RESEARCH ARTICLE

Effective and Efficient LDA+ELM Model for Supervised Classification of Brain Tumor Types Using 2D MRI Scans

Lim Jia Qi[*], Norma Alias

Department of Mathematical Sciences, Faculty of Science, 81310 UTM Johor Bahru, Johor, Malaysia

* Corresponding author email: ykj10@hotmail.com

ABSTRACT

Application of machine learning in multiclass classification of brain tumor types has contributed to the development of computer aided diagnosis (CAD) system that can potentially enhance accuracy and speed up diagnosis of the disease. LDA+ELM model with different activation functions were investigated to achieve the optimum performances in terms of accuracy, Kappa statistic, sensitivity, precision, F-measure, training time and test time. We also proposed a user-friendly GUI in characterizing brain tumor types using MR images. First, a total of 3064 slices of CE T1-weighted brain MR images with ground truth were downloaded from a free online database. The manually segmented tumor region was augmented and then undergo several feature extraction techniques. All the feature descriptors obtained were then concatenated, followed by LDA dimensionality approach. Performance of different number of LDA features and ELM activation functions were investigated by repeated training and test. The ELM output of training data for each class was used to fit GMM and these probabilistic models used to estimate posterior probabilities of test data. LDA+ELM model with 5 LDA feature input, utilizing sigmoid function as hidden nodes activation functions achieves the best generalization performance with accuracy of 98.92% and corresponding F-scores for meningioma, glioma and pituitary tumor of 97.81%, 99.1% and 99.5% respectively. The proposed method (LDA+ELM) model performs better compared to other previous works using the same dataset and performing the same classification task.

Keywords: Multi-class classification of brain tumor, Linear Discriminant Analysis (LDA), Extreme Learning Machine (ELM), activation functions, Gaussian Mixture Model (GMM).

1 Introduction

Simply speaking, brain tumor is the collection of abnormal cells in central nervous system. Early detection of brain tumor is crucial for treatment planning and prognosis, thus enhancing the chance of survival of patients. Nonetheless, tumor characterization and evaluation is difficult due to the fact that neoplastic tissues often appear heterogeneous pixel-wise in biomedical images [1]. While biopsy (extraction of tissues) remains as gold standard in cancer diagnosis, it is also invasive and could not characterize brain tumor spatial and temporal heterogeneity. Analysis of medical images data can be excellent alternative as they are non-invasive, time efficient and contain rich histopathologic information.

Nowadays, considerable amounts of medical image dataset are generated and collected during routine clinical practices [2], which open up the possibility of scientific knowledge discovery, while at the same time rendering manual analysis of medical data implausible. Current advancement of statistical modeling and computer vision enables semi-automated or even automated image processing in clinical practice, like image registration, tissue classification, volumetric analysis, and image segmentation. This gives rise to an emerging field of study, called radiomics, from which can be dated back to the work of [3]. Radiomics is high-throughput conversion of image data into mineable high-dimensional features for the purpose of

improved clinical diagnosis. Categorization of brain tumor types using medical images can be achieved through sequence of radiomics workflow: data acquisition, regions of interest (ROI) segmentation, feature extraction, feature reduction, followed by model construction and validation. This study can contribute to the development of computer aided diagnosis (CAD) system. One of the most important purposes of CAD is to provide objective and accurate diagnosis outcome that can complement radiologists in decision making. Additionally, it can shorten image analysis duration and solves the issue of variability of opinions among clinical experts. Ultimately, if proven accurate and consistent in real-time application, biopsy (extraction of tumor samples) can be bypassed or at least avoided if necessary in diagnosis procedure.

In view of the severity of brain tumor and lack of effective treatment, early diagnosis of the disease can be crucial for the patient recovery and treatment planning for the doctor. Advanced imaging protocols, like computed tomography (CT), positron emission tomography (PET), and MRI have provided unprecedented high resolution of medical images that make medical image analysis possible. In this paper, MRI will be the modality of choice because : 1) MRI is non-invasive, 2) produces multiple slices of images of the same tissue region with different contrast by applying different image acquisition protocols and parameters [4], 3) high contrast of soft tissues and high spatial resolution. In this paper, only contrast enhanced T1-weighted MR image will be considered as multi-spectral MR images analysis is expensive and increases odds of segmentation errors due to inconsistency and misalignment [5]. This paper aims to extract radiomic features from MRI scans to determine the correct type of tumor for MR image slices.

In recent years, there has been surge in research interest regarding supervised classification of brain tumor types using MRI scans. Early work performed by [1] investigated the use of machine learning schemes (LDA, kNN and nonlinear SVM) in binary classification of metastases and glioma as well as grades of glioma. In addition, the authors also performed multiclass classification by applying one versus all SVM model. The overall accuracy in distinguishing glioma grade II, III, IV and metastases is 63.3%. The same classification problem was explored by [6]. The highest leave-one-out cross-validation accuracy of 76.29% was achieved by Best First Search algorithm and VFI as classifier. [7] put forward a comprehensive radiomics workflow, in which classes involved are astrocytoma, glioblastoma multiforme, childhood tumor-medulloblastoma, meningioma (MEN), secondary tumor-metastatic and normal regions. A total sample of 856 regions of interest and 218 features were extracted. The overall accuracy achieved is 85.23% using PCA-ANN approach. Classification problem of 6 classes, glioma grade I, II, III and IV, meningioma and metastases was performed by [8] using one against one SVM model.

Recent work of [9] utilized first order and second order statistical features from brain MR images to develop SVM model. Their experiment shows that highest accuracy of 85% was attained by SVM model trained by second-order statistical features. [10] proposed a new machine learning pipeline consisting of hybrid structure descriptor and fuzzy-based RBF SVM that categorized MR images into meningioma, metastases, glioma II and III. The overall accuracy of 96.74% was reported. Back-propagation artificial neural networks (BPANN) trained with statistical features from DWT transform and Gabor filter developed by [11] in classifying meningioma, glioma and pituitary tumor achieved overall accuracy of 91.9%.

Deep learning has now become the state-of-the-art in many domains, especially in automatic image analysis, when a major breakthrough in image classification of ImageNet challenge in 2012 [12]. The reasons behind the success of deep neural networks include: 1) availability of sophisticated hardware with high computing power, 2) development of learning algorithm [13], 3) reinforcement learning, which can capture crucial semantic information pertaining class discrimination[14]. Even though deep learning model is well known for its remarkable recognition accuracy, it requires tuning of hyper-parameters, sophisticated hardware and large

amount of annotated data. Its performance in small data samples problem is still inconclusive.

To the best of our knowledge, most of the previous works regarding supervised classification of brain tumor types focus on development of novel machine learning pipelines and attaining high overall accuracy. In other word, only the discrete outputs of classifier (with test data as input) are being taken into account during evaluation phase of the proposed model. In addition, output scores (confidence) of classifier might not represent underlying real probability of certain class given the attributes of sample. In fact, modern neural networks are poorly calibrated, despite being accurate [15]. Accurate class probability estimates is crucial in real-world high stake decision making system, especially medical diagnosis [16], whereby there is difference in cost of misclassification and domain knowledge is required [17] because it provides likelihood or uncertainty of predictions. Since human have natural cognitive intuition for

probability [18], probabilistic outputs enhance reliability of the system. This study bridges the gap by applying Gaussian mixture model to transform the extreme machine learning (ELM) model outputs in reference to work of [19].

The goal of this paper is to address multi-class recognition of brain tumor types using 2D MRI scans by sequence of steps outlined in radiomics. Combination of linear discriminant analysis (LDA) feature reduction approach and ELM classifier, which we abbreviated as LDA+ELM is proposed. Apart from that, different activation functions, such as sigmoid, hard-limit, sine and radial basis function (RBF) are investigated. Last but not least, a user-friendly graphic interface (GUI) is put forward, with probability of respective brain tumor types (e.g. meningioma, glioma and pituitary tumor) as output.

2   Research Methodology

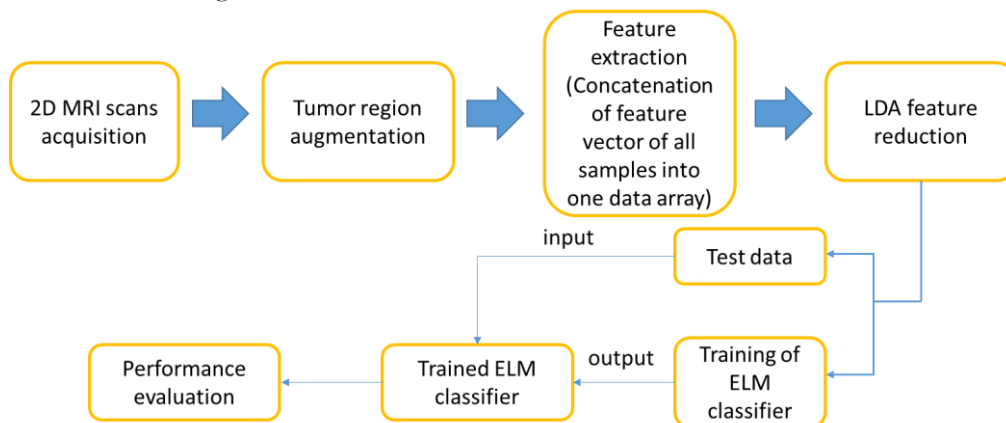The research framework proposed can be summarized in Fig. 1 and Fig. 2.

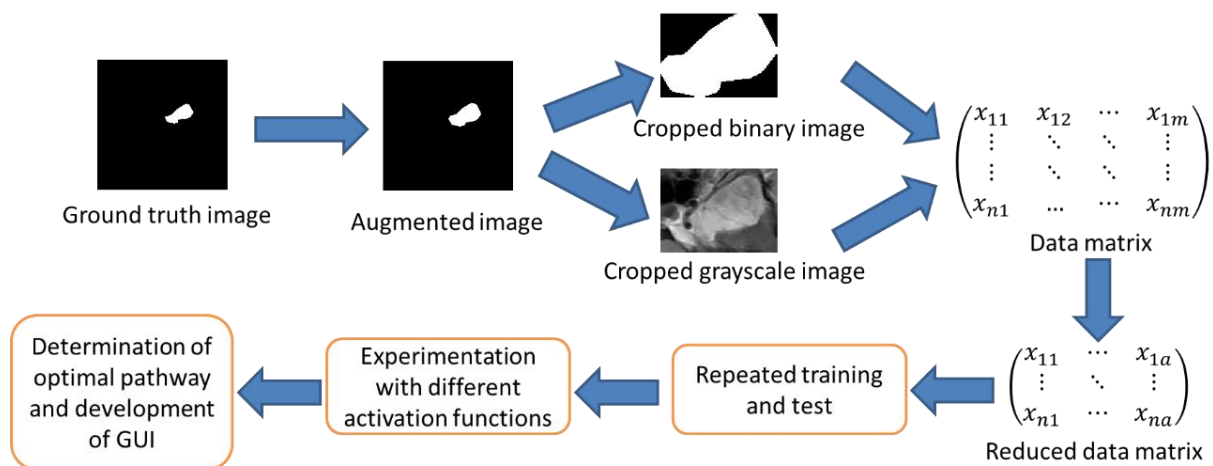

Figure 1: *Summary of research framework*

Figure 2: *Schematic diagram of research flow.*

This workflow was performed under one to 80 LDA attributes. Additionally, for each training and test data partitioning, random stratified sampling was conducted to ensure similar class distribution for both training and test dataset. 70% of original data is allocated for training, while another 30% is partitioned for performance evaluation (test). The optimal machine learning pipeline was then determined by analyzing several performance indicators and the output of the trained ELM classifier was used to fit GMM model for each class. The resulting GMM models can be used estimate the class membership probability of test data. Lastly, a user-friendly GUI capable of performing brain tumor types classification was developed.

2.1   2D MRI scans acquisition

The brain MR images were downloaded from publicly available online database. A total of 3064 slices of T1-weighted contrast enhanced MR images from 233 patients was downloaded from https://figshare.com/articles/brain_tumor_data set/1512427. There are three kinds of brain tumor in the MR images downloaded, namely meningioma (708 slices), glioma (1426 slices), and pituitary tumor (930 slices). The brain T1-weighed CE-MRI dataset was acquired from Nanfang Hospital, Guangzhou, China, and General Hospital, Tianjing Medical University, China, from year 2005 to 2010. The images have an in-plane resolution of 512×512 with pixel size 0.49×0.49 mm^2. The slice thickness is 6 mm and the slice gap is 1 mm [20, 21].

2.2   Tumor region augmentation

The 2D brain MR images comes with ground truth tumor region delineated manually by experienced radiologists. As pointed out in [22], tissues surrounding tumors can provide useful discriminative information about the types of tumor. Thus, augmentation of tumor region can be beneficial in extracting robust features. In this study, augmentation of tumor region was performed by morphological dilation with disk-shape structuring element with radius, R of 8.

2.3   Feature extraction

Simply speaking, feature extraction is a dimensionality reduction method, in which instead of utilizing pixel intensity in raw images, a set of attributes, also known as feature vector is constructed to represent a certain image. This stage is crucial as it can directly impacts the generalization performance of classification models [23]. Ideal set of feature should be relevant, non-redundant, provide intrinsic dimensional representation for visualization, and increase training and inference speed of learning algorithms [24], 2003). In this paper, several feature extraction techniques were employed, including shape parameters, geometric moment invariants [25, 26], Zernike moments [27, 28], pseudo Zernike moments [29], histogram of oriented gradients (HOG) [30], linear binary pattern (LBP) [31], and bag of words (BOW) model [32]. It is worth noting that the image features will be extracted from the augmented tumor region. The number of features extracted from each method are summarized in Table 1. Therefore, the whole feature matrix dimension is 3064 × 1553.

Table 1: *Feature extraction and its number of features.*

Feature extraction methods	Number of features
Binary shape parameters	11
Geometric moment invariants	7
Zernike moments	12
Pseudo Zernike moments	15
HOG	900
LBP	108
Bag of words model	500

2.4   LDA feature reduction

LDA is a linear dimensionality reduction method that can resolve the issue arise from high dimension data, such as high computational cost and complexity as well as undesirable modeling performance [33]. LDA is a supervised approach where class label of dataset is required. This is because, in the formulation of LDA, it tries to

compute the "directions" that will maximize the separation of training samples between classes and at the same time minimizes the separation of test samples within class.

Let define some important notation, x_i be the row vector representing features of a sample, $\overline{x_k}$ is the mean vector for k class, μ is the overall mean vector of whole data, C_k denotes the indices of samples belong to k class and N_k is the number of class available. The steps involved for feature transformation using LDA are as follow:

1) Calculate the mean for each class, $\overline{x_k}$ and overall mean, μ.

2) Calculate the within class scatter matrix, S_w and between class scatter matrix, S_b using formula below:

$$S_w = \sum_{k=1}^{N_k} \sum_{i \in C_k} (x_i - \overline{x_k})(x_i - \overline{x_k})^T \quad (1)$$

$$S_b = \sum_{k=1}^{N_k} n_k (\overline{x_k} - \mu)(\overline{x_k} - \mu)^T \quad (2)$$

3) Maximize w such that between class scatter is maximized and within class scatter is minimized. Seeking optimum vector w is equivalent to solving generalized eigenvector system below:

$$S_b w = \lambda S_w w \quad (3)$$

4) After finding w and λ, rearrange the eigenvectors according to the descending order of λ. The original data matrix is projected into new dimensional space, Z by linear transformation:

$$Z = Xw \quad (4)$$

2.5 ELM classifier

Recently, ELM has attracted a lot of interest among researchers due to its fast training convergence and similar prediction accuracy compared to state-of-the-art classification models like neural networks and support vector machine (SVM). The working principle of ELM is to train single layer feed forward neural networks (SLFN). Unlike conventional neural networks in which the hidden nodes parameters are iteratively tuned based on the training error, the hidden node parameters of ELM are generated randomly. It has been proven in the work of [34, 35] that without tuning hidden node parameters of SLFN with arbitrary number of hidden nodes, the system can still learn from training data. Since the hidden node parameters are randomly generated, they are independent of each other as well as of training data. The general architecture of ELM is depicted as in Fig. 3.

Consider training set containing n samples, (x_i, t_i), where $x_i \in \mathbb{R}^m, t_i \in \mathbb{R}^{N_k}$, whereby $t_1 = [1, -1, -1]^T$, $t_2 = [-1, 1, -1]^T$, $t_3 = [-1, -1, 1]^T$. The output function can be expressed as below:

$$f_L(x) = \sum_{i=1}^{L} \beta_i G(a_i, b_i, x) \quad (5)$$

where β_i is the output weight for the output of hidden node activation function $G(a_i, b_i, x)$, a_i and b_i are both parameters for hidden node i, L is the number of hidden node used. Several hidden node output functions, such as sigmoid, hard limit, sine and radial basis function (rbf) activation functions in which their formula are displayed in Table 2 were experimented. It should be noted that 300 hidden nodes are chosen in this study. The above equation can also be written in matrix form:

$$H\beta = T_L \quad (6)$$

where,

$$H = \begin{pmatrix} G(a_1, b_1, x_1) & \cdots & G(a_L, b_L, x_1) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_n) & \cdots & G(a_L, b_L, x_n) \end{pmatrix}_{n \times L}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_L \end{pmatrix}_{L \times N_k}$$

$$T_L = \begin{pmatrix} T_1 \\ \vdots \\ T_n \end{pmatrix}_{n \times N_k}$$

The objective of ELM is to minimize $\|H\beta - T\|$ and $\|\beta\|$. Minimization of $\|H\beta - T\|$ is equivalent to minimization of training error while minimization of $\|\beta\|$ is equivalent to minimization of norm of weight of connection between hidden nodes and output layers, β. According to Bartlett's theory [36], for feedforward neural networks reaching smaller training error, the smaller the norm of weights is,

the better generalization performance the networks tends to have. In the paper of [37], minimizing $\|\boldsymbol{\beta}\|$ can be regarded as maximizing the distance of the separating margins of the two different classes in the ELM feature space. As such in order to optimize $\boldsymbol{\beta}$, minimal norm least square formula can be employed:

$$\boldsymbol{\beta} = \boldsymbol{H}^{\dagger}\,\boldsymbol{T} \qquad (7)$$

where $\boldsymbol{H}^{\dagger}$ denotes generalized Moore-Penrose inverse matrix. The above equation can also be expressed as:

$$\boldsymbol{\beta} = (\boldsymbol{H}^{T}\boldsymbol{H})^{-1}\,\boldsymbol{H}^{T}\boldsymbol{T} \qquad (8)$$

Nonetheless, the above formulation can become computationally implausible if inverse of $\boldsymbol{H}^{T}\boldsymbol{H}$ do not exists and computationally unstable if the condition number of square matrix $\boldsymbol{H}^{T}\boldsymbol{H}$ is large. To avoid these issues, diagonal elements of $\boldsymbol{H}^{T}\boldsymbol{H}$ can be added with positive value according to ridge regression theory [38]. In our case where number of training instances are large, $\boldsymbol{\beta}$ can be solved via [39]:

$$\boldsymbol{\beta} = (\frac{I}{c} + \boldsymbol{H}^{T}\boldsymbol{H})^{-1}\,\boldsymbol{H}^{T}\boldsymbol{T} \qquad (9)$$

where c is a positive number. The author recommended the use of c in the range of $\{2^{-24}, 2^{-23}, \dots, 2^{25}\}$. 5-fold cross validation methods are employed to find the best c.
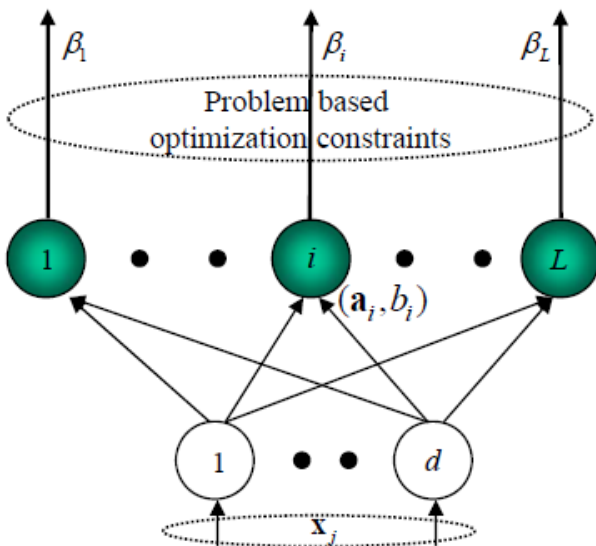

Figure 3: *General architecture of ELM [40]*

Table 2: *Formula of activation functions*

Activation functions	Formula
Sigmoid	$G(\boldsymbol{a}, b, \boldsymbol{x}_i)$ $= \dfrac{1}{1 + \exp\left(-(a x_i + b)\right)}$
Hard-limit	$G(\boldsymbol{a}, b, \boldsymbol{x}_i) = \begin{cases} 1, a x_i + b \geq 0 \\ 0, otherwise \end{cases}$
Sine	$G(\boldsymbol{a}, b, \boldsymbol{x}_i) = \sin\left(\boldsymbol{a} x_i + b\right)$
rbf	$G(\boldsymbol{a}, b, \boldsymbol{x}_i) = \exp\left(-b\|\boldsymbol{x}_i - \boldsymbol{a}\|^2\right)$

2.6 GMM

Simply speaking, GMM is a type of unsupervised soft clustering method that assign the likelihood of cluster membership for each data instance, with assumption that data are generated by finite number of Gaussian distributions. Being a parametric probability density function, the probability of certain data point is represented as:

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}|\mu_k, {\textstyle\sum}_k) \qquad (10)$$

where π_k is mixing coefficient, $\mathcal{N}$ represents multivariate probability density function, μ_k is mean vector, $\sum_k$ is covariance matrix, and K is the number of clusters. Parameters involving π_k, μ_k and $\sum_k$ are optimized by maximizing log likelihood function as expressed below:

$$\ln(p(\boldsymbol{x}))$$
$$= \sum_{i=1}^{n} \ln\left(\sum_{k=1}^{K} \pi_i \mathcal{N}(\boldsymbol{x_n}|\mu_k, {\textstyle\sum}_k)\right) \qquad (11)$$

Since there is no elegant closed form solution to equation (11), it can be solved using iterative algorithm, called expectation-maximization (EM) algorithm. To produce reliable probability estimates, separate GMM is used to fit the output vectors (from training data) of ELM that are correct for each class. Each GMM can model conditional probability of ELM output given respective class, $p(t|C)$. By applying Bayes' theorem, we can estimate the class conditional probability given a certain ELM output from test data:

$$p(C_i|t) = \frac{p(t|C_i)p(C_i)}{p(t)}$$
$$= \frac{p(t|C_i)p(C_i)}{\sum_{i=1}^{N_k} p(t|C_i)p(C_i)} \qquad (11)$$

The only arbitrary parameter is the number of clusters, K. We employ Akaike Information Criterion (AIC) to select the best K. We utilize the idea in the work of [19] by fitting GMM with only the correctly identified samples during training phase and calculate the class conditional probability based on the above formula during test phase.

2.7 Performance evaluation

In order to evaluate the performance of LDA+ELM classification model under different number of features, 7 commonly accepted performance measures were utilized, including test accuracy (%), Kappa statistic, sensitivity (%), precision (%), F-measure (%), training time (s) and test time (s). Mean and standard deviation of each performance metrics were computed for unbiased analysis and comparison among methods employed. This process is necessary as robust performance evaluation can provide clear picture of the strengths and weaknesses of the machine learning paradigms being analyzed. Table 3 shows some notations in confusion matrix. Table 4 describes the performance measures and its definitions.

Table 3: *Notations in confusion matrix.*

Predicted classes				
		A (meningioma)	B (glioma)	C (pituitary tumor)
True classes	A (meningioma)	TP_A	E_{AB}	E_{AC}
	B (glioma)	E_{BA}	TP_B	E_{BC}
	C (pituitary tumor)	E_{CA}	E_{CB}	TP_C

Table 4: *Performance measures and its respective formula.*

Performance measures	Definition	Formula
Accuracy	Overall efficiency and generalizability of classifier [41]. However, its use in performance comparison among classifiers is limited [42].	$Acc = \frac{TP}{n_T}$, $n_T = number\ of\ test\ samples$
Kappa statistic	Measure the degree of agreement between the predicted labels and the ground truth [43].	$Kappa = \frac{Acc - P(E)}{1 - P(E)}$, P(E)=expected agreement between classifier and ground truth by chance
Sensitivity	Compute the proportion of samples of class 'A' label that are correctly predicted as class 'A' in the test phase.	For class A, $Sensitivity\ of\ A = \frac{TP_A}{TP_A + E_{AB} + E_{AC}}$
Precision	Indicate the proportion of test samples that are predicted to be class 'A' that match the known true class label.	For class A, $Precision\ of\ A = \frac{TP_A}{TP_A + E_{BA} + E_{CA}}$
F-measure/F-score	Harmonic mean of precision and sensitivity.	For class A, $F - score\ of\ A = \frac{2TP_A}{2TP_A + E_{AB} + E_{AC} + E_{BA} + E_{CA}}$
Training time	Quantify the convergence speed of training.	-
Test time	The rate at which the output labels of test samples are generated.	-

3   Results

In this paper, we investigated different activation functions of ELM under different number of LDA features. Fig. 4-7 shows the performance measures (e.g. accuracy, Cohen' Kappa, training time and test time) of sigmoid, hard-limit, sine and rbf hidden node activation functions respectively. Empirical observation of these figures shows that the performances of sigmoid and hard-limit transfer function are relatively more stable and insensitive to changes in number of features (more resistant to overfitting issue). Accuracies and Kappa statistics for both sine and rbf decline after number of features increase beyond a threshold.
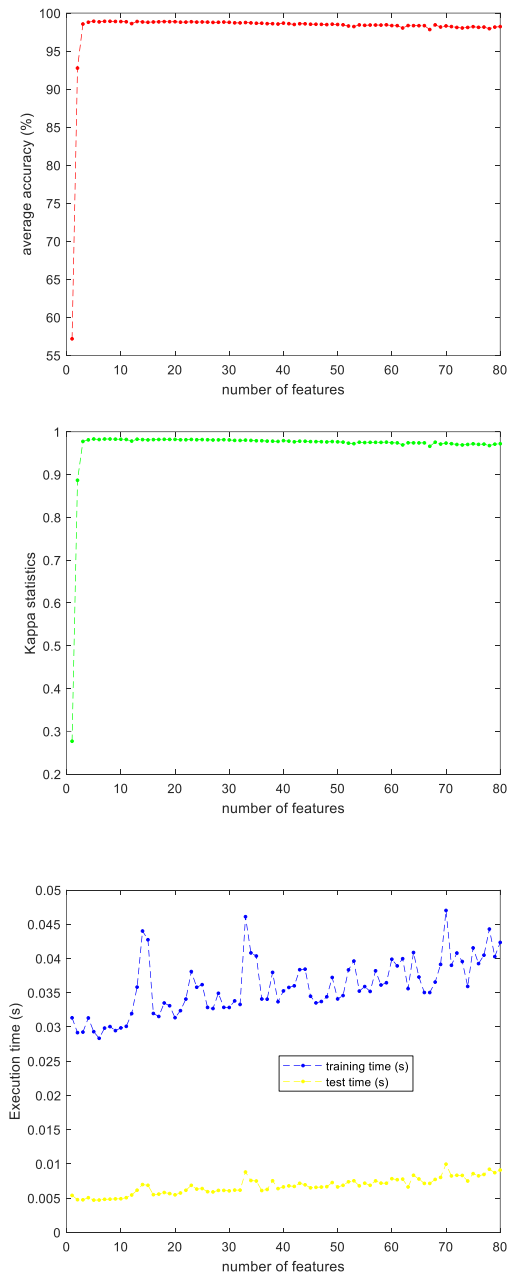

Figure 4: *Performance measures versus number of features for sigmoid activation function.*
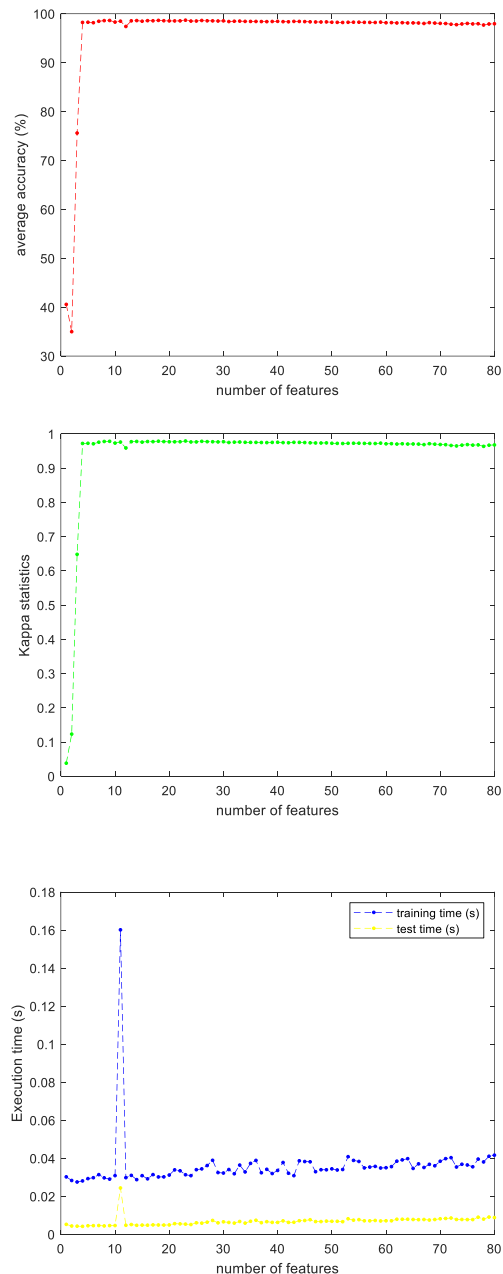

Figure 5: *Performance measures versus number of features for hard-limit activation function.*
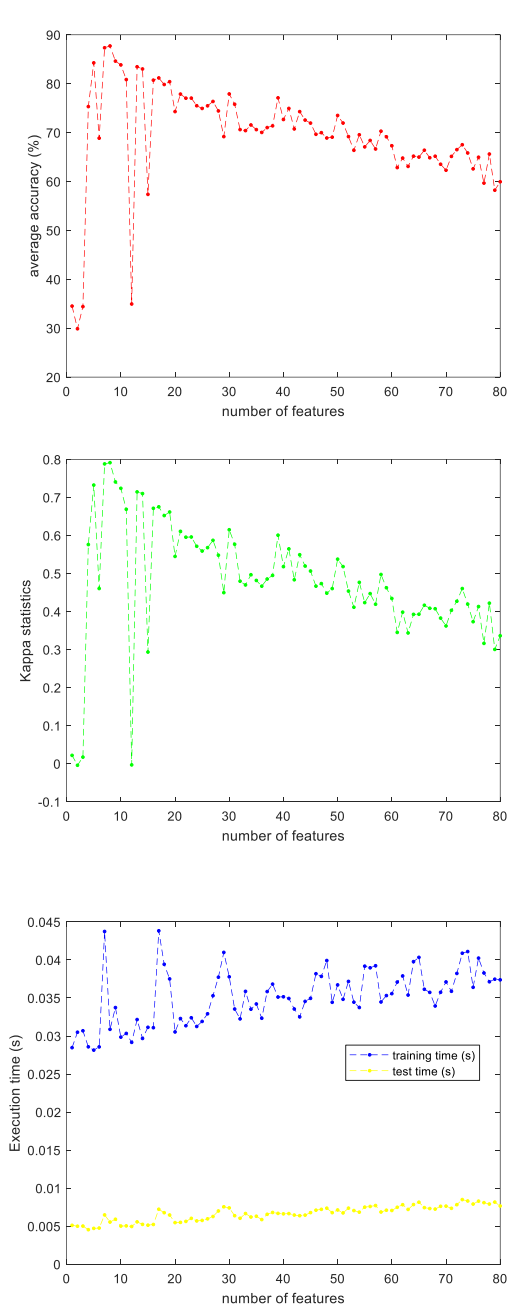
Figure 6: *Performance measures versus number of features for sine activation function.*
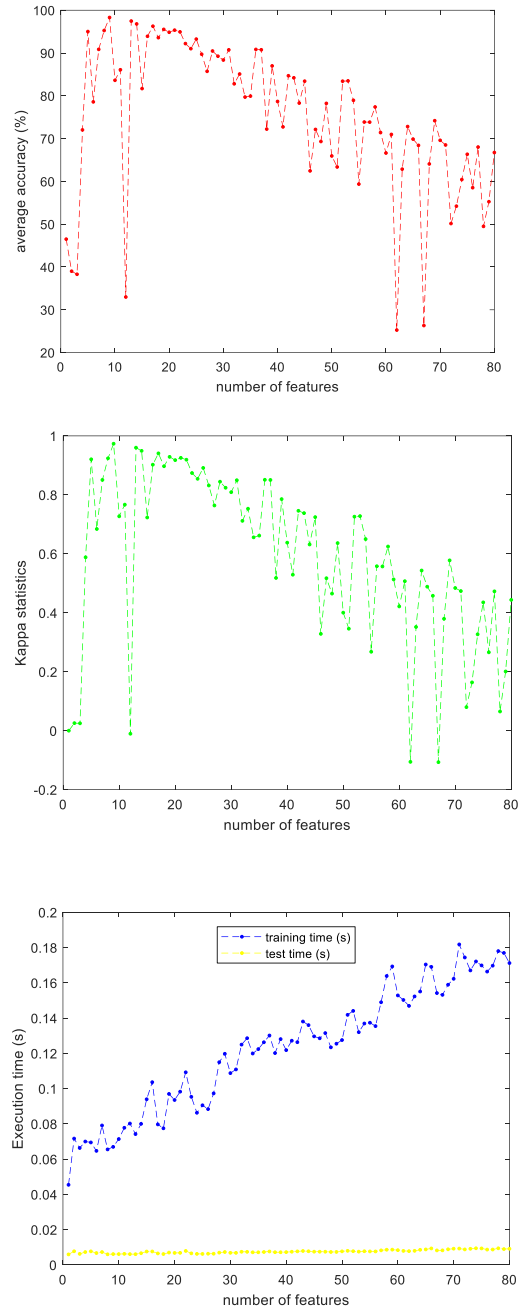
Figure 7: *Performance measures versus number of features for rbf activation function.*

Table 6 shows the performance measures of activation function under optimal number of features. The best generalization performance is achieved by sigmoid function, with accuracy of 98.92%, using just 5 LDA attributes, Interestingly, very high standard deviation in performance measures of sine activation function indicates that its performance is very sensitive to sampling of data. Table 7 shows performance metrics of sigmoid function before and after applying GMM on ELM outputs. There is no significant difference in performance measures before and after applying GMM on ELM outputs, suggesting that GMM can approximate the probability of classes well.

Fig. 8 shows the GUI developed and the function of each buttons clearly explained.

Table 6: *Performance measures of activation function under optimal number of features.*

Activation functions		sigmoid	hardlim	sine	rbf
number of features		5	23	8	9
Accuracy (%)		98.923±0.378	98.680±0.361	87.715±11.912	98.306±0.518
Kappa statistics		0.983±0.006	0.979±0.006	0.791±0.218	97.337±0.819
Sensitivity (%)	A	97.536±0.971	96.186±1.447	64.232±27.192	95.088±2.684
	B	99.509±0.360	99.509±0.360	99.953±0.113	99.392±0.595
	C	99.081±0.513	99.308±0.351	86.838±22.503	99.093±0.846
Precision (%)	A	98.081±0.905	98.861±0.687	99.280±1.618	98.060±1.488
	B	98.694±0.503	98.045±0.632	81.970±13.132	97.792±1.199
	C	99.928±0.197	99.547±0.373	99.710±0.490	99.366±0.944
F-measure (%)	A	97.805±0.740	97.498±0.763	73.947±24.859	96.514±1.085
	B	99.099±0.370	98.770±0.364	89.466±8.667	98.579±0.558
	C	99.502±0.263	99.426±0.246	90.496±20.066	99.223±0.481
training time (s)		0.029±0.003	0.031±0.003	0.031±0.002	0.067±0.004
test time (s)		0.005±0.000	0.005±0.000	0.006±0.001	0.006±0.000

A-meningioma, B-glioma, C-pituitary tumor

Table 7: *Performance metrics (mean ± standard deviation) of sigmoid function before and after applying GMM on ELM outputs.*

Performance measures		Using classifier scores	After using GMM (fitted with all training samples)	After using GMM (fitted with correctly identified training samples)
Accuracy (%)		98.923±0.378	98.669±0.431	98.799±0.439
Kappa statistics		0.983±0.006	0.979±0.007	0.981±0.007
Sensitivity (%)	A	97.536±0.971	97.552±1.088	97.474±1.074
	B	99.509±0.360	98.893±0.563	99.353±0.548
	C	99.081±0.513	99.176±0.525	98.961±0.519
Precision (%)	A	98.081±0.905	97.010±1.452	97.640±1.256
	B	98.694±0.503	99.002±0.455	98.693±0.542
	C	99.928±0.197	99.453±0.464	99.868±0.221
F-measure (%)	A	97.805±0.740	97.272±0.906	97.551±0.894
	B	99.099±0.370	98.946±0.392	99.021±0.424
	C	99.502±0.263	99.312±0.268	99.412±0.289
training time (s)		0.029±0.003	0.278±0.035	0.556±1.349
test time (s)		0.005±0.000	0.007±0.001	0.007±0.002

A-meningioma, B-glioma, C-pituitary tumor

① This push button displays the 2D brain MR image and its file pathname.

② This push button displays display ground truth tumor region.

③ This push button perform feature extraction mentioned in section 2.3.

④ This push button perform LDA linear dimensionality reduction. The dimension of resulting attribute vector is 5.

⑤ This push button feed the trained ELM model (300 hidden neurons with sigmoid activation function) with the attribute vector and output probability estimates of each brain tumor types.
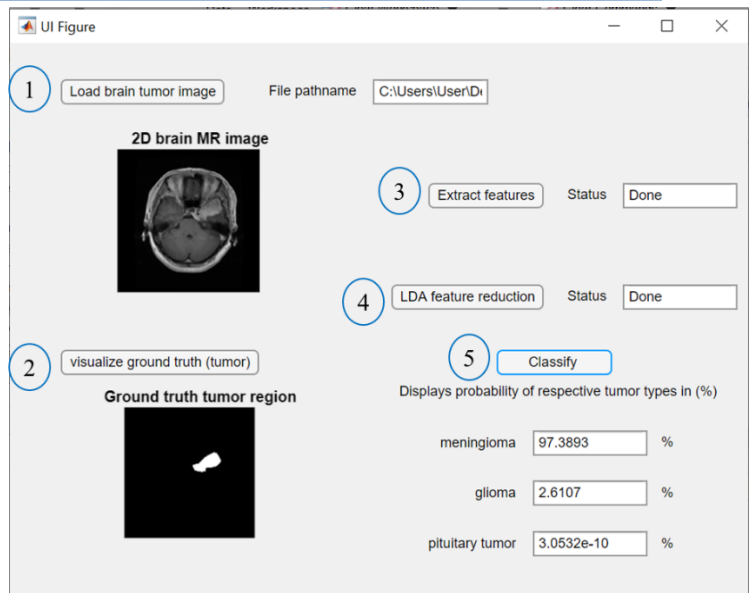

Figure 8: *Details of GUI developed.*

4   Discussion

Our experiment shows that sigmoid function achieves best performance compared to other non-linear activation functions, which agrees with the previous work of [44]. This probably can be attributed to meaning feature mappings that is helpful for brain tumor types discrimination.

It is also noted that sine and rbf activation function can become unstable if the number of features increases beyond a certain threshold. Below are likely justifications for the poor performances of these activation functions:

- According to [45], RBF/Gaussian activation function can only achieve good results if the data approximately follows Gaussian distribution. Additionally, the values of hidden nodes (a and b) can have significant impact of the elements of H (matrix of hidden nodes' outputs). If vector a is too far from x, or when b is too large, the zero percentage of H can be high.
- Activation functions typically used in neural networks is monotonic. As pointed out in [46], neurons learn to respond to particular regularity, in which

stronger positive correlation with the input corresponds to stronger (or equal) activation and vice versa. Sine function, being a sinusoidal (non-monotonic) function, activations of hidden nodes may oscillate between stronger and weaker activations. This condition may not be desirable if for example, two similar instances (close to each other in Euclidean space) that belong to same class may have very different projection in H. This helps to explain the large variation seen in performance measures when different set of data is used for training and test.

The GUI developed do not require any user input parameters, but the images loaded must be attached with segmented tumor region. The probability estimates of each class is computed based on Bayes' theorem stated in section 2.6, using the $p(t|C_i)$ calculated from fitted GMM during training phase. Table 8 shows comparison with other findings in literature *using same dataset* and *performing the same classification tasks.*

Table 8: *Previous works using same dataset and performing the same classification tasks.*

Authors	Classifiers	Feature input	Accuracy
[47]	Deep neural network	Downsampled to 64×64 window from images	95.6%
[48]	Convolutional neural networks	Original images	94.2%
[49]	Capsule networks	Downsampled 64×64 raw image patches.	86.56%
[50]	Capsule netwoks (DCNet ++)	Downsampled 64×64 raw image patches.	87.5%
[11]	ANN	Statistical features from DWT and Gabor filter	91.9%
[51]	Kernel ELM	Representation learning using CNN with 28×28 image patches as input.	93.68%
[20]	SVM	BoW model with region partitioning	91.28%
Our work	ELM	**5 LDA features** extracted augmented tumor region	**98.92%**

5   Conclusions

Multi-class brain tumor categorization was performed by LDA+ELM model, alongside multiple feature descriptors. Several activation functions were experimented and it was found that sigmoid activation function achieves the highest generalization performance with accuracy of 98.92% by using only 5 LDA attributes. The performance of our proposed model is comparable or better than previous works on the same data, performing the same classification task. By fitting GMM model on training data, posterior probability estimates for test data can be obtained. Eventually, a user-friendly GUI is developed based on the optimal pipeline.

Further study should be oriented towards development of statistical models (supervised learning models) that can perform online incremental learning, as medical image data normally come one by one or chunk by chunk.

6   Declarations

6.1   Study Limitations

This study just focuses on contrast enhanced T1-weighted MR images, thus the results may not be extended to other imaging modalities. Another limitation of this research is that no external validation dataset to evaluate performance of proposed model on unknown data from other source (e.g. different MRI machine and its parameters as well different patients).

6.2   Acknowledgements

6.3   Competing Interests

Not applicable.

How to Cite this Article:

References

[1]   E. I. Zacharaki *et al.*, "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme," (in eng), *Magnetic resonance in medicine,* vol. 62, no. 6, pp. 1609-1618, 2009.

[2]   C. Parmar, J. D. Barry, A. Hosny, J. Quackenbush, and H. Aerts, "Data Analysis Strategies in Medical Imaging," (in eng), *Clin Cancer Res,* vol. 24, no. 15, pp. 3492-3499, Aug 1 2018.

[3]   R. J. Gillies, A. R. Anderson, R. A. Gatenby, and D. L. Morse, "The biology underlying molecular imaging in oncology: from genome to anatome and back again," (in eng), *Clinical radiology,* vol. 65, no. 7, pp. 517-521, 2010.

[4]   D. Mortazavi, A. Z. Kouzani, and H. Soltanian-Zadeh, "Segmentation of multiple sclerosis lesions in MR images: a review," (in eng), *Neuroradiology,* vol. 54, no. 4, pp. 299-320, Apr 2012.

[5]   N. Nabizadeh and M. Kubat, "Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features," *Computers & Electrical Engineering,* vol. 45, pp. 286-301, 2015/07/01/ 2015.

[6]   E. I. Zacharaki, V. G. Kanas, and C. Davatzikos, "Investigating machine learning techniques for MRI-

based classification of brain neoplasms," (in eng), *International journal of computer assisted radiology and surgery,* vol. 6, no. 6, pp. 821-828, 2011.

[7] J. Sachdeva, V. Kumar, I. Gupta, N. Khandelwal, and C. K. Ahuja, "Segmentation, feature extraction, and multiclass brain tumor classification," (in eng), *Journal of digital imaging,* vol. 26, no. 6, pp. 1141-1150, 2013.

[8] P. Svolos *et al.*, "Investigating brain tumor differentiation with diffusion and perfusion metrics at 3T MRI using pattern recognition techniques," (in eng), *Magn Reson Imaging,* vol. 31, no. 9, pp. 1567-77, Nov 2013.

[9] K. M. Priya, S. Kavitha, and B. Bharathi, "Brain tumor types and grades classification based on statistical feature set using support vector machine," in *2016 10th International Conference on Intelligent Systems and Control (ISCO)*, 2016, pp. 1-8.

[10] A. Jayachandran and R. Dhanasekaran, "MULTI CLASS BRAIN TUMOR CLASSIFICATION OF MRI IMAGES USING HYBRID STRUCTURE DESCRIPTOR AND FUZZY LOGIC BASED RBF KERNEL SVM," *Iranian Journal of Fuzzy Systems,* vol. 14, no. 3, pp. 41-54, 2017.

[11] M. R. Ismael and I. Abdel-Qader, "Brain Tumor Classification via Statistical Features and Back-Propagation Neural Network," in *2018 IEEE International Conference on Electro/Information Technology (EIT)*, 2018, pp. 0252-0257.

[12] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Neural Information Processing Systems,* vol. 25, 01/01 2012.

[13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research,* vol. 15, pp. 1929-1958, 06/01 2014.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," (in eng), *Nature,* vol. 521, no. 7553, pp. 436-44, May 28 2015.

[15] Chuan Guo, Geoff Pleiss, Yu Sun, and K. Q. Weinberger, "On Calibration of Modern Neural Networks," in *Proceedings of the 34 th International Conference on Machine Learning, PMLR 70, 2017.*, Sydney, Australia, 2017.

[16] X. Jiang, M. Osl, J. Kim, and L. Ohno-Machado, "Calibrating predictive model estimates to support personalized medicine," (in eng), *Journal of the American Medical Informatics Association : JAMIA,* vol. 19, no. 2, pp. 263-274, Mar-Apr 2012.

[17] B. Zadrozny and C. Elkan, *Transforming classifier scores into accurate multiclass probability estimates*. 2002, p. 694.

[18] L. Cosmides and J. Tooby, "Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty," *Cognition,* vol. 58, no. 1, pp. 1-73, 1996/01/01/ 1996.

[19] E. Eirola *et al.*, "Extreme Learning Machines for Multiclass Classification: Refining Predictions with Gaussian Mixture Models," in *Advances in Computational Intelligence*, Cham, 2015, pp. 153-164: Springer International Publishing.

[20] J. Cheng *et al.*, "Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition," (in eng), *PLoS One,* vol. 10, no. 10, p. e0140381, 2015.

[21] J. Cheng *et al.*, "Retrieval of Brain Tumors by Adaptive Spatial Pooling and Fisher Vector Representation," (in eng), *PLoS One,* vol. 11, no. 6, p. e0157112, 2016.

[22] W. Yang *et al.*, "Content-based retrieval of brain tumor in contrast-enhanced MRI images using tumor margin information and learned distance metric," (in eng), *Med Phys,* vol. 39, no. 11, pp. 6929-42, Nov 2012.

[23] S. A. Medjahed, *A Comparative Study of Feature Extraction Methods in Images Classification*. 2015, pp. 16-23.

[24] I. Guyon, Andr, #233, and Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.,* vol. 3, pp. 1157-1182, 2003.

[25] L. Keyes and A. Winstanley, "Using moment invariants for classifying shapes on large-scale maps," *Computers, Environment and Urban Systems,* vol. 25, no. 1, pp. 119-130, 2001/01/01/ 2001.

[26] H. Zhihu and L. Jinsong, "Analysis of Hu's moment invariants on image scaling and rotation," in *2010 2nd International Conference on Computer Engineering and Technology*, 2010, vol. 7, pp. V7-476-V7-480.

[27] A. Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 12, no. 5, pp. 489-497, 1990.

[28] D. v. P. P. Bhaskara Rao, Ch.pavan Kumar, "Feature extraction using zernike moments," *International Journal of Latest Trends in Engineering and Technology* vol. 2, no. 2, March 2013 2013.

[29] C. Singh and R. Upneja, "Accurate calculation of high order pseudo-Zernike moments and their numerical stability," *Digital Signal Processing,* vol. 27, pp. 95-106, 2014/04/01/ 2014.

[30] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 1, pp. 886-893 vol. 1.

[31] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, no. 7, pp. 971-987, 2002.

[32] D. Cai, H. Bao, and X. He, "Sparse concept coding for visual analysis," in *CVPR 2011*, 2011, pp. 2905-2910.

[33] Q. Gu, Z. Li, and J. Han, "Linear Discriminant Dimensionality Reduction," in *Machine Learning and Knowledge Discovery in Databases*, Berlin, Heidelberg, 2011, pp. 549-564: Springer Berlin Heidelberg.

[34] H. Guang-Bin, "Learning capability and storage capacity of two-hidden-layer feedforward networks," *IEEE Transactions on Neural Networks,* vol. 14, no. 2, pp. 274-281, 2003.

[35] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing,* vol. 70, no. 1, pp. 489-501, 2006/12/01/ 2006.

[36] P. L. Bartlett, "The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network," *IEEE Transactions on Information Theory,* vol. 44, no. 2, pp. 525-536, 1998.

[37] G.-B. Huang, X. Ding, and H. Zhou, "Optimization method based extreme learning machine for classification," *Neurocomputing,* vol. 74, no. 1, pp. 155-163, 2010/12/01/ 2010.

[38] A. Hoerl and R. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics,* vol. 12, pp. 55-67, 04/09 2012.

[39] G. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme Learning Machine for Regression and Multiclass Classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics),* vol. 42, no. 2, pp. 513-529, 2012.

[40] G.-B. Huang, "Extreme Learning Machines (ELM); a survey. Int. J. Mach. Learn. & Cyber. 2, 107–122 (2011). https://doi.org/10.1007/s13042-011-0019-y

[41] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks,"

Information Processing & Management, vol. 45, no. 4, pp. 427-437, 2009/07/01/ 2009.

[42] F. J. Provost, T. Fawcett, and R. Kohavi, "The Case against Accuracy Estimation for Comparing Induction Algorithms," presented at the Proceedings of the Fifteenth International Conference on Machine Learning, 1998.

[43] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," (in eng), *Annals of translational medicine,* vol. 4, no. 11, pp. 218-218, 2016.

[44] W. Cao, J. Gao, Z. Ming, and S. Cai, "Some Tricks in Parameter Selection for Extreme Learning Machine," *IOP Conference Series: Materials Science and Engineering,* vol. 261, p. 012002, 10/01 2017.

[45] S. Liu, L. Feng, Y. Xiao, and H. Wang, "Robust activation function and its application: Semi-supervised kernel extreme learning method," *Neurocomputing,* vol. 144, pp. 318-328, 2014/11/20/ 2014.

[46] G. Parascandolo, H. Huttunen, and T. Virtanen, "Taming the waves: sine as activation function in deep neural networks," in *ICLR,* 2017.

[47] N. Ghassemi, A. Shoeibi, and M. Rouhani, "Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images," *Biomedical Signal Processing and Control,* vol. 57, p. 101678, 2020/03/01/ 2020.

[48] A. Kabir Anaraki, M. Ayati, and F. Kazemi, "Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms," *Biocybernetics and Biomedical Engineering,* vol. 39, no. 1, pp. 63-74, 2019/01/01/ 2019.

[49] P. Afshar, A. Mohammadi, and K. Plataniotis, *Brain Tumor Type Classification via Capsule Networks.* 2018.

[50] R. P. Sai Samarth, Apoorva Sikka, Abhinav Dhall and Deepti Bathula, "Dense and Diverse Capsule Networks:Making the Capsules Learn Better," Available: https://arxiv.org/pdf/1805.04001.pdf

[51] A. Pashaei, H. Sajedi, and N. Jazayeri, "Brain Tumor Classification via Convolutional Neural Network and Extreme Learning Machines," in *2018 8th International Conference on Computer and Knowledge Engineering (ICCKE)*, 2018, pp. 314-319.

Publish your books with AIJR publisher-
- ❖ Publish with ISBN and DOI.
- ❖ Publish Thesis/Dissertation as Monograph.
- ❖ Publish Book Monograph.
- ❖ Publish Edited Volume/ Book.
- ❖ Publish Conference Proceedings
- ❖ Retain full copyright of your books.

Submit your manuscript at books.aijr.org

Publish your research article in AIJR journals-
- ❖ Online Submission and Tracking
- ❖ Peer-Reviewed
- ❖ Rapid decision
- ❖ Immediate Publication after acceptance
- ❖ Articles freely available online
- ❖ Retain full copyright of your article.

Submit your article at journals.aijr.in