Real Time Detection and Grading of Brain Tumors from MRI using Improved HHO Optimized Deep CNN Frame Work.

[1]Deepak V.K,

[1]Research Scholar, Department of Electronics & Communication, Noorul Islam Centre for Higher Education, Thuckalay, Kumaracoil, Tamilnadu.

deepakvk2@gmail.com

[2]Dr.Sarath. R

[2]Associate Professor, Department of Electronics &Instrumentation, Noorul Islam Centre for Higher Education, Thuckalay, Kumaracoil, Tamilnadu.

sarathraveendran@gmail.com

Abstract:

Brain tumor detection and its grade identification is a crucial task to evaluate the tumors and make a treatment decision according to their classes. Among various imaging techniques MRI is commonly used due to its superior image quality due to the fact of relying on no ionizing radiation. Deep learning (DL) using convolutional neural network (CNN) is a subfield of machine learning and recently showed a remarkable performance, especially in classification and segmentation problems. In this paper, an optimized DL model based on a convolutional neural network is proposed to classify different brain tumor types using two publicly available clinical datasets. The former one classifies tumors into (meningioma, glioma, and pituitary tumor). The other one differentiates between the three glioma grades (Grade II, Grade III, and Grade IV). The datasets include 233 and 73 patients with a total of 3064 and 516 images on T1-weighted contrast-enhanced images for the first and second datasets, respectively. The proposed network structure achieves a significant performance with the best overall accuracy of 97.16% and 98.2%, respectively, for the two studies. The results indicate the ability of the model for brain tumor multi-classification purposes.

INDEX TERMS: *Brain tumor, convolutional neural network, deep learning, HHO, MRI.*

1.Introduction

Brain tumor can be defined as unnatural and uncontrolled growth in brain cells. Since the human skull is a rigid, any unexpected growth may affect a human function especially CNS [1]. Brain tumors can be classified mainly as primary and secondary tumors. The former represents about 60% of all brain tumors, while secondary tumors are the residuals 40%. This classification is determined according to tumors origin just as tumors first originate in the brain are called primary tumors. On the other side, tumors first arise in any other part of the body and then transferred to the brain are called secondary tumors, and most of them are malignant [4]. Among numerous imaging techniques used to detect and classify brain tumors, MRI is one of the most common non-invasive techniques. MRI popularity comes from the fact of using no ionizing radiation during the scan as well as its superior soft-tissue resolution plus the ability to acquire different images using various imaging modalities or by employing contrast-enhanced agents [5], [6]. The most common type of brain tumors includes Gliomas. Meningiomas & Pituitary. Gliomas are the most prevalent type of brain tumors that originate in the glial cells of the brain [7]. Gliomas classified into four grades according to the WHO starting from type I to IV [8]. Grade I tumors are benign and have a much similar texture of the normal glial cells, Grade II is a slightly different in texture, Grade III tumors are malignant with abnormal tissue appearance while grade IV is the most severe stage of gliomas and tissue abnormalities that can be visualized by naked eye [1]. Meningioma is a tumor that forms on the membrane that covers the brain and spinal cord inside the human skull and grows placidly. Most of meningioma tumors are benign [8]. However, pituitary tumor starts from the pituitary glands that control hormones and regulate functions in the body. It can be benign, benign that expands to bones, and malignant. Complications of pituitary tumors may cause permanent hormone deficiency and vision loss [1]. Machine Learning (ML) is the study of algorithms and statistical models that can be used to perform a specific task without using outright instructions, relying on patterns. Deep Learning (DL) is a subdivision of ML that is based on learning data representations and hierarchical feature learning. DL algorithms utilize arrange of numerous layers of nonlinear processing identities for feature extraction. The output of each sequential layer is the input of the next one, and that helps in data abstraction as we go deep within the network [15]. Convolutional Neural Network (CNN) is a class of DL and commonly used in analyzing visual imagery and designed to require minimal preprocessing [16]. It is inspired by biological processes in human brain [17] and utilized to handle data that come in multiple arrays [18]. The main advantages of CNNs are feature learning and providing unlimited accuracy rather than traditional machine learning and vanilla neural networks which may be achieved by increasing training samples and therefore leads to a more robust and accurate model [6]. In CNN architecture, the convolutional filters are acting as features extractors, and as we go deep, we extract more and more complex features (spatial and structural information). Feature extraction hap- pens through convolving small filters with the input patterns followed by selection of the most distinguishing features and then start to train the classification network [18]. To prevent over fitting problems and for easy convergence various optimization schemes are used to adjust weighting parameters of the network [9],[10].

2.Literature review.

Brain tumors classification has been performed using many machine learning techniques and imaging modalities over the years. In 2009, Zacharaki et al. [21] proposed a system to classify different grades of glioma besides a binary clas- sification for high and low grade using SVMs and KNN. Accuracy of 85% is obtained for multi-classification and 88% for binary classification. El-Dahshan et al. [22] introduced a method to classify 80 brain tumor normal and abnormal images using Discrete Wavelet Transform (DWT) to extract features, Principal Component Analysis (PCA) to reduce features, and then ANN and KNN to classify images with overall accuracy of 97% and 98% respectively. In 2015, Cheng et al. [23] proposed a method to enhance the brain tumor classification performance by augmenting the tumor region via image dilation and then by splitting into sub- regions. They used three approaches to extract features; inten- sity histogram, Gray Level Co-occurrence Matrix (GLCM) and Bag of Words (BOW) and finally achieved best accuracy of 91.28% by using ring form partition in addition to tumor region augmentation.

In the work proposed by Ertosun and Rubin [24], the authors used CNN to classify different grades of gliomas pathological images (Grade II, Grade III and Grade IV) and another task to classify Low-Grade Glioma (LGG) vs. High- Grade Glioma (HGG). They obtained accuracies of 71% and 96% respectively. Paul et al. [25] used axial brain tumor images to train and develop two main approaches for clas- sification (fully connected neural network and a convolu- tional neural network), CNN architecture was formed of two convolutional layers with two corresponding max-pooling layers followed by two fully connected layers and achieved maximum accuracy of 91.43%. Posteriorly, Afshar et al. [26] presented a capsule network (CapsNet) that integrates both the MRI brain image and the coarse tumor boundaries to clas- sify the brain tumor. Accuracy of 90.89% has been obtained in this study. In another study, Anaraki et al. [27] proposed a model of two combined regulations to classify brain tumor images based on CNN and Genetic Algorithms (GA-CNN), in the first case study, accuracy of 90.9% has been attained. in classifying three grades of glioma, while 94.2% accuracy for classification of glioma, meningioma and pituitary tumor have been fulfilled in the second case study. Using a K-mean clustering technique combined with the Fuzzy C-mean algorithm, Eman et al (2) presented a successful image segmentation method. To provide an effective diagnosis of brain tumor, threshold and plane segmentation levels follow. With regard to the minimum computation time, the proposed approach will take advantage of a K-mean set to segment the image.

Clara et al (3) have proposed a survey of deep learning in medical image processing using deep learning algorithms. Deep learning algorithms, particularly convolutionary networks, have quickly become the tool of choice in order to analyze medical images. The core principles of deep learning applicable to medical image analysis summaries more than 300 field contributions, most of which occurred in the past year. Here we will explore the use of deep learning for image recognition, object detection, segmentation, registration, and other activities. Many brain imaging problems have been completely solved by the DNNs.

To diagnose a brain tumor at an early stage, Prasad et al. (4) developed a computational detection method using mathematical morphological reconstruction (MMR). The segmentation of the pre-processed image is performed using mathematical morphological

operations in the pre-processed image using a median filter. Using statistical features and texture features, extraction of features first, then reduction of features using main component analysis. With GRB kernel support vector machines, the classification is completed. Experimental findings show that the segmented images are very accurate and substantially decrease the computing time at the same time.

Heba et al (5) established a classification using deep learning neural networks in order to classify brain tumors. In this article, one of the DL architectures, the Deep Neural Network classifier, was used to classify a data set of 66 brain MRIs into 4 classes. The classifier was combined with the discrete wavelet transform (DWT) with the effective feature extraction method and main component analysis (PCA) and the performance assessment was very robust for all performance measurements. In this model, Fuzzy C stands for image segmentation, feature extraction of discrete wavelet transforms (DWT) and reduction using the technique of design component analysis (PCA) and classification of DNN.

An effective classifier for the classification of brain tumors was developed by Anjali and Priya (6). Proposed scheme; The median filter is used to minimize noise. Classification uses CART and SVM classifiers, which are hybrid CART process and SVM combined, 92.31 percent of accuracy was obtained by the proposed system.

Lakshmi and Angulakshmi (7) exhibit MRI segmentation of the brain tumor using spectral clustering based on super pixels. ROI identification alleviates the spectral clustering statistical pressure. High quality clustering results are provided for brain tumor segmentation by the segmentation of ROI using spectral clustering. For oedema segmentation and Tumor Core (TC) areas, the proposed approach outperforms the other clustering approaches with competitive dice score values from MRI images.

Sanjeev et al (9) designed a classification of brain tumor MRI images: in this method, a brain tumor was performed from MR images using a hybrid approach. A discrete wavelike transformation (DWT) for use in feature extraction, a genetic algorithm to minimize the number of features, and a support vector machine (SVM) to identify brain tumors are included in this hybrid method. The results of the simulation analysis approach show that the hybrid approach provides better performance by improving accuracy and reducing RMS error compared to advanced technologies in the same context.

In this paper, an optimized CNN architecture is proposed to classify different types and grades of brain tumors. The architecture of the network is evolved using different configuration to acquire the most appropriate structure. The paper is organized as follows; in section 3, the proposed methodology is discussed in details starting from the original dataset and how manipulation occurs to adapt the CNN model to the tools and hardware resources used in this research. Section 4 and, 5 are dedicated to results and discussion respectively followed by a conclusion in Section 6.

2.Proposed Methodology

Figure. 1 shows the block diagram of the proposed method, in which the system starts to load and extract images and labels from datasets raw files and then make a preprocessing and optimization techniques just after splitting the dataset into training, validation and test sets. Then, the structure of

the proposed method is introduced, followed by setting the hyper-parameters, regularization techniques, and optimiza- tion algorithm. Finally, network training and performance computations are presented.

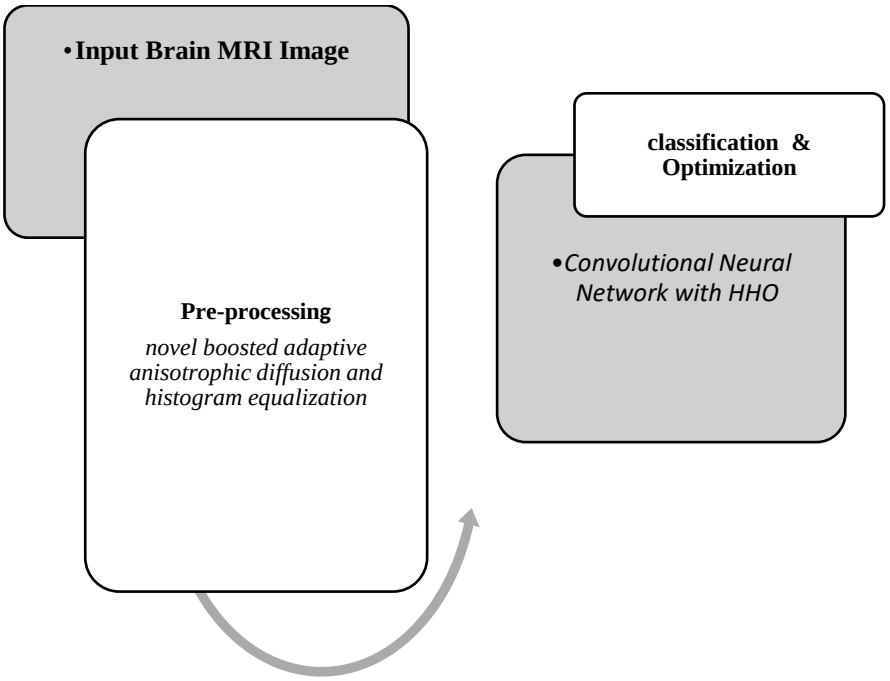
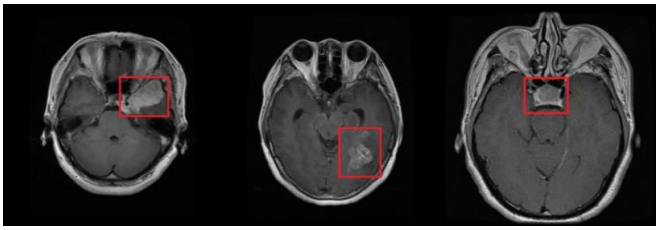


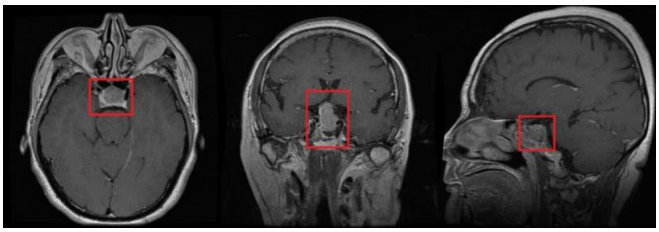
Figure1: Overall flow of proposed work

2.1. Clinical Dataset Used.

Two different clinical data set are used I this work. The first one is acquired from Pushpagiri Medical Hospital, Kerala, India and Gokulam Medical Hospital, Kerala,India from 2012 to 2020 .The database includes T1-weighted contrast-enhanced images acquired from 233 patients with three types of brain tumors that are meningioma, glioma, and pituitary tumor [28]. Brain tumors can be different in shape, location, and size according to their types and grades as figured in Figure 2(a). The dataset includes three different views: axial, coronal and sagittal views as shown in Figure 2(b). The second dataset is obtained from The Cancer Imaging Archive (TCIA) public access repository [29]. The Repository of Molecular Brain Neoplasia Data (REMBRANDT) contains MRI multi-sequence images from 130 patients with different diseases, grades, races, and ages [30]. We selected images on T1-weighted contrast-enhanced that include different grades of glioma (Grade II, Grade III, and Grade IV) as shown in Figure 3. Table 1 and Table 2 show supplementary details about the description of the two datasets respectively.



(a)



(b)

Figure 2. (a) Different three axial brain tumour types as follows; Meningioma, Glioma and Pituitary tumor from left to right respectively,

(b) Pituitary tumor is demonstrated in three different acquisition views (Axial, Coronal, and Sagittal) from left to right respectively. Tumors are localized inside a red rectangle.

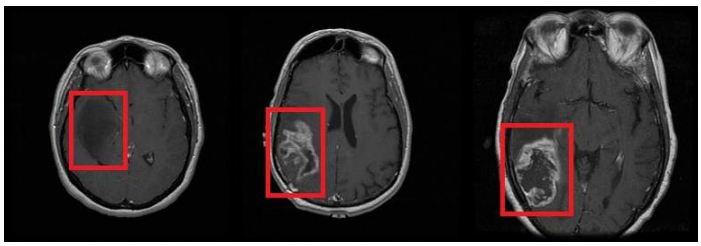


Figure 3. Different glioma grades included in REMBRANDT dataset (Grade II, Grade III and Grade IV from left to right respectively). Tumours are localized inside ared rectangle.

Table 1. Number of slices for each brain tumor type (meningioma, glioma and pituitary) in dataset I and their corresponding number of patients.

Tumor Category	Number of Patients	Number of Slices
Meningioma	82	708
Glioma	91	1426
Pituitary	60	930
	233	**3064**

Table 2. Number of slices for glioma grades in dataset II and their corresponding number of patients.

Tumor Category	Number of Patients	Number of Slices
Grade II	33	205
Grade III	19	129
Grade IV	21	182
	73	**516**

2.2. Pre-processing stage

Before feeding the images into the proposed structure, a pre- processing step is performed. The first process is to down- size the original image from 512 * 512*1 pixels into 128 *128* 1 pixels in order to decrease dimensionality, computations and help the network to show a better performance in lower time and more straightforward calculations. These data is then preprocessed using a combined approach in which boosted anisotropic diffusion filtering (BADF) is combined with contrast limited histogram equalization, (CLAHE) Then, data is shuffled before splitting them to maintain the system to train on unsorted data and prevent focusing on a narrow band of the entire dataset. Data is divided into three sections; training, validation, and test sets all with their individual target labels (68% for training and 32% for system test and validation).

2.2.1 Adaptive Histogram Equalization

The main purpose of image enhancement is to improve certain characteristics of an image to improve its visual quality.

This is an extension of the conventional technique of Histogram Equalization. By transforming the values in the intensity image, it improves the image contrast. It is obvious that each bar on the equalized histogram is of the same height, according to the physical sense of the histogram. That is, that is,

$$p_s(s)ds = p_r(r)\mathrm{dr} \qquad (1)$$

Suppose $s = T(r)$A monotonically increasing function and its inverse function, in the interval,$r = T^{-1}(s)$ is a monotonic function also. According to (1), we can deduce

$$p_s(s) = \left[p_r(r)\frac{1}{ds/dr}\right]_{r=T^{-1}(s)} = p_r(r)\frac{1}{p_r(r)} = 1 \qquad (2)$$

Mapping relationship for traditional graph equation algorithm: The relationship between i (the gray value of the pixels in the original image) and f_i(the gray value of the pixels in the optimized image) in separate conditions.

$$f_i = (m-1)T(r) = (m-1)\sum_{k=0}^{i}\frac{q_k}{Q} \qquad (3)$$

Where m is the number of gray levels displayed in the original image, q_k is the number of pixels in the image with the gray level of k^{th}; Q is the total number of pixels in the image.

Suppose that an image has n different levels of gray and the probability of occurrence of the i^{th} gray level is p_i, hence the entropy of the gray level

$$e(i) = -p_i log p_i \quad (4)$$

The entropy of the whole image is

$$E = \sum_{i=0}^{n-1} e(i) = -\sum_{i=0}^{n-1} p_i log p_i \quad (5)$$

2.2.2 Boosted Anisotropic diffusion filter

By maintaining fine details in the picture, it primarily focuses on removing noise. After generating the scattered image, this BADF adds the Partial Differential Equation (PDE), which adds an additional benefit to the existing anisotropic scattering filter. Smoothing can also be accomplished by a diffusion process that is absent at the edges and boundaries.
Anisotropic filtering is a conventional smoothing technique in which the smoothing process is based on the following PDE governing process.

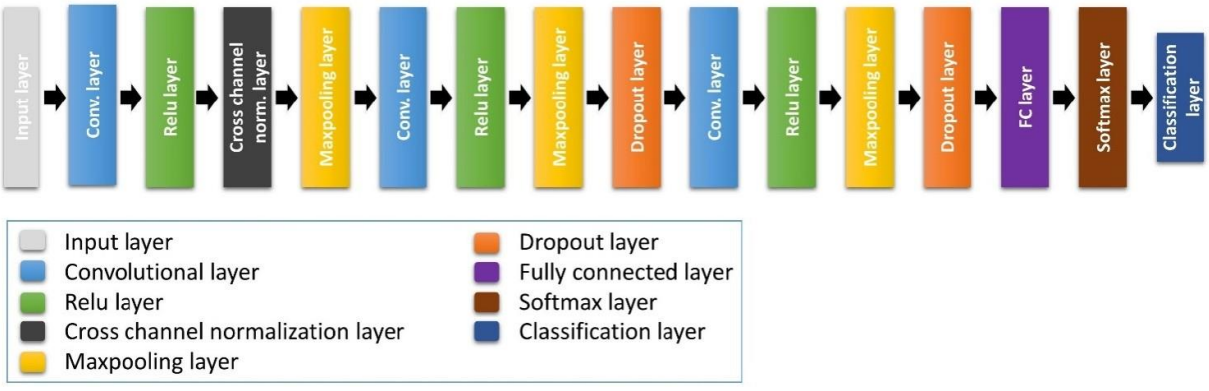
$$\frac{\partial I_m}{\partial t} = div(T \nabla I_m) \quad (6)$$

where I_m is the image intensity in weighting direction m, ∇ is a gradient operator, *div* is a divergence operator, t is the time. $\boldsymbol{T}$ is a *structure tensor* that provides the directionality of smoothing. It is constructed from a common *gradient tensor* $\boldsymbol{G}$, which is obtained by convolving the sum of outer products of ∇I_m over all weighting directions with a Gaussian kernel K_ρ:

$$G = k_p * \sum_m (\nabla I_m \oplus \nabla I_m) \quad (7)$$

where $\bigotimes$ represents the outer product operator. The parameter ρ is the standard deviation of the Gaussian kernel, which determines the spatial scale of the gradient tensor.

3.Proposed CNN Architecture.



The convolutional neural Network (CNN) is an artificial neural network that Kunihiko Fukushima [30] and Yann LeCan et al . introduced in 1980[35][36]. CNN consists of 6-layer types: input layer, wrapping layer, nonlinear layer, pool layer, output layer and completely connected layer. ResNet50 is a 50-layer residual network with 26-million parameters. A deep convolution model for neural networks is the residual network. We are not learning from

features in the residual network, but from residuals, which are subtractions of features learned from the inputs of the plane. The skip link was used by ResNet to relay information through layers. ResNet directly links the nth layer input to some (n+x) th layer, which allows stacking of additional layers and forming a deep network. Fig. 4 represent the proposed architecture for CNN.

A collection of convolutional nuclei in which each neuron behaves as a nucleus is made up of the convolutional layer. The convolution operation becomes a correlation operation, however, if the kernel is symmetrical (Ian Goodfellow et al. 2017). By slicing the image into small slices, usually known as receptive fields, the convolutional kernel works. You may express the convolution operation as follows:

$$f_1^k(p,q) = \sum_c \sum_{x,y} i_c(x,y) \,.\, e_1^k(u,v) \quad (8)$$

Where, $i_c(x,y)$is an element of the input image tensor I_c , which is element wise multiplied by $e_1^k(u,v)$ index of the k[th] convolutional kernel k_1 of the l[th] layer. Whereas output feature-map of the k[th] convolutional operation can be expressed as

$$F_1^k = [f_1^k(1,1), \dots (f_1^k(p,q), .. f_1^k(P,Q))]. \quad (9)$$

When properties are retained in the pool layer, as long as their approximate position relative to others is maintained, their exact location becomes less significant. An interesting local activity is pooling, or down sampling. It aggregates similar data in the receiving field neighbourhood and generates the dominant response within that local area.

$$Z_l^k = g_p(F_l^k) \quad (10)$$

Equation (2) shows the pooling operation in which Z_l^k represents the pooled feature-map of l[th] layer for k[th] input feature-map F_l^k , whereas $g_p(.)$defines the type of pooling operation. The use of pooling operation helps to extract a combination of features, which are invariant to translational shifts and small distortions. The activation function for a convolved feature-map is defined in equation (32).

$$T_l^k = g_a(F_l^k) \quad (11)$$

the above equation, $F_l^k$is an output of a convolution, which is assigned to activation function $g_a(.)$that adds non-linearity and returns a transformed output $T_l^k$for lth layer.Batch normalisation is used within feature-maps to resolve the problems associated with the internal covariance change. A change in the distribution of hidden unit values is the internal covariance shift, which slows down convergence (by shifting the learning rate to a small value) and requires careful parameter initialization. Batch normalization for a transformed feature-map $F_l^k$is shown in equation (34).

$$N_l^k = \frac{F_l^k - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (12)$$

In equation (4), N_l^k represents normalized depict mean andμ_B and σ_B^2feature-map, F_l^k is theinput feature-map, B variance of a feature-map for a mini batch respectively. Inside the network, Dropout introduces regularisation, which gradually enhances generalisation by randomly omitting certain units or links with a certain probability. At the end of the network, the completely connected layer is mainly used for classification. It is a global process, in contrast to grouping and convolution. Accepts input from extraction phases of the function and measures data from all previous stages globally. Using different optimization approaches in deep neural networks by changing parameters such as weights and learning speeds to minimise losses. Here, the Harish Hawks Optimizer (HHO), proposed by Heidariet al.[20] in 2019, is used. Harris Hawks Optimization (HHO) is a modern, gradient-free, population-based optimization algorithm inspired by nature that imitates the bird-chasing style of Harris Hawks.

4.Experiments and Results

For study I, Figure 5(a,b) shows both the accuracy progress and loss during the validation phase for our proposed net- work. Figure 5(a) shows that almost 100% accuracy has been achieved right after 5000 iterations. After the 8550th iteration, the accuracy shows a plateau of nearly 100%, and finally, the best overall accuracy obtained during the test phase is 96.13%. While in the mini-batches loss graph Figure 5(b), it is clear that the curve first starts to drop sharply, but some fluctuations appear due to using a small mini-batch size of 32 images. These fluctuations tend to disappear after 6400 iterations and the loss curve almost hits zero.

For study II, Figure 6(a,b) shows both the accuracy progress and loss during the validation phase. We can see from Figure 6(a) that accuracy of 100% has been achieved right after 100 iterations. Hence, the best overall accuracy obtained during the test phase is 98.7%. From the mini-batches loss graph in Figure 6(b), the curve first starts to drop sharply. This slope tends to disappear after 100 iterations and the loss curve almost hits zero.

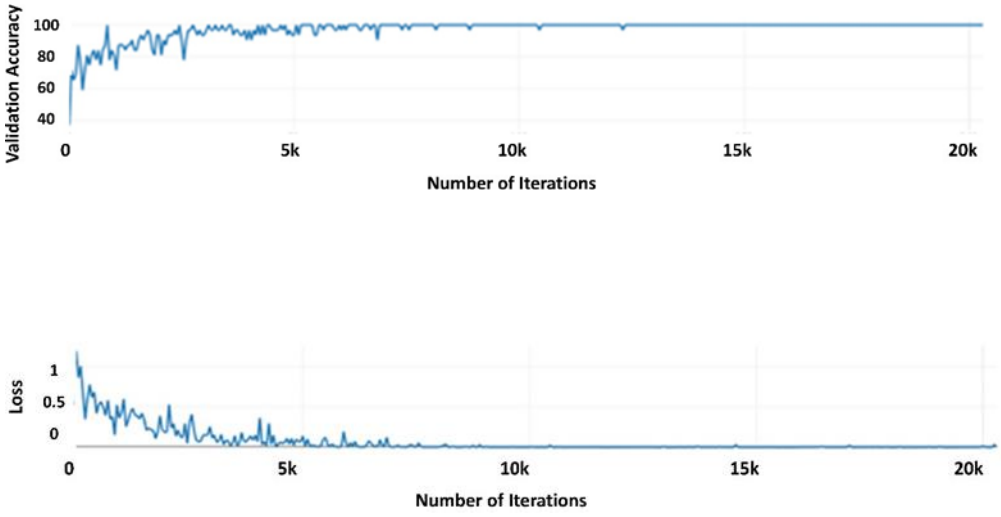


Figure 5 (a,b). Validation accuracy and loss over the whole training iterations of study I: (a) Validation accuracy(higher is better), and (b) Loss (lower is better).

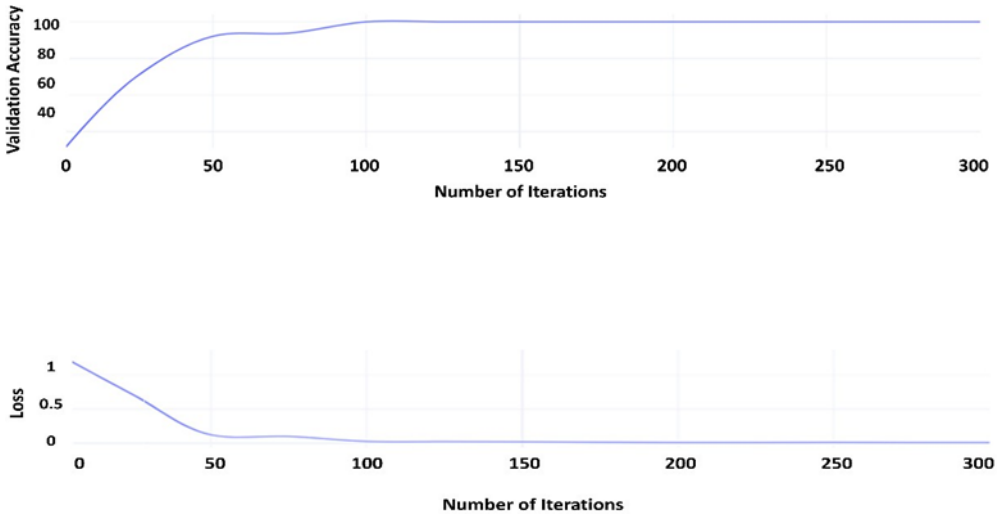


Figure 6(a,b). Validation accuracy and loss over the whole training iterations of study II: (a) Validation accuracy(higher is better), and (b) Loss (lower is better).

For preprocessing three parameters are considered, peak signal to noise ratio (PSNR), mean square error (MSE) & structural similarity index (SSIM). Also, for classification four measures are considered. The formulas are shown in Table 3 and the obtained values are shown in Table 4 for various grade images

Table 3: performance metrics

PREPROCESSING	$\text{PSNR} = 10\log10\left(\frac{\text{MAX}^2}{\text{MSE}}\right)$ (13)
	$\text{MSE} = \sum_{i=1}^{K} p_i\,(x_i - t)^2$ (14)
	$\text{SSIM} = \frac{(2\mu_x\mu_y+l_1)\,(2\sigma_{xy})l_2)}{(\mu_x^2+\mu_y^2+l_1)\,(\sigma_x^2+\sigma_y^2+l_2)}$ (15)
CLASSIFICATION	$\text{Sensitivity} = \frac{TP}{TP+FN}\times 100\%$ (16)
	$\text{Specificity} = \frac{TN}{TN+FP}\times 100\%$ (17)
	$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}\times 100\%$ (18)
	$\text{Precision} = \frac{TP}{TP+FP}\times 100\%$ (19)
	$F-\text{measure} = \frac{2*\text{Precision}*\text{Recall}}{\text{Pr ecision}+\text{Recall}}$ (20)

Table 4: Performance evaluation for various filters in different grades of images

Metrics	PSNR		MSE		SSIM	
methods images	***AHE***	***BADF***	***AHE***	***BADF***	***AHE***	***BADF***
Grade 1	31.5247	39.4470	3.369335	0.022112	0.7435	0.8293
Grade 2	31.7810	40.6386	3.176197	0.016806	0.6864	0.7163
Grade 3	34.3646	39.3762	1.752057	0.022475	0.7157	0.9921
Grade 4	30.0447	39.1034	4.737437	0.023933	0.6403	0.8206

For pre-processed brain cancer MRI images, performance measurements such as PSNR, MSE, and SSIM are analysed and values are obtained. The PSNR is significantly high in all image samples when examining this table 4, relative to MSE values for sample images. Normally the SSIM values lies between -1 to 1, here the values of SSIM approximately equal to 1 this denotes our proposed pre-processing outperforms well compared to other technique. Comparing to this 2 technique BADF gives best performance.

Table 5: performance value of classifiers

Metrics	*SVM*	*HHO-SVM*	*BOVW*	*HHO-BOVW*	*CNN*	*HHO-CNN*
Accuracy	0.8500	0.9150	0.93000	0.9400	0.95	0.97
Error	0.1500	0.0850	0.0700	0.0600	0.05	0.03
Sensitivity	0.8500	0.9150	0.9300	0.9400	0.95	0.97
Specificity	0.9500	0.9717	0.9767	0.9800	0.9833	0.99
Precision	0.8542	0.9158	0.9318	0.9416	0.95027	0.9699
F1-Score	0.8495	0.9150	0.9300	0.9400	0.95005	0.96997

Table 5 shows the scheme of accuracy, error, sensitivity, specificity, and accuracy of the proposed classifiers. From the bar chart, HHO based CNN classifiers get higher resolution, sensitivity, accuracy and specification; The error value is less than the proposed classifiers, the results prove that the performance of the proposed algorithm is excellent. It has been shown that the implementation of the proposed approach for early tumour detection increases the quality and accuracy of clinical practise. In order to help pathologists discern the exact tumour region and type of tumour, the implementation of the proposed method for tumour monitoring

has been demonstrated.The cofusion matrix indicating true and predicted labels for differet types and grades for brain cancer is shown in Figure 7.

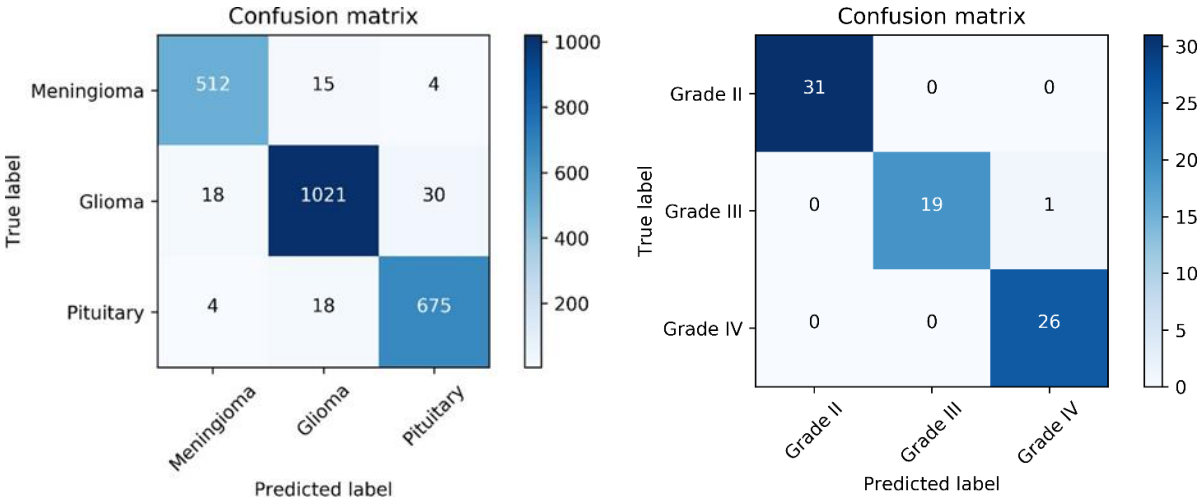
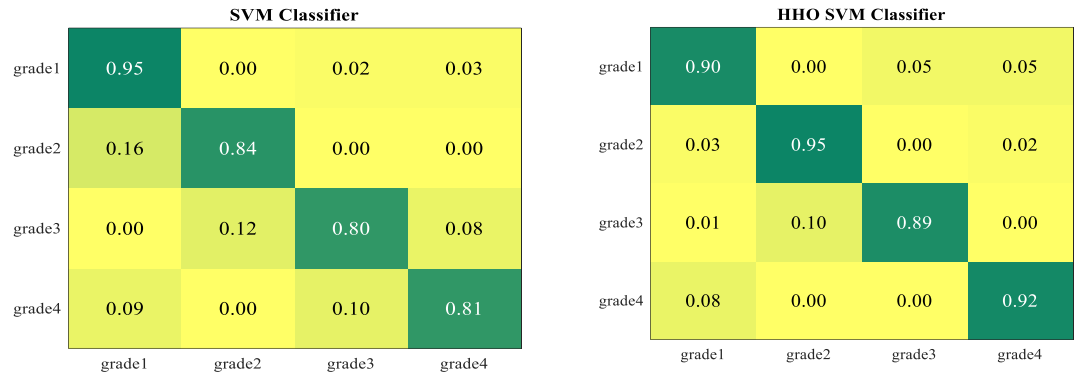


Figure 7. The confusion matrix of the proposed model: (a) for study I, and (b) for study II

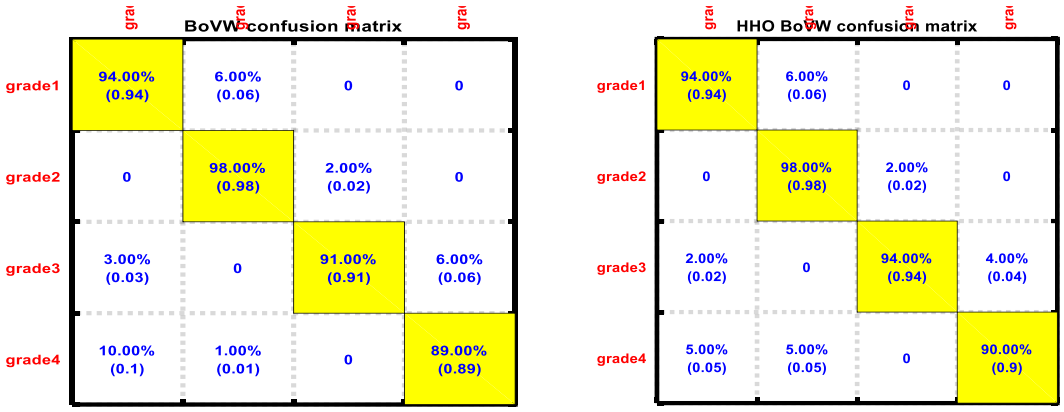
The classified parametric values for the two cases of study are show below in Table. 6 ad a comparative performance evaluation based on confusion matrix is shown in figures 8 (a,b,c)

Table 6. Accuracy metrics in terms of TP, TN, FP, FN, precision, sensitivity, specificity, and accuracy.

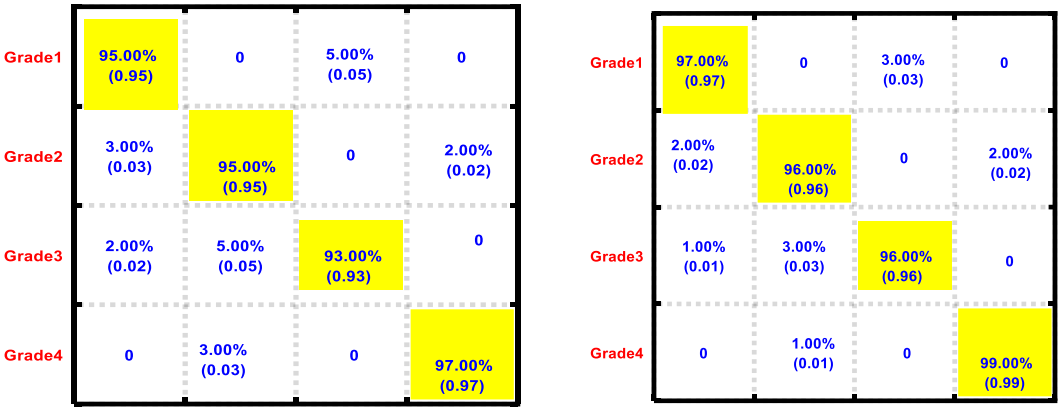
Metrics / Method	Tumor Type	TP	TN	FP	FN	Precision	Sensitivity	Specificity	Accuracy	Total no.
Proposed Model for Study I	Meningioma	512	1744	22	24	0.958	**0.955**	**0.987**	**97.54%**	547
	Glioma	1021	1195	33	61	**0.972**	0.944	0.951	95.81%	1069
	Pituitary	675	1566	34	48	0.952	0.934	0.97	96.89%	697
Proposed Model for Study II	Grade II	31	46	0	0	**1**	**1**	**1**	**100%**	31
	Grade III	19	57	0	1	**1**	0.95	**1**	95%	20
	Grade IV	26	50	1	0	0.963	**1**	0.98	**100%**	26



(a)



(b)



(c)

Figure 8: (a-c) confusion matrix of proposed Classifiers

The Confusion Graph generates a Graph of the Confusion Matrix from True Labels True Labels and Expected Labels Expected Labels and returns a Graph of the Confusion Matrix. The rows of the confusion matrix match the actual class and the columns match the expected class. Diagonal and off-diagonal cells lead, respectively, to cases classified correctly and incorrectly. The uncertainty matrix of SVM, BOVW and CNN classifiers with and without optimization is shown in Figure 5, in which classifiers classify brain images according to tetrolate and GLCM properties. In each cell, the uncertainty matrix displays the total number of observations. The rows of the confusion matrix match the actual class and the columns match the expected class.

Diagonal and off-diagonal cells lead, respectively, to cases classified correctly and incorrectly. The diagonal area of the confusion matrix shows the perfect identification.

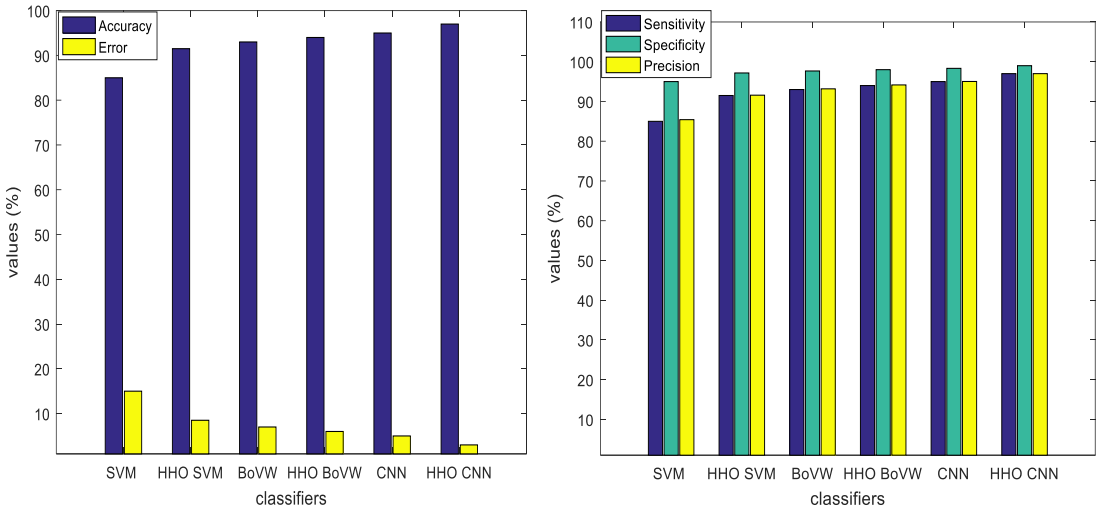
Table 4. Different architectures and hyper-parameters tested and tried before reaching the final model.

Factor(s)	Values
Number of convolutional + ReLU layers	1, 2, 3, 4
Number of cross-channel normalization layers	1, 2, 3
Maximum epochs	20,40,50,60,80,100
Number of drop out layers	1, 2, 3
Number of fully connected layers	1, 2, 3
Number of convolutional kernels	8, 16, 32, 64, 128, 256
Kernel size	2, 3, 5, 7, 9, 10, 11
pooling layer	Max pooling, average pooling
pooling layer window size	2, 3, 4, 5
Optimizers	SGD, Adam, RMSprop, HHO
Mini-batch size	1, 4, 8, 16, 32, 64, 128
Dropout rate	0.1, 0.15, 0.2, 0.25, 0.5
Initial learning rate	0.01, 0.001, 0.0001
Learning rate drop factor	0.1, 0.2, 0.3

5.Tools and Time Consumption.

The proposed deep neural network structure is trained on Intel i7-7700HQ CPU (2.8 GHz), NVIDIA GTX 1060 (6 GB) GPU, 16GB RAM, Matlab 2018b and Python. The training time was 289 minutes for (10,417 images) in study I and 2.5 minutes for (350 images) in study II. The average test execution time was 8.5 and 9.6 milliseconds per image for study I and II respectively.

It has been shown that the implementation of the proposed method for early detection of a Brain tumour increases clinical practice quality and precision, thus reducing the risk of misdiagnosis and mismanagement. Helping doctors to diagnose brain tumours more easily for further treatment. The proposed system can help doctors know the type of brain tumour, and the present stage for further treatment. Figure 9(a,b) indicate the comparative accuracy and precision of our method.



(a) (b)

Figure 9: (a & b) comparison plot of various classifiers

6.Conclusion

In this work, presented a CAD system for the brain tumor detection from MR images into three types (meningioma, glioma, and pituitary) in one study, and further classifying of gliomas into different grades (Grade II, Grade III and Grade IV) using a custom deep neural network. The proposed network is constructed from 16 layers starting from the input layer which holds the pre-processed images passing through the convolution layers and their activation functions (3 convolution, 3 ReLU, normalization and3 Maxpooling layers). Additionally, two dropout layers are used to prevent overfitting followed by a fully connected layer anda SoftMax layer to predict the output and finally a optimized classification layer that produces the predicted class. Although the dataset is relatively not big (due to the variety of imagingviews), data optimization helped well to show better resultsand hence overcome this problem. Our proposed architecturehas achieved the highest accuracy of 96.13% and 98.7% concerning the two datasets used in this paper.

7.Refarance

[1] L. M. De Angelis, ‘‘Brain tumors,’’ New England J. Med., vol. 344, no. 2, pp. 114–123, Jan. 2001.

[2] B. W. Stewart and C. P. Wild, World Cancer Report 2014. Lyon, France: IARC, 2014.

[3] Brain, Other CNS and Intracranial Tumours Statistics. Accessed:

May 2019. [Online]. Available: https://www.cancerresearchuk.org/

[4] A. Behin, K. Hoang-Xuan, A. F. Carpentier, and J.-Y. Delattre, ‘‘Pri- mary brain tumours in adults,’’ Lancet, vol. 361, no. 9354, pp. 323–331, 2003.

[5] A. Drevelegas, Imaging of Brain Tumors With Histological Correlations. Berlin, Germany: Springer, 2002.

[6] VK, D. (2022). Classification of brain tumours in MRI images using convolutional neural network through Cat Swarm Optimization. *Expert Systems*, *39*(9), e13021.

[7] M. L. Goodenberger and R. B. Jenkins, ‘‘Genetics of adult glioma,’’ Cancer Genet., vol. 205, no. 12, pp. 613–621, Dec. 2012.

[8] VK, D. (2022). An intelligent brain tumor segmentation using improved Deep Learning Model Based on Cascade Regression method. *Multimedia Tools and Applications*, 1-20.

[9] C. Bishop, Pattern Recognition and Machine Learning. Berlin, Germany: Springer-Verlag, 2006.

[10] T. Rajesh and R. S. M. Malar, ‘‘Rough set theory and feed forward neural network-based brain tumor detection in magnetic resonance images,’’ in Proc. Int. Conf. Adv. Nanomaterials Emerg. Eng. Technol. (ICANMEET), Jul. 2013, pp. 240–244.

[11] K. Machhale, H. B. Nandpuru, V. Kapur, and L. Kosta, ‘‘MRI brain cancer classification using hybrid classifier (SVM-KNN),’’ in Proc. Int. Conf. Ind. Instrum. Control (ICIC), May 2015, pp. 60–65.

[12] M. Shasidhar, V. S. Raja, and B. V. Kumar, ‘‘MRI brain image segmen- tation using modified fuzzy C-means clustering algorithm,’’ in Proc. Int. Conf. Commun. Syst. Netw. Technol. (CSNT), Jun. 2011, pp. 473–478.

[13] S. Goswami and L. K. P. Bhaiya, ‘‘Brain tumour detection using unsuper- vised learning based neural network,’’ in Proc. Int. Conf. Commun. Syst. Netw. Technol. (CSNT), Apr. 2013, pp. 573–577.

[14] S. Khalid, T. Khalil, and S. Nasreen, ‘‘A survey of feature selection and feature extraction techniques in machine learning,’’ in Proc. Sci. Inf. Conf., Aug. 2014, pp. 372–378.

[15] L. Deng and D. Yu, ‘‘Deep learning: Methods and applications,’’ Found. Trends Signal Process., vol. 7, nos. 3–4, pp. 197–387, Jun. 2014.

[16] Y. LeCun. (2015). Lenet-5, Convolutional Neural Net- works. Accessed: May 2019. [Online]. Available: http://yann.lecun.com/exdb/lenet

[17] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, ‘‘Subject independent facial expression recognition with robust face detection using a convo- lutional neural network,’’ Neural Netw., vol. 16, nos. 5–6, pp. 555–559, Jul. 2003.

[18] Y. LeCun, Y. Bengio, and G. Hinton, ‘‘Deep learning,’’ Nature, vol. 521, no. 7553, p. 436, 2015.

[19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, ‘‘Gradient-based learn- ing applied to document recognition,’’ Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ‘‘ImageNet classification with deep convolutional neural networks,’’ in Proc. Adv. Neural Inf. Pro- cess. Syst. (NIPS), Jan. 2012, pp. 1097–1105.

[21] E. I. Zacharaki, S. Wang, S. Chawla, D. S. Yoo, R. Wolf, E. R. Melhem, and C. Davatzikos, ‘‘Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme,’’ Magn. Reson. Med., vol. 62, no. 6, pp. 1609–1618, Dec. 2009.

[22] E. S. A. El-Dahshan, T. Hosny, and A. B. M. Salem, ‘‘Hybrid intelligent techniques for MRI brain images classification,’’ Digit. Signal Process., vol. 20, pp. 433–441, Mar. 2010.

[23] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang, and

Q. Feng, ‘‘Enhanced performance of brain tumor classification via tumor region augmentation and partition,’’ PloS ONE, vol. 10, no. 10, Oct. 2015, Art. no. e0140381.

[24] M. G. Ertosun and D. L. Rubin, ‘‘Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks,’’ in Proc. AMIA Annu. Symp. Proc., vol. 2015, Nov. 2015, pp. 1899–1908.

[25] J. S. Paul, A. J. Plassard, B. A. Landman, and D. Fabbri, ‘‘Deep learning for brain tumor classification,’’ Proc. SPIE, Med. Imag., Biomed. Appl. Mol., Struct., Funct. Imag., vol. 10137, Mar. 2017, Art. no. 1013710. doi: 10.1117/12.2254195.

[26] P. Afshar, K. N. Plataniotis, and A. Mohammadi, ‘‘Capsule net- works for brain tumor classification based on MRI images and course tumor boundaries,’’ 2018, arXiv:1811.00597. [Online]. Available: https://arxiv.org/abs/1811.00597

[27] A. K. Anaraki, M. Ayati, and F. Kazemi, ‘‘Magnetic resonance imaging- based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms,’’ Biocybernetics Biomed. Eng., vol. 39, no. 1, pp. 63–74, Jan./Mar. 2019.

[28] J. Cheng. ‘‘Brain Tumor Dataset.’’ Apr. 2, 2017. Distributed by Figshare. Accessed: May 30, 2019. [Online]. Available: https://figshare.com/articles/brain_tumor_dataset/1512427/5

[29] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore,

S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, ‘‘The cancer imaging archive (TCIA): Maintaining and operating a public information repository,’’ J. Digit. Imag., vol. 26, no. 6, pp. 1045–1057, 2013.

[30] L. Scarpace, A. Flanders, R. Jain, T. Mikkelsen, and D. Andrews, ‘‘Data from REMBRANDT. The cancer imaging archive,’’ 2015. doi: 10.7937/K9/TCIA.2015.588OZUZB.

[31] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, ‘‘Under- standing data augmentation for classification: When to warp?’’ 2016, arXiv:1609.08764. [Online]. Available: https://arxiv.org/abs/1609.08764

[32] S. Ioffe and C. Szegedy, ‘‘Batch normalization: Accelerating deep network training by reducing internal covariate shift,’’ 2015, arXiv:1502.03167. [Online]. Available: https://arxiv.org/abs/1502.03167

[33] V. Nair and G. E. Hinton, ‘‘Rectified linear units improve restricted boltzmann machines,’’ in Proc. 27th Int. Conf. Mach. Learn. (ICML), 2010, pp. 807–814.

[34] D. Scherer, A. Müller, and S. Behnke, ‘‘Evaluation of pooling operations in convolutional architectures for object recognition,’’ in Artificial Neural Networks. Berlin, Germany: Springer, 2010, pp. 92–101.

[35] J. Nagi, F. Ducatelle, G. A. D. Caro, D. Ciresan, U. Meier, A. Giusti,

F. Nagi, J. Schmidhuber, and L. M. Gambardella, ‘‘Max-pooling con- volutional neural networks for vision-based hand gesture recognition,’’ in Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA), Nov. 2011, pp. 342–347.

[36] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and

R. Salakhutdinov, ‘‘Dropout: A simple way to prevent neural networks from overfitting,’’ J. Mach. Learn. Res., vol. 15, no. 1, pp. 1929–1958, 2014.

[37] N. M. Nasrabadi, ‘‘Pattern recognition and machine learning,’’ J. Electron. Imag., vol. 16, no. 4, 2007, Art. no. 049901.

[38] D. Cireşan, U. Meier, and J. Schmidhuber, ‘‘Multi-column deep neu- ral networks for image classification,’’ 2012, arXiv:1202.2745. [Online].

Available: https://arxiv.org/abs/1202.2745

[39] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan,

V. Vanhoucke, and A. Rabinovich, ‘‘Going deeper with convolutions,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 1–9.

[40] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT press, 2016.

[41] L. Bottou, ‘‘Large-scale machine learning with stochastic gradient descent,’’ in Proc. COMPSTAT. Physica-Verlag HD, 2010,