

## Segmentation and classification of brain tumor images using statistical feature extraction and deep neural networks

Gaurish Joshi<sup>\*</sup>, Rajeev Kumar and Amit Kumar Singh Chauhan

Assistant Professor, Department of Electronics Engineering, Rajkiya Engineering College, Kannauj, India

Received: 14-August-2021; Revised: 20-December-2021; Accepted: 22-December-2021

©2021 Gaurish Joshi et al. This is an open access article distributed under the Creative Commons Attribution (CC BY) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

*Cancer happens to be the second leading cause of death globally. One of the deadliest types of cancers with a very high mortality rate is brain cancer, which is characterized by malignant tumors or neoplasm. The general approach to diagnosis is the analysis of brain image scans. However, mere manual inspection of the images can be non-conclusive, and the diagnosis may need to be preceded by a biopsy, after a surgery. The entire process of prognosis and a biopsy may result in the loss of valuable time and advancement of cancer. Hence, automated tools which can analyze brain scan images and classify them may aid the prognosis and speed up the actual course of action. This paper presents an approach based on segmentation, statistical feature extraction and classification based on deep neural networks. The images used in this work is brain Magnetic Resonance Imaging (MRI) scans. The two-dimensional discrete wavelet transform has been employed for denoising the raw data, followed by statistical feature extraction. Two deep neural network architecture has been proposed which are the Deep Bayes Network (DBN) and the Residual Network (ResNet). The performance of the deep neural network has been evaluated in terms of classification accuracy. The classification of brain MRI images has been done in three categories of images which are meningioma, glioma and pituitary tumors. It has been shown that the proposed method attains an average classification accuracy of 98.13% and 96.89 using DBN and the ResNet respectively, for the used dataset. A comparative analysis with existing work shows improvement in classification accuracy with respect to existing techniques.*

### Keywords

*Brain tumor classification, Discrete wavelet transform, Statistical feature extraction, Deep bayes net, Residual network (ResNet-50), Classification accuracy.*

## 1. Introduction

Cancer cases are increasing worldwide, primarily due to unhealthy lifestyles, poor dietary choices, lack of physical activity, increasing pollution levels and exposure to radiation [1]. Early and accurate diagnosis of brain cancers is critical for saving the life of patients. Brain cancer has one of least survival rates for the next 5-10 years from the time of detection of malignant tumours. The cancer progresses from initially being a benign lesion to becoming a cancerous tumor. Cancer causes severe mental trauma, financial strain and social stigma mainly in developing countries.

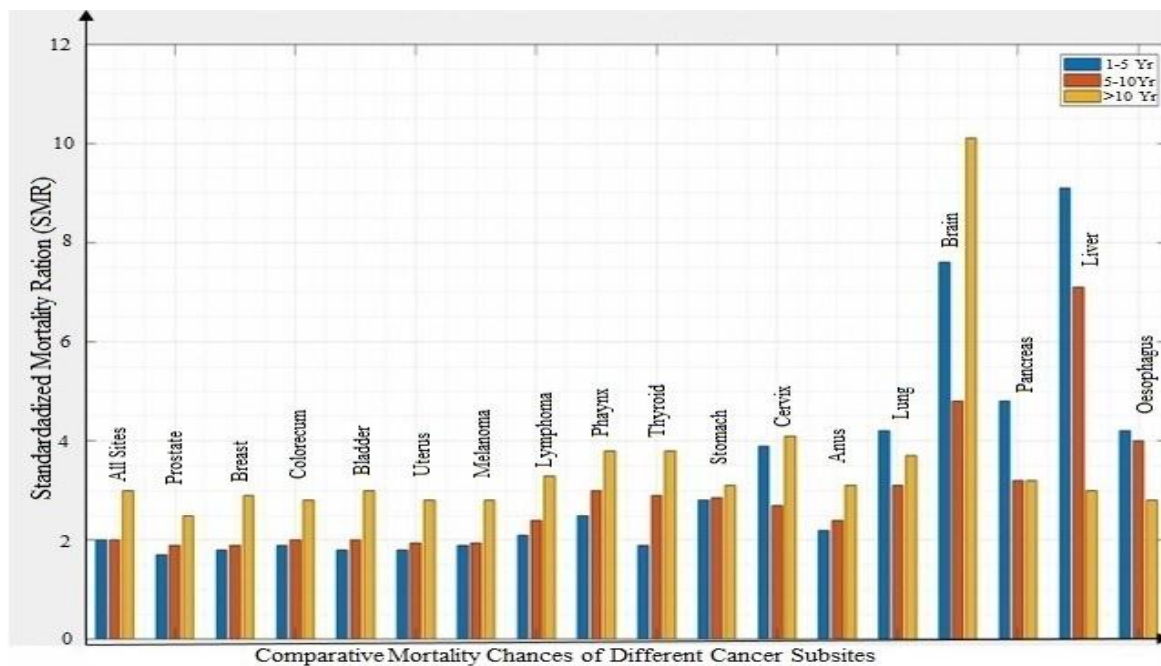
Brain cancer has one of the highest mortalities among all cancers with very low five-year and ten-year survival rates [2].

The fact is evident from *Figure 1* which depicts the Standardized Mortality Ratio (SMR) for different cancer sub-sites. Thus, early detection and correct diagnosis of brain cancer is mandatory for possible chances of recover. This serves as a fundamental motivation for several researchers to develop techniques which could augment early and accurate detection of brain tumors along with a classification of the type of tumor. Early detection is often hindered due to patients overlooking initial symptoms of the disease such as headaches, muscle twitches, occasional unconsciousness and seizures. This causes a delay in the start of diagnosis, which may prove to be lethal owing to the aggressive nature of progress of malignant brain tumors [3]. Typically, the benign tumors exhibit a clear or sharp boundary with respect to the neighboring regions and have a relatively slower spreading rate compared to the malignant tumors. The malignant tumors are much more invasive and exhibit a blurred boundary or distinction and have a much faster spreading rate. The brain

<sup>\*</sup> Author for correspondence

tumors may have different origins viz. originating from the brain (termed as primary tumors) or originating somewhere else in the body and then traversing to the brain (termed as secondary tumors). The growth rate of tumors is fundamentally governed by the oncogenes termed as proto-oncogenes. These genes govern the mitosis (cell-division) and further growth of the tumor. Typically, malignant tumors are categorized into four stages [3]. In stage 1, the tumor is mostly localized and grows at a relatively slower rate. In stages 2 and 3, the growth rate accelerates and the carcinoma spreads rapidly. Stage 4 is characterized by the spread of carcinoma to other locations of the body through the lymph nodes [4].

Tumors are often analyzed using the Fine Needle Aspiration Cytology (FNAC) or biopsy tests. The FNAC is prone to errors since the result of the test often depends on the location and the type of tissue collected during the sampling process. The biopsy is often regarded as the gold standard for detecting malignancy or carcinoma in tumors. Conducting the FNAC or the biopsy is generally much more challenging in case of brain tumors as compared to tumors in other parts of the body such as on the skin, breasts or throat [5]. Hence, imaging is the most prevalent and effective technique to initialize prognosis in cases of brain tumors.



**Figure 1** Mortality ratio for different cancer sub-sites

Source: Database “SEER 18 Regs Research Data+Hurricane Katrina Impacted Louisiana Cases, Nov 2017 Sub (1973–2015 varying, <https://seer.cancer.gov/data-software/documentation/seerstat/nov2017>)

The most commonly used scans to identify tumors are Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) scans [4]. Hyperspectral Imaging (HSI) is also employed in cases where the tumor is thought to be highly invasive and can spread rapidly. The HSI based technique uses a comparatively wider electromagnetic spectrum for capturing the image, thereby rendering much more detail [5, 6]. However, a radiologist or oncologist does the prognosis regarding the category of tumor. The decision regarding the stage of the tumor depends upon the experience and expertise of medical practitioner. The subsequent direction of

treatment critically depends on this phase and hence augmenting the analysis stage is crucial.

To augment the prognosis, machine learning based models are being developed which can detect and classify Magnetic Resonance (MR) images into benign and malignant categories [7]. Achieving high accuracy of classification using machine learning approaches is challenging due to several factors such as existing noise effects in the raw data, overlapping image parameters among the different categories of the tumors and divergences within the data samples of a particular category [8]. The objective of the proposed work is to therefore devise a mechanism for

automated detection and classification of brain tumors which would yield high accuracy of classification. This would include an image pre-processing followed by feature extraction and classification. In case of deep learning techniques, separate handpicked feature extraction is not needed. Automated tools for brain tumor detection and classification are developed with the aim of providing the physician with a strong second opinion regarding the presence and type of tumor. Present approaches suffer from two major challenges, one being inaccuracies in classification results for relatively small datasets, and the other being large computational complexity of existing algorithms which may make the method practically infeasible in real life situations [9].

Section 1 of the paper presents the introduction to automated detection of brain tumours along with the objectives and motivation behind the work. Section 2 presents the literature review citing noteworthy contribution in the domain. Section 3 presents the methodology. Section 4 presents the experimental results and section 5 presents the detailed discussion on the obtained results. Section 6 of the paper presents the concluding remarks and the avenues for future enhancement.

## 2.Literature review

Several machine learning approaches have been explored for automated brain tumor classification with the primary aim of low computational complexity, need for lesser training data and high accuracy. Unsupervised machine learning methods such as k-means clustering have been used to separate or segment the affected tumor area from the rest of the image after grayscale conversion. Fuzzy c-means and k-means clustering techniques were compared pertaining to their effect on classification accuracy [10]. The Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) based approaches are used in [11] and it is shown that the KNN approach outperforms the linear SVM in terms of classification accuracy. A Kernel Support Vector Machine (K-SVM) optimized using the Particle Swarm Optimization (PSO) based approach is presented in [12] and it was shown to perform better than baseline techniques. The Gray Level Co-occurrence Matrices (GLCM) based feature extraction and subsequent classification using Back Propagation Neural Networks (BPNN) was proposed in [13] which was shown to achieve high accuracy of classification. A transfer learning-based approach using Convolutional Neural Networks (CNN) was

proposed in [14], where a three-category classification was performed among glioma, meningioma and pituitary tumors. A Region based Convolutional Neural Network (RCNN) was implemented in [15] which was shown to perform faster than the conventional CNN in terms of execution time for automated brain tumor classification. An approach for brain tumor classification using Adaptive Neuro Fuzzy Inference Systems (ANFIS) was proposed in [16]. Prior to classification, the approach uses the Adaptive Histogram Equalization (AHE) Modified-Fuzzy C Means (MFCM) for pre-processing. Feature optimization is performed using Adaptive Elephant Herd Optimization (AEHO) algorithm. A curvelet transform based feature extraction followed by classification using different classifiers such as the Probabilistic Neural Network (PNN), SVM and ANFIS was proposed in [17]. It was shown that the PNN outperforms the SVM and ANFIS in terms of classification accuracy. Deep CNN features were extracted and an SVM was used for classification in [18]. An ada-boost ensemble neural classifier was used with fast boosting and pre-trained network for brain tumor classification [19]. Generative Adversarial Networks (GANs) and pre-trained models have been explored for brain tumor classification. In this approach, fictitious training datasets are generated by GANs to increase the training data without increasing the effort for data collection and labelling [20]. The method is effective in the absence of large training datasets, but may suffer in terms of classification accuracy for diverse datasets.

Deep neural networks and deep learning exhibit better performance compared to techniques such as SVM or expert systems which typically do not show significant improvement in performance after adding more data to the data set [21]. Deep nets, however suffer from the challenges of extracting large datasets to provide high accuracy of classification which may be extremely tedious and computationally ineffective [22].

A comparative analysis of the recent literature in the domain renders the following insights: brain tumor classification is challenging due to the high similarity among different categories of tumor images whose differences may manifest at very late stages. Moreover, noise effects make the feature computation more challenging. Image filters such the median filter, Wiener filter or transform domain image enhancement proves to be useful for further

feature calculation and classification. Several machine learning-based techniques have been used such as Neural Networks, SVM, expert systems, ANFIS, KNN etc. The performance of all these classifiers depends upon the accuracy with which the image features are calculated, the correct choice of image features which would result in least overlap and the efficacy of the classifier to separate datasets with overlapping features. SVM and expert systems suffer from performance saturation with increasing large datasets. On the contrary, deep learning models rule out the need for handpicked feature calculation and are much deeper compared to the conventional machine learning algorithms. The possible pitfalls of the approach, however are the chances of overfitting and loss of control over feature selection. This may adversely affect the classification accuracy for some datasets. The computational complexity of deep learning approaches may also be high. To circumvent the problem of tedious data acquisition and labelling, GANs are becoming popular. The performance of GANs varies, though with the type of images being classified. For non-high risk classification, GANs have proven to be effective. However, rigorous validation needs to be done prior to implementing GANs on medical image classification especially brain tumor images. This creates the need for a study, which would investigate both feature selection followed by machine learning based classification as well as deep learning based classification. It is judicious to devise a machine learning based approach, which can separate overlapping datasets. In addition, it is necessary to devise deep learning models, which can avoid the challenges of overfitting and computational complexity. Thus, the proposed work investigates two separate approaches:

- 1) A machine learning based model using the Deep Bayes Network (DBN).
- 2) A Residual Network (ResNet)-50 CNN structure.

The first approach is helpful in extracting handpicked features from the images and investigating the performance of the classifier, while the second approach is a deep learning based approach not requiring separate feature selection and computation. Additionally, the second approach is an improved version of the conventionally used CNN with lesser chances of overfitting.

### 3.Methods

The data used in the study are MR images acquired from two datasets viz Kaggle and figshare datasets [23, 24]. The image is that of a composite MR image containing the tumor region. Prior to training and

classification of the images, the region of interest needs to be separated which is termed as segmentation. The images are first labelled into three categories which are meningioma, glioma and pituitary tumors, with the aim of empirical classification of a new case. In general, images are affected by noise effects during capturing, storage and transmission, which leads to degradation in the quality of the image. Hence, pre-processing of the images prior to feature calculation and classification is done [25].

#### 3.1Image pre-processing

Images are often affected by noise and disturbance. Effects of noise come into the picture with its different origins, such as noise in electronic devices, noise in the analog to digital conversion process and spikes of voltages and currents. Noise in electronic devices is termed as Gaussian noise. Noise in analog to digital conversion process causes multiplicative speckle noise. Spikes of voltages and currents cause salt and pepper noise. Capturing lesser pixels or pixels with low resolution results in a particular type of noise called Poisson noise [26]. Noise removal is critically important to attain accuracy in feature calculation driving the classification process. Several noise removal techniques have been explored such as using image filters and noise removal in the transform domain. One of the most effective techniques for noise removal from brain MR images is the wavelet transform. The wavelet transform has the distinctive advantage of noise removal in the transform domain employing high pass and low pass filtering. The wavelets unlike the conventional Fourier methods do not have smooth base functions. Non-smooth kernel functions of the transform make them effective for abruptly changing signals, such as images affected by noise [27]. The discrete version of the wavelet transform termed as the Discrete Wavelet Transform (DWT) is used as an iterative filter in this work for de-noising images. The wavelet transform applied on a discrete sequence  $x(k)$  is given in Equation 1.

$$X(k) \xrightarrow{DWT} X_{LPF}, X_{HPF} \quad (1)$$

Here,  
*DWT* denotes the discrete wavelet transform operation  
 $X_{LPF}$  denotes the coefficient values obtained through low pass filtering  
 $X_{HPF}$  denotes the coefficient values obtained through high pass filtering

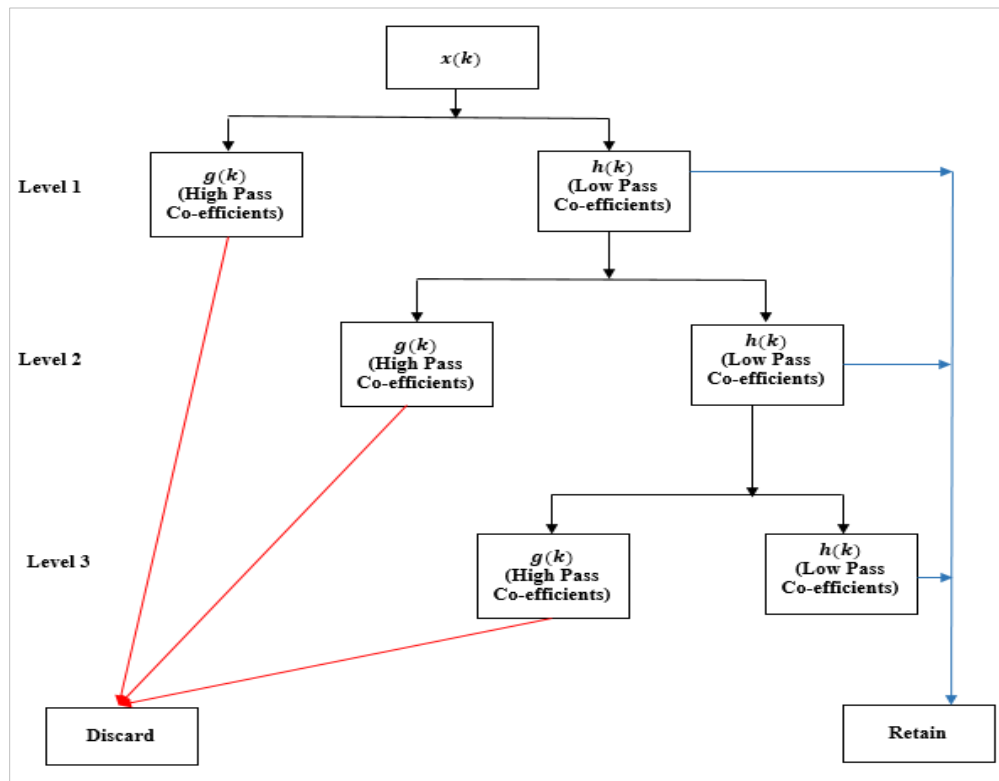
The co-efficient values obtained through low pass filtering contain the maximum information content of the signal while the that obtained from the high pass filtering contain the additional details of the signal. The DWT thus decomposes the signal into two components with different attributes. The low pass filtering coefficients are also termed as approximate coefficients  $C_A$  while the high pass coefficients are termed as detailed coefficients  $C_D$ . The coefficient values can be computed by using Equations 2 and 3.

$$X_{HPF}(n) = X(k)g(2l - k) \quad (2)$$

$$X_{LPF}(n) = X(k)h(2l - k) \quad (3)$$

The Equations 2 and 3 imply that a down-sampling or decimation of the signal has occurred by a factor of 2 i.e., every other sample of the signal has been left out, thereby decimating the signal. The opposite effect can be observed in the frequency domain in which the frequency doubles compared to that of the original signal. Thus, the frequency domain

resolution of the signal increases. An iterative application of the DWT would keep decreasing the time resolution, but at the same time keeps increasing the frequency resolution of the signal. Typically, the lower frequency coefficients or the approximate coefficients would contain the maximum useful information and discarding the high pass or the detailed coefficients would not result in significant information loss. The decomposition can be truncated based on the amount of frequency resolution needed, as successive decomposition would increase the complexity of the process. Noise typically happens to exist in the high pass coefficients of the image decomposition and hence can be filtered out using iterative DWT decomposition of the image. Based on the various base functions of the wavelets, the DWT has multiple families such as Haarlet, Coiflet etc. [28]. A typical decomposition of images using the DWT is depicted in *Figure 2*.



**Figure 2** Iterative DWT decomposition of data

The process thus iterates over the  $C_{A,i}$  values and discard the  $C_{D,i}$  values to remove the noise effects. Here the number of levels is denoted by  $n$ . Increasing the number of levels reduces the noise effects, but increases the computation

complexity of the algorithm. Thus, a careful choice of the level should be made. In case of images, the coefficients can be computed horizontally and vertically thereby rendering the horizontal coefficients ( $C_H$ ) and the vertical coefficients ( $C_V$ ). The iterative decomposition of the set of 'n' images

and retention of the approximate coefficient is implemented using the following logic:

$$\begin{aligned} & \text{for } (i = 1, i \leq n, i++) \\ & \{ \\ & \quad \text{Image}_i \xrightarrow{DWT2} CA_{i,1}, CD_{i,1} \\ & \quad CA_{i,1} \xrightarrow{DWT2} CA_{i,2}, CD_{i,2} \dots \\ & \quad CA_{i,k-1} \xrightarrow{DWT2} CA_{i,k}, CD_{i,k} \\ & \} \end{aligned}$$

Here,

$i$  denotes the  $i^{\text{th}}$  image

The subscript index 1,2,..k denotes the level of decomposition

$CA$  denotes the approximate coefficients of a particular level

$CD$  denotes the detailed coefficients of a particular level

DWT2 denotes the 2-dimensional discrete wavelet transform

The validity of the multi-level decomposition of the image using the discrete wavelet transform for noise removal can be validated by the histogram analysis of the original image, the approximate coefficients and the detailed coefficients. The normal histogram is computed using Equation 4.

$$h = \sum_{k=1}^l s_l \quad (4)$$

Here,

$h$  denotes the normal histogram

$k$  denotes the histogram index

$l$  denotes the total number of bins in the distribution

$s$  denotes the histogram function

The total number of bins to be assigned are computed using Equation 5.

$$l = \frac{\max(X) - \min(X)}{l_w} \quad (5)$$

Here,

$\max(X)$  denotes the maximum value of the random variable ' $X$ '

$\min(X)$  denotes the minimum value of the random variable ' $X$ '

$l_w$  is the bin width typically equal to the number of distinct values in the random variable

In case of an effective decomposition, the histogram of the original image would be identical to that of the approximate coefficients while the dissimilarity would be observed in case of the detailed coefficients. Another critical operation prior to feature extraction is segmentation or separation of the

area of interest from the composite image. Segmentation is necessary to differentiate the attributes of different categories of images, which in this case happen to be normal, benign and malignant nature of the samples [29]. The segmentation of the region containing tumors can be done based on the sudden change in the gradient of image parameters with respect to the pixel location. This category of segmentation considered in this case is often termed as bi-level segmentation with the assumption that the image histogram can be split into two distinct categories viz the segmented region and the background. Considering the image under interest to have an area ' $A$ ' with a central reference ' $C_0$ ', the radial gradient is computed as Equation 6:

$$g_r = \max(r, C_0) |G_\sigma(r) \frac{\partial}{\partial r} \oint_{r, C_0}^R \frac{I(x,y)}{2\pi r} dA| \quad (6)$$

Here,

$g_r$  denotes the radial gradient

$I(x, y)$  denotes the image under interest

$C_0$  denotes the central reference

$G_\sigma$  denotes the Gaussian kernel

$r$  denotes the radial distance from the central reference

$R$  denotes the maximum radial distance from the central reference

$\max$  denotes the operation to find the maxima

$dA$  denotes the differential area

This segmentation technique is effective in case of a relative sharp change in the gradient in the contour of a composite image. However, the segmentation will be less accurate in case of a blurred or fuzzy boundary between the area of interest and the background. An alternative segmentation based on the mutual information between the different regions to be separated can be implemented by computing the entropy corresponding to the grayscale histogram. The entropy of an image of  $m \times n$  pixels with a histogram  $h_n$  corresponding to  $n$  grayscale levels can be computed using Equation 7 [30]:

$$h_l = \frac{1}{1-l} \log_2 \sum_{i=1}^n p_i^l \quad (7)$$

Here,

$l$  denotes the order of entropy, and  $l > 1$ .

$p_i^l$  denotes a discrete probability distribution corresponding to the histogram  $h$

As  $\lim_{l \rightarrow 1} h_l$  approaches the Shannon's entropy. The segmentation of the image into  $n$  levels would yield an additive entropy which can be computed using Equation 8.:

$$E_l(t) = \text{Arg max} \{ \sum_{i=1}^n h_l c_i \} \quad (8)$$

Here,

$E_i(t)$  denotes the additive entropy

$c_i$  denotes the number of categories of segmentation corresponding to the value of  $n$

The threshold value which segments the image can be computed using Equation 9.

$$T_{k1,k2} = \frac{\sum_{i=k1}^{k2-1} iP(i)}{\sum_{i=k1}^{k2-1} P(i)} \quad (9)$$

Here,

$P(i)$  denotes the discrete probability distribution

$k1, k2$  denote the indices of thresholding for the threshold value  $T$ .

The concept can be extended to a multi-level thresholding by increasing the order of  $T$ . The entropy based segmentation. In the present case, the segmentation can be implemented by a two-fold segmentation with the area of the tumor being separated from the rest of the image based on the following algorithm:

1. Load image of interest.
2. Initialize.  $r = 0$ , corresponding to the central reference  $C_0$
3. Increment the value of  $r$  up to  $R$  and compute the gradient  $g_r$  given by Equation 6
4. Based on Equation 9, obtain the value of  $T$ , where either  $g_r$  or  $E_i(t)$  attains a maxima  $\forall r, 0: R$
5. Designate different binary values to mask the 2 regions using Equation 10 and 11.

$$M_1 = \sum_{x,y}^{k1} P_{x,y} \quad (10)$$

and

$$M_2 = \sum_{x,y}^{k2} P_{x,y} \quad (11)$$

Here,

$M_1$  and  $M_2$  are the masking value typically chosen as 1 and 0 respectively

The process of feature extraction is critically important for classification as pixel wise analysis may be extremely overlapping in mature thereby jeopardizing the classifier used.

### 3.2 Feature extraction

Once the segmentation is performed, the next step is the computation of critical features based on which the decision regarding the category of the image can be taken. Statistical features are computed in the proposed approach since statistical features can be computed for a wide range of images without loss of generality [31]. Statistical feature computation can be augmented with the computation of GLCM features which try to estimate the co-occurrence of pairs of pixels in a defined spatial closeness. The GLCM

matrix is computed based on the joint probability of occurrence or joint probability distribution given by  $P_j$ . The GLCM features are computed for a given distance and angle denoted by ' $d$ ' and ' $\theta$ ' respectively for each of the pixels and its corresponding neighbour. The GLCM matrix is divided using a normalizing factor ' $N$ ' to obtain the normalized GLCM matrix. The features computed in the proposed work are [32, 33]:

1. Mean or first moment: It is the average value or first moment which can be computed using Equation 12.

$$Mean(\mu) = \frac{1}{N} \sum_i^N f_i X_i \quad (12)$$

2. Standard Deviation: It is a measure of the difference between the instantaneous value and the mean, and can be computed using Equation 13.

$$3. Standard Deviation(\sigma) = \sqrt{\frac{1}{N} \sum_i^N (X_i - \mu)^2} \quad (13)$$

4. Variance: It is the square of the standard deviation which is defined by Equation 14.

$$var = \sigma^2 \quad (14)$$

5. Skewness: It denotes the amount of asymmetry of the probability distribution curve with respect to the y-axis of the curve given by Equation 15.

$$skewness = \frac{\sum_i^N (X_i - \mu)^3}{(N-1)\sigma^3} \quad (15)$$

6. Kurtosis: The kurtosis is also termed as the fourth standard moment and is given in Equation 16.

$$Kurtosis = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] \quad (16)$$

7. Energy: The energy is also termed as the angular secondary moment and is defined by Equation 17.

$$Energy = \sum_{i,j}^N |p_{i,j}|^2 \quad (17)$$

8. Contrast: It is the degree of difference among the average and differential change in illuminance and is mathematically given by Equation 18.

$$Contrast = \sqrt{\frac{1}{mn} \sum_{i,j}^{m,n} [X(i,j) - Mean(i,j)]^2} \quad (18)$$

9. Entropy: It is the average information content associated with a random variable having a probability function  $P$ , and is computed using Equation 19.

$$E = -P(I_{x,y}) \log_2 I_{x,y} \quad (19)$$

10. Homogeneity: It is a statistical measure pertaining to the similarity of distributed values of a random variable, and is mathematically defined in Equation 20.

$$H = \sum_{i,j}^{m,n} \frac{P_{i,j}}{1-|i-j|^2} \quad (20)$$

11. Correlation: It is a measure of the amount of similarity among the pixel values computed using Equation 21.

$$Corr_{i,j} = \sum_{i,j}^{m,n} \frac{(i-\mu_x)(j-\mu_y)P_{j,x,y}}{\sigma_x\sigma_y} \quad (21)$$

12. Inverse difference moment: It is computed using Equation 22.

$$IDM = \sum_{i,j}^{m,n} \frac{1}{1+(i-j)^2} p_{i,j} \quad (22)$$

The GLCM normalizing factor is computed using Equation 23.

$$N = \frac{X_{i,j}}{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} X_{i,j}} \quad (23)$$

Here,

$X_i$  denotes the instantaneous value of the random variable  $X$

$I_{x,y}$  denotes an image which is a function of spatial co-ordinates  $(x, y)$

$m, n$  denote the pixels along  $x$  and  $y$  axes

$mean$  denotes the average illuminance of the image

$X$  denote the distinct values in the set

$f$  denotes the frequency distribution of the values

$N$  denotes the total number of levels in the normalized GLCM matrix

$p_{i,j}$  denotes the normalized GLCM matrix

$P_j$  denotes joint probability

$P$  denotes probability

### 3.3 Classification

The classification is performed based on the initial training and subsequent testing of the classifier. The system model is depicted in *Figure 3*. Two classifiers are designed in this work, which are the DBN and the ResNet. The two neural networks have distinctively different classification approaches. The Bayesian Network (BayesNet) tries to classify data samples based on the Bayesian theorem. Regularization has been employed in the proposed BayesNet [34]. The ResNet is a modified version of the CNN, which tries to circumvent the issues of overfitting and vanishing gradient. The ResNet can be seen as an improvement over the existing CNN models such as the AlexNet and VGGNet [35]. The BayesNet performs based on the conditional probability model wherein the training dataset is considered to be true as the apriori probability. The classification is done based on the maximum probability of the sample belonging to a particular class. The network is based on the assumption that the weights of the network are random variables and would change based on the data set used to train the network. The BayesNet is particularly useful for classification problems which

exhibit close similarity among feature values of multiple classes or overlapping feature values [36]. This is typically true in case of brain tumor classification as the region of the tumor and the composite image exhibit varying levels of fuzziness in the boundary depending upon the tumor case. The weight of the network is to be chosen in such a way that it maximizes the conditional probability of a data belonging to a particular class. The probability function can be computed using Equation 24.

$$P\left(\frac{X}{X_i, k_1, k_2, M}\right) = \frac{P\left(\frac{X_i}{X, k_2, M}\right)P\left(\frac{X_i}{k_1, M}\right)}{P\left(\frac{X}{k_1, k_2, M}\right)} \quad (24)$$

Here,

$P$  denotes probability

$X_i$  denotes the set of weight and bias

$X$  denotes the training data set

$M$  denotes the network architecture in terms of the hidden layers and neurons

$k_1$  and  $k_2$  are the regularization parameters for the network.

Generally, the term  $\rho = \frac{k_1}{k_2}$  is called the regularization ratio [36]. The regularization parameter is adopted in this case to limit the variations in the weights by introducing a penalty factor to the learning algorithm's cost function or objective function  $J$ . The regularization is different from early stopping or convergence in the sense that the earlier truncates the iterations prior to convergence to a minimum value of  $J$  whereas the latter tries to restrict the values of weights and number of parameters by modifying the cost function. Thus, regularization allows a much steeper decrease in the cost function and eventually lesser values as compared to early stopping. This significantly helps to reduce the time complexity of the algorithm. The regularization parameter is responsible for the number of iterations needed to reach convergence and typically shows an inverse relationship with the number of iterations  $n$ . The number of hidden layers in the network has been taken as 20, with the activation function being the sigmoid function. The Deep BayesNet is typically a deep neural network trained using the Bayesian Rule.



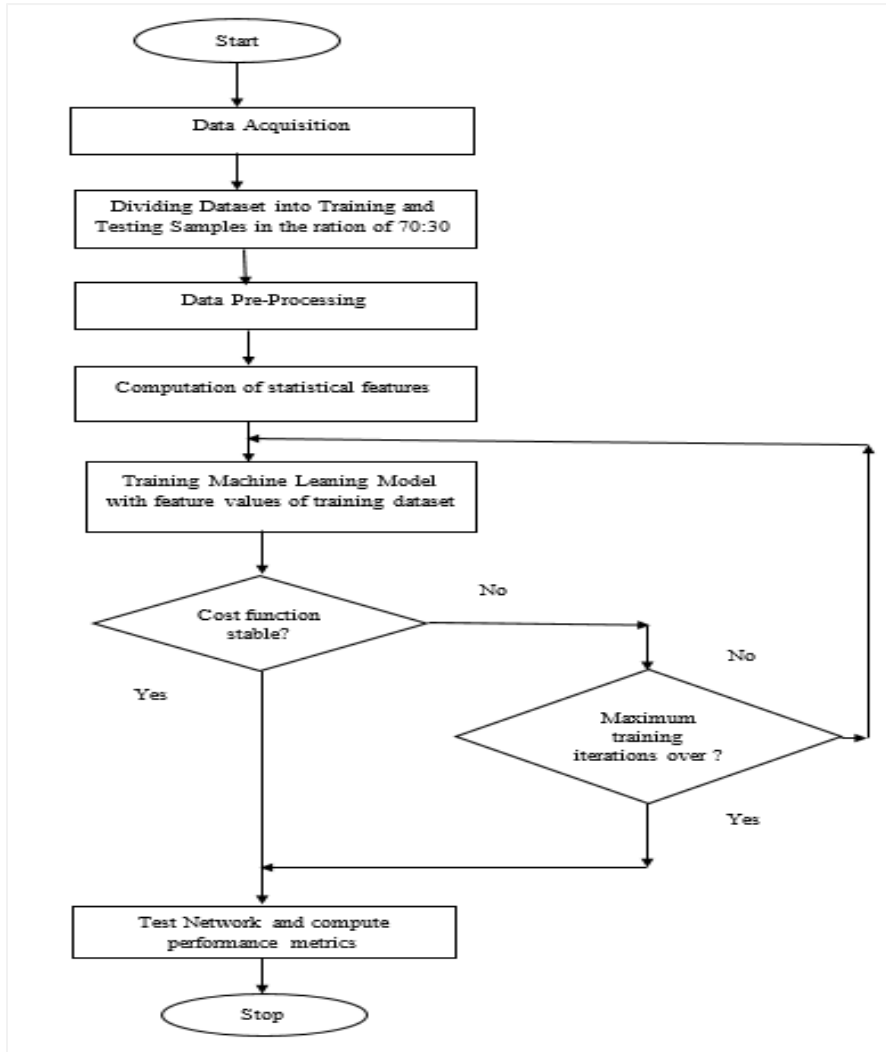


Figure 3 System model

**Algorithm:**

The training algorithm adopted in this work is given by:

Step.1: Initialize weights (w) randomly.

Step.2: Fix the maximum number of iterations (n)

and compute  $\rho = \frac{k_1}{k_2}$

Step.3: Update weights using gradient descent with an aim to minimize the objective function J given by Equation 25:

$$J = \frac{1}{m} \sum_{i=1}^m (v_i - v'_i)^2 \tag{25}$$

Step.4: Compute the Jacobian Matrix J has given by:

$$J = \begin{bmatrix} \frac{\partial^2 e_1}{\partial w_1^2} & \dots & \frac{\partial^2 e_1}{\partial w_m^2} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 e_n}{\partial w_1^2} & \dots & \frac{\partial^2 e_n}{\partial w_m^2} \end{bmatrix} \tag{26}$$

Here,

The error for iteration ‘i’ designated by  $e_i$  is computed as Equation 27:

$$e_i = (y_i - y'_i) \tag{27}$$

Here

$y_i$  is the actual value

$y'_i$  is the predicted value

Step.5: Iterate steps (1-4) till the cost function J stabilizes or the maximum number of iterations set in step 2 is reached, whichever occurs earlier.

Overlapping feature values with fuzzy boundaries cannot be classified accurately based on hard boundary conditions. Hence the BayesNet is applied. Another classifier employed in this work is the ResNet, which a modified version of the ubiquitous

CNN [37]. The ResNet has multiple convolution layers, but unlike typical convolutional networks, it has with skip connections between the layers [38]. The architecture of the ResNet doesn't allow the direct cascade of the weights in the hidden layer. This serves two important purposes:

- 1) Reduces the chances of overfitting the network.
- 2) Avoiding the chances of vanishing gradient commonly encountered in conventional CNNs.

The number of convolution layers in the network are 48, with one max-pool layer. The activation function used is the Rectified Linear (ReLU), with a stride of 2. The addition of more hidden layers in conventional CNNs often leads to saturation in the performance with high chances of performance saturation, which is mitigated by the ResNet architecture with the skip connections and addition of identity layers [39]. The performance of the two training algorithms is evaluated subsequently in terms of classification accuracy. The concept of skip connections in the ResNet is depicted in *Figure 4* [38]. The ResNet architecture used in the proposed work has an input size of  $243 \times 243 \times 3$  for the separate R, G and B channels of the image. A max pooling of  $2 \times 2$  with a stride of 2 has been used. The feature layer of *Fc1000* was employed with 1000 feature vectors.

The training rule and layer details can be mathematically represented as

Let,  $Y_L$  designate the output of the  $Y^{th}$  residual block, and  $m_L$  be the mapping function of the residual block and  $r_B$  be a Bernoulli random variable, then the output of the training block is computed using Equation 28.

$$Y_L = f(r_B \times m_L(Y_{L-1}) + m_L d(Y_{L-1})) \quad (28)$$

Here,

$d$  denotes a temporal delay

$f$  denotes the ReLu (Rectified Linear) activation function defined in Equations 29 and 30.

$$f(x) = x; x \geq 0 \quad (29)$$

$$f(x) = 0, x < 0 \quad (30)$$

Two cases arise out of equation (28), one when  $r_B$  takes a value of 0 and when it takes a value of 1. In case  $r_B$  takes a value of 1, the residual block remains active in operation. In case  $r_B$  takes a value of 0, the residual block becomes inactive and input to the output with a temporal delay of  $d$ . It is customary to consider the survival probability of the residual blocks as either a linear or decaying exponential given by Equation 31.

$$p_s = 1 - \frac{L}{K}(1 - p_K) \quad (31)$$

Here,

$K$  denotes the last residual block.

$p_K$  denotes the survival probability of the last residual block.

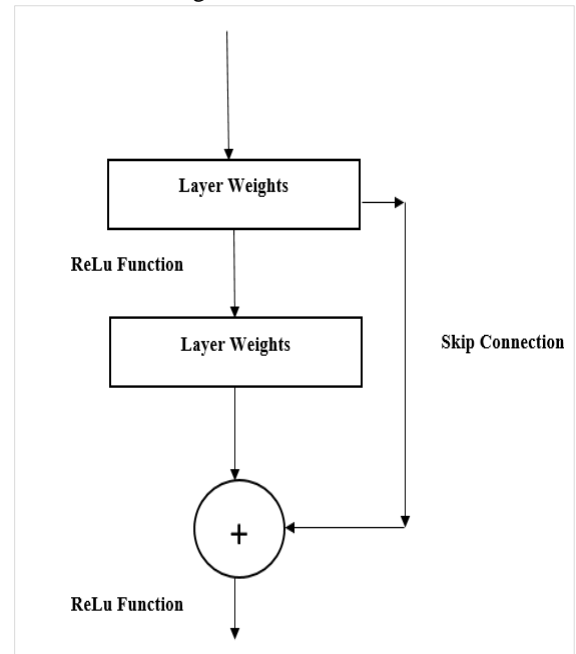
$L$  is the total number of survival blocks.

$p_s$  is the survival probability of a block.

The survival probability of a residual block is not made zero immediately after feature computation as a deeper layer would later use the feature value computed at a shallow (lower) layer after some time. The skip connection among the layers in the ResNet is depicted in *Figure 4*. The layer properties are listed in *Table 1*.

The repeated convolution layers are not mentioned separately for the sake of brevity.

In the testing approach, the trained network is given 30% of the data (which was reserved for testing) and not used for training. In case of the DBN, the new image is pre-processed, and its features are computed to subsequently feed the network. In case the ResNet, the folder containing the testing images is directly fed to the ResNet. The performance metrics are computed after the testing is completed. Although 30% of the data has been kept for testing, 20%, 25% or 35% of the entire data could have also been utilized for testing.



**Figure 4** Skip connections in ResNet

**Table 1** Layer properties

S.No.	Layer	Attribute
1	Input	224×224×3
2	Conv 1	7×7, 64, stride 2
3	MaxPool	3×3 MaxPool, stride 2
4	Conv2	1×1 Cov, 64
5	Conv3	3×3 Conv, 64
6	Conv4	1×1 Conv, 256
7	Conv5	1×1 Conv 128/2
8	Conv6	3×3 Conv 128
9	Conv7	1×1, Conv 512
10	Conv8	1×1, Conv 128
11	Conv9	1×1, Conv 256/2
12	Conv10	3×3 Conv 256
13	Conv11	1×1, Conv1024
14	Conv12	1×1, Conv 256
15	Conv13	1×1, Conv512/2
16	Conv14	3×3, Conv 512
17	Conv15	1×1, Conv 2048
18	Conv16	1×1, Conv 512
19	Pooling	Average Pooling
20	Fully Connected	Fully Connected $f_c$ 1000

The classification accuracy is given by Equation 32.

$$c = \frac{TP+TN}{TP+TN+FP+FN} \quad (32)$$

The sensitivity or recall is given by Equation 33.

$$Se \text{ or } Re = \frac{TP}{TP+FN} \quad (33)$$

The precision is given by Equation 34.

$$Pr = \frac{TP}{TP+FP} \quad (34)$$

The specificity is computed by Equation 35.

$$Sp = \frac{TN}{TN+FP} \quad (35)$$

The F-measure or F-Score is given by Equation 36.

$$F - \text{Measure} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (36)$$

The F-Measure can also be computed using Equation 37.

$$F - \text{Measure} = \frac{TP}{TP + \frac{1}{2}(FP+FN)} \quad (37)$$

Here,

- (TP): True Positive
- (TN): True Negative
- (FP): False Positive
- (FN): False Negative

## 4.Results

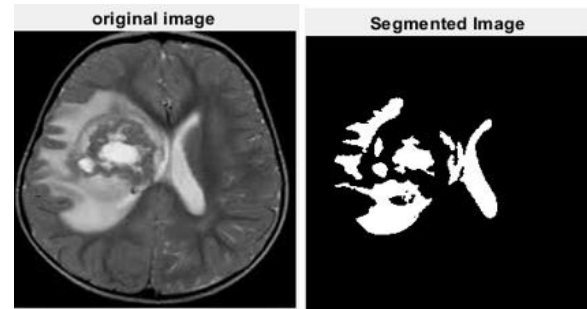
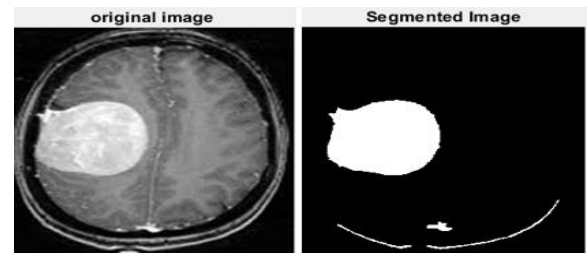
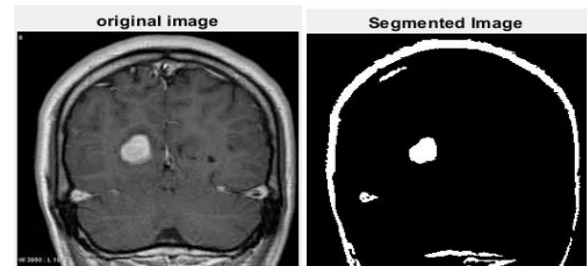
The experiment has been performed in Matlab 2020a. The hardware configuration of the system used is: Intel Core i5 9300H Central Processing Unit (CPU) with base clock frequency of 2.4GHz and RAM of 8GB. The data set used has been cited in [24, 25]. The tumor classes in the training dataset are meningioma (708 images), glioma (1426 images) and

pituitary tumors (930 images). The image properties on importing to Matlab workspace are (Table 2):

**Table 2** Image properties

S. No.	Attribute	Value
1.	Size	224×224×1 (Grayscale)
2.	Initial Class	Unsigned integer (uint8)
3.	Class Converted	double

The images are first loaded, subsequently the pre-processing, segmentation and feature extraction is implemented. The input images are first converted from RGB to grayscale images, de-noised using the discrete wavelet transform after which the segmentation is performed. Figure 5 to Figure 7 depict the original and segmented images for all the three categories, which are meningioma, glioma and pituitary tumors.

**Figure 5** Original and segmented images for the meningioma category**Figure 6** Original and segmented images for the glioma category**Figure 7** Original and segmented images for the pituitary tumor category

The segmentation of the images is followed by the computation of the features of the image. Feature extraction may be done separately for images so as to train classifiers, not with the pixel information directly, but rather with statistical values unique to a class of images. This approach may lead to lesser computational time compared to approaches which directly use image pixels to train a classifier. The features computed from the one set of images containing three classes i.e., meningioma, glioma and pituitary tumors are shown in *Table 3*.

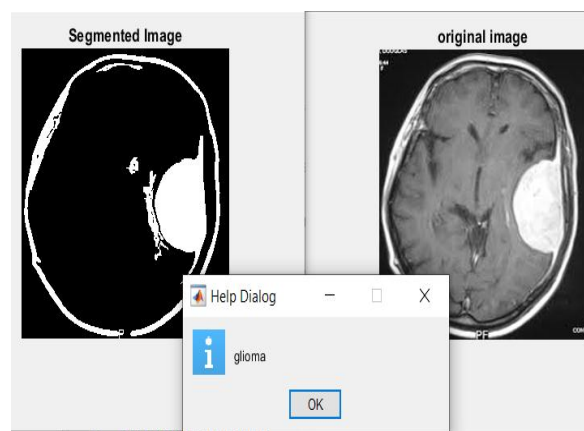
Feature for all the images in the dataset has been computed similarly. An important observation which can be made is the fact that the features for all the three classes bear close values. This necessitates a classifier which can accurately classify data sets with overlapping or fuzzy boundaries. *Table 3* tabulates the feature values for a sample image for each of the three categories, for the sake of brevity.

**Table 3** Image feature values

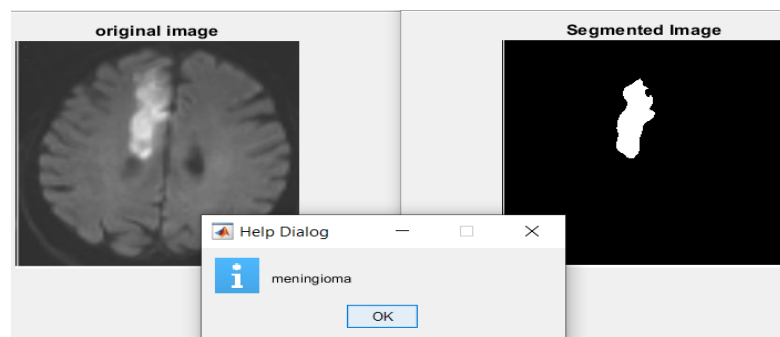
Feature	Meningioma	Glioma	Pituitary tumor
Contrast	0.380681818181818	0.318750000000000	0.381250000000000
Energy	0.671219653925620	0.654314953512397	0.670699251033058
Mean	0.00379912401078097	0.00273752315016644	0.00478474995626005
Standard Deviation	0.106561473755997	0.106594046533438	0.106521747166905
Entropy	3.33424331718665	3.63208178823710	3.63027532445053
Inverse Difference	0.106600358177805	0.106600358177805	0.106600358177805
Variance	0.0113137529923128	0.0112772768179399	0.0112708149837308

After the computation of the image features as per Equations 12-23, the feature vector is used to train two categories of classifiers viz the DBN and the ResNet-50. The choice of the DBN incorporating regularization helps in avoiding overfitting. The architecture of the ResNet inherently avoids overfitting due to no direct cascade of weight of adjacent layers. The performance of the system has been evaluated based on the classification accuracy based on the calculations of true positive, true negative, false positive and false negative rates. A comparative analysis of the results has been presented subsequently. A new set of test images distinct from that used for segmentation in *Figure 5* to *Figure 7* have been used as an illustration in *Figure 8* to *Figure 10*. The final interface for implementation has been depicted in *Figure 8*, where the case of glioma has been classified. A similar

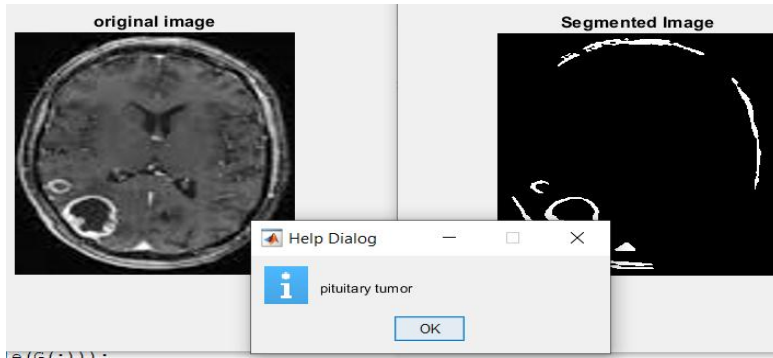
process is implemented for meningioma and pituitary tumors and depicted in *Figures 9* and *10*.



**Figure 8** Classification of new image as glioma

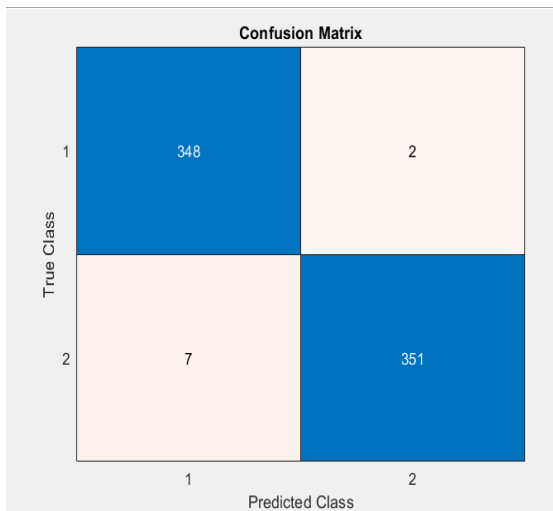


**Figure 9** Classification of new image as meningioma

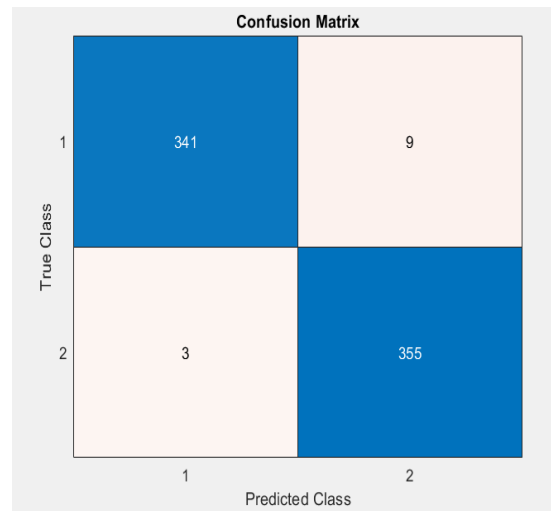


**Figure 10** Classification of new image as pituitary tumor

The subsequent *Figures (11 to 16)* depict the confusion matrix for the proposed work.



**Figure 11** Confusion Matrix for meningioma (BayesNet)

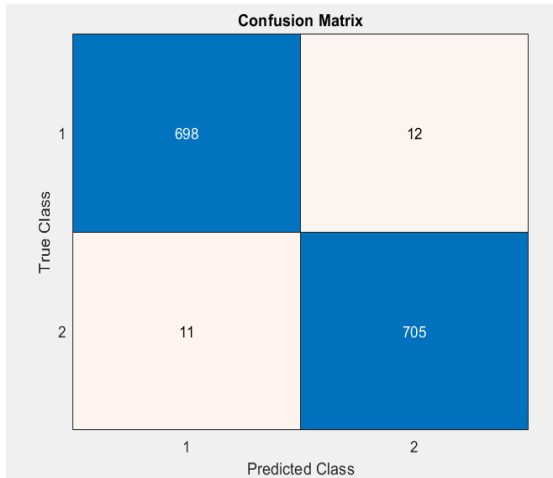


**Figure 12** Confusion Matrix for meningioma (ResNet)

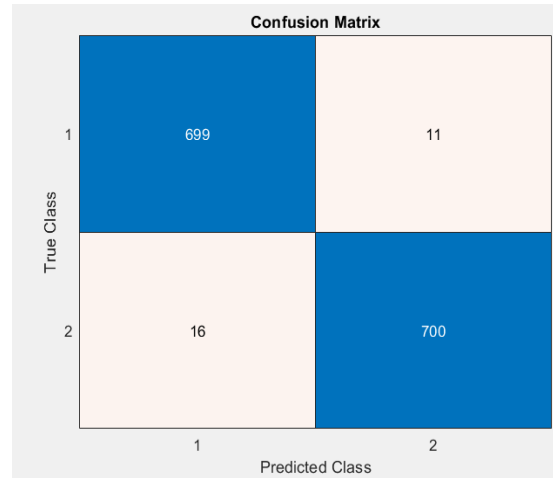
A similar approach has been adopted for feature calculation for all the images in the dataset. The features are used to train the BayesNet. Separate feature extraction is not needed in the case of the ResNet where lower level features are computed at the initial layers and higher-level features are computed at the denser layer. Subsequently, both the BayesNet and ResNet have been tested with the testing dataset and the results of the testing have been presented in *Figures 8 to 10*. Again, for the sake of brevity, only single image illustrations have been presented for each of the three categories. To illustrate the overall performance of both the

classifiers for all the three image classes, the confusion matrices have been presented.

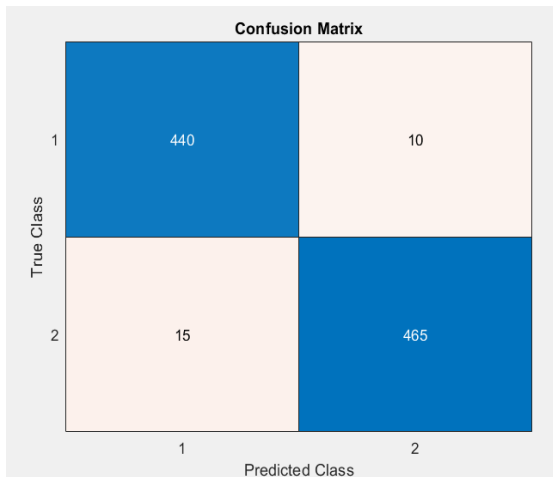
The confusion matrices for each of the three categories of tumors for both the classifiers BayesNet and ResNet depict the number of true and false classifications of the 3 tumor cases. The confusion matrix also allows the ready computation of the performance metrics such as accuracy, sensitivity or recall, specificity, precision and f-measure. A detailed analysis of the results obtained follows in the subsequent section.



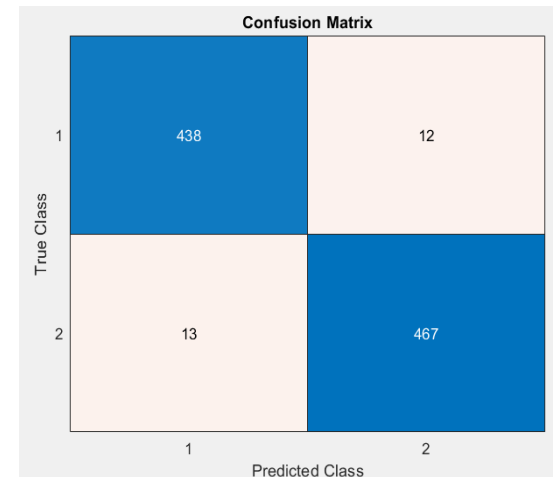
**Figure 13** Confusion Matrix for glioma (BayesNet)



**Figure 14** Confusion Matrix for glioma (ResNet)



**Figure 15** Confusion Matrix for pituitary tumor (BayesNet)



**Figure 16** Confusion matrix for pituitary tumor (ResNet)

## 5. Discussions

The proposed approach aims at addressing the problems identified in the existing literature. Two distinct approaches have been proposed and implemented, which are the BayesNet (machine learning model) and the ResNet (Deep Learning Model). The major contributions of the proposed work are:

- 1) The approach successfully applies segmentation and wavelet based denoising in the image enhancement part.
- 2) An exhaustive set of features is computed to aid pattern recognition and subsequent classification.
- 3) A DBN (with regularization) has been designed and trained with the image features. The regularization term helps to avoid overfitting. Moreover, the BayesNet is effective for classifying overlapping datasets with fuzzy boundaries.

- 4) A deep learning approach based on the ResNet-50 deep neural network has also been trained and tested. The choice of ResNet stems from the fact that several CNN architectures encounter the vanishing gradient and the overfitting problem. The skip connections in the ResNet architecture allows to avoid both.

Thus, the proposed approach puts forth two different approaches, one employing feature extraction and machine learning and the other using deep learning. A comparative analysis of the proposed system with respect to the existing techniques show that the proposed system outperforms the exiting techniques in terms of accuracy of classification. It can be observed from *Table 4* that that the proposed system attains an average accuracy 98.13% and 96.89% for the three categories of the dataset, employing the DBN and ResNet respectively. After the testing of

all the images of the three classes of the dataset, the performance metrics are evaluated and tabulated in *Table 4* and *Table 5*. *Table 4* highlights the performance of the proposed system when the DBN has been used. It can be observed that the system attains highest accuracy in case of meningioma, followed by glioma and pituitary tumors. *Table 5* highlights the performance of the proposed system with ResNet 50.

It can be seen that the system attains highest accuracy in case of meningioma, followed by glioma and

pituitary tumors. It can also be observed that that the DBN performs marginally better than the ResNet 50, in terms of classification accuracy. A comparative analysis with exiting work in the domain has been presented in *Table 6*.

It can be observed from *Table 6* that both the DBN and the ResNet architectures outperform the benchmark techniques cited in *Table 6*.

A complete list of abbreviations is shown in *Appendix I*.

**Table 4** Performance parameters for DBN

Class	Accuracy%	Sensitivity Or Recall%	Specificity%	Precision%	F-Measure
meningioma	98.72	99.42	98.04	98.02	98.71%
glioma	98.38	98.30	98.46	98.44	98.36%
pituitary tumors	97.31	97.77	96.87	96.70	97.23%

**Table 5** Performance parameters for ResNet 50

Class	Accuracy%	Sensitivity Or Recall%	Specificity%	Precision%	Recall
meningioma	98.30	97.42	99.16	99.12	98.26%
glioma	98.10	98.45	97.76	97.76	98.10%
pituitary tumors	94.29	97.33	97.29	97.11	97.21%

**Table 6** Comparative Analysis w.r.t. existing work

Method	No. of Images Used	Accuracy
Phaye et al. [40]	3064	95.05
Anaraki et al. [41]	989 (Axial)	94.2
Tahir et al. [42]	3064	86
Ghassemi et al. [43]	3064	93.01 (introduced split) 95.6 (random split)
<b>Proposed</b>	<b>3064</b>	<b>98.13 (average) (DBN)</b> <b>96.89 (ResNet-50)</b>

## 6. Conclusion and future work

This paper presents a deep neural network based model for brain tumor classification. The proposed scheme presents two network architectures viz. the DBN and the ResNet-50 for multi-class classification of brain MR images. The proposed model presents a rigorous image pre-processing paradigm prior to feature computation, which makes the system more robust and immune to noise effects. The statistical feature extraction presented in this paper would be applicable to a wide range of images, thereby rendering flexibility in the choice of data source and image formats. The performance metrics of the proposed system have been chosen as accuracy, sensitivity, specificity and precision. The two models attain an average accuracy of 98.13% and 96.89% along with significantly higher values of the other performance metrics. A comparative analysis has also

been done with latest existing work in the domain. It has been shown that the proposed technique outperforms the existing work in terms of classification accuracy. As manual labelling the data for large datasets is extremely cumbersome and prone to errors, hence further directions of research can be explored self-supervised learning models and techniques to reduce the complexity in labelling large datasets. The limitation of the proposed approach can be thought of as the requirement for large training datasets and tedious labelling procedure.

Future directions of research in the domain can explore the utility and efficacy of transfer learning. This would allow pre-trained models to be tested on new diverse datasets which may reduce the time spent in labelling of new datasets and make use of the

existing pre-trained models for time critical applications of machine learning in the medical field.

### Acknowledgment

The authors would like to acknowledge the critical suggestions of the faculty members of the Department of Electronics Engineering, Rajkiya Engineering College, Kannauj, India. The suggestions and constructive criticism has helped in polishing the paper and making it more comprehensible.

### Conflicts of interest

The authors have no conflicts of interest to declare.

### References

- [1] [https://www.who.int/health-topics/cancer#tab=tab\\_1](https://www.who.int/health-topics/cancer#tab=tab_1). Accessed 21 July 2021.
- [2] <https://seer.cancer.gov/data-software/documentation/seerstat/nov2017/>. Accessed 21 July 2021.
- [3] Roy PK, Bhui S. Multi-objective quasi-oppositional teaching learning based optimization for economic emission load dispatch problem. *International Journal of Electrical Power & Energy Systems*. 2013; 53:937-48.
- [4] Comoglio PM, Trusolino L, Boccaccio C. Known and novel roles of the MET oncogene in cancer: a coherent approach to targeted therapy. *Nature Reviews Cancer*. 2018; 18(6):341-58.
- [5] Wani AA, Wani MA, Ramzan AU, Nizami FA, Malik NK, Shafiq S, et al. Combination of needle aspiration and core needle biopsy: a new technique of stereotactic biopsy. *Asian Journal of Neurosurgery*. 2016; 11(2):94-7.
- [6] Iqbal S, Khan MU, Saba T, Rehman A. Computer-assisted brain tumor type discrimination using magnetic resonance imaging features. *Biomedical Engineering Letters*. 2018; 8(1):5-28.
- [7] Park C, Took CC, Seong JK. Machine learning in biomedical engineering. *Biomedical Engineering Letters*. 2018; 8(1):1-3.
- [8] Imani M, Ghassemian H. An overview on spectral and spatial information fusion for hyperspectral image classification: current trends and challenges. *Information Fusion*. 2020; 59:59-83.
- [9] Ayadi W, Elhamzi W, Charfi I, Atri M. Deep CNN for brain tumor classification. *Neural Processing Letters*. 2021; 53(1):671-700.
- [10] Arunkumar N, Mohammed MA, Abd GMK, Ibrahim DA, Abdulhay E, Ramirez-gonzalez G, et al. K-means clustering and neural network for object detecting and identifying abnormality of brain tumor. *Soft Computing*. 2019; 23(19):9083-96.
- [11] Kumari N, Saxena S. Review of brain tumor segmentation and classification. In *international conference on current trends towards converging technologies 2018* (pp. 1-6). IEEE.
- [12] Ajai AR, Gopalan S. 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 2020* (pp. 1-10). Springer, Singapore.
- [13] Song G, Huang Z, Zhao Y, Zhao X, Liu Y, Bao M, et al. A noninvasive system for the automatic detection of gliomas based on hybrid features and PSO-KSVM. *IEEE Access*. 2019; 7:13842-55.
- [14] Jabber B, Rajesh K, Haritha D, Basha CZ, Parveen SN. An intelligent system for classification of brain tumours with GLCM and back propagation neural network. In *international conference on electronics, communication and aerospace technology (ICECA) 2020* (pp. 21-5). IEEE.
- [15] Deepak S, Ameer PM. Brain tumor classification using deep CNN features via transfer learning. *Computers in Biology and Medicine*. 2019.
- [16] Kaldera HN, Gunasekara SR, Dissanayake MB. Brain tumor classification and segmentation using faster R-CNN. In *advances in science and engineering technology international conferences 2019* (pp. 1-6). IEEE.
- [17] Aarthi R, Prabha KH. Classification of brain neoplasm from multi-modality MRI with the aid of ANFIS classifier. *Multidimensional Systems and Signal Processing*. 2021; 32(3):933-57.
- [18] Krishnammal PM, Raja SS. Medical image segmentation using fast discrete curvelet transform and classification methods for MRI brain images. *Multimedia Tools and Applications*. 2020; 79(15):10099-122.
- [19] Deepak S, Ameer PM. Automated categorization of brain tumor from MRI using CNN features and SVM. *Journal of Ambient Intelligence and Humanized Computing*. 2020; 12:5357-69.
- [20] Jayaprada S, Jayalakshmi G, Kanyakumari L. Fast hybrid adaboost binary classifier for brain tumor classification. In *IOP conference series: materials science and engineering 2021* (pp. 1-9). IOP Publishing.
- [21] Yerukalareddy DR, Pavlovskiy E. Brain tumor classification based on MR Images using GAN as a pre-trained model. In *ural-siberian conference on computational technologies in cognitive science, genomics and biomedicine 2021* (pp. 380-4). IEEE.
- [22] Shafique M, Naseer M, Theocharides T, Kyrkou C, Mutlu O, Orosa L, et al. Robust machine learning systems: challenges, current trends, perspectives, and the road ahead. *IEEE Design & Test*. 2020; 37(2):30-57.
- [23] Nazir M, Shakil S, Khurshid K. Role of deep learning in brain tumor detection and classification (2015 to 2020): a review. *Computerized Medical Imaging and Graphics*. 2021.
- [24] [https://figshare.com/articles/dataset/brain\\_tumor\\_data/1512427](https://figshare.com/articles/dataset/brain_tumor_data/1512427). Accessed 21 July 2021.
- [25] <https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri>. Accessed 21 July 2021.
- [26] Manjón JV. MRI preprocessing. In *imaging biomarkers 2017* (pp. 53-63). Springer, Cham.



- [27] Dhruv B, Mittal N, Modi M. Analysis of different filters for noise reduction in images. In recent developments in control, automation & power engineering 2017 (pp. 410-5). IEEE.
- [28] Choi H, Jeong J. Despeckling images using a preprocessing filter and discrete wavelet transform-based noise reduction techniques. *IEEE Sensors Journal*. 2018; 18(8):3131-9.
- [29] Zhang Y, Liu Z, Huang M, Zhu Q, Yang B. Multi-resolution depth image restoration. *Machine Vision and Applications*. 2021; 32(3):1-5.
- [30] Sarkar S, Das S, Chaudhuri SS. Multi-level thresholding with a decomposition-based multi-objective evolutionary algorithm for segmenting natural and medical images. *Applied Soft Computing*. 2017; 50:142-57.
- [31] Guido RC. A tutorial review on entropy-based handcrafted feature extraction for information fusion. *Information Fusion*. 2018; 41:161-75.
- [32] Srivastava D, Rajitha B, Agarwal S, Singh S. Pattern-based image retrieval using GLCM. *Neural Computing and Applications*. 2018; 32:10819-32.
- [33] Zhou W, Yu L, Qiu W, Zhou Y, Wu M. Local gradient patterns (LGP): an effective local-statistical-feature extraction scheme for no-reference image quality assessment. *Information Sciences*. 2017; 397:1-14.
- [34] Vlachas PR, Pathak J, Hunt BR, Sapsis TP, Girvan M, Ott E, et al. Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics. *Neural Networks*. 2020; 126:191-217.
- [35] Abraham B, Nair MS. Computer-aided detection of COVID-19 from X-ray images using multi-CNN and Bayesnet classifier. *Biocybernetics and Biomedical Engineering*. 2020; 40(4):1436-45.
- [36] Zhu D, Cai C, Yang T, Zhou X. A machine learning approach for air quality prediction: model regularization and optimization. *Big Data and Cognitive Computing*. 2018; 2(1):1-15.
- [37] Sarwinda D, Paradisa RH, Bustamam A, Anggia P. Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer. *Procedia Computer Science*. 2021; 179:423-31.
- [38] Al-haija QA, Adebajo A. Breast cancer diagnosis in histopathological images using ResNet-50 convolutional neural network. In international IOT, electronics and mechatronics conference 2020 (pp. 1-7). IEEE.
- [39] Vejjanugraha P, Kotani K, Kongprawechnon W, Kondo T, Tungpimolrut K. Automatic screening of lung diseases by 3D active contour method for inhomogeneous motion estimation in CT image pairs. *Walailak Journal of Science and Technology*. 2021; 18(12).
- [40] Phaye SS, Sikka A, Dhall A, Bathula D. Dense and diverse capsule networks: making the capsules learn better. arXiv preprint arXiv:1805.04001. 2018.
- [41] Anaraki AK, Ayati M, Kazemi F. Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *Biocybernetics and Biomedical Engineering*. 2019; 39(1):63-74.
- [42] Tahir B, Iqbal S, Usman GKM, Saba T, Mehmood Z, Anjum A, et al. Feature enhancement framework for brain tumor segmentation and classification. *Microscopy Research and Technique*. 2019; 82(6):803-11.
- [43] Ghassemi N, Shoeibi A, Rouhani M. Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images. *Biomedical Signal Processing and Control*. 2020.



**Gaurish Joshi** obtained his Bachelors in Engineering from Z.H.C.E.T, Aligarh Muslim University (AMU) Aligarh and Masters in Technology from Motilal Nehru National Institute of Technology, (MNIT) Allahabad. His areas of interest are Image Processing, Data Science, Machine Learning and

Wireless Networks. Presently he is an Assistant Professor in the Department of Electronics Engineering, Rajkiya Engineering College, Kannauj, India.

Email: gaurishjoshi@reck.ac.in



**Rajeev Kumar** obtained his Bachelors in Technology from UIET, Kanpur and Masters in Technology from Indian Institute of Technology, Kanpur (IITK). He is presently a PhD scholar at IIT, Kanpur. His areas of interest are Signal Processing, Wireless Sensor Networks, IoT and Future Generation

Wireless Networks. Presently he is an Assistant Professor in the Department of Electronics Engineering, Rajkiya Engineering College, Kannauj, India.

Email: rajeev@reck.ac.in



**Amit Kumar Singh Chauhan** obtained his Bachelors in Technology from BIET Jhansi and Masters in Technology from Motilal Nehru National Institute of Technology, (MNIT) Allahabad. His areas of interest are Statistical Signal Processing, Image, Wireless Networks and

Communication Systems. Presently he is an Assistant Professor in the Department of Electronics Engineering, Rajkiya Engineering College, Kannauj, India.

Email: amitchauhan@reck.ac.in

**Appendix I**

S. No.	Abbreviation	Description
1	$A_c$	Accuracy
2	AHE	Adaptive Histogram Equalization
3	AEHO	Adaptive Elephant Herd Optimization
4	ANFIS	Adaptive Neuro Fuzzy Inference Systems
5	BayesNet	Bayesian Network
6	BPNN	Back Propagation Neural Networks
7	$C_A$	Approximate Co-Efficient
8	$C_D$	Detailed Co-Efficient
9	$C_H$	Horizontal Co-Efficient
10	$C_V$	Vertical Co-Efficient
11	Conv	Convolution
12	CNN	Convolutional Neural Networks
13	CPU	Central Processing Unit
14	CT	Computed Tomography
15	DBN	Deep Bayes Network
16	DWT	Discrete Wavelet Transform
17	FN	False Negative
18	FNAC	Fine Needle Aspiration Cytology
19	FP	False Positive
20	GAN	Generative Adversarial Network
21	GLCM	Gray Level Co-Occurrence Matrices
22	HSI	Hyperspectral Imaging
23	IDM	Inverse Difference Moment
24	J	Jacobian Matrix
25	KNN	K-Nearest Neighbor
26	K-SVM	Kernel Support Vector Machine
27	MFCM	Modified-Fuzzy C Means
28	max	Maximum
29	MR	Magnetic Resonance
30	MRI	Magnetic Resonance Imaging
31	PNN	Probabilistic Neural Network
32	PSO	Particle Swarm Optimization
33	$P_r$	Precision
34	RAM	Random Access Memory
35	RCNN	Region Based-Convolutional Neural Network
36	ReLU	Rectified Linear
37	ResNet	Residual Network
38	RGB	Red, Green, Blue
39	SMR	Standardized Mortality Ratio
40	$S_e$	Sensitivity
41	$S_p$	Specificity
42	SVM	Support Vector Machine
43	TN	True Negative
44	TP	True Positive