Send Orders for Reprints to reprints@benthamscience.net

Current Medical Imaging, **2021**, *Volume, Page Enation* **1**

REVIEW ARTICLE

Recent Advances in Classification of Brain Tumor from MR Images – State of the Art Review from 2017 to 2021

Ghazanfar Latif [a, b], Faisal Yousif Al Anezi [c], D.N.F. Awang Iskandar [d], Abul Bashar [a], and Jaafar Alghazo [e]

[a] *College of Computer Engineering and Sciences, Prince Mohammad bin Fahd University, Khobar, Saudi Arabia.*
[b] *Université du Québec a Chicoutimi, 555 boulevard de l'Université, Chicoutimi, QC, G7H2B1, Canada*
[c] *Management Information Department, Prince Mohammad bin Fahd University, Khobar, Saudi Arabia.*
[d] *Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Malaysia.*
[e] *Department of Electrical and Computer Engineering, Virginia Military Institute, Lexington, VA*
*Corresponding Author *: Ghazanfar Latif, Department of Computer Science, Prince Mohammad bin Fahd University, Al-Khobar, 31952, Saudi Arabia, Tel: +966-501057422, Email: glatif@pmu.edu.sa*

ARTICLE HISTORY

Received:
Revised:
Accepted:

DOI:

Abstract: ***Background:*** The task of identifying a tumor in the brain is a complex problem that requires sophisticated skills and inference mechanisms to accurately locate the tumor region. The complex nature of the brain tissue makes the problem of locating, segmenting, and ultimately classifying Magnetic Resonance (MR) images a complex problem. The aim of this review paper is to consolidate the details of the most relevant and recent approaches proposed in this domain for the binary and multi-class classification of brain tumors using brain MR images.

Objective: In this review paper, a detailed summary of the latest techniques used for brain MR image feature extraction and classification is presented. A lot of research papers have been published recently with various techniques proposed for identifying an efficient method for the correct recognition and diagnosis of brain MR images. The review paper allows researchers in the field to familiarize themselves with the latest developments and be able to propose novel techniques that have not yet been explored in this research domain. In addition, the review paper will facilitate researchers, who are new to machine learning algorithms for brain tumor recognition, to understand the basics of the field and pave the way for them to be able to contribute to this vital field of medical research.

Results: In this paper, the review is performed for all recently proposed methods for both feature extraction and classification. It also identifies the combination of feature extraction methods and classification methods that when combined would be the most efficient technique for the recognition and diagnosis of brain tumor from MR images. In addition, the paper presents the performance metrics particularly the recognition accuracy, of selected research published between 2017-2020.

Keywords: Brain Tumor Detection, Feature Extraction, Magnetic Resonance (MR) Image Classification, Convolutional Neural Networks, Deep Learning, Glioma Tumor Classification

1. INTRODUCTION

Identifying a brain tumor is a complex task that requires specialized skills and interpretation techniques to pinpoint the tumor's location. In addition, capturing the internal structure of the brain in high resolution with all relevant features is of utmost importance in this process. Doctors rely on mainly three essential imaging modalities in diagnosing brain tumors, namely Computed Tomography (CT), Positron Emission Tomography (PET), and Magnetic Resonance Imaging (MRI) [1]. MRI scans use powerful magnets, radio waves, and computing machines to capture a detailed image of the internal structure of the brain. The MRI scans provide better picture details, contrast, and brightness than other methods due to the tissue relaxation properties (T1 and T2) and are preferred by doctors for diagnosis [2]. The manual diagnosis of a brain tumor is still a slow and lengthy process. The MRI machine can produce different types of scans (T1, T2, T1c, and flair) based on the contrast and brightness value, repetition time, and time to echo [3]. The different types of MRI scans are widely used in various image processing

1573-4056/19 $58.00+.00

approaches for the diagnosis and classification of brain tumors by the process of feature extraction from images [4].

The task of classifying a brain MRI as either tumorous or non-tumorous is a challenging problem that is further complicated by accurately pinpointing the tumorous region. A number of methods are proposed in previous studies but they are still unable to provide an optimal solution. The classification approach from the Machine Learning (ML) domain seeks to identify and categorize the input dataset into various classes [5]. Classification approaches are broadly categorized into two types, namely, supervised classification and unsupervised classification. In supervised classification, a labeled sample data is used to train the classifier prior to testing, whereas unsupervised classification is used when labeled training datasets are not available. There are many unsupervised and supervised classification techniques that have been utilized recently for the classification of brain images into benign (non-tumorous) and malignant (tumorous) categories. It is difficult to compare brain MRI classification results claimed in different studies due to the unavailability of various small-sized datasets.

Figure 1 shows the steps and techniques used for brain tumor classification from MR images. Brain MR images are first taken through the preprocessing phases, which include noise reduction, image segmentation, image resizing, and augmentation. The next phase is feature extraction. Feature extraction can consist of various methods, including texture features, statistical features, and deep learning-based features. The features extraction then goes through a feature reduction phase to find the best features that would later be used in the classification stage. The feature reduction can be done using Principal Component Analysis (PCA), Autoencoders, and Linear Discriminant Analysis (LDA) [6]. The classification phase can be completed by using well-known classification techniques such as Convolutional Neural Networks (CNN), Support Vector Machines (SVM), Multilayer Perceptron (MLP), K-Nearest Neighbor (KNN), Random Forest, Genetic Algorithms, Naïve Bayes, and Ensemble methods. There are two options for the classification of the MRI brain Images. The first option is referred to as a Binary Classification that classifies the images into tumorous (malignant) and non-tumorous (benign) images. The second option is Multiclass Classification, which classifies the images into one of four modalities, namely, Necrosis, Edema, Enhancing, and Non-Enhancing [7]. The classification performance is then measured using various metrics such as Accuracy, Precision, Recall, F1 Score, Sensitivity, Specificity, and Confusion matrix.

The rest of this paper is organized as follows. Section 2 describes the structure of the brain tumor. Section 3 details the feature extraction techniques used for brain MRI images. Section 4 details the various classification techniques used for binary and multi-class classification of the brain MRI images. Section 5 compares the various existing results and their relevant discussion to the scope of our review work. Section 6 presents the key conclusions and pointers to possible future work.

2. STRUCTURE OF THE BRAIN TUMOR

Brain tumors (or brain cancer) occurs when different cells inside the brain start reproducing in an irregular manner. The author in [8], categorized this unusual growth into two types i.e., benign, and malignant tumors. Figure 2 depicts a sample brain MR image distinguishing the healthy and tumorous cells. The benign tumors are less harmful which remain isolated from the primary brain tissues. However, if not detected at an early stage, benign tumors can cause severe complications. The malignant tumors affect the core brain tissues and are very harmful which may even lead to death. Malignant brain tumors can grow fast and destroy primary brain tissues.

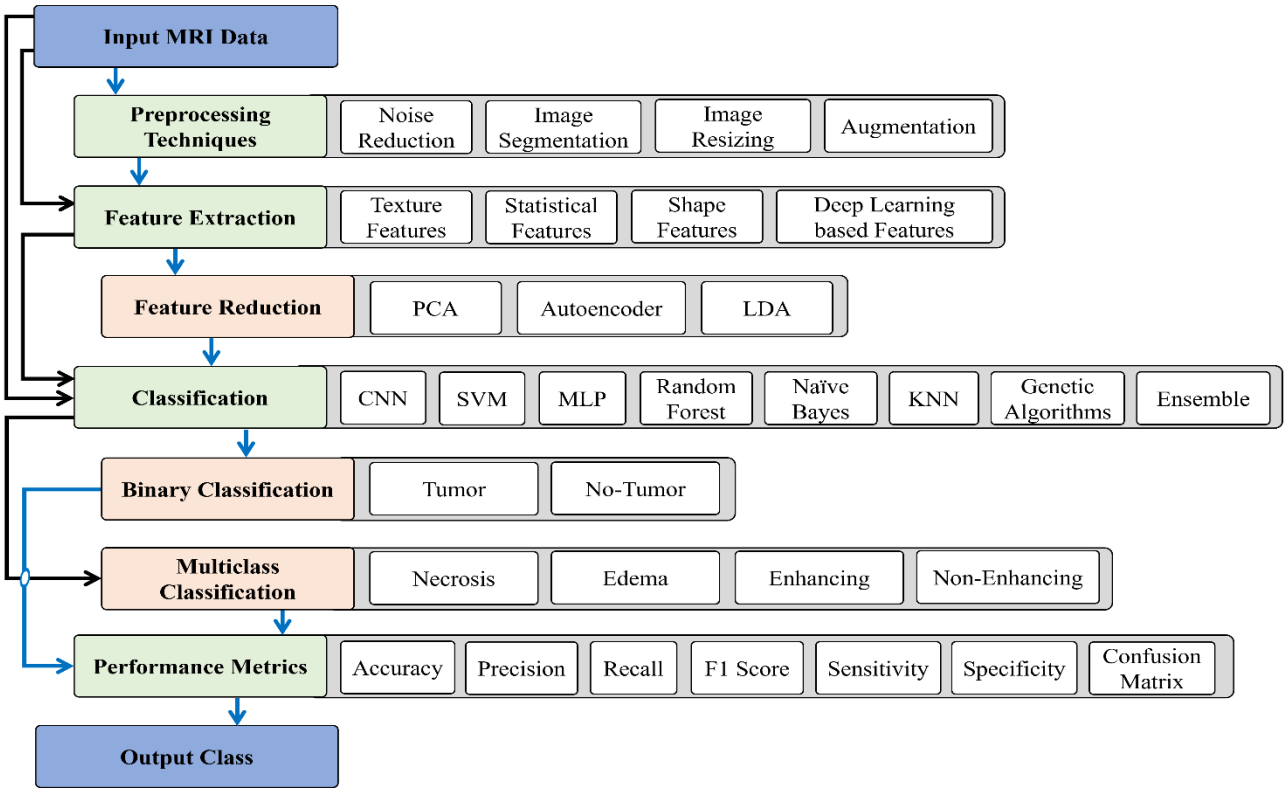

Figure 1. Overview of the steps and methods involved in brain tumor classification using MRI

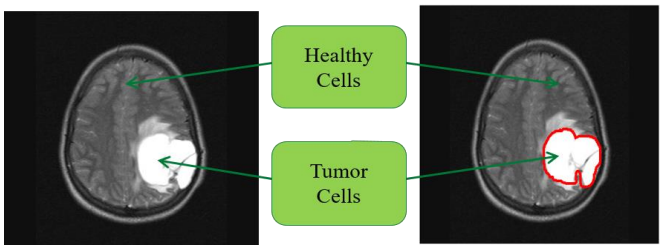

Figure 2. Brain MR image depicting healthy and tumorous brain cells.

MRI system works on the concept of passing radio waves through a magnetic field to visualize the organs and their structure by recording data on a computer system. MRI procedure can be used on the head to detect brain tumors, on the chest to detect breast or lung cancer. Also, it can be helpful in detecting irregularities in blood vessels and finding information about different soft tissues in the human body. It uses magnets that utilize the dipole moment of neutrons and protons to generate the internal structure of the human body. Figure 3 shows the samples of brain images captured from different imaging modalities. The modalities basically refer to the various views of the brain structure which the doctors usually observe for diagnosing the status of the patient's brain. It can be seen here that manual inspection of these various image modalities is a physically tiring task for the medical practitioners. Misdiagnosis may be possible even after the doctors and technicians go through these images. Another possibility is that two doctors can make different diagnosis based on the study of these images, as this decision is subjective in nature. Finally, this manual diagnosis process is time consuming. In order to address these issues faced during the manual diagnosis process, an automated, rapid, accurate and computer assisted process is recommended. One of the popular solutions is to resort to Machine Learning based image identification and classification of MR images. This solution requires a dataset of MR images which can be used to train a model using the image features and then utilize the trained model for making decisions. The process of feature extraction from images is described in detail in the following section.

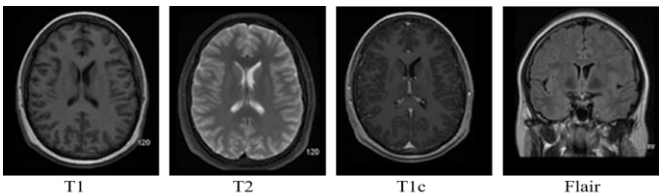

Figure 3. Different brain MR image modalities

3. FEATURE EXTRACTION TECHNIQUES FOR BRAIN TUMOR DETECTION

The accuracy of the classification techniques is dependent on the criteria used for feature selection. Numerous techniques related to feature extraction from MR images have been proposed by researchers, namely, Gabor features, wavelet transform-based features, discrete cosine features, discrete Fourier transform features, and texture features [9-10]. Since some of these methods produce a large set of features, Feature Reduction techniques have also been proposed in recent studies, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) [11-12].

3.1. Discrete Cosine Transform (DCT)

Image visual features from isolated partial images are used as the basis to extract textural features using DCT [13]. Feature vectors are composed on top low-frequency coefficient. Both DCT and DFT transform images from the spatial domain to the low-frequency domain. Two Dimensional DCT (2D-DCT) has a unique compaction property that makes it suitable for image and signal processing. Equation 1 defines N data points of a one-dimensional DCT, where F is the linear combination of the basis vectors.

$$F(\mu) = \left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_{i=0}^{N-1} A(i).\cos\left[\frac{\pi.\mu}{2.N}(2i+1)\right] f(i) \quad (1)$$

For each column and row of F a one-dimensional DCT is applied in order to produce a 2D-DCT of n x m as shown in Equation 2.

$$F(\mu,v) = \left(\frac{2}{N}\right)^{\frac{1}{2}}\left(\frac{2}{M}\right)^{\frac{1}{2}} \sum_{i=0}^{N-1}\sum_{j=0}^{M-1} \Delta(i).\Delta(j).\cos\left[\frac{\pi.\mu}{2.N}(2i+1)\right]\cos\left[\frac{\pi.\mu}{2.M}(2j+1)\right].f(i,j) \quad (2)$$

where $0 \le v \le N$, $0 \le u \le N$ and $F^{-1}(u,v)$ is the inverse of 2D-DCT transform based on Equation 3.

$$\Delta(\varepsilon) = \begin{cases} \frac{1}{\sqrt{2}} & for\ \varepsilon = 0 \\ 1 & otherwise \end{cases} \quad (3)$$

DCT features of MR images of the brain are computed using 2D-DCT. The natural logarithm is applied to each absolute value of the corresponding element. After applying the 2D-DCT to obtain the DCT features, the top left low-frequency coefficients are selected as the final features and are shown in Figure 4. In a representative research work, DCT features were used as an input to SVM classifier, which achieved 98.82% accuracy but only a dataset of 255 images was used in the experiments [14]. Similarly, in another research work, DCT and DWT combined features were used with the MLP classifier [15]. In order to achieve maximum classifier accuracy, the top 100 low frequencies of DCT and top 64 low frequencies are selected. The combined DCT features-based classification of MR images into tumorous and non-tumorous categories achieved 86.98% accuracy but the dataset used for the experiments comprised of only 3064 MR images from 233 patients.

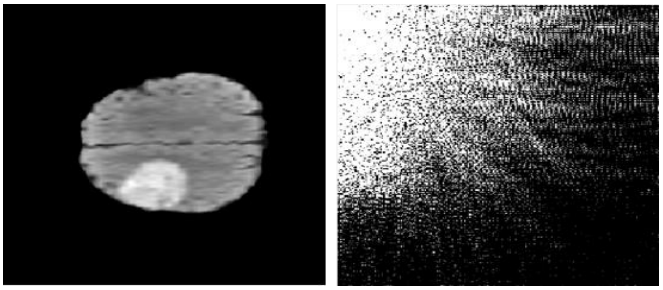

Figure 4. 2D-DCT of brain MR image (log abs). Original MRI (left image), and the 2D-DCT (right image) with low frequency coefficients visible on the top left corner.

3.2. Discrete Wavelet Transform (DWT)

In terms of cost and computation, DWT is considered a more efficient feature extraction process [16]. Wavelet is used in image processing to be implemented as a multi-resolution process. In this method, a third level decomposition is used to extract the wavelet features. During the classification phase, wavelet coefficients are used as feature vectors.

Continuous Wavelet Transform (CWT) is achieved by performing a summation of a finite energy image signal x(t) multiplied with its scaled version of the wavelet function (Ψ) as shown in Equation 4 [17].

$$WT_x(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\ \Psi^*\left(\frac{t-\tau}{a}\right) dt \quad (4)$$

$$WT_x(a,\tau) = < x(t), \varphi_{a\tau}(t) > \quad (5)$$

Positioning and scaling information such as the wavelet coefficients are contained in the function $WT_x(a,\tau)$, as given by Equation 5.

Discrete points and scale sets need to be selected in order to proceed with the DWT. DWT of the image $x(t)$ is calculated using a shifted version of the scaling function ($\emptyset_{j,k}$) and a shifted version of the wavelet function $\Psi_{j,k}$ as given in Equation 6.

$$x(t) = \sum_{k\in Z} u_{j_0,k}\emptyset_{j_0,k}(t) + \sum_{j=-\infty}^{j_0}\sum_{k\in Z} w_{j,k}\Psi_{j,k}(t) \quad (6)$$

where $w_{j,k}\ (j < j_0)$ are the wavelet coefficients and $u_{j,k}\ (j < j_0)$ are the scaling coefficients.

$$u_{j,k} = < x, \emptyset_{j,k} >,\ w_{j,k} = < x, \Psi_{j,k} > \quad (7)$$

Equation 8 is used to compute the family of scalar functions ($\emptyset_{j,k}(t)$) while Equation 9 is used to compute the family of wavelet functions ($\Psi_{j,k}(t)$).

$$\emptyset_{j,k}(t) = 2^{-j/2}\emptyset\left(2^{-j/2}t - k\right) \quad (8)$$

$$\Psi_{j,k}(t) = 2^{-j/2}\Psi\left(2^{-j/2}t - k\right) \quad (9)$$

Since the resulting image contains redundant data, it is a standard practice to refine it using feature extraction of the

complete signal space. By using this method, the data extracted will include only discriminatory features, which result in a significantly lower dimension feature vector. Thus, DWT with the successive use of high and low pass filters can extract characteristics at different levels. Wavelet coefficients are thus considered as a continuation of the detailed and approximation coefficients. Obtaining the detailed and approximation coefficients is performed by decomposing the signal into three levels with the help of Daubechies wavelet filtering [7]. The features can then be extracted from the resulting coefficients.

Function and numerical analyses in image processing are done using DWT, and they include signal processing and data compression. Feature vectors are the wavelet coefficients produced through analysis of image texture in image processing. Using linear transformation, an image is divided into various frequencies. This division produces what is referred to as sub-bands; detailed coefficients LL, HL, HH, and approximate coefficients LL. Figure 5 shows the approximate and detailed coefficients resulting from the 3 level DWT process. A representative research work proposed DWT feature-based classification of MR images into normal (tumorous) and abnormal (non-tumorous) and achieved 93.5% accuracy using SVM as a classifier but the dataset used for experiments is small consisting of only 100 MR images. The model proposed uses level 1 to 5 DWT decomposition to extract the brain MR image features [18].

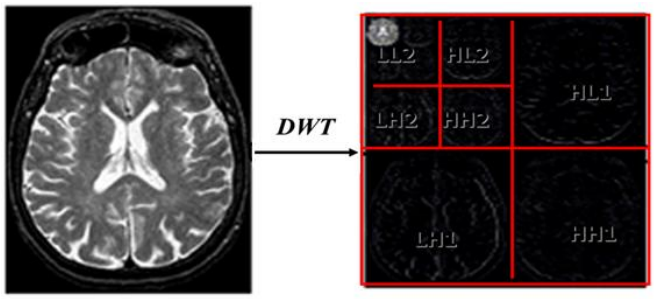

Figure 5. Schematic of DWT 3 level for brain MRI

3.3. Discrete Fourier Transform (DFT)

The Discrete Fourier Transform (DFT) method is used to transform the finite time-based data into discrete frequency-based data [19]. The dataset input to DFT results in a complex number within a specific range of frequencies. Equation 10 defines DFT that transforms complex time-domain numbers of size N into complex frequency domain numbers of size N.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{\frac{-i2\pi nk}{N}} \quad (10)$$

Calculating DFT and Inverse DFT can be done efficiently using Fast Fourier Transform (FFT). Using the imaginary and real portions of a complex number produces the magnitude of the DFT (A).

$$A = \sqrt{R^2 + I^2} \quad (11)$$

where A is magnitude value, R is real part and I imaginary part of the complex number.

2D FFT is used to calculate the DFT of the MR image of the brain [20]. The output is then rearranged so that the zero frequency is centered in the array. The natural logarithm is then applied to every absolute value of the corresponding element. 2D FFT of the brain MR image result is shown Figure 6.

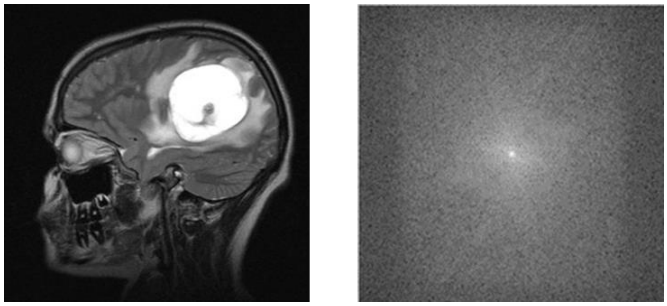

Figure 6. Result of 2D FFT on brain MR image

3.4. Texture Features

Texture analysis describes the various sections of the image based on their texture content [21]. This analysis technique attempts to measure the image texture's instinctive qualities such as smooth, silky, rough, or bumpy by finding a function of the spatial changes by measuring pixel intensities. This means differences in the intensity values or gray levels refer to a texture content such as bumpiness or roughness.

Texture analysis has been used in various applications including but not limited to automated inspection, remote sensing, and medical image processing [22-24]. Texture segmentation refers to finding the texture boundaries. Texture analysis is most useful when image objects are defined by their texture rather than their intensities and when the traditional thresholding approaches are less effective.

A representative research work applying texture and statistical feature-based MR image classifications using MLP classifier method is proposed in [25]. An accuracy of 91.9% is achieved using MRI dataset of 3,064 images. The major problem, however, with MLP is the deep connection between the hidden layers that makes this deep learning network difficult to scale.

3.5. First-order histogram-based features

The image histogram provides the statistical information contained in it. Therefore, the first-order statistical information of an image is given by the first-order histogram [26-27]. Equation 12 provides the way to calculate the probability density of occurrence of the intensity levels by dividing the histogram intensity level over the total number of pixels in the image.

$$P(i) = \frac{h(i)}{M\,N}, \quad i = 0,1,\ldots G-1 \quad (12)$$

where M and N refer to the number of resolution cells in the vertical and horizontal spatial domain, respectively. The grey level of the image is represented by G.

As mentioned previously, histograms provide features of an image by finding the first-order statistical features. Equations 13-18 are used to measure a variety of statistical information. The average of the intensity of the image is the *mean*. The

intensity variations near the mean produce the *variance*. The symmetry of the histogram around its mean is a measurement called *skewness*. The flatness of the histogram is a measurement called *kurtosis*. *Entropy* is the measure of the uniformity of the histogram [27].

Mean: $$\mu = \sum_{I=0}^{G-1} ip(i) \quad (13)$$

Variance: $$\sigma^2 = \sum_{I=0}^{G-1} (i-\mu)^2 p(i) \quad (14)$$

Skewness: $$\mu_3 = \sigma^{-3} \sum_{I=0}^{G-1} (i-\mu)^3 p(i) \quad (15)$$

Kurtosis: $$\mu_4 = \sigma^{-4} \sum_{I=0}^{G-1} (i-\mu)^4 p(i) - 3 \quad (16)$$

Energy: $$E = \sum_{I=0}^{G-1} [p(i)]^2 \quad (17)$$

Entropy: $$H = -\sum_{I=0}^{G-1} p(i) \log_2 [p(i)] \quad (18)$$

3.6. Co-occurrence Matrix based Features

The features provided by the histogram are local in nature, which do not take the spatial information into account. Thus, grey level spatial co-occurrence matrix hd (i,j) based features are usually computed, which actually provide the second-order histogram-based features. The basis of these features is the calculation of joint probability distribution of pairs of pixels. To calculate this probability, the angle θ and the distance d within a given neighborhood are needed [27-28]. A sample co-occurrence matrix is shown in Figure 7, which is used to describe these texture features obtained from the calculations using Equations 19-25.

Angular second moment: $$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} [p(i,j)]^2 \quad (19)$$

Correlation: $$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{ij\, p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (20)$$

Inertia: $$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i-j)^2\, p(i,j) \quad (21)$$

Absolute Value: $$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} |i-j|\, p(i,j) \quad (22)$$

Inverse Difference: $$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{p(i,j)}{1+(i-j)^2} \quad (23)$$

Entropy: $$H = -\sum_{I=0}^{G-1} \sum_{j=0}^{G-1} p(i,j) \log_2 [p(i)] \quad (24)$$

Maximum Probability: $$\max_{i,j} p(i,j) \quad (25)$$

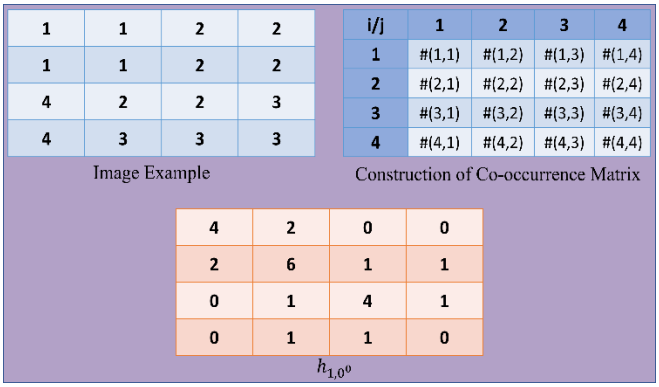

Figure 7. The spatial co-occurrence matrix calculation

3.7. Convolutional Neural Network (CNN) based Features

To perform an automated extraction of features from the images, a Deep Learning approach termed as Convolutional Neural Networks, is found to perform better than other existing feature extraction methods [29]. The key element in the CNN is the filters that slide over the input images producing a feature map. The type of filter used affects the types of feature maps produced. The number of filters also affects the number of image features which in turn affects the image recognition process. Parameters controlling the size of the feature map include depth (number of filters), stride (number of pixels the filter covers), and zero padding (bordering the image with zeroes) [30-33].

A normalization layer is needed to speed up the training process and minimize the sensitivity to network initialization. The process employed in the normalization layer is to subtract the mean of the mini-batch and divide it by the standard deviation of the mini-batch [34]. Further, the input is shifted by a learnable offset β and also scaled by a factor of γ. For input of mini-batch $B = \{x_{1,\ldots,m}\}$ with parameters, β, and γ produce the output layer $\{y_i = BN_{\gamma,\beta}(x_i)\}$ based on Equations 26-29.

$$\mu_B \leftarrow \frac{1}{m} \sum_{i=1}^{m} x_i \quad (26)$$

$$\sigma_B^2 \leftarrow \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_B)^2 \quad (27)$$

$$x_i' = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (28)$$

$$y_i \leftarrow \gamma x_i' + \beta \equiv BN_{\gamma,\beta}(x_i) \quad (29)$$

Computation complexity and spatial dimensions are reduced using a pooling layer. The layer also takes care of the overfitting problem. Different pooling functions can be used in this layer, with max-pooling being the most popular. Usually, a pooling layer with a filter size of 2x2 is used.

In a representative research work [35], the authors proposed a system that can automatically recognize multi-class Glioma brain tumors. A simple deep CNN was used in this study using T1 weighted CE-MR images from BraTS dataset. They achieved a training accuracy of 98.51% and a validation accuracy of 84.19%. Though the proposed method is an attempt to reduce overfitting, yet overfitting and lack of data continue to be a problem in these types of studies.

4. CLASSIFICATION TECHNIQUES FOR BRAIN TUMOR DETECTION

There are various techniques that have been used by researchers to solve the automated image classification problem. In this section, a review is performed for some of the most popular classification techniques, which are Random Forests (RF), Multilayer Perceptron (MLP), Support Vector Machines (SVM), Naïve Bayes classifier, K-nearest neighbor (KNN), Ensemble Classifiers, Genetic Algorithms (GA) and Convolutional Neural Network (CNN) [36-41].

4.1. Random Forest (RF)

Random Forest (RF) as a classifier is is very popularly used in classifying MRI of brain images into tumorous and non-tumorous classes [42]. The term implies that a collection of several decision trees are initially created and combined in a random manner to build a forest. Further, the constituent trees

are trained using a random sample of data from the training dataset. In order to avoid the overfitting of the trained model, RF performs a Bagging step that uses bootstrap aggregation [43]. The test features gathered after model creation are used to predict the outcome of each decision tree. The final outcome of the algorithm is achieved by taking the prediction with maximum votes. For testing, an unknown input is supplied to the classifier, and the desired output is calculated based on majority votes.

In the training part of the random forest algorithm, the bagging technique is applied to the decision trees. Consider we have input training data set $X = x_1, x_2 \ldots, , x_n$. Bagging iteratively fits the decision trees to a random sample from the training set. Once the training is complete, an unknown input with feature $x'$is passed to the algorithm. The output is calculated using Equation 30, by taking the average of the votes from the constituent decision trees. In the equation, T_i is the i^{th} decision tree and B is the total number of decision trees in the forest.

$$Y_i = \frac{1}{B} \times \sum_{i=1}^{B} T_i(x') \tag{30}$$

A classifier $f(x, v_i)$ for tree T_i in the random forest is defined by generating a vector v_i that is independent of previous vectors $v_1, \ldots, v_{i-1}$ but has the same distribution. The classifier is generated by growing this tree with the random vector v_i , and the training sequence. A collection of these classifiers forms the random forest classifier i.e. $\{f(x, v_i), i = 1, \ldots, B\}$. The classification decisions are made based on the desired conditions i.e. $x_i < H$ is met where H represents the threshold.

The margin function for an RF classifier is defined as the degree of votes for correct classification minus the maximum proportion of votes for any other class in the dependent variable. Let us assume that V is a random vector sampled from training data, and Y is the response of the classifier. If $\theta_k = \theta_{k1}, \theta_{k2}, \ldots, \theta_{kB}$ are the parameters of a decision tree classifier $f_k(x)$, the margin function can be calculated by using Equation 31.

$$mg(V, Y) = \frac{\sum_{i=1}^{B} I(f_k(V)=Y)}{B} - \max_{j \neq y} \left[\frac{\sum_{i=1}^{B} I(f_k(V)=j)}{B}\right] \tag{31}$$

$I(.)$ is the Indicator Function as shown in Equation 32.

$$I_A(x) = \begin{cases} 1 & if\ x\ \in A \\ 0 & otherwise \end{cases} \tag{32}$$

The classification results can be interpreted as correct if the value of the margin function is positive, and it is treated as wrong if it is negative. The measure of confidence in the results is directly proportional to the magnitude of the margin function value achieved.

PE^* is defined as the misclassification rate over the space V, Y and is given in Equation 33.

$$PE^* = P_{V,Y}(mg(V, Y) < 0) \tag{33}$$

For a large random forest, i.e., $B \rightarrow \infty$, the generalization error can be calculated from Equation 34.

$$PE^* \rightarrow P_{V,Y}\big(P_\theta(h(V, \theta) = Y)) - \max_{j \neq y} P_\theta(h(V, \theta) = j)) < 0\big) \tag{34}$$

The misclassification rate has a limiting value due to no overfitting in the Random Forest classifier. The strength *s* of a classifier in a random forest is the *Expectation* of the misclassification rate.

$$s = E_{X,Y}(mr(X, Y)) \tag{35}$$

In a representative study using RF as a classifier, a variety of features are extracted, which include histogram calculated using a set of Gabor filters with various sizes and orientations as well as first-order intensity statistical feature [44]. The BRATS 2013 dataset is used in this study and compared with an in-house clinical dataset consisting of 11 multimodal images of patients. The in-house dataset produced an average detection sensitivity of 86% compared with 96% for BRATS, while the segmentation results against ground truth were 0.84 for the in-house dataset and 0.89 for BRATS 2013 [80]. However, their approach, in general, has a limitation of segmenting small volumes with additional limitation at the tissue boundary, causing overlap with other tissue types.

4.2. Multilayer Perceptron (MLP)

The most basic version of neural networks, called the Multilayer Perceptron (MLP), consists of hidden layers where each element in a particular layer is connected to the elements in the next layer [45]. The purpose of the MLP is to perform denoising of images and their classification. It is to be noted that in brain MRI the tumorous and non-tumorous data is not linearly separable. Therefore, in MLP, all the nodes in the hidden layers function as neurons to model a non-linear activation function. An example of an MLP is shown in Figure 8.

In an MLP network, the matrices W_i encode the conversion from one layer of the network to another using matrix multiplication. If you have *m* neurons in one layer connected to *n* neurons in the next layer, then the transformation matrix has the dimension as shown in Equation 36.

$$W \epsilon R^{m*n} \tag{36}$$

This transformation layer can map an input *X* to an output *Y* according to the relation given in Equations 37-39.

$$X \epsilon R^{m*1} \tag{37}$$

$$Y \epsilon R^{1*n} \tag{38}$$

$$Y = X^T W \tag{39}$$

Each column in the transformation matrix W converts the edges moving from one layer to the next. However, the major problem with MLP is that the deep connection between the hidden layers makes this deep learning network difficult to scale. It makes the training of the MLP a difficult and time-consuming process. This challenging process ultimately affects the sensitivity of the network if the features of the image keep changing since the MLP will need to re-learn and training becomes harder. In a representative research work [15], the authors have combined DCT and DWT features and used MLP classifier-based classification of MR images into tumorous and non-tumorous classes, which achieves 86.98%

accuracy, but the dataset used for the experiments is small, consisting of only 3064 MR images from 233 patients.

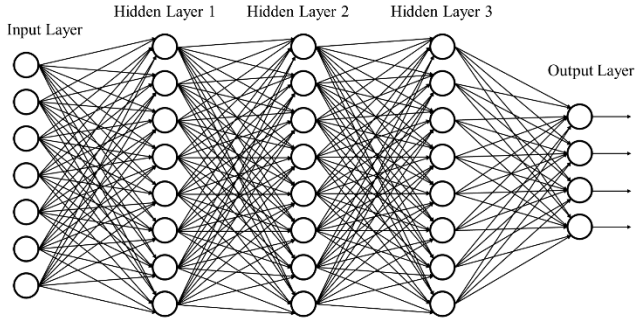

Figure 8. Architecture of the Multilayer Perceptron (MLP)

4.3. Support Vector Machine (SVM)

Support Vector Machine (SVM) is another supervised machine learning approach, which provides classification results in a binary form [46]. SVM classifies the input data into different classes by forming a hyperplane in a higher dimension space of the feature set. This technique transforms the input data, which is divided into separate classes non-linearly, by applying kernel functions that transform the input into hyperplanes, as shown in Figure 9. It uses structural error minimization for classification and maximizes the margins between the classes in the hyperplane [47]. SVM is one of the most popular methods used in image classification, text classification, voice recognition, and gene data study [48-50].

Given a set of *n* training images of two different classes indicated by

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots\ldots (x_n, y_n) \qquad (40)$$
$$i = 1, 2, 3, \ldots\ldots\ldots\ldots n$$

$$x_i \in R_d \quad \textit{D-Dimensional feature space}$$
$$y_i \in \{-1, +1\} \quad \textit{Class label} \qquad (41)$$

The linear separation of images into any hyperplanes can be described based on Equation 42.

$$\begin{aligned} (w, x_i + b) &\geq 1, \quad && if\ y_i = 1 \\ (w, x_i + b) &\leq -1, && if\ y_i = -1 \\ (w, x_i + b) &= 0 \end{aligned} \qquad (42)$$

Here *w* is the weight vector, and *b* indicates the bias weight. The SVM tries to maximize the difference between the classes by decreasing the weight vector i.e.

$$\min\left(\tfrac{1}{2}\|w\|^2\right) \qquad (43)$$

The given formulation works when the data is linearly separable. This can be converted into the Lagrangian formulation, as shown in Equation 44.

$$\min_{w} \max_{\alpha} L_D \qquad (44)$$

where L_D is the dual Lagrangian given by Equation 45.

$$L_d = \sum_{i=1}^{n} \alpha_i - \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i x_j \qquad (45)$$

with the constraints given in Equation 46.

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \quad where \quad 0 \leq \alpha_i \leq C \qquad (46)$$

The generalization ability of the SVM is dependent on two attributes, $\alpha_i \geq 0$, which is the Lagrange multiplier and a penalty parameter C. The input training points x_i which satisfies the equation $y(w, x_i + b) \leq -1$ are termed as the support vectors which have non-zero Lagrange multipliers. Once the training is complete, and the Lagrange multipliers are calculated, image classification can be performed for the linearly separable classes.

For non-linear classification, the separating hyperplanes can be generated by using kernel functions. Since all the supporting vectors that appear in the Lagrangian formulation are scalar products, various kernel functions can be applied to transform the inputs into high dimension Hilbert Feature space, as shown in Equation 47.

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \ \ where \ \ \varphi: R_d \rightarrow H \qquad (47)$$

If the kernel satisfies Equation 47, then it can be used for the transformation of data into higher dimensional space. For the kernel to be symmetric, it must be positive definite, i.e., for any given data x_i and real number λ_i as shown in Equation 48.

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j K(x_i, x_j) \geq 0 \qquad (48)$$

There are various kernel functions that satisfy the above criteria and can be used for the classification of non-linear data, e.g., isotropic Gaussian kernel, polynomial kernel, hyperbolic tangent kernel, etc. To increase the accuracy of the classification, a slack variable is usually introduced, which keeps track of the frequency of errors that happen during the classification process.

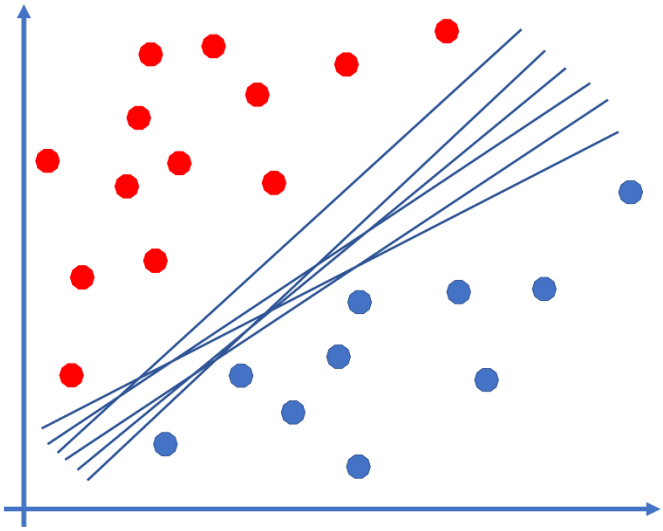

Figure 9. Support Vector Machine (finding the optimal line separator)

In one of the sample research work, DCT features, with SVM as a classifier, are used, which achieved 98.82% accuracy, but the experiments are conducted using very small datasets of 255 images only [14]. In another article, the authors used level

1 to 5 DWT decomposition to extract the brain MR image features [18]. The DWT feature-based classification of MR images into normal (tumorous) and abnormal (non-tumorous) achieves 93.5% accuracy using SVM as a classifier, but the dataset used for experiments is small, covering only 100 MR images.

4.4. Naïve Bayes Classification

Bayes's theorem is the basis for implementing the Naïve Bayes classifier, which relies on a probabilistic statistical model [51]. It is called Naïve because it assumes strong independence between the features of the image. In the basic Bayes classifier, the conditional probability that data belongs to a class can be calculated by finding out the conditional and unconditional probabilities of the data belonging to each class in the data set. The challenging part in Naïve Bayes classifier is to find the class of data having the same number of attributes with strong independence.

Let *x* be an event belonging to a random variable *X* containing *n* objects, i.e., $x_1, x_2, \ldots, x_n$, representing features. Every object in the random variable *X* has *l* different attributes, i.e., $t_1, t_2, \ldots, t_l$. If *C* is a random variable representing *m* possible classes, i.e., $c_1, c_2, \ldots, c_m$, the probability that an event *x* belongs to class *k* can be calculated by using the Bayes' Maximum a Posteriori (MAP) decision rule as shown in Equation 49 [52].

$$\mathrm{P}(c_k|X) = \frac{\mathrm{P}(X|c_k)\mathrm{P}(c_k)}{\mathrm{P}(\mathrm{X})} \quad (49)$$

where $\mathrm{P}(c_k)$ represents the class prior probabilities that are calculated by Equation 50.

$$\mathrm{P}(c_k) = \frac{s_i}{s} \quad (50)$$

In Equation 50, the number of training samples is denoted as s_i which have a class as c_i , whereas *s* represents the total number of training samples. Equation 49 can be interpreted as a classification problem in which a given information *x* can be classified into class *k* if the probability of *X* belonging to class *k* is higher than all the other class probabilities. To do that, we need to calculate $\mathrm{P}(c_k|X)$ for the *n*-dimensional random variable *X*. Assuming every object in vector *X* is statistically independent on the class c_k from each other, the estimate can be calculated by Equation 51.

$$\mathrm{P}(X|c_k) = \prod_{i=1}^{n} \mathrm{P}(x_i|c_k) \quad (51)$$

The probabilities $\mathrm{P}(x_i|c_k)$ values are estimated from the training samples from the input MR image set with an attribute t_k that can take on the value x_{1k} as shown in Equation 52.

$$\mathrm{P}(x_i|c_k) = \frac{s_{ik}}{s} \quad (52)$$

where the total training sample is denoted as s_{ik} in class c_k having a value of x_k for an attribute t_k, and the number of training samples are represented as s_i. The Naïve Bayes classifier that finds the class *k* with maximum probability can be specified in Equation 53.

$$\underset{k}{\operatorname{argmax}} \ \mathrm{P}(c_k) \prod_{i=1}^{n} \mathrm{P}(x_i|c_k) \quad (53)$$

In one of the representative research work [53], the authors did the experiments using brain MR images to classify them into tumorous and non-tumorous using Naïve Bayes along with DCT features which achieved an average accuracy of 82.86%. It has been observed that Naïve Bayes algorithm did not perform well for the brain tumor classification compared to other classifiers.

4.5. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) is one more supervised classifier that uses the distance metric among its nearest neighbor to classify the input data into various classes. Various distance metrics can be used, e.g., Euclidean distance, cosine similarity, Manhattan distance, city block distance, correlation or hamming distance, etc. [54]. The Euclidean distance between test data points and every other data point in the training set can be calculated by Equations 54 and 55.

$$d = \sqrt{(x_1 - x_1')^2 + (x_2 - x_2')^2 + \cdots + (x_N - x_N')^2} \quad (54)$$

$$d = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \quad (55)$$

Where $\boldsymbol{x_i}$ are data points in the training set. Given a positive value for the parameter K, a distance metric ***d*** (e.g., Euclidean distance) and let the ***K*** nearest data points to ***x*** be called ***A***. The algorithm works by computing the metric ***d*** between ***K*** nearest data points for every training data. The algorithm then calculates the conditional probability of all the points in the dataset ***A*** that have the given class label as shown in Equation 56.

$$P(y = m|X = x) = \frac{1}{K} \sum_{i \in A} I(y^i = m) \quad (56)$$

Here ***I(x)*** is the indicator function that results in a value of one when the argument ***x*** is true and zero otherwise. The input ***x*** is then allocated to the class having the maximum probability.

Figure 10 shows an example of a KNN classifier with the value of K as 5. It can be seen that the black color object is an input instance that has been classified in the same class where the red color object belongs to.

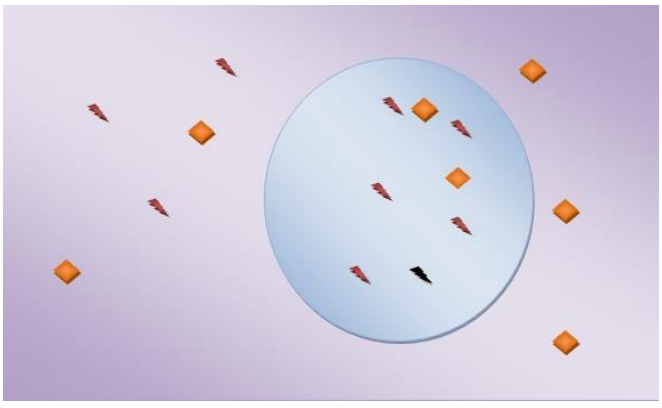

Figure 10. K-Nearest Neighbor example (k=5)

In one sample research work [55], the authors used KNN classifier for the brain MR images classification into benign

and malignant. They achieved an average accuracy of 85.71% with 300 iterations using OASIS brain MRI dataset and compared the results with the SVM classifier.

4.6. Ensembles of Classifiers (AdaBoost)

Ada-boost classifier is an ensemble of weak classification algorithms that combine to form a strong classifier [56]. In an ensemble of classifiers, the weights of each classifier are reassigned in each iteration until the right value of weights are calculated and iterations stop. Boosting is the process of building this ensemble of classifiers. If $H_1(x)$, $H_2(x)$, $H_3(x), \ldots\ldots, H_N(x)$ are the weak classifiers that are used, then the boosting process assigns a certain contribution weight of each classifier, i.e., $w_1(x)$, $w_2(x)$, $w_3(x), \ldots\ldots, w_N(x)$ to build a strong classifier given by Equation 57.

$$H(x) = sign\left(\sum_{i=1}^{N} w_i H_i(x)\right) \quad (57)$$

Boosting is an iterative method that uses random sampling of the data without replacement. In boosting, errors in the prediction of earlier models are removed by predictions of future models. This step is different from the bagging process used in Random Forest classifiers that use an ensemble of independently trained classifiers.

The adaptive boosting (Ada-boost) algorithm uses weighted versions of the same training data instead of randomly selecting each weak classifier's dataset. If the initial base learner is already strong with high classification accuracy, it will leave only outliers and noisy instances with high weights to be learned in future iterations. Every weak classifier H_k in the ensemble tries to obtain a family of classifier H that fulfills Equation 58.

$$\sum_{i=1}^{N} w_i h\left(-y_i H_k(x_i)\right) \leq \frac{1}{2} - \varepsilon \quad (58)$$

Different variations of Ada-boost algorithm can be used for binary or multi-class supervised learning problems. Most of the multiclass variations formulate the problem as a concatenation of many binary class problems. AdaBoost.M1 and AdaBoost.M2 are two variations of the original Ada-Boost algorithm for multiclass classification [57]. In a representative research work [58], an accuracy of 80.02% was reported, which proposed using Intensity and Wavelet features with RusBoost Classifier for the multi-class classification of brain tumors.

4.7. Genetic Algorithms

Genetic Algorithms are designed based on the biological evolutionary theory of survival of the fittest. These are based on stochastic processes that can provide a robust and effective image classification. In a genetic algorithm, the problem states are represented in the form of strings, and a fitness function is used to apply three operations, i.e., reproduction, crossover, and mutation, iteratively. In each repetition, the population of points is updated with a new and efficient selection that can minimize the probability of points entering the local extrema [59]. The updating of points in each iteration is done by applying the three genetic operations, i.e., reproduction, crossover, and mutation. The process stops when the iterations converge, or a basic criterion is satisfied. The string of points with the highest fitness value is used as the final result of the algorithm. The string of points can be indicated by Equation 59.

$$S = s_1, s_2, s_3, s_4, \ldots\ldots\ldots s_n \text{ where } s_i (1 \leq i \leq n) \quad (59)$$

In one of the sample research work, the authors proposed a modified genetic algorithm approach for the classification of normal and abnormal MR images for tumor detection [60]. In the proposed approach, the authors tried to minimize the random nature of conventional GA and achieved 98% accuracy but the dataset used for the experiments only consisted of only 550 MR images.

4.8. Convolutional Neural Network (CNN)

Deep learning is a subset of machine learning techniques inspired by the structure of the human brain that enables machines to mimic human behaviors. Deep learning is the future of automated classification without the need for human intervention, but this automation comes at a cost of higher volume of data requirements and high computing power. Training the deep learning models requires high computational power using Graphical Processing Units (GPUs) having thousands of cores for processing. One of the deep learning classifiers is a CNN which is a variant of a deep neural network that is commonly used for the classification of images.

In one of the representative research works [61], the authors propose the use of deep learning for multi-class brain tumor classification, using augmented BraTS data for training followed by a pre-trained CNN model for classification and reported an accuracy of 94.5%. This method addresses the problem of lack of data and overfitting, however, the limitations of the proposed method are complexity, computation time, and memory usage. In another study, the authors proposed a system that can automatically recognize multi-class Glioma brain tumors [35]. A simple deep CNN was used in this study with the use of T1 weighted CE-MR images from BraTS dataset. They achieved a training accuracy of 98.51% and validation accuracy of 84.19%. Though the proposed method is an attempt to reduce overfitting, yet overfitting and lack of data continues to be a problem in these types of studies.

5. DISCUSSIONS

Extensive research has been accomplished in tackling the challenge of automatic recognition and diagnosis of brain tumors using brain MRI images. Recent research has applied various methods as outlined in this review paper. Table 1 shows a comparison of the results reported in the recent literature of various methods applied to brain MRI images for the binary classification of the images into tumorous and non-tumorous images applied on the BraTS dataset. The recent results are collected from the research papers disseminated from the year 2018-2021. The method proposed by Raja et al. consists of seven phases and requires high computational time [62]. Kumar et al. used optimization driven Dolphin-SCA based Deep CNN for brain tumor classification, however, the authors only utilized a portion of the BraTS dataset (65 MR images) [63]. The also failed to provide the justification for

selective data rather than using the complete BraTS dataset. Even with a portion of the dataset, their reported accuracy is relatively less than other methods listed in Table 1. The highest accuracy was reported at 98.77% [64] using CNN based features with SVM classifier. The lowest accuracy was reported at 91.31% [68] using texture and Gabor wavelets features with SVM-PKU kernels.

Table 1. Comparison of the binary classification (Tumor, No-tumor) techniques proposed in recent literature using BraTS brain MR images dataset

Reference	Methodology	Results	
		Acc.	Recall
Raja et al. 2020 [62]	Hybrid deep autoencoder with Softmax regression	98.5 %	96%
Kumar et al. 2020 [63]	Optimization driven Dolphin-SCA based Deep CNN Classification	95.3%	97.7%
Latif et al. (2020) [64]	Convolutional Neural Network based features with SVM classifier	98.77%	-
Mzoughi et al., 2020 [65]	Multi-Scale 3D CNN for brain MR Image Classification	96.49%	-
Arasi et al. 2019 [66]	GLCM and Gabor feature with BSVM classifier	97.69%	94%
Srinivas et al. 2019 [67]	Hybrid CNN features and KNN based brain MRI classification	96.25%	95.8%
Chinnam et al. 2019 [68]	Texture and Gabor Wavelets features with SVM-PKU kernel	91.31%	95.9%
Latif et al. 2018 [28]	Hybrid statistical and wavelets features with MLP classifier	96.73%	96.7%
Seetha et al. 2018 [69]	Five CNN layers-based Model for brain MRI classification.	97.5%	
Sriramakrishnan et al., 2018 [70]	Block Based Feature Extraction and Random Forest Classifier	95%	94%
Latif et al. 2017 [53]	Discrete Cosine Transform (DCT) features with KNN classifier	96.91%	-

Other researchers chose to apply their proposed binary classification techniques on different datasets of varying sizes which are smaller than the BraTS Brain MR images dataset. Table 2 shows the results of recent literature proposing various techniques for the binary classification of brain MR images (smaller datasets) into tumorous and non-tumorous images. The lowest accuracy reported in Table 2 is 72.50% [73], which proposed the use of GLCM features and the KNN classifier. The highest accuracy reported is 99.61% [14] which proposed DWT +SURF +BoW features with SVM and RBF classifiers. The higher accuracy is expected due to the small size of the dataset compared with the BraTS dataset. The experimental results listed in Table 2 are observed not to be reliable due to the fact that experiments were performed on very small and selective datasets. Though Ayadi et. A. [14] and Latif et. al. [72] reported accuracies of more than 99%, yet a slight change in the MRI dataset can largely impact the accuracies.

Table 2. Comparison of the binary classification (Tumor, No-tumor) techniques proposed in recent literature using different small MR images datasets

Reference	Methodology	Dataset (No. of images)	Results	
			Acc.	Recall
Amin et al. 2020 [71]	Shape, texture, and intensity features with SVM	100	97.1%	91.9%
Ayadi et al. 2019 [14]	DCT features with SVM	255	98.82%	-
Ayadi et al. 2019 [14]	DWT + SURF + BoW features with SVM and RBF classifier	255	99.61%	100%
Latif et al. 2018 [72]	Level 3 DW Features with MLP classifier	2920	99.43%	-
Wasule et al., 2017 [73]	GLCM Features, SVM	251	85%	76%
Wasule & Sonar, 2017 [73]	GLCM Features, KNN Algorithm	251	72.50%	73.33%
Bahadure et al. 2017 [74]	Biologically Inspired BWT features and SVM	135	96.51%	97.72%
Mohankumar, 2016 [18]	DWT + SVM	100	93.5%	-

Selected research published between 2018-2021 for multi-class brain tumor classification is shown in Table 3 applied on the CE-MRI dataset, which consists of 3064 images. The lowest accuracy of 84.19% was reported in [35], which proposed hyper-parameter optimization-based CNN technique for multi-class classification. The highest accuracy of 94.82% was reported in [75], which proposed the use of block-wise transfer learning for CNN classifier for the multi-class classification problem. Very little research targeted multiclass tumor classification, thus, this area of research is still open as a scope for future research.

Table 3. Comparison of the Multiclass (Glioma, Meningioma, Pituitary) tumor classification techniques proposed in recent literature using CE-MRI Dataset of 3064 images

Reference	Methodology	Results	
		Acc.	Recall
Fasihi et al. 2020 [15]	DCT+DWT combined	86.98%	-
Swati et al. 2019 [75]	Block wise transfer learning for CNN Classifier	94.82%	94.25%
Abiwinanda et al. 2019 [35]	Hyper-parameters Optimization based CNN	84.19%	-
Ismael et al. 2018 [25]	DWT + Gabor +Statistical Features with MLP classifier	91.9%	91.07

Selected publications from 2018 to 2021 are shown in Table 4, which tackled the multi-class brain tumor classification using the BraTS MR image dataset. The lowest accuracy of 80.02% was reported in [58], which proposed the use of Intensity and Wavelet features with RusBoost Classifier for the multi-class classification of brain tumors. Among the manuscripts reviewed, the highest accuracy was reported in [76] which 96%. In [76], the authors proposed the use of CNN model based on fused features of (Fair +T1+T1c+T2), however, this technique is highly dependent on the selection and combination of features. Feature selection complicate the process and thus self-learning (CNN based techniques) along with model optimization is preferred. Literature review of CNN based approaches for Multiclass Glioma Tumor classification showed that optimal accuracies have not been reached (Table 4) and further research using CNN models is required to reach and optimal accuracy that would be

acceptable to adopt these techniques in actual hospital and medical setting.

Table 4. Comparison of the Multi-class (edema, necrosis, enhancing, non-enhancing) Glioma tumor classification techniques proposed in recent literature using BraTS brain MR images dataset.

Ref no.	Methodology	Results	
		Acc.	Recall
Amin et al. 2020 [76]	CNN Model based on the Fused Features of (Fair+T1 + T1c + T2)	96%	98%
Khan et al. 2020 [77]	VGG19 features with ELM classifier	92.5%	-
Xue et al., 2020 [78]	Dual path Residual Convolutional Neural Network	84.9%%	-
Sajjad et al., 2019 [61]	Deep CNN with extensive data augmentation	94.58%	88.41%
Amjad et al. 2019 [79]	3D CNN features with the FNN classifier	92.67%	-
Soltaninejad et al., 2018 [80]	Texture Features from Supervoxels and Random Forest as Classifier	90.67%	86%
Khalid et al. 2017 [58]	Intensity and Wavelet features with RusBoost classifier	80.02%	-

It is noticed that the selected publications reported in the results section for the binary classification and multi-class classification of the brain MR images were all published recently between 2017 and 2021. Yet, the accuracy and performance measures are still far from perfect, and there are many factors that need to be accounted for in order to adopt the automatic classification and diagnosis of brain tumors using image-processing techniques. The severity and consequences of misdiagnosis of a dangerous medical case such as brain tumor entail that the accuracy must reach nearly 100% and that all performance metrics and other factors must be accounted for before bringing the machine learning algorithms within the hospital setting to assist doctors and medical staff in the detection and diagnosis of such a severe illness.

This makes this review paper and other review papers in the field very important as a one-stop reference for researchers in the field to be up to date with the latest proposed algorithms and to be able to start where others have ended. The future of machine learning in hospital settings is getting closer to reality and is even closer than many realize. The vast benefits of including machine learning algorithms in the precise diagnosis of medical problems are apparent to all stakeholders and benefit everyone.

6. CONCLUSION

In this paper, a detailed review is presented based on the approach that researchers are following in tackling the binary classification and multi-class classification of brain MR images. The approach consists of various phases to include preprocessing, feature extraction, and classification of the images. The most important techniques for different stages are explained, and the results of the selected recent literature are reported for both binary and multi-class classification of the brain MR images for both the BraTS MR images dataset and other smaller datasets. The details of the stages along with the result of recent literature, will assist researchers in the field as they can benefit from a paper that reviews all the latest developments in the field. Such review will allow them to start where others have ended and continuously strive to achieve the highest accuracy and performance metrics in the important medical field, such as the automatic diagnosis of brain tumors.

Future work will include a review of the literature for other medical conditions such as skin cancer, breast cancer, and extremely vital medical problems. The researchers also intend to continue their own research on developing innovative machine learning algorithms to automatically identify and diagnose brain tumors, skin cancer, breast cancer, and other medical-related research.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest in this research.

ACKNOWLEDGEMENTS

The authors would like to thank the College of Computer Engineering and Sciences, Prince Mohammad bin Fahd University for providing computing resources for experiments. The authors would like to thank Dr. Runna Alghazo for proofreading the manuscript and editing to eliminate language issues and enhance readability wherever warranted.

REFERENCES

[1]. Overcast, W.B., Davis, K.M., Ho, C.Y., Hutchins, G.D., Green, M. A., Graner, B. D., ... & Veronesi, M.C. (2021). Advanced imaging techniques for neuro-oncologic tumor diagnosis, with an emphasis on PET-MRI imaging of malignant brain tumors. *Current Oncology Reports*, *23(34)*, 2021.

[2]. Lizio, G., Salizzoni, E., Coe, M., Gatto, M. R., Asioli, S., Balbi, T., & Pelliccioni, G. A. (2018). Differential diagnosis between a granuloma and radicular cyst: effectiveness of magnetic resonance imaging. *International endodontic journal*, *51*(10), 1077-1087.

[3]. Agravat, R. R., & Raval, M. S. (2018). Deep learning for automated brain tumor segmentation in mri images. In *Soft Computing Based Medical Image Analysis* (pp. 183-201). Academic Press.

[4]. Biratu E.S., Schwenker F., Ayano Y.M, Debelee T.G. A Survey of Brain Tumor Segmentation and Classification Algorithms. *Journal of Imaging*. 2021; 7(9):179.

[5]. Kaur P., Singh G., Kaur P., A Review of Denoising Medical Images Using Machine Learning Approaches. *Current Medical Imaging*. 2018; 14(5), pp 675-685.

[6]. Maxime C., Erika P., Sila G., Kate D., Maxime D., Greg D. P., Chantal M.W., Derek K. J., Dimensionality reduction of diffusion MRI measures for improved tractometry of the human brain, NeuroImage,Vol. 200, 2019, pp. 89-100,

[7]. Magadza T, Viriri S. Deep Learning for Brain Tumor Segmentation: A Survey of State-of-the-Art. *Journal of Imaging*. 2021; 7(2):19.

[8]. Singh, R., Goel, A., & Raghuvanshi, D. K. (2020). Computer-aided diagnostic network for brain tumor classification employing modulated Gabor filter banks. *The Visual Computer*, 1-15.

[9]. Latif, G., Iskandar, D. A., & Alghazo, J. (2018, September). Multi-class Brain Tumor Classification using Region Growing based Tumor Segmentation and Ensemble Wavelet Features. In *Proceedings of the 2018 International Conference on Computing and Big Data* (pp. 67-72).

[10]. Kavin Kumar, K., Devi, M., & Maheswaran, S. (2018). An efficient method for brain tumor detection using texture features and SVM

classifier in MR images. *Asian Pacific journal of cancer prevention: APJCP*, *19*(10), 2789.

[11]. Morais, C. L., Lima, K. M., & Martin, F. L. (2019). Uncertainty estimation and misclassification probability for classification models based on discriminant analysis and support vector machines. *Analytica chimica acta*, *1063*, 40-46.

[12]. Kaplan, K., Kaya, Y., Kuncan, M., & Ertunç, H. M. (2020). Brain tumor classification using modified local binary patterns (LBP) feature extraction methods. *Medical hypotheses*, *139*, 109696.

[13]. Bazine, R., Wu, H., & Boukhechba, K. (2019). Spatial Filtering in DCT Domain-Based Frameworks for Hyperspectral Imagery Classification. *Remote Sensing*, *11*(12), 1405.

[14]. Ayadi, W., Elhamzi, W., Charfi, I., & Atri, M. (2019). A hybrid feature extraction approach for brain MRI classification based on Bag-of-words. *Biomedical Signal Processing and Control*, *48*, 144-152.

[15]. Fasihi, M. S., & Mikhael, W. B. (2020, August). MRI brain tumor classification Employing transform Domain projections. In *2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS)* (pp. 1020-1023). IEEE.

[16]. Goel, A., & Vishwakarma, V. P. (2017). Fractional DCT and DWT hybridization based efficient feature extraction for gender classification. *Pattern Recognition Letters*, *95*, 8-13.

[17]. Lapins, S., Roman, D. C., Rougier, J., De Angelis, S., Cashman, K. V., & Kendall, J. M. (2020). An examination of the continuous wavelet transform for volcano-seismic spectral analysis. *Journal of Volcanology and Geothermal Research*, *389*, 106728.

[18]. Mohankumar, S. (2016). Analysis of different wavelets for brain image classification using support vector machine. *International Journal of Advances in Signal and Image Sciences*, *2*(1), 1-4.

[19]. Barigye, S. J., Freitas, M. P., Ausina, P., Zancan, P., Sola-Penna, M., & Castillo-Garit, J. A. (2018). Discrete Fourier transform-based multivariate image analysis: application to modeling of aromatase inhibitory activity. *ACS combinatorial science*, *20*(2), 75-81.

[20]. Saeed, S., Abdullah, A., & Jhanjhi, N. Z. (2019). Implementation of fourier transformation with brain cancer and CSF images. *Indian Journal of Science and Technology*, *12*, 37.

[21]. Seifi Majdar, R., & Ghassemian, H. (2017). A probabilistic SVM approach for hyperspectral image classification using spectral and texture features. *International Journal of Remote Sensing*, *38*(15), 4265-4284.

[22]. Kuess, P., Andrzejewski, P., Nilsson, D., Georg, P., Knoth, J., Susani, M., ... & Nyholm, T. (2017). Association between pathology and texture features of multi parametric MRI of the prostate. *Physics in Medicine & Biology*, *62*(19), 7833.

[23]. Zhang, J., Geng, W., Liang, X., Li, J., Zhuo, L., & Zhou, Q. (2017). Hyperspectral remote sensing image retrieval system using spectral and texture features. *Applied optics*, *56*(16), 4785-4796.

[24]. Hameed, S. A. A., Radi, M. A. H., & Gaata, M. T. (2018). Medical Image Classification Approach Based on Texture Information. *Iraqi Journal of Information Technology*. *V*, *8*(3).

[25]. Ismael, M. R., & Abdel-Qader, I. (2018, May). Brain tumor classification via statistical features and back-propagation neural network. In *2018 IEEE international conference on electro/information technology (EIT)* (pp. 0252-0257). IEEE.

[26]. Kandemirli, S. G., Chopra, S., Priya, S., Ward, C., Locke, T., Soni, N., ... & Bathla, G. (2020). Presurgical detection of brain invasion status in meningiomas based on first-order histogram based texture analysis of contrast enhanced imaging. *Clinical Neurology and Neurosurgery*, *198*, 106205.

[27]. Qurat-Ul-Ain, G. L., Kazmi, S. B., Jaffar, M. A., & Mirza, A. M. (2010). Classification and segmentation of brain tumor using texture analysis. *Recent advances in artificial intelligence, knowledge engineering and data bases*, 147-155.

[28]. Latif, G., Iskandar, D. A., Alghazo, J. M., & Mohammad, N. (2018). Enhanced MR image classification using hybrid statistical and wavelets features. *IEEE Access*, *7*, 9634-9644.

[29]. Shaheen, F., Verma, B., & Asafuddoula, M. (2016, November). Impact of automatic feature extraction in deep learning architecture. In *2016 International conference on digital image computing: techniques and applications (DICTA)* (pp. 1-8). IEEE.

[30]. Shaikh, E., Mohiuddin, I., Manzoor, A., Latif, G., & Mohammad, N. (2019, October). Automated grading for handwritten answer sheets using convolutional neural networks. In *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)* (pp. 1-6). IEEE.

[31]. Butt, M. M., Latif, G., Iskandar, D. A., Alghazo, J., & Khan, A. H. (2019). Multi-channel Convolutions Neural Network Based Diabetic Retinopathy Detection from Fundus Images. *Procedia Computer Science*, *163*, 283-291.

[32]. Alghmgham, D. A., Latif, G., Alghazo, J., & Alzubaidi, L. (2019). Autonomous traffic sign (ATSR) detection and recognition using deep CNN. *Procedia Computer Science*, *163*, 266-274.

[33]. Lee, W. Y., Park, S. M., & Sim, K. B. (2018). Optimal hyperparameter tuning of convolutional neural networks based on the parameter-setting-free harmony search algorithm. *Optik*, *172*, 359-367.

[34]. Latif, G., Alghazo, J., Alzubaidi, L., Naseer, M. M., & Alghazo, Y. (2018, March). Deep convolutional neural network for recognition of unified multi-language handwritten numerals. In *2018 IEEE 2nd International workshop on Arabic and derived script analysis and recognition (ASAR)* (pp. 90-95). IEEE.

[35]. Abiwinanda, N., Hanif, M., Hesaputra, S. T., Handayani, A., & Mengko, T. R. (2019). Brain tumor classification using convolutional neural network. In *World congress on medical physics and biomedical engineering 2018* (pp. 183-189). Springer, Singapore.

[36]. Soltaninejad, M., Zhang, L., Lambrou, T., Yang, G., Allinson, N., & Ye, X. (2017, September). MRI brain tumor segmentation and patient survival prediction using random forests and fully convolutional networks. In *International MICCAI Brainlesion Workshop* (pp. 204-215). Springer, Cham.

[37]. Bhattacharjee, K., & Pant, M. (2019). Hybrid particle swarm optimization-genetic algorithm trained multi-layer perceptron for classification of human glioma from molecular brain neoplasia data. *Cognitive Systems Research*, *58*, 173-194.

[38]. Deepa, A. R. (2018, May). MRI brain tumor classification using cuckoo search support vector machines and particle swarm optimization based feature selection. In *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 1213-1216). IEEE.

[39]. Kaur, T., Saini, B. S., & Gupta, S. (2019). An adaptive fuzzy K-nearest neighbor approach for MR brain tumor image classification using parameter free bat optimization algorithm. *Multimedia Tools and Applications*, *78*(15), 21853-21890.

[40]. Brunese, L., Mercaldo, F., Reginelli, A., & Santone, A. (2020). An ensemble learning approach for brain cancer detection exploiting radiomic features. *Computer methods and programs in biomedicine*, *185*, 105134.

[41]. Anaraki, A. K., Ayati, M., & Kazemi, F. (2019). Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *biocybernetics and biomedical engineering*, *39*(1), 63-74.

[42]. Yap, F. Y., Varghese, B. A., Cen, S. Y., Hwang, D. H., Lei, X., Desai, B., ... & Duddalwar, V. A. (2020). Shape and texture-based radiomics signature on CT effectively discriminates benign from malignant renal masses. *European Radiology*, 1-11.

[43]. Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197-227.

[44]. Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T. L., Barrick, T. R., ... & 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.

[45]. Bisong, E. (2019). The Multilayer Perceptron (MLP). In *Building Machine Learning and Deep Learning Models on Google Cloud Platform* (pp. 401-405). Apress, Berkeley, CA.

[46]. Chen, S. G., & Wu, X. J. (2018). A new fuzzy twin support vector machine for pattern classification. *International Journal of Machine Learning and Cybernetics*, *9*(9), 1553-1564.

[47]. Xie, S., Li, Z., & Hu, H. (2018). Protein secondary structure prediction based on the fuzzy support vector machine with the hyperplane optimization. *Gene*, *642*, 74-83.

[48]. Foody, G. M., & Mathur, A. (2006). The use of small training sets containing mixed pixels for accurate hard image classification: Training

on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment*, *103*(2), 179-189.

[49]. Mahmoud, A. A., Alawadh, I. N. A., Latif, G., & Alghazo, J. (2020, April). Smart Nursery for Smart Cities: Infant Sound Classification Based on Novel Features and Support Vector Classifier. In *2020 7th International Conference on Electrical and Electronics Engineering (ICEEE)* (pp. 47-52). IEEE.

[50]. Shukla, A. K., Singh, P., & Vardhan, M. (2019). A new hybrid wrapper TLBO and SA with SVM approach for gene expression data. *Information Sciences*, *503*, 238-254.

[51]. Kharya, S., Agrawal, S., & Soni, S. (2014). Naive Bayes classifiers: a probabilistic detection model for breast cancer. *International Journal of Computer Applications*, *92*(10), 0975-8887.

[52]. De Campos, C. P. (2011, June). New complexity results for MAP in Bayesian networks. In *IJCAI* (Vol. 11, pp. 2100-2106).

[53]. Latif, G., Iskandar, D. A., Jaffar, A., & Butt, M. M. (2017). Multimodal brain tumor segmentation using neighboring image features. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *9*(2-9), 37-42.

[54]. Malini Devi, G., Seetha, M., & Sunitha, K. V. N. (2016). A Novel K-Nearest Neighbor Technique for Data Clustering using Swarm Optimization. *International Journal of Geoinformatics*, *12*(1).

[55]. Ajai, A. R., & Gopalan, S. (2020). Analysis of Active Contours Without Edge-Based Segmentation Technique for Brain Tumor Classification Using SVM and KNN Classifiers. In *Advances in Communication Systems and Networks* (pp. 1-10). Springer, Singapore.

[56]. Győrfi, Á., Kovács, L., & Szilágyi, L. (2019, October). Brain tumor detection and segmentation from magnetic resonance image data using ensemble learning methods. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (pp. 909-914). IEEE.

[57]. Farias, G., Dormido-Canto, S., Vega, J., Martinez, I., Alfaro, L., & Martinez, F. (2017). Adaboost classification of TJ-II Thomson Scattering images. *Fusion Engineering and Design*, *123*, 759-763.

[58]. Usman, K., & Rajpoot, K. (2017). Brain tumor classification from multi-modality MRI using wavelets and machine learning. *Pattern Analysis and Applications*, *20*(3), 871-881.

[59]. Lahoz-Beltra, R., & Rodriguez, R. J. (2020). Modeling a cancerous tumor development in a virtual patient suffering from a depressed state of mind: Simulation of somatic evolution with a customized genetic algorithm. *Biosystems*, *198*, 104261.

[60]. Hemanth, D. J., & Anitha, J. (2019). Modified Genetic Algorithm approaches for classification of abnormal Magnetic Resonance Brain tumour images. *Applied Soft Computing*, *75*, 21-28.

[61]. Sajjad, M., Khan, S., Muhammad, K., Wu, W., Ullah, A., & Baik, S. W. (2019). Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *Journal of computational science*, *30*, 174-182.

[62]. Raja, PM Siva. "Brain tumor classification using a hybrid deep autoencoder with Bayesian fuzzy clustering-based segmentation approach." *Biocybernetics and Biomedical Engineering* 40.1 (2020): 440-453.

[63]. Kumar, S., & Mankame, D. P. (2020). Optimization driven Deep Convolution Neural Network for brain tumor classification. *Biocybernetics and Biomedical Engineering*, *40*(3), 1190-1204.

[64]. Latif, G., Iskandar, D. N. F. A., Alghazo, J., & Butt, M. M. (2020). Brain MR Image Classification for Glioma Tumor detection using Deep Convolutional Neural Network Features. *Current medical imaging*.

[65]. Mzoughi, H., Njeh, I., Wali, A., Slima, M. B., BenHamida, A., Mhiri, C., & Mahfoudhe, K. B. (2020). Deep multi-scale 3D convolutional neural network (CNN) for MRI gliomas brain tumor classification. *Journal of Digital Imaging*, *33*, 903-915.

[66]. Arasi, P. R. E., & Suganthi, M. (2019). A clinical support system for brain tumor classification using soft computing techniques. *Journal of medical systems*, *43*(5), 1-11.

[67]. Srinivas, B., & Rao, G. S. (2019). A Hybrid CNN-KNN model for MRI brain tumor Classification. *International Journal of Advanced Science and Technology (IJAST)*, *127*, 20-25.

[68]. Chinnam, S., Sistla, V. P. K., & Kolli, V. K. K. (2019). SVM-PUK Kernel Based MRI-brain Tumor Identification Using Texture and Gabor Wavelets. *Traitement du Signal*, *36*(2), 185-191.

[69]. Seetha, J., & Raja, S. S. (2018). Brain tumor classification using convolutional neural networks. *Biomedical & Pharmacology Journal*, *11*(3), 1457.

[70]. Sriramakrishnan, P., Kalaiselvi, T., Nagaraja, P., & Mukila, K. (2018). Tumorous slices classification from MRI brain volumes using block based features extraction and random forest classifier. *Int J Comput Sci Eng May*, *6*(4), 191-6.

[71]. Amin, J., Sharif, M., Yasmin, M., & Fernandes, S. L. (2017). A distinctive approach in brain tumor detection and classification using MRI. *Pattern Recognition Letters*.

[72]. Latif, G., Iskandar, D. A., Alghazo, J., & Jaffar, A. (2018). Improving brain MR image classification for tumor segmentation using phase congruency. *Current Medical Imaging*, *14*(6), 914-922.

[73]. Wasule, V., & Sonar, P. (2017, May). Classification of brain MRI using SVM and KNN classifier. In *2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS)* (pp. 218-223). IEEE.

[74]. Bahadure, N. B., Ray, A. K., & Thethi, H. P. (2017). Image analysis for MRI based brain tumor detection and feature extraction using biologically inspired BWT and SVM. *International journal of biomedical imaging*, *2017*.

[75]. Swati, Z. N. K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., & Lu, J. (2019). Brain tumor classification for MR images using transfer learning and fine-tuning. *Computerized Medical Imaging and Graphics*, *75*, 34-46.

[76]. Amin, J., Sharif, M., Gul, N., Yasmin, M., & Shad, S. A. (2020). Brain tumor classification based on DWT fusion of MRI sequences using convolutional neural network. *Pattern Recognition Letters*, *129*, 115-122.

[77]. Khan, M. A., Ashraf, I., Alhaisoni, M., Damaševičius, R., Scherer, R., Rehman, A., & Bukhari, S. A. C. (2020). Multimodal brain tumor classification using deep learning and robust feature selection: A machine learning application for radiologists. *Diagnostics*, *10*(8), 565.

[78]. Xue, Y., Yang, Y., Farhat, F. G., Shih, F. Y., Boukrina, O., Barrett, A. M., ... & Roshan, U. W. (2019, October). Brain tumor classification with tumor segmentations and a dual path residual convolutional neural network from MRI and pathology images. In *International MICCAI Brainlesion Workshop* (pp. 360-367). Springer, Cham.

[79]. Rehman, A., Khan, M. A., Saba, T., Mehmood, Z., Tariq, U., & Ayesha, N. (2021). Microscopic brain tumor detection and classification using 3D CNN and feature selection architecture. *Microscopy Research and Technique*, *84*(1), 133-149.

[80]. Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T. L., Barrick, T. R., ... & 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.