$w_i = \frac{B(p, l) + C(p, l)}{A} \quad (3)$$

$$w_b = \frac{I(p, l) + B(p, l)}{A} \quad (4)$$

$$A = 2(I(p, l) + B(p, l) + C(p, l)) \quad (5)$$

3.3.2. Features extraction using PCA

PCA is a method of feature extraction to find the eigenvectors of a covariance matrix with the highest eigenvalues that are used for dimensionality reduction of large datasets. It identifies the similarities and differences from a large dataset to recognize the tumor area obtaining the maximum variance of data. The algebraic definition of PCA is as follows:

- For data framework D , compute the mean of D as follows:

$$\mu = E(D) \quad (6)$$

And then, calculate the covariance of D as follows:

$$CV = Cov(D) = E[(D - \mu)(D - \mu)^T] \quad (7)$$

- Count the eigenvalue λ_i and eigenvector $e_1, e_2, \dots, e_N, i = 1, 2, \dots, N$ of the covariance CV , sorting the eigenvalue in descending order; For the covariance CV , solving the equation as follows:

$$\delta \lambda I - CV \delta = 0 \quad (8)$$

where I denote the identity matrix as the similar dimension of CV . By using singular value decomposition (SVD), gain the eigenvalue, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0$ and respectively eigenvector $e - k (k = 1, 2, \dots, M)$. Where Λ represents the diagonal matrix of eigenvalues of $D^T D$. In order to select the λ_k for getting principal components, we count the proportion of data covered by the first M eigenvalues. Weight matrix W_m whose columns are eigenvectors of $D^T D$ is calculated as follows:

$$W_m = \frac{\sum_{i=1}^M \lambda_i}{\sum_{i=1}^N \lambda_i} \quad (9)$$

We select the primary M eigenvalue which made the combined extent bigger as 85% as the principal segments.

- Extended the information to a lower measurement subspace

$$P = W^T X \quad (10)$$

where T denotes the transpose matrix and X denotes the scaled version of original data. By utilizing the primary M individually eigenvector, we can lessen the number of factors or measurements from n to M ($M < n$).

Fig. 1. Flowchart of proposed methodology for the detection of tumor.

3.4. Segmentation and tumor detection

Segmentation of images is more important for image processing and detection of an object in the MRI image. Image segmentation refers to the segregation of a given image into various non-covering locales. Segmentation represents the partitioning image into different portions that are more enormous and less complex for analysis (Janani & Meena, 2013; Mehrotra et al., 2020). It is applied to approximately locate the boundaries of objects for segmenting; the resulting segments collectively cover the whole segmentation. The segmentation algorithms work on one in all the two characteristics of image intensity; similarity and discontinuity (Acharya et al., 2013). However, there are available methods widely used for image segmentation as follows random forest method, region growing, FCM, thresholding, wavelet transform, SVM and K-means clustering (Husein & Mahmoud, 2012; Mathew & Anto, 2017; Singh et al., 2015; Tang et al., 2015). Clustering methods are the most promising technologies that run on selected clusters (Panda & Patra, 2008). In our proposed scheme, we used the TK-means clustering algorithm for the segmentation with $K=8$ cluster to segment the image and efficiently detected the tumor from MR image.

3.4.1. Template based K-means clustering algorithm

The template-based K-means clustering algorithm is chosen instead of the K-means clustering algorithm for segmentation of MR images. Since a template in the template-based K-means clustering algorithm is selected, based on segmentation is performed that provides a better and robust result. The proposed template-based K-means clustering algorithm operates like a traditional K-means algorithm based on the desired clustering template. Like K-means, template-based K-means clustering is a simple unsupervised learning algorithm. It offers a simple way to classify a given dataset into a certain number of clusters i.e. a set of data such as x_1, x_2, \dots, x_n are grouped into K clusters. Firstly, the concept behind this algorithm is to define K clusters with c centers; secondly, to pick a template-based window; and finally, to calculate the template based on their desired scheme (Alsabti et al., 1997; Kanungo et al., 2002). Centers for the K cluster are chosen randomly. The calculation of distance plays a vital role in this algorithm's success. For this algorithm, various distance measurement techniques revach

However, the choice of a proper distance calculation technique depends entirely on the type of data that we are going to cluster (Loohach & Garg, 2012; Singh et al., 2013). In our proposed scheme, the Euclidean distance technique is used to calculate distance metrics. This technique provides more computational efficiency. Furthermore, it is best suited for multidimensional vectors..

Step by step template based K -means clustering Algorithm 1 is described as follows:

Algorithm 1 Template based K -means clustering algorithm for tumor detection .

Assume that the set of data points is $X = x_1, x_2, x_3, \dots, x_n$, and the set of centers is $V = v_1, v_2, v_3, \dots, v_c$.

- 1: Define cluster number ' K '.
- 2: Defining cluster centers arbitrarily ' c '.
- 3: Select the template based window T_{mn} for clustering

$$T_{mn} = \prod_{i=1}^M \prod_{j=1}^N P(x_i, y_j) \times \prod_{k=1}^U \prod_{l=1}^V P(x_k, y_l); \quad k \in M, l \in N \quad (11)$$

where $P(x_i, y_j)$ represents the image matrix with M row and N column and $P(x_k, y_l)$ represents the temper dependent image with gray level intensity numbers U and V number of bins

- 4: Calculate the desired template as follows:

$$B(x_i, y_j) = \prod_{i=i+1}^M \prod_{j=j+1}^N P(x_i, y_j) \times T_{mn} \quad (12)$$

- 5: Calculate the distance between the centers of the cluster and each data point.
- 6: Assign the data points to the cluster whose distance is minimum from the center of the cluster among all of the cluster center.
- 7: We then compute the new center of the cluster as follows:

$$V_i = \frac{1}{c_i} \sum_{j=1}^{c_i} x_i \quad (13)$$

where c_i represents data points number of i^{th} cluster.

- 8: It recalculates the distance between each data point and the newly acquired cluster centers.
- 9: If data points are not available to reassign to the cluster then exit, otherwise goto steps from 3 to 8.

The template is chosen along with the traditional K -means algorithm, based on temperature or gray level intensity in the brain image. The Eq. (14) express the segmentation of MR image based on template based K -means clustering algorithm as follows:

$$J = \prod_{i=i+1}^M \prod_{j=j+1}^N B(x_i, y_j) \times \prod_{i=1}^K \prod_{j=1}^C P_{ij} \bar{n} x_i - c_j \bar{n}^2 \quad (14)$$

where P_{ij} represents a binary image matrix with M rows, N columns; $B(x_i, y_j)$ is a coarse image. Also, K and C represents the number of data points and the number of clusters respectively and $||\cdot||$ is L_2 norm of P_{ij} . The desired template is described in Eq. (15) .

$$B(x_i, y_j) = \prod_{i=i+1}^M \prod_{j=j+1}^N P(x_i, y_j) \times T_{mn} \quad (15)$$

The selected template based window T_{mn} is calculated as follows:

$$T_{mn} = \prod_{i=1}^M \prod_{j=1}^N P(x_i, y_j) \times \prod_{k=1}^U \prod_{l=1}^V P(x_k, y_l); \quad k \in M, l \in N \quad (16)$$

In Eq. (16), T_{mn} represents the temper based image matrix with gray level intensity U and V bins that are used to determine the temper of rough image matrix $P(x_i, y_j)$. The aim of the image and convolution of the temper-based image matrix T_{mn} is to achieve the K -means clustering algorithm template.

Fig. 2. Database of brain tumor consists of 40 MR images.

3.4.2. Morphological operation

Morphological operation are sets of important functional methods that are used for extracting image component based on region of shape or features in image (Natarajan et al., 2012; Zhang et al., 2016). In the morphological operation, every pixel of the image is changed dependent on the estimation of other neighborhood pixels. There are two types of morphological operation namely dilation and erosion used in image processing (Sharma et al., 2012, 2018). In our scheme, erosion and dilation method are used for extracting tumor from segmented image. Moreover, it performs based on the feature that are extracted using PCA. In addition, it is much better because it can use any structure elements of shape and more fast.

4. Results and discussion

The experimental results and the associated discussion are presented in this section. We created our database by collecting more complex MR images from kaggle.com (2018) and (Alam et al., 2019). Our database contains 40 MR images shown in Fig. 2. The image of the database contains tumor is so critical that it is too hard to detect it by the common people so easily. These images need to pre-process for a better result in our algorithmic application. We evaluate the performance of our proposed algorithm to detect brain tumor in MR images through simulation experiments. The simulation is executed using MATLAB 2016a with Intel Core i5-8250U CPU processor and 8 GB of RAM.

The input images relate to brain tumors are taken from the database is prepared using the median filter. Superpixels based and PCA based features are extracted that helped in segmentation to get a better and robust result. Moreover, the proposed scheme will be fruitful for the multi-sequence MRI data which has different properties based on acquisition parameter such as pulse sequence and contrast mechanisms. Multimodal magnetic resonance imaging (mMRI) is used to detect the heterogeneous patterns gliomas tumors (Raschke et al., 2019). In the case of mMRI data, superpixels and PCA algorithm perform an important role for extracting metabolic information which leads to accurate classification. Furthermore, our proposed TK-means enable to create $K=8$ cluster so that it can easily detect heterogeneous patterns brain tumor. In the simulation, the TK-means algorithm used for

Fig. 3. Experimental result of tumor detection for MR image number 25 in database.

segmentation and separated the tumor region from other parts of the image. Besides, morphological operation extracts tumor portion only with the background image. Based on segmentation and morphological operations, affected tumor area is detected and marked by red color.

Fig. 3 shows the overall tumor detection procedure of our proposed scheme for the MR image number 25 of the database shown in Fig. 2. In Fig. 3(a), the input image taken from the database contains blare and artifacts. In Fig. 3(b), the input image is filtered by the median filter before image conversion. In Fig. 3(c), the image is converted to a grayscale shadow image for further processing. In Fig. 3(d), the image quality also enriched so that the image segmented correctly. In Fig. 3(e), the image is segmented according to the template-based K-means algorithm. Finally, in Fig. 3(f), the tumor is detected from the segmented image and marked by red color.

Fig. 4 shows the simulation results for the proposed scheme of the MR image number 8, 3, 4, 30, 18, and 25 in the database shown in Fig. 2. In Fig. 4(a), the input MR image selected from the database for the detection of a tumor. Then, the median filter and histogram filter used for image enhancement. The enhancement image is important for the next step of image processing. The enhancement image of input MR image shown in Fig. 4(b). Segmentation is the heart of tumor detection in MR images. Fig. 4(c) shows the segmented image done by the template-based K-means clustering algorithm. Finally, Fig. 4(d) shows the detected tumor marked by red color.

4.1. Inference and forecasting

The MR image pre-processing and feature extraction are crucial for perfect image segmentation. The MR images are transmitted through a subsystem for pre-processing and feature extraction. Then the region of interest (ROI) is segmented by using the TK-means clustering method with $k = 8$. Then the suggested scheme is enforced to predict the brain tumor. The accuracy of tumor detection from MR images of our proposed algorithm is much better than other existing methods.

4.2. System performance

The tumor detection efficiency of our proposed algorithm is established based on sensitivity, specificity, and accuracy. The values of sensitivity, specificity, and accuracy are determined based on the true positive value, true negative value, false positive value and false negative value. In this paper, 40 MR images are used to evaluate the performance of our proposed superpixels, PCA & TK-means scheme.

Fig. 4. (a) Input MR images (b) Enhancement of filtered image (c) Segmented image by template based K-means Algorithm (d) Detected tumor marked by red color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To validate and testing of our proposed scheme, we compared the proposed scheme with some existing schemes like Thresholding, Region Growing, FCM, etc. that had been performed based on another 2D MR dataset. The detection performance of our proposed scheme and other existing scheme are presented in Table 2.

- TP (True Positive): The actual observation indicates that there is a tumor and the experiment detects the brain tumor from the MR image (i.e., the detection result is positive).
- TN (True Negative): The actual observation indicates that there is a tumor whereas the experiment cannot detect the brain tumor from the MR image (i.e., the detection result is negative).
- FP (False Positive): The actual observation indicates that there is no tumor and the experiment result indicates that there is no tumor in the MR image (i.e., the detection result is positive).
- FN (False Negative): The actual observation indicates that there is no tumor whereas the experiment detects the brain tumor from the MR image (i.e., the detection result is negative).

Sensitivity is the true positive rate. It measures how frequently the experiment detects the tumor from the MRI image when the MRI image really contains a tumor.

It is the ability to test correctly classify, and individual are affected by diseases. The sensitivity, α is defined as follows:

$$\alpha = \frac{TP}{TP + TN} \times 100 \quad (17)$$

Specificity is the ability of an experiment to correctly classify the MRI image that does not contain a tumor.

Table 1

The performance analysis among existing scheme (Alam et al., 2019) and proposed Superpixels, PCA & TK-means scheme.

Algorithms	TP	FP	TN	FN
Thresholding	22	5	9	4
Region growing	27	2	10	1
FCM	24	1	10	5
Second order + ANN	21	2	16	1
Texture combined + ANN	21	1	17	1
Proposed method	37	0	1	2

Table 2

Comparisons among the existing scheme (Alam et al., 2019) and proposed Superpixels, PCA & TK-means scheme.

Algorithms	α	β	η
Thresholding	76.9	64.25	72.5
Region growing	92.8	83.3	90.0
FCM	82.7	90.9	85.0
Second order + ANN	95.5	88.8	92.5
Texture combined + ANN	95.45	94.44	95.0
Proposed method	97.36	100	95.0

Table 3

Comparison of computational time t of the existing scheme (Alam et al., 2019) and the proposed Superpixels, PCA & TK-means scheme.

Techniques	Computational time
Thresholding	3 min
Region growing	10 min
FCM	130 140 s
Second order + ANN	7 15 min
Texture combined + ANN	2 min
Proposed method	35 60 s

It is the ability to test correctly classify, and individuals are diseases free that means not affected by diseases. The specificity, β is defined as follows:

$$\beta = \frac{TN}{TN + FP} \times 100 \quad (18)$$

Accuracy is defined as the measurement of actual classification. It is the ability to test correctly classification of diseases. The accuracy, η is defined as follows:

$$\eta = \frac{\text{No of Correctly Classified Record}}{\text{Total Record in the test set}} \times 100 \quad (19)$$

$$= \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

The performance metrics is shown in Table 1 for the proposed scheme and other existing schemes. The performance analysis is done by the analyzing of performance parameter TP, TN, FP and FN. Table 2 demonstrates the comparison among existing scheme and our proposed scheme for the detection of tumors. In Table 2, three attribute named as sensitivity, specificity, and accuracy play a vital role to measure the performance in the detection of brain tumor from MR images. From Table 2, we observed that the proposed superpixels, PCA & TK-means schemes achieved a better detection accuracy when compared to other existing detection scheme. Very few of them schemes achieved equal or better accuracy over our proposed scheme while different specification are not so good. Table 3 signifies the computational time to detect brain tumor from the MR image for our proposed scheme and other existing schemes. From Table 3, we noticed that most of the algorithms have taken a long time when compared to the proposed superpixels, PCA & TK-means scheme of 35 60 s. Therefore, the proposed superpixels, PCA & TK-means scheme will be more suitable for the detection of tumor and applicable in future medical research. Furthermore, we can compare our proposed method with the recently proposed attention mechanism (Jin et al., 2021; Liu et al., 2020, 2019; Wu et al., 2021; Yang et al., 2020; Zhou et al., 2020) when our method is built on a deep learning system with 3D images.

5. Conclusions and future work

In this paper, the achievement of the proposed superpixels, PCA & TK-means scheme is better than other existing detection schemes. In brain tumor detection, the proposed scheme achieved an accuracy of 95.0%, the sensitivity of 97.36%, and the specificity of 100% which are better when compared to other existing detection schemes. Moreover, superpixels and PCA are played a vital role in feature extraction that reduced the dimensions and complexity of the MR images. As a result, the brain tumors are detected accurately from MR images using our proposed scheme in a short execution time of 35 60 s.

The drawback of this study, it is implemented on small dataset and does not use the 2019 WHO international classification of diseases (ICD-11) system with real clinical MRI data. In the future, we will analyze the real clinical MR image with high dimension features extraction to improve the detection scheme. Furthermore, we will enhance the accuracy of detection and classification with stage identification of tumor. In addition, we will make the scheme more compatible based on a deep learning system which can be applied multi sequential BRATS 3D images as well as other radio imaging techniques such as MRI, CT, PET, and SPECT.

- Borole, V. Y., Nimbhore, S. S., & Kawthekar, D. S. S. (2015). Image processing techniques for brain tumor detection: A review. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCs)*, 4(2), 1–14.
- Byale, H., Lingaraju, G., & Sivasubramanian, S. (2018). Automatic segmentation and classification of brain tumor using machine learning techniques. *International Journal of Applied Engineering Research*, 13(14), 11686–11692.
- cancer. net (2020). <https://www.cancer.net/cancer-types/brain-tumor/statistics>, Brain Tumor: Statistics.
- Chavan, N. V., Jadhav, B., & Patil, P. (2015). Detection and classification of brain tumors. *International Journal of Computer Applications*, 112(8).
- Chithra, P., & Dheepa, G. (2018). An analysis of segmenting and classifying tumor regions in MRI images using CNN. *International Journal of Pure and Applied Mathematics*, 118(24), 2018.
- Coatrieux, G., Huang, H., Shu, H., Luo, L., & Roux, C. (2013). A watermarking-based medical image integrity control system and an image moment signature for tampering characterization. *IEEE Journal of Biomedical and Health Informatics*, 17(6), 1057–1067.
- Gondal, A. H., & Khan, M. N. A. (2013). A review of fully automated techniques for brain tumor detection from MR images. *International Journal of Modern Education and Computer Science*, 5(2), 55.
- Gopal, N. N., & Karnan, M. (2010). Diagnose brain tumor through MRI using image processing clustering algorithms such as Fuzzy C Means along with intelligent optimization techniques. In *2010 IEEE international conference on computational intelligence and computing research* (pp. 1–4). IEEE.
- Hebli, A., & Gupta, S. (2017). Brain tumor prediction and classification using support vector machine. In *2017 international conference on advances in computing, communication and control (ICAC3)* (pp. 1–6). IEEE.
- Husein, E. M., & Mahmoud, D. M. A. (2012). *Brain tumor detection using artificial neural networks*. Sudan University of Science and Technology.
- Islam, M. K., Ali, M., Das, A., Duranta, D. U., & Alam, M. (2020). Human brain tumor detection using K-means segmentation and improved support vector machine. 11, (p. 6).
- Janani, V., & Meena, P. (2013). Image segmentation for tumor detection using fuzzy inference system. *International Journal of Computer Science and Mobile Computing*, 2(5), 244–248.
- Jin, Y., Yang, G., Fang, Y., Li, R., Xu, X., Liu, Y., & Lai, X. (2021). 3D PBV-net: an automated prostate MRI data segmentation method. *Computers in Biology and Medicine*, 128, Article 104160.
- kaggle. com (2018). <https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection>, kaggle datasets.
- Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 881–892.
- Kapoor, L., & Thakur, S. (2017). A survey on brain tumor detection using image processing techniques. In *2017 7th international conference on cloud computing, data science & engineering-confluence* (pp. 582–585). IEEE.
- Liu, J., Li, M., Wang, J., Wu, F., Liu, T., & Pan, Y. (2014). A survey of MRI-based brain tumor segmentation methods. *Tsinghua Science and Technology*, 19(6), 578–595.
- Liu, Y., Yang, G., Hosseiny, M., Azadikhah, A., Mirak, S. A., Miao, Q., Raman, S. S., & Sung, K. (2020). Exploring uncertainty measures in Bayesian deep attentive neural networks for prostate zonal segmentation. *IEEE Access*, 8, 151817–151828.
- Liu, Y., Yang, G., Mirak, S. A., Hosseiny, M., Azadikhah, A., Zhong, X., Reiter, R. E., Lee, Y., Raman, S. S., & Sung, K. (2019). Automatic prostate zonal segmentation using fully convolutional network with feature pyramid attention. *IEEE Access*, 7, 163626–163632.
- Loohach, R., & Garg, K. (2012). Effect of distance functions on k-means clustering algorithm. *International Journal of Computer Applications*, 49(6), 7–9.
- Mathew, A. R., & Anto, P. B. (2017). Tumor detection and classification of MRI brain image using wavelet transform and SVM. In *2017 international conference on signal processing and communication (ICSPC)* (pp. 75–78). IEEE.
- Mehrotra, R., Ansari, M., Agrawal, R., & Anand, R. (2020). A transfer learning approach for AI-based classification of brain tumors. *Machine Learning with Applications*, 2, Article 100003.
- Natarajan, P., Krishnan, N., Kenkre, N. S., Nancy, S., & Singh, B. P. (2012). Tumor detection using threshold operation in MRI brain images. In *2012 IEEE international conference on computational intelligence and computing research* (pp. 1–4). IEEE.
- Panda, M., & Patra, M. R. (2008). Some clustering algorithms to enhance the performance of the network intrusion detection system. *Estudios de Economia Aplicada*, 26(2).
- Rai, H. M., & Chatterjee, K. (2020). Detection of brain abnormality by a novel Lu-Net deep neural CNN model from MR images. *Machine Learning with Applications*, 2, Article 100004.
- Raschke, F., Barrick, T. R., Jones, T. L., Yang, G., Ye, X., & Howe, F. A. (2019). Tissue-type mapping of gliomas. *NeuroImage: Clinical*, 21, Article 101648.
- Rathi, V. G. P., & Palani, S. (2015). Brain tumor detection and classification using deep learning classifier on MRI images. *Research Journal of Applied Sciences, Engineering and Technology*, 10(2), 177–187.
- Sachdeva, J., Kumar, V., Gupta, I., Khandelwal, N., & Ahuja, C. K. (2013). Segmentation, feature extraction, and multiclass brain tumor classification. *Journal of Digital Imaging*, 26(6), 1141–1150.
- Sandhya, G., Giri, K., & Savitri, S. (2017). A novel approach for the detection of tumor in MR images of the brain and its classification via independent component analysis and kernel support vector machine. *Imaging in Medicine*.
- Seerha, G. K., & Kaur, R. (2013). Review on recent image segmentation techniques. *International Journal on Computer Science and Engineering*, 5(2), 109.
- Sharma, P., Diwakar, M., & Choudhary, S. (2012). Application of edge detection for brain tumor detection. *International Journal of Computer Applications*, 58(16).
- Sharma, K., Kaur, A., & Gujral, S. (2014). Brain tumor detection based on machine learning algorithms. *International Journal of Computer Applications*, 103(1), 7–11.
- Sharma, M., & Mukharjee, S. (2013). Brain tumor segmentation using genetic algorithm and artificial neural network fuzzy inference system (ANFIS). In *Advances in computing and information technology* (pp. 329–339). Springer.
- Sharma, M., Purohit, G., & Mukherjee, S. (2018). Information retrieves from brain MRI images for tumor detection using hybrid technique K-means and artificial neural network (KMANN). In *Networking communication and data knowledge engineering* (pp. 145–157). Springer.
- Shree, N. V., & Kumar, T. (2018). Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network. *Brain Informatics*, 5(1), 23–30.
- Singh, A., Yadav, A., & Rana, A. (2013). K-means with three different distance metrics. *International Journal of Computer Applications*, 67(10).
- Singh, A., et al. (2015). Detection of brain tumor in MRI images, using combination of fuzzy c-means and SVM. In *2015 2nd international conference on signal processing and integrated networks (SPIN)* (pp. 98–102). IEEE.
- Sinha, K., & Sinha, G. (2014). Efficient segmentation methods for tumor detection in MRI images. In *2014 IEEE students' conference on electrical, electronics and computer science* (pp. 1–6). IEEE.
- Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T. L., Barrick, T. R., Howe, F. A., & Ye, X. (2017a). Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI. *International Journal of Computer Assisted Radiology and Surgery*, 12(2), 183–203.
- Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T. L., Barrick, T. R., Howe, F. A., & Ye, X. (2018). Supervised learning based multimodal MRI brain tumour segmentation using texture features from supervoxels. *Computer Methods and Programs in Biomedicine*, 157, 69–84.
- Soltaninejad, M., Zhang, L., Lambrou, T., Yang, G., Allinson, N., & Ye, X. (2017b). MRI brain tumor segmentation and patient survival prediction using random forests and fully convolutional networks. In *International MICCAI brainlesion workshop* (pp. 204–215). Springer.
- Srivastava, H. B., Kumar, V., Verma, H., & Sundaram, S. (2009). Image pre-processing algorithms for detection of small/point airborne targets. *Defence Science Journal*, 59(2).
- Tang, H., Lu, H., Liu, W., & Tao, X. (2015). Tumor segmentation from single contrast MR images of human brain. In *2015 IEEE 12th international symposium on biomedical imaging (ISBI)* (pp. 46–49). IEEE.
- Tian, Z., Liu, L., Zhang, Z., Xue, J., & Fei, B. (2017). A supervoxel-based segmentation method for prostate MR images. *Medical Physics*, 44(2), 558–569.
- Uddin, M. P., Mamun, M. A., & Hossain, M. A. (2019). Effective feature extraction through segmentation-based folded-PCA for hyperspectral image classification. *International Journal of Remote Sensing*, 40(18), 7190–7220.
- Wang, M., Liu, X., Gao, Y., Ma, X., & Soomro, N. Q. (2017). Superpixel segmentation: A benchmark. *Signal Processing: Image Communication*, 56, 28–39.
- Wu, Y., Hatipoglu, S., Alonso-Álvarez, D., Gatehouse, P., Li, B., Gao, Y., Firmin, D., Keegan, J., & Yang, G. (2021). Fast and automated segmentation for the three-directional multi-slice cine myocardial velocity mapping. *Diagnostics*, 11(2), 346.
- Wu, Z., Hu, Z., & Fan, Q. (2012). Superpixel-based unsupervised change detection using multi-dimensional change vector analysis and SVM-based classification. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 7, 257–262.
- Xue, Y., Tang, T., Pang, W., & Liu, A. X. (2020). Self-adaptive parameter and strategy based particle swarm optimization for large-scale feature selection problems with multiple classifiers. *Applied Soft Computing*, 88, Article 106031.
- Xue, Y., Xue, B., & Zhang, M. (2019). Self-adaptive particle swarm optimization for large-scale feature selection in classification. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 13(5), 1–27.
- Yang, G., Chen, J., Gao, Z., Li, S., Ni, H., Angelini, E., Wong, T., Mohiaddin, R., Nyktari, E., Wage, R., et al. (2020). Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention. *Future Generation Computer Systems*, 107, 215–228.
- Zhang, S., Xu, G., et al. (2016). A novel approach for brain tumor detection using MRI images. *Journal of Biomedical Science and Engineering*, 9(10), 44–52.
- Zhang, W., Yang, G., Huang, H., Yang, W., Xu, X., Liu, Y., & Lai, X. (2021). ME-net: Multi-encoder net framework for brain tumor segmentation. *International Journal of Imaging Systems and Technology*.
- Zhou, X., Ye, Q., Jiang, Y., Wang, M., Niu, Z., Menpes-Smith, W., Fang, E. F., Liu, Z., Xia, J., & Yang, G. (2020). Systematic and comprehensive automated ventricle segmentation on ventricle images of the elderly patients: A retrospective study. *Frontiers in Aging Neuroscience*, 12.
- Zulpe, N., & Pawar, V. (2012). GLCM Textural features for brain tumor classification. *International Journal of Computer Science Issues (IJCSI)*, 9(3), 354.