Stacked Sparse Autoencoder and Softmax Classifier Framework to Classify MRI of Brain Tumor Images

Ahmed Aldhahab[1]*   **Sarmad Ibrahim[1]**   **Wasfy B. Mikhael[2]**

[1] Department of Electrical Engineering, University of Babylon, Babylon, Iraq
[2] Department of Electrical and Computer Engineering, University of Central Florida, Orlando, Florida, USA
* Corresponding author's Email: aldhahab2012@knights.ucf.edu

Abstract: Classification of a brain tumor is a critical step in the design of computer-aided diagnosis systems for Magnetic Resonance Image (MRI) analysis. This work presents an efficient algorithm to classify a tumor in brain MRI images using statistical-based features and deep neural network. Data, within the region of interest, is transformed into two-dimensional discrete Gabor filter and wavelet transform. These filters are combined in this algorithm as directional transformation methods for utilizing all information in all orientations of the MRI input image. MRI Features are extracted based on the first and second order statistics from both domains. Two types of neural network classifiers are employed: Stacked Sparse Autoencoder (SSA) and Softmax Classifier (SMC). Two regularization functions are used in the training of the SA, sparsity regularization and L2-weight regularization. Sparsity regularization controls the firing of the neurons in the hidden layer, whereas L2-weight regularization reduces the effect of the overfitting and improves the performance of the SA. Two datasets are used to evaluate the proposed algorithm. The first dataset consists of 3,064 of T1-weighted MRI slices with three kinds of tumors: Pituitary, Glioma, and Meningioma. The second dataset consists of 200 MRI slices with low-grade and high-grade Glioma tumor collected from the BRATS dataset. The performance of the proposed algorithm is validated using the experimental results in terms of accuracy, specificity, and sensitivity compared to the existing algorithms. For the first dataset, the accuracy obtained is 94.0%, the sensitivity of Meningioma, Glioma, and Pituitary is 87.44%, 97.29%, and 94.27%, respectively, and the specificity of Meningioma, Glioma, and Pituitary is 98%, 96.89%, and 96.78%, respectively. For the BRATS dataset, the accuracy, the specificity, and the sensitivity achieved are 98.8%, 100%, and 100%, respectively.

Keywords: Autoencoder, Brain tumor, Gabor filter, Softmax classifier, Statistical features.

1. Introduction

Based on the American Cancer Society, 23,890 malignant tumors in the Central Nervous System (CNS) will be diagnosed in 2020. In addition, 18,020 patients will die due to a CNS tumor [1]. Consequently, scientists in the field of medicine, computer science, and engineering are working on developing new techniques to diagnose and treating brain tumors effectively [2]. Two approaches are used to diagnose a brain tumor, invasive and non-invasive. In invasive, the radiologist takes a sample from the affected area of the brain for the exam using biopsy or spinal tap methods. The most common modality for non-invasive diagnoses of brain tumors

is the Magnetic Resonance Imaging (MRI) due to its high sensitivity to local changes in tissue water, its high resolution especially in differentiating soft tissues, and its ability to create multiple images with different contrast visualization when examining the same tissue.

This is how it helps physicians and radiologists to study the scanned tissue more precisely [3]. Manual diagnosis is based on the radiologist's viewing, which is time consuming due to the large amount of MRI slices to be analyzed [4]. Usually, physicians and radiologists have faced some difficulties in some sophisticated cases in the classification stage. These sophisticated cases require specialists having expertise in order to find and localize the tumor, test the tumor tissue with the

surrounding region, employ filters if necessary, adjust the tumor's region to be more clear for physicians and radiologists vision, and finally conclude whether there is a tumor or not along with specifying the tumor type and grade level if available. These tasks to be accomplished will consume time and might end up with wrong diagnosis. Therefore, other tools are needed to support both physicians and radiologists in their diagnosis [5]. Naturally, an invasive based diagnosis such as spinal tap and biopsies are risky, time-consuming, expensive, and painful [6]. All these drawbacks and concerns encourage the radiologists to take advantage of any aided diagnosis tools that are available. Computer-aided diagnosis system (CAD) is aimed to support the radiologists and not to replace them. Indeed, the proposed algorithm can be considered as a second opinion for radiologists. The information obtained from the MRI images is also considered as a second diagnosis. Moreover, a well-designed CAD system will minimize or prevent human errors, such as missing readings that are caused by fatigue, overlooking, and data overloading when analyzing the MRI slices, and allow for better health care [7].

The main contribution of this paper is the implementation of the Stacked Sparse Autoencoder (SSA) and the Softmax classifier (SMC) for classifying the type and the grade of the MRI brain tumors. Unlike the BPNN classifier, the proposed classification technique trains each layer separately, controls the firing of each neuron in the hidden layer, and reduces the overfitting. This ability is resulted from using sparsity regulation in the training of the SA. Also, the Softmax function used by the SMC is designed for the multiclass logistic regression (LR), while the sigmoid function used by the BPNN classifier is restricted to two-class LR.

Brain tumors included in this study are: Meningioma, Glioma including the high-grade and low-grade, and Pituitary tumors. Meningioma is a homogeneous tumor with high intensity region. Glioma is a heterogeneous tumor with low intensity region and the tumor border is brighter than inside region. Pituitary is a heterogeneous tumor ranging from high to low intensity [8].

The main image attributes used to describe these types and grades of the brain tumors are the intensity and texture of the tumor regions. Therefore, statistical features that can be extracted from these two main image attributes are proposed in this algorithm along with other measures to quantify the image attributes. Features are generated using first and second order statistics in addition to other measures such as energy and entropy. The features allow the identification of these types/grades of brain tumors via machine learning algorithms. To enhance classification performance, the high level of accountability of all tumors' characteristics can be accomplished by performing efficient model and feature representation. The proposed algorithm is not only extracting features from the spatial domain, but also from other domains to achieve high recognition rates that yield to better class discrimination. Hence, another set of features obtained by other domains, Wavelet and Gabor, is proposed. Based on the results shown later in the Tables 4-7, the proposed algorithm shows significant improvements in the classification rates, specificity, as well as sensitivity in comparison with the existing algorithms.

The remainder of this paper is organized as: a brief background is presented in section 2, a description of the methodology used is described in Section 3. Section 4 presents the experimental setup and the simulation results. Concluding remarks are presented in Section 5.

2. Background

CAD system is an application of pattern recognition that aims to help physicians and radiologists to make a proper diagnosis decision taking into consideration that the final opinion about the examined case is made by the radiologists [9]. Different methods have been proposed to develop a CAD system that can detect or classify abnormal tissues from brain MRI images [10]. These methods have three main steps: feature extraction, dimensionality reduction, and classification. Two-Dimensional Discrete Wavelet Transform (2D DWT) and texture features are the most common techniques used to extract features from brain MRI images. The wavelet transform is a useful tool for image analysis since it provides significant information with less time [11]. Texture features are extracted from the image using either statistical features or the Gabor filter [12]. Statistical features are determined by first or second order statistics. The histogram of an image is used to extract the first order statistical features [13], while the second order features are determined by using Gray-Level Co-occurrence matrix (GLCM), Gray Level Run Length Matrix (GLRM) [14], or Neighborhood Gray Tone Difference Matrix (NGTDM) [15] are employed to extract the second order statistical features. Statistical features can be extracted from the sub-band images that resulted from the 2D DWT. In the literature, some authors extract these features from the low-low frequency (LL) sub-band [12, 16] while other works suggested using the low-high (LH) and high-low frequencies (HL) sub-bands images for statistical features

calculations [6, 17]. The Gabor filter has been widely used to classify the brain tumor by the texture features [12]. The Gabor filter has good characteristics in both time and frequency domains. Hence, statistical features can be obtained in the spatial frequency domain using the images resulting from the 2D Gabor filter implemented with different frequencies and orientations [18]. Dimensionality of the images is the main challenge that faces the classification process. The extra features increase the time and memory storage required for the CAD system implementation [7]. The most common technique used in the literature for the feature reduction in brain tumor classifications is the Principal Component Analysis (PCA) [4, 19]. PCA is an efficient tool that transforms a high dimensional input feature with correlated variables into a low dimensional uncorrelated feature vector using orthogonal transformation [3, 4].

There are other techniques for feature selection and dimensionality reduction that are employed in the classification of brain tumors, such as the Genetic Algorithm (GA) [20] and the Cumulative Variance Method (CVM) [12]. Classification is the most significant step in the CAD system since it gives the final decision about the required class based on the extracted features of the MRI image [4, 7]. To classify the brain tumor in MRI images, authors have used many different techniques. The most widely used classification techniques are neural networks (NN) and support vector machines (SVM). The NN has been proved as an efficient tool for classification purposes due to its ability to adjust data and approximate any function with arbitrary accuracy. Few architectures were used to implement NN classifications, such as the feedforward NN, the multilayer perceptron NN trained with backpropagation algorithm (BPA) and the Probabilistic NN (PNN) [21]. SVM is a new type of classifier based on a statistical learning technique that minimizes the error by maximizing the margin between the separating hyperplane and the data [22]. Three kernels were used to implement the SVM classifier: Gaussian radial basis function, inhomogeneous polynomial, and homogenous polynomial [13, 19]. J. Cheng H. Wei and C. Shuangliang [23] proposed a technique to extract features using three methods, which are GLCM, the Bag of Words model, and the intensity histogram, and the SVM for recognition. In [24], Vani and Geetha proposed a classification algorithm for brain tumor using three types of classifiers: SVM, KNN, and Decision Tree. Wasule and Sonar in [25] suggested an algorithm based on SVM and KNN
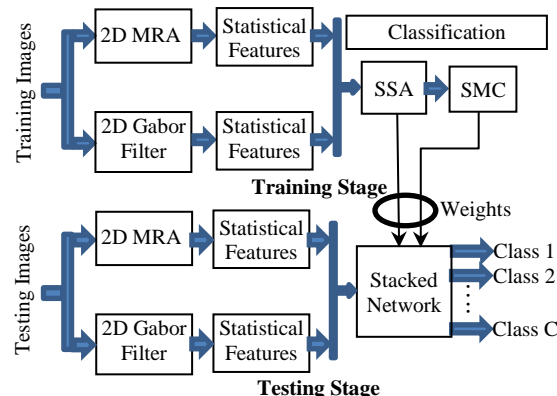

Figure. 1 Block diagram of the proposed algorithm

classifiers for brain tumor using the GLCM. In [26], Farhi and Yusuf presented a survey on MRI image classification using machine learning techniques. Furthermore, five different classifiers are used: ANN, Decision Tree, KNN, Nave Bayes, and SVM. Also, in [26], GLCM and PCA were employed for feature selection and dimensionality reduction.

3. Proposed algorithm

The proposed algorithm contains two phases: the training phase and the testing phase as shown in Fig. 1. In each phase, there are two main steps: feature extraction and classification. The feature extraction step is implemented by using 2D DWT in conjunction with the 2D Gabor filter followed by a statistical calculation. In the classification step, the SSA is trained and stacked with the SMC during the training step. Afterward, SSA is used to classify the generated features to its classes during the testing phase. The image in each phase is either the tumor region or ROI. ROI is extracted either manually by a radiologist or automatically by using a preprocess automated segmentation algorithm.

3.1 Feature extraction

The choice of the feature extraction technique plays a vital role in the performance of the classifier. High order statistical features that are related to the spatial domain tumor attributes are obtained from the 2D DWT and the 2D Gabor filter. The 2D DWT is an efficient tool for image representation, and the Gabor filter is a powerful tool in measuring the heterogeneity and the textural analysis. The 2D DWT decomposes the input image into four sub-band images: the approximation sub-band (LL), the horizontal detail sub-band (LH), the vertical detail sub-band (HL), and the diagonal detail sub-band (HH). As a result, three-directional features (i.e.,

271

Table 1. A list of selected statistical features

Feature	Description	Formula
Mean (m)	The average intensity of the image	$m = \sum_{i=0}^{L-1} i\, p(i)$
Variance (μ)	The second moment about the mean	$\mu = \sum_{i=0}^{L-1} (i - m)^2\, p(i)$
Skewness	The third moment about the mean	$Skewness = \sum_{i=0}^{L-1} (i - m)^3\, p(i)$
Kurtosis	The fourth moment about the mean	$Kurtosis = \sum_{i=0}^{L-1} (i - m)^4\, p(i)$
Contrast	A measure of the local variations in the GLCM	$Contrast = \sum_{i=1}^{A} \sum_{j=1}^{A} (i - j)^2\, p(i,j)$
Correlation	A measure of the correlation of a pixel with its neighbor	$Correlation = \sum_{i=1}^{A} \sum_{j=1}^{A} (i - m1)(j - m2)\, \dfrac{p(i,j)}{\sigma_1 \sigma_2}$
Energy	A measure of uniformity	$Energy = \sum_{i=1}^{A} \sum_{j=1}^{A} \left(p(i,j)\right)^2$
Homogeneity	Measures the spatial closeness of GLCM elements to the diagonal	$Homogeneity = \sum_{i=1}^{A} \sum_{j=1}^{A} \dfrac{p(i,j)}{1 + |i - j|}$
Entropy	A measure of the randomness in the GLCM	$Entropy = -\sum_{i=1}^{A} \sum_{j=1}^{A} p(i,j)\, log\left(p(i,j)\right)$
Maximum probability	The maximum probability in GLCM	$Maximum\ probability = MAX_{i,j}\ p(i,j)$

horizontal, vertical, and diagonal features) are three directions are not enough for expressing all the directional features in the images, especially for the medical images. Other types of transformation such as Gabor filter are needed for better directional features' representation [27]. The Gabor filter analyzes the edges of the input image and produces several images with different wavelengths and orientations. In addition, Gabor filter captures visual properties of the image represented by spatial localization, orientation selectivity, and spatial frequency [28].

In this paper, to utilize all the directional features of the MRI image, 2D DWT and 2D Gabor filters are combined as directional transformation methods. Then, the statistical features are obtained from the resulting images for the classification. Combining the 2D DWT is able to obtain better accuracies in comparison with those obtained using each filter separately.

3.1.1. Gabor filter and multi-resolution analysis

Gabor filter is a linear band pass filter whose kernel function is generated by modulating a Gaussian function with sinusoidal wave, and it is described by a specific frequency and orientation. The Gabor filter has an advantageous in computer vision and image processing, especially for texture characterization. This is due to its frequency and orientation are like human visual system as well as its optimal localization properties in both spatial and frequency domains [29, 30]. The general form of the 2D Gabor filter is defined as:

$$G(x, y, \lambda, \theta, \psi, \sigma, \gamma) = e^{-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}}\, e^{i\left(2\pi \frac{x'}{\lambda} + \psi\right)} \quad (1)$$

where G is the output of the Gabor filter, x and y are the indices of the input image, , ψ is the phase offset, θ is the orientation of the Gabor function, λ is the wavelength of the sinusoidal form, σ is the standard deviation of the Gaussian envelope, γ is the spatial aspect ratio, and $x' = x\, cos(\theta) + y\, sin(\theta)$ and $y' = -x\, sin(\theta) + y\, cos(\theta)$ [18].

The wavelet transform is used for image representation due to its ability to analyze the time and frequency contents of an image simultaneously
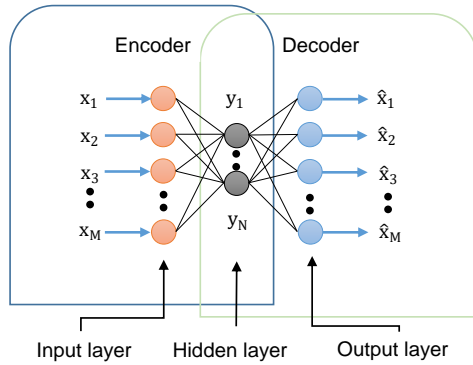
Figure. 2 The basic autoencoder

convolved with a function called the mother wavelet $\psi_{a,b}(x)$, where $a \in \mathbb{R}^+\backslash\{0\}$ and $b \in \mathbb{R}$ are the dilation and the translation of the $\psi_{a,b}(x)$. This is expressed as:

$$W_\psi(a,b) = \int_{-\infty}^{\infty} f(x)\,\psi_{a,b}(x)\,dx \qquad (2)$$

where $W_\psi(a,b)$ is the wavelet coefficients of the function $f(x)$. For the discrete input signal, the DWT, multi-resolution analysis (MRA), is obtained by the dilation and the translation of the two functions: the mother wavelet $\psi(x)$ and the scaling function $\varphi(x)$. These are defined by:

$$\varphi(x) = \sqrt{2}\sum_{n=-\infty}^{\infty} h_o[n]\,\varphi(2x-n), n \in \mathbb{Z} \qquad (3)$$

$$\psi(x) = \sqrt{2}\sum_{n=-\infty}^{\infty} h_1[n]\,\varphi(2x-n), n \in \mathbb{Z} \qquad (4)$$

where $h_o[n]$ and $h_1[n]$ are the scaling and the wavelet functions coefficients, respectively [32]. For image analysis, the 2D DWT, which is an extension of the 1-D DWT, is applied to the row variable and then to the column variable [27]. The detail and the approximation coefficients are utilized for better texture representation.

3.1.2. Statistical features

In the classification of the biomedical images, statistical features are used to describe the texture of the input image [17]. First-order statistical features are obtained by employing the histogram of an image, whereas the second-order statistical features are calculated from the GLCM [32]. In this algorithm, four first-order statistics, which are kurtosis, skewness, variance, and mean, are used. Also, six second-order statistics, which are Contrast, Homogeneity, Energy, Correlation, Maximum

probability, and Entropy are employed as shown in Table 1.

In Table 1: $p(i)$ is the image histogram, $p(i,j)$ is the element of GLCM, $(i,j) \in \mathbb{Z}^+$, m_1 and m_2 are the mean of rows and columns, respectively, σ_1 and σ_2 are the standard deviation of rows and columns, respectively, A is the GLCM size, and L is the number of intensity levels.

3.2 Classification

For classification, the advantage of a classifier model that comprises the two types of NNs, the SSA and SMC, is taking into consideration. The SSA is designed for feature learning while the SMC is designed for multiclass classification.

3.2.1. Autoencoder

The basic autoencoder is a NN that is trained to learn deep NN for concise feature representation. It consists of three layers, the input layer, the hidden layer, and the output layer as shown in Fig. 2. The hidden layer is called the code layer since it describes a code used to represent the input. The training is employed using the BPA with unsupervised learning to make the output of the network akin to the input. The autoencoder has an encoder-decoder architecture, where the encoder maps or encodes the input vector (X) into another vector (Y) with less dimensions. The drawback of the basic autoencoder is the problem of inconsiderable solutions that affect the performance of the network. In addition, it is unable to learn useful features when the size of the hidden layer is greater than the size of the input layer. To avoid this problem is to introduce a regularization function in the hidden layer and control the firing of neurons in that layer. In this case, each neuron in the hidden layer will fire to a small number of training examples or vectors [33, 34].

For training set $\{X(i), D(i)\}$, $X(i)$ is the input features and $D(i)$ is the labels. The cost function $\mathbb{C}(i)$ of sparse autoencoder is given as [35]:

$$\mathbb{C}(i) = \frac{1}{S}\sum_{i=1}^{S}\sum_{k=1}^{M}(X_k(i) - \hat{X}_k(i))^2 + a_1 SR$$
$$+ a_2 LR \qquad (5)$$

The first term is the cost function of the basic autoencoder, $X_k(i)$ is the input vector, $\hat{X}_k(i)$ is the estimated input vector, $(i,k) \in \mathbb{Z}^+$, S is the number of samples, and M is the size of the input vector. The second and third terms are the sparsity regularization (SR) and L2-weight regularization (LR), respectively,

273

Meningioma   Glioma   Pituitary   High-grade Glioma   Low-grade Glioma

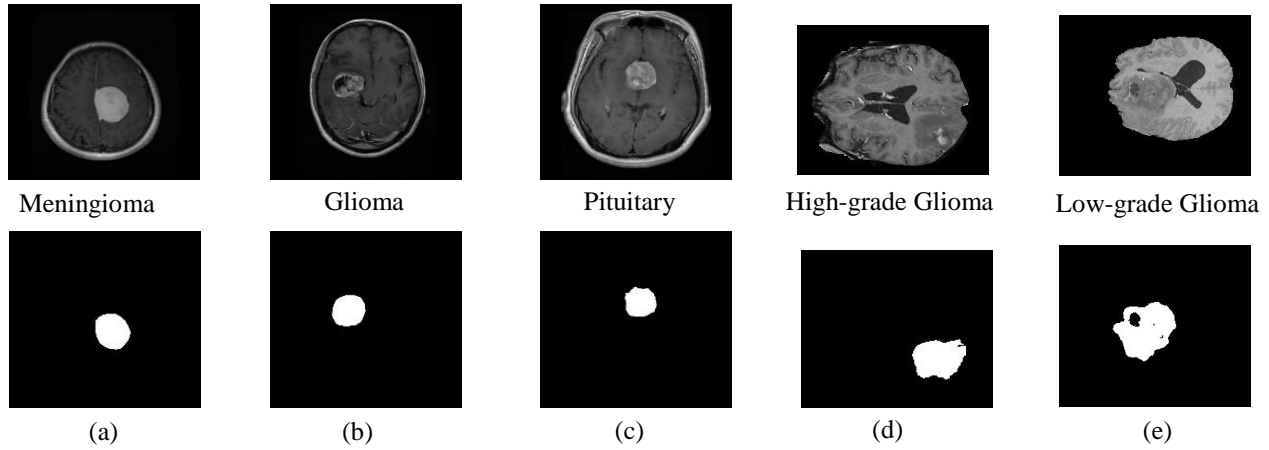(a)   (b)   (c)   (d)   (e)

Figure. 3 Sample images from the databases, (a), (b), (c), (d), (e), and its tumour region respectively

where a_1 and a_2 are the coefficients of the second and third terms, respectively. The SR is calculated using the Kullback-Leibler divergence method, which is a measure of how two distributions are different. It is expressed as [34]:

$$SR = \sum_{j=1}^{N} \eta \, log\left(\frac{\eta}{\hat{\eta}_j}\right) + (1 - \eta) \, log\left(\frac{1 - \eta}{1 - \hat{\eta}_j}\right) \quad (6)$$

where N is the number of neurons in the hidden layer, η is the sparsity parameter that is close to 0, and $\hat{\eta}_j$ refers to the average of activation of neuron in the hidden layer, and $j \in \mathbb{Z}^+$. $\hat{\eta}_j$ is calculated as:

$$\hat{\eta}_j = \frac{1}{S} \sum_{i=1}^{S} f\left(\sum_{k=1}^{N} (w_{jk} \, x_k(i)) + b_j\right) \quad (7)$$

where w_{jk} is the weight between neuron k in layer l_1 and neuron j in layer l_{1+1}, $l \in \mathbb{Z}^+$, $x_k(i)$ is the kth training sample, and b_j is the jth entry of the bias vector. The third term is introduced herein to decrease the magnitude of the weight and reduce the effect of overfitting. LR is determined as [34, 36]:

$$LR = \frac{1}{2} \sum_{l=1}^{L-1} \sum_{i=1}^{N_l} \sum_{j=1}^{N_{l+1}} (W_{ji}^{(l)})^2 \quad (8)$$

where $L \in \mathbb{Z}^+$ is the number of layers, N_l and N_{l+1} are the number of neurons in the layers l, $l+1$, respectively, and $W_{ji}^{(l)}$ is the weight between neuron i in layer l_1 and neuron j in layer l_{1+1}.

3.2.2. Softmax classifier (SMC)

The Softmax classification layer is a multi-class generalized logistic regression. It is a type of NNs designed in such a way such that the activation function of the output layer guarantees that the outputs sum up to unity and lie within the range [0 1]. The Softmax activation function is given as [9]:

$$\sigma_k = \frac{e^{y_k}}{\sum_{j=1}^{C} e^{y_j}} \, , \, k = 1, \dots, C, y = \{y_1, .., y_C\} \quad (9)$$

where σ_k is the output of the Softmax activation function, i.e. σ_k is the output of the output layer, y_k is an element in the input vector y, which is the entry to the Softmax activation function σ_k, to the output layer, C is the number of neurons in the output layer, i.e. the number of classes C.

4. Experimental results

Two databases are used to train and test the algorithm proposed. 233 patients with 3,064 brain MRI slices are in the first dataset. The first dataset contains three types of brain tumors: Pituitary, Glioma, and Meningioma. These slices were acquired from Nanfung Hospital, Guangzhou city, and Tianjin Medical University General Hospital in China during the period from 2005 to 2010. These slices were generated using a T_1-weighted MRI sequence with 512×512 size pixels and were manually segmented by three expert radiologists to generate a tumor region. Each tumor has a number of slices, which are: 930 slices for Pituitary tumor, 708 slices for Meningioma, and 1,426 slices for Glioma [23].
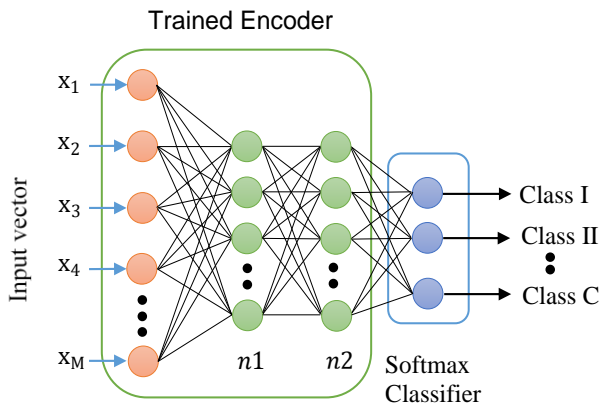
Figure. 4 Architecture of the proposed classifier model

Figure. 5 Classification accuracies for different types of wavelet filters

The second database was obtained from the BRATS (Multimodal Brain Tumor Image Segmentation) database, which has high-grade and low-grade MRI images of patients [37]. For classification purpose, 200 MRI slices were collected from 30 patients with Glioma tumors, 20 patients with high-grade Glioma tumors and 10 patients with low-grade Glioma tumors. Samples of the two databases along with their tumor regions are shown in Fig. 3.

The first dataset is partitioned into two sets: the training set represented by 70 % of the dataset, which contains 633, 493, and 1,019 slices for Pituitary, Meningioma, and Glioma tumors, respectively. The testing set represented by 30% of the dataset, which is results in 297, 215, and 407 slices for Pituitary, Meningioma, and Glioma tumors, respectively. The second dataset is divided to: the training set with 120 slices in which there are 64 high-grade Glioma tumors and 56 low-grade Glioma tumors, and the testing set with 80 slices in which there are 36 high-grade Glioma tumors and 44 low-grade Glioma tumors.

The 2D DWT is applied to the MRI tumor image using three levels of decompositions employing Symlets4 filter. Hence, 12 images are resulted from using all wavelet sub-bands features of each level. The 2D Gabor filter is applied to the MRI tumor image using three wavelengths, which are 2, 4, and 8, and five orientations, which are 0°, 45°, 90°, 135°, and 180° resulting in 15 images. The 10 statistical features mentioned in Table 1 are applied to each image resulted from the 2D DWT and Gabor filter. Thus, the total number of statistical features of each tumor slice obtained from the 2D DWT, the 2D Gabor filter, and the combination are 120, 150, and 270, respectively.

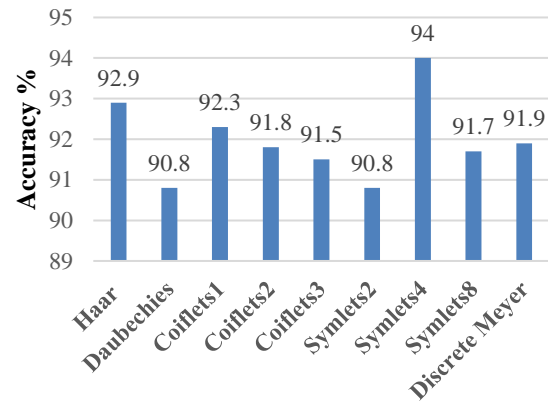Proposed classifier model is constructed by stacking two neural networks, which are SSA and SMC. The SSA network comprises four layers: an input layer, two hidden layers, and an output layer.

The size of the input and output layers is M, which is the same size of the input features that are 270 neurons. The size of the first hidden layer is n_1 with 150 neurons, while the size of the second hidden layer is n_2 with 75 neurons. The sparsity regularization coefficient and the L2-weight regularization coefficient are set to be 1 and 0.001, respectively. The SMC network has two layers: input and output layer as shown in Fig. 4. The input layer is of size 75 neurons. The output layer is of size C. where C is either 3 for the first dataset or 2 for the BRATS dataset.

Three stages are utilized to train the classifier. In the first stage, the autoencoder is trained by the BPA to optimize the cost function in an unsupervised manner. In the second stage, the SMC is trained using a supervised learning method, which is based on BPA with a cross-entropy cost function. The third stage is performed after training the autoencoder. The encoder part is separated and combined with the SMC to generate a multilayer deep NN, which is trained by fine-tuning its weights using the BPA.

The classification capability is quantified by three commonly used performance criteria: accuracy, sensitivity, and specificity [38].

$$Accuracy~\% = \frac{N_{TP} + N_{TN}}{N_{FP} + N_{FN} + N_{TP} + N_{TN}} \quad (10)$$

$$Sensitivity~\% = \frac{N_{TP}}{N_{FN} + N_{TP}} \quad (11)$$

$$Specificity~\% = \frac{N_{TN}}{N_{FP} + N_{TN}} \quad (12)$$

275

Table 2. Performance analysis for the first dataset using the wavelet features

Classifier Performance		LL%	LH and HL%	All sub-bands%
Accuracy	Meningioma	89.55	75.95	92.38
	Glioma	94.23	81.06	96.08
	Pituitary	89.66	85.53	91.73
Sensitivity	Meningioma	80	50.23	83.72
	Glioma	94.1	83.05	96.56
	Pituitary	81.48	70.37	85.86
Specificity	Meningioma	92.47	83.81	95.03
	Glioma	94.34	79.49	95.7
	Pituitary	93.57	92.76	94.53

Table 3. Performance analysis for the BRATS dataset using the wavelet features

Classifier Performance	LL sub-band%	LH and HL sub-bands%	All sub-bands%
Accuracy	92.5	92.5	95
Sensitivity	97.22	100	100
Specificity	88.64	86.36	90.9

where N_{TP} and N_{TN} are number of true positive and negative, respectively, N_{FP} and N_{FN} are number of false positive and negative, respectively.

To examine the effect of the wavelet filter types on the performance of the proposed algorithm, different types of wavelet filters are used. These filters are: Haar, Daubechies 2, Coiflets1, Coiflets2, Coiflets3, Symlets2, Symlets4, Symlets8, and Discrete Meyer. Fig. 5 summarizes the accuracies of

the proposed algorithm employing each type of the wavelet filters. Using the wavelet transform, the statistical features can be extracted from either the LL sub-band [12, 16] or from the LH and HL sub-bands [6, 17]. Combining the detail and approximation coefficients can enhance the discrimination capability of the recognition [17]. The proposed algorithm is implemented using the features obtained from the LL sub-band, the LH and HL sub-bands, and all the wavelet sub-bands of the three levels of the 2D DWT decompositions. The performance of the proposed algorithm of the two databases is shown in the Tables 2 and 3.

The intensity and the texture of the MRI tumor image are mostly described using the LL sub-band, whereas the LH and HL sub-bands capture either the edges or the high-frequency components. Since the intensity and the texture are the main concerns in classifying the tumor types/grades in the MRI images, the features extracted from the LL sub-bands are better than those obtained using the LH and HL sub-bands. Combining all the wavelet sub-bands features improves the performance of the classifier compared to other sub-bands since it has all the features contained in the MRI image.

The classifier is trained using: 1. Wavelet statistical features, 2. Gabor statistical features, and 3. An integration of Gabor and Wavelet statistical features. The proposed algorithm is assessed by sensitivity, accuracy, and specificity for each type/ grade of the tumor. In Tables 4 and 5, the proposed

Table 4. Results of the proposed algorithm for the first dataset (100%)

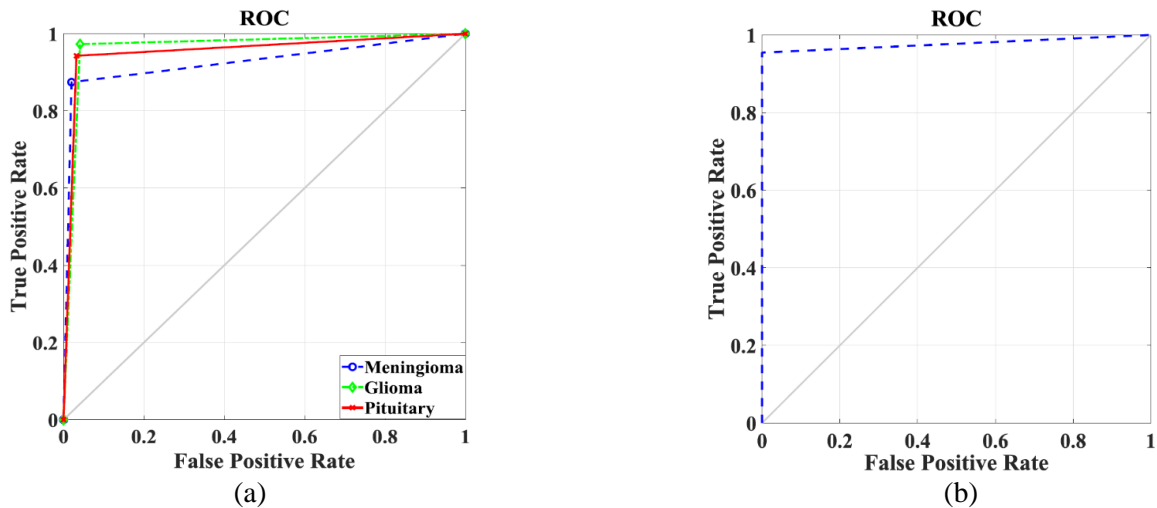| Classifier Performance			Wavelet statistical Features	Gabor statistical Features	Wavelet and Gabor statistical Features
The Proposed Algorithm Using the SSA and SMC	Accuracy	Meningioma	92.38	93.03	95.54
		Glioma	96.08	93.14	96.52
		Pituitary	91.73	94.23	95.97
	Sensitivity	Meningioma	83.72	85.12	87.44
		Glioma	96.56	93.36	97.29
		Pituitary	85.86	89.56	94.27
	Specificity	Meningioma	95.03	95.45	98
		Glioma	95.7	92.97	95.89
		Pituitary	94.53	96.46	96.78
The Proposed Algorithm Using the BPNN Classifier	Accuracy	Meningioma	90.1	91.62	93.91
		Glioma	94.89	91.73	95.75
		Pituitary	90.21	92.27	94.23
	Sensitivity	Meningioma	82.33	83.72	86.97
		Glioma	94.1	91.65	95.1
		Pituitary	82.5	85.52	91.24
	Specificity	Meningioma	92.47	94.03	96
		Glioma	95.51	91.8	96.29
		Pituitary	93.89	95.5	95.66

Figure. 6 ROC Curves for the classification model, (a) first dataset, (b) BRATS dataset
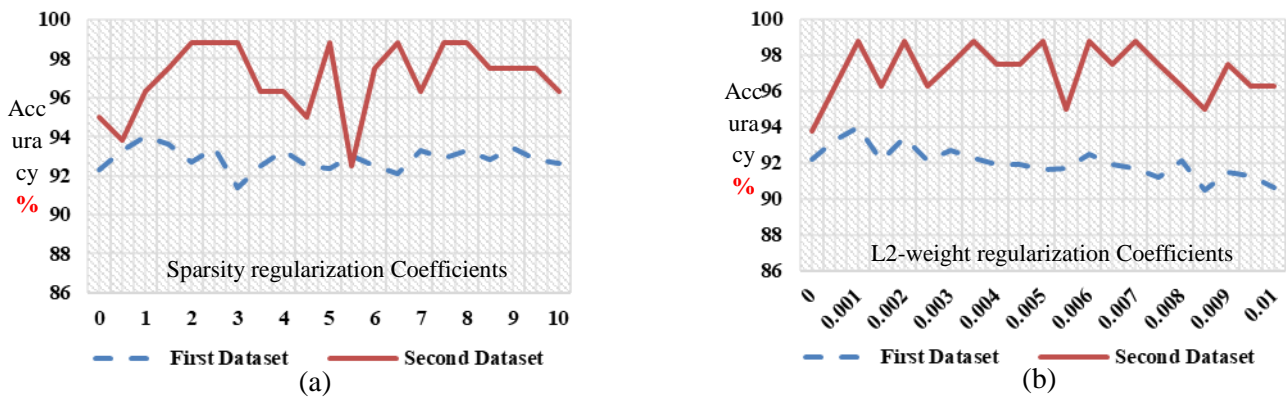

Figure. 7 The effect of regularization coefficients on the accuracy: (a) Sparsity and (b) L2-wight

Table 5. Results of the proposed algorithm for BRATS dataset (100%)

Classifier Performance		Wavelet statistical Features	Gabor statistical Features	Wavelet and Gabor statistical Features
The Proposed Algorithm Using the SSA and SMC	Accuracy	95	96.3	98.8
	Sensitivity	100	100	100
	Specificity	90.9	93.18	100
The Proposed Algorithm Using the BPNN Classifier	Accuracy	93.8	95	96.3
	Sensitivity	100	100	100
	Specificity	88.64	90.9	93.18

algorithm is performed using the SSA and SMC and the BPNN classifier. The BPNN classifier is employed using the multilayer perceptron, which comprises the 3 layers: 270 neurons for the input layer, 90 neurons for the hidden layer, and 3 neurons for the output layer. As can be seen in Tables 4 and 5,

the results achieved by employing the proposed classifier are higher than those obtained using the BPNN classifier in terms of accuracy, specificity, as well as sensitivity.

In Fig. 6 (a), the receiving operating characteristics (ROC) curves are showing the performance of the proposed algorithm and the relation between the true positive rate and the false positive rate for both datasets. The area under the ROC curve of the first dataset is 0.927, 0.966, and 0.955 for Meningioma, Gliomas, and Pituitary tumors, respectively. In Fig. 6 (b), the area under the ROC curve for the BRATS dataset is 0.9886.

Regularization constraints increase the performance of the autoencoder network by controlling the firing of the neurons in the hidden layer and reducing the magnitude of the weight. This regularization is introduced in the design of the autoencoder network using two regularization terms, sparsity regularization and L2-weight regularization. For each term, there is a coefficient that controls the functionality of its regularization inside the cost

Table 6. Comparison Results for first dataset

Classifier Performance	The Proposed Algorithm%	Method [23]%
Accuracy	94	91.28
Sensitivity-Meningioma	87.44	86.0
Sensitivity - Glioma	97.29	96.4
Sensitivity - Pituitary	94.27	87.3
Specificity-Meningioma	98	95.5
Specificity-Glioma	96.89	96.3
Specificity-Pituitary	96.78	95.3

Table. 7 Comparison Results for BRATS dataset

Classifier Performance	The Proposed Algorithm %	Method [24]%	Method [25]%	Method [26]%
Accuracy	98.8	85.45	85	97.6
Sensitivity	100	88.89	76	95.6
Specificity	100	——	100	100

of each coefficient, the accuracy of the algorithm was determined using different values of sparsity regularization and L2-weight regularization coefficients as shown in Fig. 7. The overall accuracy of the proposed algorithm is calculated using the values of the sparsity regularization coefficient ranging from 0 to 10 in the step of 0.5, while the value of the L2-weight regularization coefficient is fixed at 0.001. For the first dataset, the maximum and minimum accuracy achieved is 94.0% at a coefficient of 1 and 91.4% at a coefficient of 3, respectively. For the BRATS dataset, the maximum and minimum accuracy achieved is 98.8% and 92.5%, respectively. Similarly, the L2-weight regularization coefficient is tested with different values ranging from 0 to 0.01 in the step of 0.0005. The value of the sparsity regularization coefficient gets fixed at 1. For the first dataset, the maximum accuracy of 94.0% occurred at a coefficient value of 0.001, while the minimum accuracy attained to 90.5% with a coefficient value of 0.0085. For the second dataset, the maximum accuracy achieved is 98.8%, while the minimum accuracy is 93 .8%.

To show the superiority of the algorithm proposed, the performance of the proposed algorithm is compared with the existing algorithm using the same datasets. For the first dataset, the proposed algorithm is compared with the algorithm proposed by J. Cheng H. Wei and C. Shuangliang [23]. For the second dataset, BRATS dataset, the performance of the proposed algorithm is compared with the existing algorithm presented by [24-26].

The existing techniques used either 2D DWT or PCA or Gabor filter to extract features along with applying different classifiers. Three-directional features (i.e., horizontal, vertical, and diagonal features) are obtained by utilizing 2D DWT to the MRI image. Eigenvalues and eigenvectors are determined by applying PCA to the MRI images. The Gabor filter analyzes the edges of the MRI image and produces several images with different wavelengths and orientations. For the medical image analysis, directional features or Eigen decomposition features or Gabor features don't express all the substantial features of the MRI image. Therefore, in this paper, the 2D DWT and 2D Gabor filters are combined herein as directional transformation methods to take advantage by taking into consideration all features of all orientations of the MRI image. Also, some authors used the statistical features from the raw images, which lacks the directional information contained in the transformation methods (2D DWT and 2D Gabor filters). Therefore, robust algorithm that extracts the statistical features from the combined of the two transformation methods is proposed.

The features extracted using the algorithm proposed, statistical features obtained from the combination of 2D DWT and Gabor filter, are more efficient than those features extracted by using each filter alone and the algorithm presented by [23-26]. Furthermore, the classification technique implemented by stacking SSA and SMC shows better performance compared with BPA classifier and those classification techniques discussed in [23-26]. Therefore, the accuracies obtained by the algorithm proposed are higher than those accomplished by the existing algorithms proposed in [23-26] as shown in Tables 6 and 7 and higher than those accuracies obtained by using the proposed algorithm with BPNN classifier as shown in Tables 4 and 5.

Table 6 shows the accuracies of the algorithm proposed in comparison with the existing one for the first dataset in terms of the total accuracy, sensitivity, and specificity for Pituitary, Glioma, and Meningioma tumors. The accuracies of the algorithm proposed compared with the existing algorithms for the BRATS dataset in terms of accuracy, specificity, and sensitivity are illustrated in Table 7.

Algorithm proposed is executed using MATLAB R2017a. The proposed algorithm is implemented using a computer having Intel Core i7-6700HQ processor@ 2.60GHz and 16GB RAM. For example, in the training phase of the first dataset, 161 seconds

278

are the elapsed time to train 2,145 slices. In the testing phase, the time is computed based on the maximum tumor segment of size 228×346 in order to know the time required in the real application to process each ROI separately. In the algorithm proposed, the time required to classify the slice with maximum tumor region is 1.2969 seconds.

5. Conclusion

The brain tumor is usually examined by radiologists employing either noninvasive or invasive techniques or a combination of both. Noninvasive diagnosis relies on using Magnetic Resonance Imaging (MRI). In this work, a new algorithm is presented to recognize three types of brain tumors, which are Pituitary, Glioma with both high and low grade, and Meningioma. Two datasets are used to evaluate the algorithm proposed. The first dataset has all the three types of tumors, while the second dataset, which is call BRATS, has only the Glioma brain tumor. The 2D DWT, the 2D Gabor filter, and the first and the second order statistics of the transform domain data are employed to create a features pool that represents all possible individual tumors attributes. The classifier model is built by combining two models of NNs, which are SSA and SMC. Simulation results show the effectiveness of the overall performance of the algorithm proposed implemented by stacking the SMC with the SSA to accomplish, in instant, the accuracy of 94.0% for the first dataset and 98.8% for the BRATS dataset. The performance of the adopted algorithm can be improved by adding features corresponding to the patient's gender and age.

References

[1] American Cancer Society, "What are the Key Statistics About Brain and Spinal Cord Tumors?", 2019. [Online].

[2] M. S. H. Al-Tamimi and G. Sulong, "Tumor Brain Detection Through MR Images: A Review of Literature", *Journal of Theoretical & Applied Information Technology*, Vol. 62, No. 2, 2014.

[3] M. Hameurlaine and A. Moussaoui, "Survey of Brain Tumor Segmentation Techniques on Magnetic Resonance Imaging", *Nano Biomedicine and Engineering* , Vol. 11, No. 2, pp. 178-191, 2019.

[4] M. K. Abd-Ellah, A. I. Awad, A. A. M. Khalaf, and H. F. A. Hamed, "Design and implementation of A Computer-Aided Diagnosis System For Brain Tumor Classification", In: *Proc of the 28th International Conf. on Microelectronics*, pp. 73–76, 2016.

[5] H. H. Sultan, N. M. Salem, and W. Al-Atabany, "Multi-Classification of Brain Tumor Images Using Deep Neural Network", *IEEE Access*, Vol. 7, pp. 69215-69225, 2019.

[6] P. John and others, "Brain Tumor Classification Using Wavelet And Texture Based Neural Network", *International Journal of Scientific & Engineering Research*, Vol. 3, No. 10, pp. 1–7, 2012.

[7] E.-S. A. El-Dahshan, H. M. Mohsen, K. Revett, and A.-B. M. Salem, "Computer-Aided Diagnosis Of Human Brain Tumor Through MRI: A Survey And A New Algorithm", *Expert Systems with Applications*, Vol. 41, No. 11, pp. 5526–5545, 2014.

[8] S. Bauer, R. Wiest, L. Notle and M. Reyes, "A Survey of MRI-Based Medical Image Analysis For Brain Tumor Studies", *Physics in Medicine and Biology*, Vol. 58, No. 13, pp. 97-129, 2013.

[9] R. O. Duda, *Pattern Recognition*, Third Edition, John Wiley & Sons, 2020.

[10] S. Iqbal, M U. G. Khan, and T. Saba, and A. Rehman, "Computer-Assisted Brain Tumor Type Discrimination Using Magnetic Resonance Imaging Features", *Biomedical Engineering Letters*, Vol. 8, No. 1, pp. 5-28, 2018.

[11] A. Aldhahab and W. Mikhael, "Face Recognition Employing DMWT Followed by FastICA", *Circuits, System, and Signal Processing*, Vol.37, pp. 2045–2073, 2018.

[12] A. Vidyarthi and N. Mittal, "Performance Analysis of Gabor-Wavelet Based Features in Classification of High Grade Malignant Brain Tumors", In: *Proc. of the 39th National Systems Conf. (NSC)*, pp. 1-6, 2015.

[13] K. Priya, S. Kavitha, and B. Bharathi, "Brain Tumor Types and Grades Classification Based on Statistical Feature Set Using Support Vector Machine", In: *Proc. of the 10th International Conf. on Intelligent Systems and Control*, pp. 1–8, 2016.

[14] K. K. Kumar, T. M. Devi, and S. Maheswaran, "An Efficient Method for Brain Tumor Detection Using Texture Features and SVM Classifier in MR Images", *Asian Pacific journal of cancer prevention*, Vol. 19, No. 10, pp. 2789-2794, 2018.

[15] U. Javed, M. M. Riaz, A. Ghafoor, and T. A. Cheema, "MRI Brain Classification Using Texture Features, Fuzzy Weighting and Support Vector Machine", *Progress in Electromagnetics Research B*, Vol. 53, pp. 73–88, 2013.

[16] M. Nazir, M. A. Khan, T. Saba, and A. Rehman, "Brain Tumor Detection from MRI Images

Using Multi-level Wavelets", In: *Proc. of International Conf. on Computer and Information Sciences*, pp. 1-5, 2019.

[17] S. Lahmiri and M. Boukadoum, "Classification of Brain MRI Using The LH and HL Wavelet Transform Sub-Bands", In: *Proc. of IEEE International Symposium of Circuits and Systems*, pp. 1025-1028, 2011.

[18] Y. Zhang, W. Li, L. Zhang, X. Ning, L. Sun, and Y. Lu, "Adaptive Learning Gabor Filter For Finger-Vein Recognition", *IEEE Access*, Vol. 7, pp. 159821-159830, 2019.

[19] V. Saraswathi and D. Gupta, "Classification of Brain Tumor using PCA-RF in MR Neurological Images", In: *Proc. of International Conf. on Communication Systems Networks*, pp. 440-443, 2019.

[20] V. K. Lakshmi, C. A. Feroz, and J. A. J. Merlin, "Automated Detection and Segmentation of Brain Tumor Using Genetic Algorithm", In: *Proc. of International Conf. on Smart Systems and Inventive Technology*, pp. 583-589, 2018.

[21] G. P. Zhang, "Neural Networks for Classification: A Survey", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Vol. 30, No. 4, pp. 451-462, 2000.

[22] H. Byun and S.-W. Lee, "Applications of Support Vector Machines For Pattern Recognition: A Survey", In: *Proc. of International Workshop on Support Vector Machines*, 2002.

[23] J. Cheng H. Wei and C. Shuangliang, "Enhanced Performance of Brain Tumor Classification Via Tumor Region Augmentation and Partition", *PLoS One*, Vol. 10, No. 10, p. e0140381, 2015.

[24] V. Vijayarangan and M. K. Geetha, "Automatic Tumor Classification of Brain MRI Images Using DWT Features", *International Journal of Advanced Research in Computer Science*, Vol. 7, No. 5, pp.93–99, 2016.

[25] V. Wasule and P. Sonar, "Classification of Brain MRI Using SVM and KNN Classifier", In: *Proc. of the Third International Conf. on Sensing, Signal Processing and Security*, pp. 218-223, 2017.

[26] L. Farhi and A. Yusuf, "Comparison of Brain Tumor MRI Classification Methods Using Probabilistic Features", In: *Proc. of the 13th IASTED International Conf. on Biomedical Engineering*, pp. 55-62, 2017.

[27] S. Lahmiri, M. Boukadoum, S. Lahmiri, and M. Boukadoum, "Hybrid Discrete Wavelet Transform and Gabor Filter Banks Processing for Features Extraction from Biomedical Images", *Journal of Medical Engineering*, Vol. 2013, pp. 1–13, 2013.

[28] Nitish, A. K. Singh, and R. Singla, "Different Approaches of Classification of Brain Tumor in MRI Using Gabor Filters for Feature Extraction", *In Soft Computing: Theories and Applications*, Vol. 1035, pp. 1175-1188, 2020.

[29] V. S. N. Prasad and J. Domke, "Gabor filter Visualization", *Journal of the Atmospheric Sciences*, Vol. 13, 2005.

[30] T. S. Lee, "Image Representation Using 2D Gabor Wavelets", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 10, pp. 959-971, 1996.

[31] A. Aldhahab, T. Alobaidi, and W. Mikhael, "Applying Multiresolution Analysis to Vector Quantization Features for Face Recognition", In: *Proc. of IEEE 62nd International Midwest Symposium on Circuits and Systems*, pp. 598–601, 2019.

[32] R. C. Gonzalez and R. E. Woods, *Digital ImageProcessing*, Pearson, NY, 2018.

[33] P. K. Mallick, S. H. Ryu, S. K. Satapathy, S. Mishra, G. N. Nguyen, and P. Tiwari, "Brain MRI Image Classification for Cancer Detection Using Deep Wavelet Autoencoder-Based Deep Neural Network", *IEEE Access*, Vol. 7, pp. 46278-46287, 2019.

[34] A. Ng, "Sparse Autoencoder", *CS294A Lecture notes*, Vol. 72, pp. 1–19, 2011.

[35] M. H. Beale, M. T. Hagan, and H. B. Demuth, "Neural Network ToolboxTM User's Guide", *Mathworks Inc.*, 2010.

[36] M. G. F. Coutinho, M. F. Torquato, and M. A. C. Fernandes, "Deep Neural Network Hardware Implementation Based on Stacked Sparse Autoencoder", *IEEE Access*, Vol. 7, pp. 40674-40694, 2019.

[37] B. H. Menze, and others, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging*, Vol. 34, No. 10, pp. 1993–2024, 2015.

[38] M. Kubat, *An Introduction to Machine Learning*, Springer, 2015.