



Deep CNN for Brain Tumor Classification

Wadhah Ayadi¹ · Wajdi Elhamzi^{2,3} · Imen Charfi³ · Mohamed Atri⁴

Accepted: 27 November 2020 / Published online: 6 January 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract

Brain tumor represents one of the most fatal cancers around the world. It is common cancer in adults and children. It has the lowest survival rate and various types depending on their location, texture, and shape. The wrong classification of the tumor brain will lead to bad consequences. Consequently, identifying the correct type and grade of tumor in the early stages has an important role to choose a precise treatment plan. Examining the magnetic resonance imaging (MRI) images of the patient's brain represents an effective technique to distinguish brain tumors. Due to the big amounts of data and the various brain tumor types, the manual technique becomes time-consuming and can lead to human errors. Therefore, an automated computer assisted diagnosis (CAD) system is required. The recent evolution in image classification techniques has shown great progress especially the deep convolution neural networks (CNNs) which have succeeded in this area. In this regard, we exploited CNN for the problem of brain tumor classification. We suggested a new model, which contains various layers in the aim to classify MRI brain tumor. The proposed model is experimentally evaluated on three datasets. Experimental results affirm that the suggested approach provides a convincing performance compared to existing methods.

Keywords Deep learning · MRI · Brain tumor · Classification · CNN

1 Introduction

The brain tumor is considered as the most common brain diseases. It is an uncontrolled and unnatural growth of brain cells [1]. It represents one of the most lethal cancers and life-threatening. According to cancer statistics in the USA, it is about 23,000 patients was diagnosed with a brain tumor in 2015. After 2 years, based on statistics, this sort of tumor is considered as the leading cause of cancer mortality around the world both in children

✉ Wajdi Elhamzi
Wajdi.Elhamzi@essths.rnu.tn

¹ Laboratory of Electronics and Microelectronics, University of Monastir, Monastir, Tunisia

² College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia

³ Higher School of Sciences and Technology of Hammam Sousse, University of Sousse, Sousse, Tunisia

⁴ College of Computer Science, King Khalid University, Abha, Saudi Arabia

and in adults [2]. In the aim to detect the tumor, the radiologist widely exploits medical imaging techniques [3]. Among the various available techniques, MRI is more favored for brain tumors according to its harmless nature. In daily routine, the radiologist identifies the brain tumors manually. The tumor classification process needs an extremely time consuming and it is based on the skills and experience of the radiologist. With the increase of patient number, the amount of data to be daily analyzed is large which make the readings based on visual interpretation expensive and inaccurate. Furthermore, the classification of brain tumors to various pathological types is more challenging compared to binary classification. The related challenges are attributed to some factors such as the high variations with respect to shape, size, and intensity for the same tumor type [4] also the similar appearances for varied pathological types [5]. A wrong diagnosis of a brain tumor can lead to a serious problem and decrease the chance of survival for the patient. In order to surpass the manual diagnosis drawbacks, there is a surge of interest in designing automated image processing systems [6–8]. Many researchers have suggested several techniques to ameliorate the CAD system that can classify some tumors in brain MRI images. Traditional machine learning methods used in the classification process are usually based on different steps such as preprocessing, dimension reduction, feature extraction, feature selection, and classification. The feature extraction represents the crucial phase in an effective CAD system [9]. It is a challenging task and requires prior knowledge about the problem domain since the classification accuracy based on the good features extracted. The traditional techniques for feature extraction can be sorted into three types: Spatial domain features, Wavelet and frequency features and Contextual and Hybrid features. The new CAD methods yield an improved performance due to the uses of deep learning (DL).

DL represents a subset of machine learning which does not require handcrafted features [10, 11]. It has been successfully proven to minimize the gap between human vision and computer vision in pattern recognition and can provide higher performance than traditional techniques [12]. It surpassed state-of-the-art schemes in several fields as generating text [13], face verification [14], image description [15], the game of Go [16], and grand challenges [17]. The higher performance in several fields encouraged the exploitation of DL in the medical image for classification, detection, and segmentation [18–24]. According to [25] and only in 2016, there are about 220 works based on deep learning at medical images are reported and this number will increase in the next years. Around 190 of them used CNN. DL permits the exploitation of a pre-trained CNN model for medical images especially for brain tumor classification, which was developed for other applications as AlexNet [26], GoogLeNet [4], ResNet-34 [5].

CNNs have gained higher performance on huge, labeled datasets as ImageNet [17] that contains more than one million images. However, it is hard to exploit such deep CNNs in the medical field. First, the size of the medical datasets is generally small because such datasets need the availability of expert radiologists to manually examine and label the images, which is time-consuming, laborious, and costly. Second, training deep CNN is a complicated task for a small dataset because of over-fitting and convergence problems. Third, domain expertise is needed to repeatedly revise the model and adjust the learning parameters to provide better performance. Therefore, training deep CNN from scratch represents a challenging task that is tedious and demands much diligence and patience.

A new model for brain tumor classification based on CNN is discussed in this paper. It contains various layers as convolution, Rectified Linear Unit (ReLU), and a pooling. Our new approach does not involve any segmentation in the pre-processing step, in contrast to some previous methods, which require prior segmentation of tumors. We validated our algorithm on three public datasets.

Our contribution in this work comprises of the following key points:

- A novel and robust model is presented for automated classification of brain tumors, which is effective in the extraction of important features on the MRI dataset.
- The model suggested exploiting 3×3 kernels for all convolutional layers with a small stride in the aim to learn the small texture of tumors in brain images, unlike other models, which use 11×11 or 9×9 as kernels size with higher strides.
- The novel model achieves good accuracy in brain tumor classification with few pre-processing compared to other techniques, which need a tumor segmentation before the classification step.
- Our model provides an acceptable classification accuracy compared to new methods, despite the few training data samples.

This paper is structured into eight sections. Some previous works are briefly given in Sect. 2. A background is outlined in Sect. 3. The proposed model is discussed in Sect. 4. The performance evaluation is outlined in Sect. 5. Section 6 is provided to describes the dataset exploited. Section 7 summarizes the results. Section 8 discussion and Sect. 9 concludes.

2 Related Work

Different methods had been suggested in the past years for classification and segmentation. These techniques used traditional machine learning [27–29] and recently exploited deep learning models [30–44]. We have looked into this section of the works exploited for brain tumor classification.

A new technique for MRI brain tumor classification which is proposed by Hemanth et al. [45] used a modified Neural Network. 540 MR brain images were exploited to test the suggested method. The dataset consists of four tumors class which are namely astrocytoma, meningioma, metastase, and glioma. The used images are of 256×256 size. Normalization is performed as a preprocessing step. Eight features are acquired based on the first-order histogram and GLCM. The suggested method provides promising results that can reach 95% as sensitivity, 98% as specificity, and 98% as accuracy.

Other work is proposed by Lin et al. [46] to classify meningioma tumor in a different grade. Grade-I contains the non-cancerous, which are slow-growing tumors. Grade-II contains cancerous and noncancerous tumors. The grade-III contains cancerous tumors, which can grow quickly. Different features are exploited as contextual and radiological features. No segmentation or preprocessing process is performed. In the classification step, the authors used multiple logistic regression. The proposed scheme is tested using 120 patients MRI images, were 90 with Grade I and 30 with Grade II or III. They exploited several sequences as FLAIR T1 and T2. DWI transformation is utilized to extract features. The results are acceptable with the exploited dataset. However, this kind of method requires many large datasets to ensure its validity.

Other work aimed to classify multi-grade brain tumors is presented in [47]. The proposed method is based on a pre-trained CNN model and segmented images. The model is tested based on three datasets. Various techniques for data augmentation are exploited in order to enhance accuracy. The technique is experimentally evaluated on the original and the augmented dataset. The provided results are convincing compared to previous works.

In order to help the radiologists in MRI classification, Sachdeva et al. [48] have suggested a semi-automatic classification scheme that contains varied steps. To detect tumor areas, a content based active contour system that allows the radiologist to manually indicate the region of interest (ROI), which is saved as a segmented ROI (SROI) is applied in the first step. Then, 71 texture and intensity features are extracted using the SROI. Optimal feature selection is performed with the application of the Genetic Algorithm (GA). The last phase classifies the chosen features using two classifier SVM and ANN. The suggested scheme is tested on different datasets. The first dataset contains 428 MR images and the second includes 260. The first set of images contains six tumors categories as Glioblastoma Multiforme (GBM), Meningioma (MEN), Astrocytoma (AS), childhood tumor-Medulloblastoma (MED), and secondary tumor-Metastatic (MET). The second dataset contains only three tumor categories, which are AS, MEN, and Low-Grade Glioma. The suggested GA-SVM aims to find a preliminary probability for the tumor category, while GA-ANN aims to confirm the accuracy. The performance calculated on the first group of images shows that GA based approach enhanced SVM accuracy to 91.7% while ANN accuracy to 94.9%. SVM accuracy has raised to 89% and 94.1% for ANN in the second group of images. The results demonstrate that the classifier GA-ANN offered the highest results compared to GA-SVM. Besides, the GA-SVM yields the speed while GA-ANN yields the accuracy. According to results, the suggested scheme has acceptable performance and can assist radiologists in taking a better decision to classify brain tumors.

Cheng et al. [27] are the first authors exploited the famous dataset [49]. The suggested system takes advantage of the manually delineated tumor border for feature extraction. They utilized the augmented tumor region as a region of interest (ROI), which was spliced into subregions based on the adaptive spatial division method. The feature was extracted in three manners, which are the gray-level co-occurrence matrix (GLCM), intensity histogram, and the bag of words (BoW). SVM achieved the highest accuracy. The experiments followed a standard five-fold cross-validation procedure. Accuracy, sensitivity, and specificity are the performance measures calculated. The highest accuracy is about 91.28%.

Ismael et al. [28] used Gabor filter and discrete wavelet transform (DWT) in the aim to extract statistical features for the classification. This algorithm exploits the tumor segmented as input and the multi-layer perceptron (MLP) as a classifier. A random division of database images into 70% and 30% was done to form the training set and validation set, respectively. The accuracy achieved is about 91.9%.

Different preprocessing manners are investigated by Tahir et al. [29] to ameliorate the classification result. They grouped these techniques into three groups: noise removal, edge detection, and contrast enhancement. The possible combinations are applied to different image sets. The authors affirm that the combination of various preprocessing techniques is more beneficial than applying a single technique. The SVM classifier was exploited and reported 86% as the highest accuracy on Figshare dataset.

According to the results, the available tumor detection systems are not providing satisfactory output. For this reason, there is a big need to get robust automated CAD systems. The conventional machine learning requires domain-specific expertise and experience. It needs efforts for manual extraction of features, which can decrease the efficiency of the system. The deep learning-based techniques surpass these drawbacks due to automatic feature extraction, which are robust for classification purposes based on the convolutional layers.

In the aim to ameliorate the classification accuracy in this dataset, Paul et al. [33] applied three different classifiers: CNN, fully connected neural network, and random forest.

CNN provides the highest accuracy rate, which attained 90.26%. The proposed model contains various layers as convolutional, MaxPool, and fully connected.

In this regard, Afshar et al. [34] propose a modified CNN architecture called capsule network (CapsNet) for the brain tumor classification. The proposed CapsNet exploits the spatial relationship between the tumor and its surrounding tissues. 86.56% present the highest accuracy provided based on segmented tumors and 72.13% based on raw brain images.

Other work used the deep belief network (DBN) in the aim to discriminate between healthy controls and patients with schizophrenia presented by 83 and 143 patients respectively from Radiopaedia dataset [50]. The proposed DBN provides 73.6% as accuracy compared to SVM which provides 68.1%.

Zhou et al. [35] proposed a holistic brain tumor classification technique. The features are extracted from the axial view using an auto-encoder and classified based on the Long Short Term Memory (LSTM). The proposed technique is tested on selected slices (989, axial only) and it reported 92.13% as the best accuracy.

Similarly, Pashaei et al. [36] developed a new architecture for the brain tumors classification. The proposed model contains five layers to extract features. A kernel Extreme Learning Machines (KELM) is used to classify images based on these extracted features. The accuracy achieved is about 93.68%, which exceeds Support Vector Machine, Radial Base Function, and some other classifiers.

Abiwinanda et al. [37] investigated the application of CNN for this data set and designed seven various neural networks. The second model provided the highest performance, which contains two convolutional and one fully connected layers. This simple model and without any previous segmentation achieve 98.51% as training accuracy and 84.19% as test accuracy.

Another reported use of CNN on this dataset is by Ghassemi et al. [38] where they proposed a new model for multi-class brain tumors classification. The model is pre-trained firstly as a discriminator in a generative adversarial network (GAN) in order to extract important features. Then, the last fully connected layer was replaced with a SoftMax classifier in the aim to differentiate three tumors. The proposed model contains six layers. It was used with various data augmentation techniques. It achieved 93.01% and 95.6% as accuracy on introduced and random split respectively.

Recently, various new architectures have been suggested in the goal to generalize the CNN to the graph domain, especially in the medical image classification [51].

Different authors choose the graph CNN (GCNN) as a solution for tumor classification [51–53]. In [52], Song et al. exploit a GCNN model in order to classify Alzheimer's disease (AD) into four categories. The proposed network contained eleven layers: nine convolutional and two FC. The ReLU activation is exploited after each layer. A Softmax is employed as the final layer in order to compute the class probabilities. The proposed scheme is tested based on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The original dataset contains 12 images per class. Due to the few data volume, various data augmentation is applied. The dataset is increased from 12 to 132 images per class. In the aim to generate a robust assessment, the author exploits a 10-fold CV. The average accuracy provided is about 65% for the SVM classifier and 89% based on GCNN.

Another work for AD classification is proposed by Guo et al. [53]. The author exploits GCNN in the goal to classify AD into 2 and 3 classes. The proposed model is tested based on the ADNI dataset. For the 2 class classification, the proposed GCNN attain 93% compared to the known ResNet architecture and SVM classifier where they reached 95% and 69% respectively. In the classification of the 3 classes scenario, the proposed GCNN attain 77%, and 65% and 57% for ResNet and SVM respectively.

A modified GCNN architecture is used in [51] for the early detection of AD. 160 images of patients from the ADNI dataset are used to test the suggested scheme. 5-fold CV is exploited to calculate the performance of the suggested model. The accuracy results surpass 90%.

Despite the various proposed schemes for the classification of brain tumors, these techniques suffer from several limitations that can be summarized as follows. The accuracy provided by the state-of-the-art schemes is inadequate considering the importance of MRI classification in the medical area. Several methods used the manually delineated tumor regions for the classification that prevented them from being fully automated. The algorithms utilized CNN and its variants could not provide an influential improvement in performance. Hence, performance evaluation based on various metrics other than accuracy becomes significant. Besides, the CNN models provide generally poor performance with small data sets, which is the case for the medical image database.

A new scheme is suggested to exceed these drawbacks. The suggested system provides the highest classification performance, compared to previous works, using three open datasets. Despite the use of a smaller number of training samples, the proposed method provides acceptable results.

3 Background

In the last years, DL has shown promising performance in several domains. DL models have the ability to learn automatically multiple levels of information from a large set of data. They have a huge advantage compared to traditional machine learning that needs a lot of effort for tuning the features and expert knowledge. Several architectures are proposed for DL. CNN represents the most exploited in image processing field due to its ability to recognize patterns in images [54].

CNN model can contain several types of layers. The frequently used are convolutional, pooling, and fully connected layers. The convolutional layer represents the main layer in a CNN scheme. It is used for feature extraction as edges and colors of the image. The pooling layer is exploited to decrease the dimensionality of extracted features, which leads to minimizing the complexity and the computational time. The fully connected layer represents the last step in CNN model, which aims to achieve linearity in the networks.

An optimizer algorithm is exploited by each CNN model in the training phase with the aim to update the weights. The model utilizes the classification loss as input and back propagate the error into the network to update the filters and the weights. At the final step, a SoftMax activation function is exploited in order to normalize the output sum. According to literature, a deeper CNN model can solve more complex tasks and improve accuracy.

In the medical image area, especially for brain tumor classification, several works have been suggested. In the last years, researchers have proposed multi-class brain tumor classification [33–38] since the binary classification [55] is insufficient for the doctor to choose the suitable treatment for the patient.

Despite the different techniques proposed in the literature, the brain tumor classification method still has such limitations that need to be considered. The binary classification leaves various ambiguities for the doctors and it is not enough to decide the good treatment. For a clear understanding of the doctor, the multi-classification is needed. Furthermore, the several datasets used, represent an obstacle for researchers to achieve precise comparison.

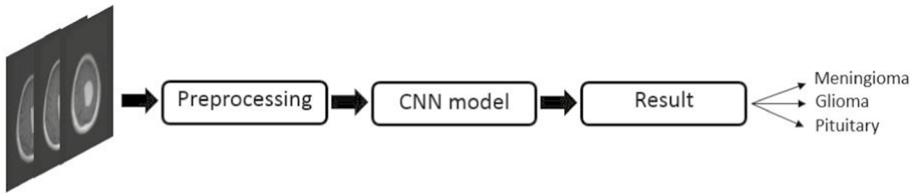


Fig. 1 The suggested approach

To overcome these limitations, we suggested a new deep CNN scheme for multi-class brain tumor classification using three publicly datasets.

A new scheme was proposed in the aim to generate an accurate classification system. The suggested approach is illustrated in Fig. 1. Our technique comprises various steps. The input images are obtained from the dataset. Normalization and contrast enhancer are exploited to ameliorate image quality. A new CNN scheme is suggested to extract the important feature. The test image is classified in one of input classes based on Soft-max activation function in the last step.

4 Proposed Model

Recently, CNN is widely exploited in all types of medical image processing applications particularly in MRI brain tumor classification and segmentation. In this work, a new CNN model is suggested for brain tumor multi-class classification.

The general architecture of the proposed sequential model is outlined in Fig. 2 and details are described in Table 1. It is consisted of several layers each having its own functionality. Image with 256×256 size represents the input model. Ten convolutional layers are exploited to extract the important feature. Max-pooling layer after every two convolutional layers is employed to reduce the data size. Each convolutional layer use 3×3 filters while 2×2 are applied in pooling layers. A non-linearity layer is added to ameliorate the fitting ability of CNN. Furthermore, a batch Normalization is used after

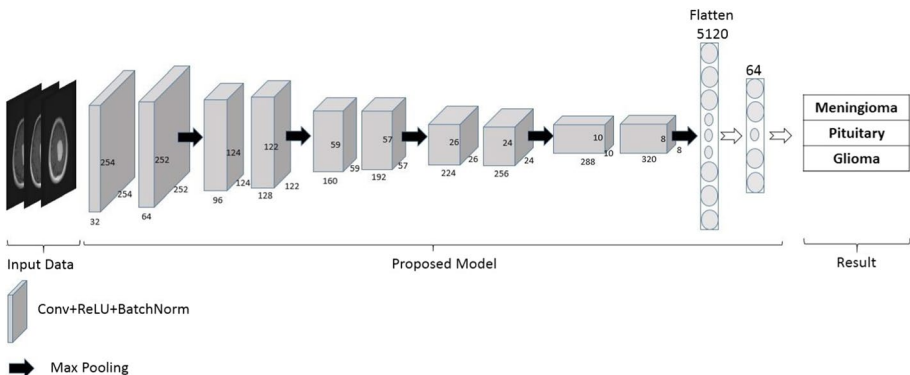


Fig. 2 The proposed architecture

Table 1 CNN proposed structure

Layer	Output shape			Kernel size
	Width	Height	Depth	
Conv/relu/BatchNorm	254	254	32	3
Conv/relu/BatchNorm	252	252	64	3
MaxPool	126	126	64	2
Conv/relu/BatchNorm	124	124	96	3
Conv/relu/BatchNorm	122	122	128	3
MaxPool	61	61	128	2
Conv/relu/BatchNorm	59	59	160	3
Conv/relu/BatchNorm	57	57	192	3
MaxPool	28	28	192	2
Conv/relu/BatchNorm	26	26	224	3
Conv/relu/BatchNorm	24	24	256	3
MaxPool	12	12	256	2
Conv/relu/BatchNorm	10	10	288	3
Conv/relu/BatchNorm	8	8	320	3
MaxPool	4	4	320	2
Flatten	1	1	5120	–
Dense	1	1	64	–
Sotmax	1	1	3	–

each convolution layer to get the best-optimized results and speed up the network convergence. Fully connected layers with 64 neurons is employed. The output layer exploits softmax classifier. We will briefly introduce these layers in this section.

4.1 Convolution Layer

This layer represents the most important part and the core component in the CNN model, which is also the origin of the name “Convolution Neural Network”. It is exploited as the first layer, in the aim to extract different features from input data. The first convolutional layer aims to extract the low-level features, while the complex features can be extracted based on more convolutional layers [56, 57]. The convolution layer is calculated as flow (Eq. 1):

$$\hat{F}_j = \sum_i F_i \otimes K_{ij} + b_j \quad (1)$$

where F_i represent the i -th input feature map, \hat{F}_j is the j -th output feature map, K_{ij} represents the convolutional kernel, b_j is the bias, and \otimes denotes the 2-D convolution operation. Both K_{ij} and b_j are learnable parameters in CNNs.

4.2 Non-linearity Layer

The second layer of the model represents the non-linearity layer. The nonlinear factor is added to ameliorate the fitting ability of CNN. It is achieved by using activation functions as Sigmoid, ReLU, leaky ReLU, ELU, etc. The ReLU function is the most used due to its simplicity and easy to apply since it does not need much calculation [58]. It is expressed as follows (Eq. 2), where x is the input value.

$$\text{ReLU}(x) = \begin{cases} x, & \text{if } x > 0, \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

4.3 Batch Normalization Layer

The suggested model used the batch normalization layer after each convolution layer. For getting the best-optimized results, this layer is utilized. It is exploited also to speed up the network convergence and to train the data perfectly.

4.4 Pooling Layer

The features obtained from the convolutional layer are still very large. If directly used, the training phase will be prone to overfitting and very time-consuming. To deal with this problem, this layer adopts a downsampling manner to compress the image and minimize the parameters. Several forms of subsampling are exploited in literature as mean pooling, max-pooling. In the proposed model, the dimension of feature maps is reduced by performing max-pooling operation since it is very easy to apply and it is achieving the highest results [59].

4.5 Fully Connected Layer (FC)

The FC layer is adopted at the end of the network. Since the features must be one-dimensional (1D) data before training with the classifier, the purpose of this layer is to flatten the output of the previous layer. When it is utilized as the last layer, the output is fixed as the number of classes used [60–62].

5 Performance Metrics

Following the same performance metrics in the previous references, the efficiency of images classification into three classes is measured based on Accuracy, Specificity, Sensitivity, Precision, and F1 Score. The formula for calculating these performance measures are written as below in Eqs. 3–7 respectively:

$$\text{Accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)} \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4)$$

$$\text{Sensitivity (Recall)} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$F1 \text{ Score} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (7)$$

where TP identifies the True Positive, TN identifies the True Negative, FP identifies the False Positive and FN identifies the False Negative. These parameters are estimated from the confusion matrix, which provides the details about the false and successful classification of images from all categories.

6 Image Data

6.1 Figshare Dataset

We exploited the publicly available brain tumor dataset Figshare [49], to evaluate our proposed model. It was collected from Nanfang Hospital, Guangzhou, China, and General Hospital, Tianjin Medical University, China, from 2005 to 2010. The dataset contains 233 patients suffering from brain tumors, wherein three classes of tumors are apparent, namely, meningioma, pituitary, and glioma tumors. They represent about 15%, 15%, and 45%, respectively, among all tumors [63]. The images have a size of 512×512 pixels. The dataset comprises 3064 slices, including 708 meningioma, 1426 gliomas, and 930 pituitary tumors [64]. Table 2 explains the details of the images data exploited in the experiments.

For every patient, three experienced radiologists initially examined independently the MRI image to obtain the pathology type. This data set is complicated, and some tumors are similar in color, shape, position... Some examples of MRI images are given in Fig. 3. Figure 3a, b present two patients having the same category of tumors, where they exhibit different appearances. Conversely, a visual similarity is shown in Fig. 3c, d, where are

Table 2 Summary of the image dataset

Tumor category	Number of patients	Number of slices	View	Number of slices
Meningioma	82	708	Transverse	209
			Sagittal	231
			Coronal	268
Pituitary	62	930	Transverse	291
			Sagittal	320
			Coronal	319
Glioma	89	1426	Transverse	494
			Sagittal	495
			Coronal	437

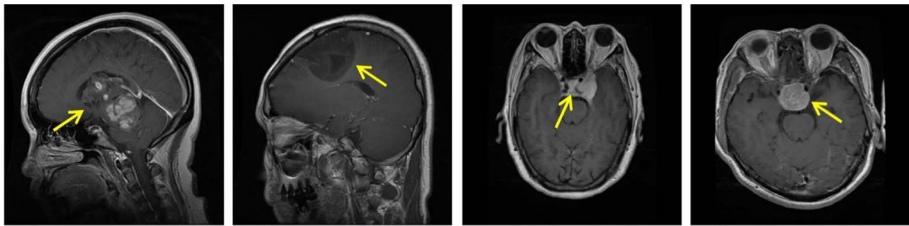


Fig. 3 Some examples of brain tumors (indicated by the yellow arrows). **a** and **b** presents a Gliomas in different subjects. **c** Present a meningioma and **d** present a pituitary tumor from different subjects [3]

Table 3 Radiopaedia dataset before and after using data augmentation

Tumor grades	Tumor types	Number of images	
		Before aug-mentation	After augmentation
Grade I	Menigiomas	36	648
Grade II	Gliomas	32	576
Grade III	Gliomas	25	450
Grade IV	Glioblastmoas	28	504

containing different pathological categories. In this study, two images include the same tumor category are considered as similar; else, they are considered dissimilar.

6.2 Radiopaedia Dataset

Radiopaedia [65] represents the second dataset exploited in this paper. It contains 121 MR images correspond to four different grades as shown in Table 3. This dataset suffers from a few numbers of images in each grade. Whereas, big data with various examples represent a key for effective deployment of various deep learning models [66]. This dataset lacks a satisfactory volume of data to use deep learning models and earn good accuracy. In order to attain higher performance, we augmented the original data by using four various augmentation methods which are detailed in Table 4. For geometric transformations, we exploit rotation and flipping. Gaussian blur and sharpening are applied to reduce the noise. About 17 parameters and four augmentation techniques are exploited, which can generate 17 other samples for each image of the dataset. The augmented dataset is detailed in Table 3. The performances are evaluated based on a five-fold CV.

6.3 REMBRANDT Dataset

The last dataset used in this paper is the Repository of Molecular Brain Neoplasia Data (REMBRANDT), which was exploited in various previous works [67]. It contains 130 brain tumor patients with an average survival rate of 47 months. The dataset consists of various brain tumor types, which are astrocytoma (AST), oligodendroglioma (OLI),

Table 4 Various data augmentation methods with the used parameters

Technique number	Augmentation technique	Parameters
1	Rotation	45°
		90°
		135°
		180°
		225°
		270°
2	Flip	LEFT_RIGHT
		TOP_BOTTOM
3	Gaussian Blur (Sigma value)	0.25
		0.50
		1.0
		2.0
4	Sharpen (lightness value)	0.50
		1.00
		1.50
		2.0

Table 5 Rembrandt details

Tumor type	Total patients	Grade	Total patients in grade
AST	47	Grade 2	30
		Grade 3	17
OLI	21	Grade 2	14
		Grade 3	7
GBM	44	Grade 4	44
Unknown tumor	18	–	18

glioblastoma multiforme (GBM), and other patients with unknown tumor. The AST type includes 47 patients divided into 2 grades: grade 2 and grade 3 which contain respectively 30 and 17 patients. The OLI type includes 21 patients divided into 2 grades: grade 2 and grade 3 which contain respectively 14 and 7 patients. The GBM type consists of 40 patients with grade 4. The data set is summarized in Table 5. In order to augmented data, we exploit 4 data augmentation techniques, which are Flip LEFT_RIGHT, Flip TOP_BOTTOM, Gaussian Blur (2.0), and Sharpen (2.0).

This dataset consists of various MRI protocols as FLAIR, T1W, T2W, and diffusion-weighted imaging (DWI). The data set slices were manually labeled by experts. In this paper, the slice is considered a negative sample if the tumor lesion was not visible. If the tumor lesion is visible, the slice is labeled as a positive sample. Based on this dataset, five sub-datasets can be designed, which are detailed below.

Table 6 C3 dataset detail

Tumor grades	Number of images	
	Before augmentation	After augmentation
Normal	1041	5205
LGG	484	2420
HGG	631	3155
Total samples	2156	10,780

Table 7 C4 dataset detail

Tumor grades	Number of images	
	Before augmentation	After augmentation
Normal	1041	5205
AST (G2 + G3)	557	2785
OLI (G2 + G3)	219	1095
GBM	339	1695
Total samples	2156	10,780

6.3.1 Two-Class Data (C2)

The main objective of this dataset is to classify the brain MRI into tumorous and normal classes. The dataset was divided into these two categories. Each slice from the dataset is included in the tumor class if the tumor is visible, while, it is included in the normal class. In this paper, we exploit 1041 and 1091 MRI slice as normal and tumorous samples, respectively. The original data set include 2132 MRIs, which are increased to 10,660 using the data augmentation technique.

6.3.2 Three-Class Data (C3)

The main objective of this dataset is to discriminate the brain MRI into three classes: normal, LGG, and HGG. As mentioned in the previous dataset, the normal class contains 1041 images. The LGG dataset included AST grade 2 and OLI grade 2, while the HGG dataset included grade 3 of AST and OLI and the GBM samples. In general, C3 includes 1041 samples in a normal class, 484 samples in LGG class, and 631 samples in HGG class. C3 details are shown in Table 6.

6.3.3 Four-Class Data (C4)

The main objective of this dataset is to classify the brain MRI into four classes, namely normal, AST, OLI, and GBM. The AST class includes AST grade 2 and grade 3. OLI class also includes OLI grade 2 and grade 3. These classes, normal, AST, OLI, and GBM contain 1041, 557, 219, and 339 samples, respectively. The C4 data set is summarized in Table 7.

Table 8 C5 dataset detail

Tumor grades	Number of images	
	Before augmentation	After augmentation
AST (G2)	356	1780
AST (G2)	201	1005
OLI (G2)	128	640
OLI (G3)	91	455
GBM	339	1695
Total samples	1115	5575

6.3.4 Five-Class Data (C5)

This dataset is designed to classify brain MRI into different classes, which are AST (grade 2), AST (grade 3), OLI (grade 2), OLI (grade 3), and GBM (grade 4). These classes contain respectively 356, 201, 128, 91, and 339 samples. The dataset details are shown in Table 8.

6.3.5 Six-Class Data (C6)

The last proposed dataset aims to discriminate brain MRI into six classes, which are normal, AST (grade 2), AST (grade 3), OLI (grade 2), OLI (grade 3), and GBM (grade 4). This dataset includes 1041, 356, 201, 128, 91, and 339 samples for these classes, respectively. The C6 data set is summarized in the next table (Table 9).

7 Experimental Results

The suggested model was trained using a desktop computer equipped with Intel Xeon Processor E5-2620 v4 and 64 GB RAM. It was implemented on Python 2.7, using Keras library and Tensor Flow. In this work, we have proposed a new model for brain tumor classification. It is tested on three datasets.

Table 9 C6 dataset detail

Tumor grades	Number of images	
	Before augmentation	After augmentation
Normal	1041	5205
AST (G2)	356	1780
AST (G2)	201	1005
OLI (G2)	128	640
OLI (G3)	91	455
GBM	339	1695
Total samples	2156	10,780

Table 10 Comparison between different optimizers and learning rates

Optimizer	Learning rate					
	0.1	0.01	0.001	0.002	0.003	0.004
Adam	66.5	70.69	79.37	79.04	81.85	76.30
SGD	85.31	91.09	86.71	89.26	88.54	89.91
Adadelata	86.40	83.35	76.69	79.56	81.07	81.33
Adagrad	81.07	88.83	92.98	93.21	94.74	92.46

Table 11 Comparison between various numbers of epochs

Number of epochs	10	20	30	40
Accuracy	94.35	94.74	94.32	94.64

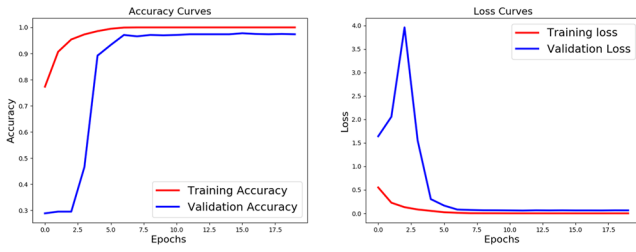
Our model consists of a 10-convolution layer in order to extract features from MRI images followed by 10 batch normalization layers in order to speed up the learning. In the aim to add more nonlinearity to the network and reduce the size of the layers, five max-pooling layers are exploited. To make Convolution output suitable for FC layers, the feature is reshaped to one dimension. An FC layer with 64 filters is exploited. The SoftMax classifier layer is used for classification [68]. It has three or four output neurons as the number of classes. All convolution layers in this model employ filters of size 3×3 . The ReLu activation function is used, as it is already the standard activation function in the image classification models. The size of the max pool kernel is 2×2 .

7.1 Hyper-parameters

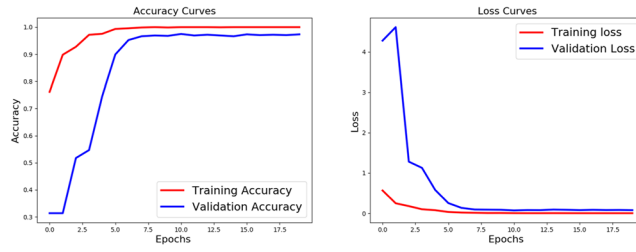
The Figshare dataset images are in the size of 512×512 . Because of computational cost reasons, this size is reduced to 256×256 . To enhance image quality, we exploit intensity normalization and contrast enhancer as a preprocessing step. We use the 5-fold proposed by the dataset in the aim to make valid and comparable results. The method is evaluated using 70% from train fold as training and 30% as validation. Each run is repeated ten times and calculated the average as a result in order to improve the soundness of the outcome. Various tests are important before the final model evaluation in order to assure the hyper-parameters value choices. Based on Tables 10 and 11, the best optimizer is Adagrad with a 0.003 learning rate. The number of epochs is 20 with 16 as batch size.

The performance of our suggested model is evaluated based on five-fold cross-validation. The results are presented in the next tables. Figure 4 shows the training-validation process of the network. In these figures, the red and blue lines illustrate respectively the training and validation process.

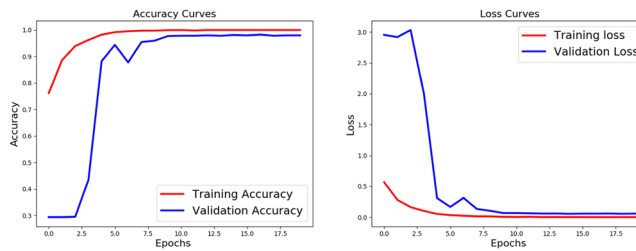
The benefit of the suggested scheme is to speed the convergence and reduce the overfitting, which is very clear from the accuracy and the loss history in Fig. 4. As outlined in these figures, our suggested CNN model reached very fast its maximum performance and there is a consistency between the training-validation accuracy and loss.



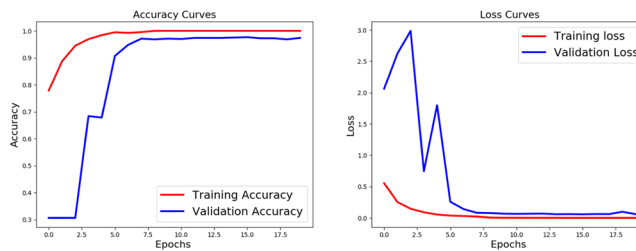
(a) Cross-Validation set 1



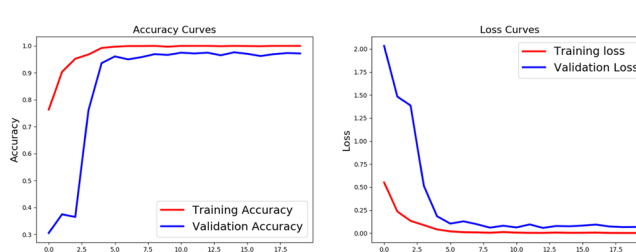
(b) Cross-Validation set 2



(c) Cross-Validation set 3



(d) Cross-Validation set 4



(e) Cross-Validation set 5

Fig. 4 Accuracy and loss history of the proposed model based on fivefold cross-validation. **a–e** represent each set Accuracy and Loss history

7.2 Complexity Analysis

Computational complexity represents one of the important criteria for measuring the quality of the CNN model. We know that the time complexity of the CNN model depends on several parameters as the number of layers, the training phase. In Table 1, we list the different layers used in the proposed model. We use 10 convolution layers, which is a medium size for a CNN model. The complexity of the CNN model will be increased also with the increase of input size. Therefore, size reduction can be used to minimize the complexity of time. We start data preprocessing by reducing the input image size for all data used. The pooling layer can be another important technique to reduce computational complexity of the model. In this regard, our model exploits 5 layers of pooling. The main goal behind these operations is to reduce the computational complexity that can be helpful to generate a low power, and less complexity model. According to the results presented in Table 12 and to balance between accuracy and used time per epochs, we use 256×256 as input size and 3×3 as filter size. The number of parameters used is about 3.3 M. Therefore, we can say that the complexity of our scheme is acceptable.

7.3 Experimental Results for Figshare Dataset

A summary of predictions provided by the proposed model can be presented as a confusion matrix where each row illustrates the actual class and each column illustrates the predicted class [69]. Table 13 shows the confusion matrix. From this table, we can note that the proposed technique classifies 2903 cases correctly and 161 cases incorrectly. It yields

Table 12 Comparison between various numbers of epochs

		Time (s)	Accuracy
Image size (with filter size 3×3)	64×64	25	82.07
	128×128	125	89.29
	256×256	550	94.74
Filter size	5×5	896	88.02
	9×9	2500	83.9

Table 13 Confusion matrix for the suggested model

True/Auto	Meningioma	Glioma	Pituitary tumor
Meningioma	635	49	24
Glioma	64	1347	15
Pituitary tumor	9	0	921

Table 14 Performance of the proposed method on the test dataset

Class	Accuracy	Sensitivity	Specificity	Precision	F1-score
Meningioma	95.23	89.68	96.9	89.68	89.68
Glioma	95.43	94.46	97	96.48	95.45
Pituitary tumor	98.43	99.03	98.17	95.93	97.45

acceptable results, which attain 94.74% as the whole accuracy. We can note also that the meningioma tumor class gains the highest misclassifications rate. The misclassification is due to the fewer images number furnished for this class in the original dataset and no augmentation technique was used.

Generally, many previous works utilized accuracy to evaluate model performance. In this study, we have exploited other measures as sensitivity, specificity, precision, and f1-score, in order to have a clear indicator of the generalization of the model. Table 14 summarizes the classification performance of each tumor type and the average of best performed proposed CNN model.

From Table 14, we note that the proposed model provides 95.23%, 95.43%, and 98.43% as accuracy for Meningioma, Glioma, and Pituitary tumor, respectively. The sensitivity is another factor of classification performance, which attains respectively 89.68%, 94.46%, and 99.03%. For all the classes, the specificity values are high which means that the model correctly recognizing the samples without a specific disease.

The performances of our model are compared with new classification methods that exploit the same dataset. As shown in Table 15, our model surpasses the newest methods based on classification accuracy. The table contains only accuracy as a performance metric since it is the common metric that is exploited in all the related works. Previous works [27, 29] exploited the traditional machine learning technique which needs manual feature extraction that is a tedious task and time consuming. Other models based on CNN have exploited shallower networks that caused a limitation to provide high accuracy [33, 34, 37]. In this study, the choice of architecture and the hyper-parameters for the brain tumor classification has proven to process the MRI images effectively and attain higher accuracy results.

The MRI images classification of brain tumor represent a challenging problem owing to the variety in intensities, size, orientation, and shape. In addition to images contrast and the noise perturbations. Furthermore, the medical datasets exploited are often limited in size and hard to access. Various inferences are made through performance analysis. From the confusion matrix in Table 13, we noticed that most of the

Table 15 Comparison with previous works

Author	Manual segmentation	Number of images used	Best accuracy
Guo et al. [53]	No	327	2 classes: 93% 3 classes: 77%
Abiwinanda et al. [37]	No	2100 (700 from each tumor type)	84.19
Paul et al. [33]	No	989 (axial only)	84.52
Paul et al. [33]	No	989 (axial only)	90.26
Tahir et al. [29]	No	3064	86
Afshar et al. [34]	Both	3064	86.56 using segmentation 72.13 using raw images
Afshar et al. [70]	Bounding box	3064	90.89
Cheng et al. [27]	Yes	3064	91.28
Ismael et al. [28]	Yes	3064	91.9
Zhou et al. [35]	No	989 (axial only)	92.13
Ghassemi et al. [38]	No	3064	93.01
Pashaei et al. [36]	Not mentioned	3064	93.68
Proposed	No	3064	94.74

Table 16 Confusion matrix for Radiopaedia dataset before using data augmentation

Class	Accuracy	Sensitivity	Specificity	Precision	F1-score
Grade I	95.61	88.23	98.75	96.77	92.3
Grade II	92.98	93.33	92.85	82.35	87.49
Grade III	93.85	84	96.62	87.5	85.71
Grade IV	98.24	96	98.87	96	96

Table 17 Confusion matrix for Radiopaedia dataset after using data augmentation

Class	Accuracy	Sensitivity	Specificity	Precision	F1-score
Grade I	96.32	90.79	98.75	96.94	93.76
Grade II	95.31	95.66	95.18	87.87	91.59
Grade III	96.18	90.84	97.61	91.05	91.05
Grade IV	99.61	98.22	100	100	99.1

misclassifications pertain to the class meningioma. This is due to the fewer samples in the dataset from this class and we don't use any technique for class-specific data augmentation to balance the dataset. The handling of smaller amounts of training data represent the other part that our work concentrated on. According to Fig. 4, we notice that the training loss decreases, and the validation loss increases during the first iterations. This indicates that the phenomenon of overfitting took place, which leads to lower classification accuracy. This overfitting could be avoided by data augmentation. This aspect offers a scope for future works.

7.4 Experimental Results for Radiopaedia Dataset

7.4.1 Experimental Results Before Data Augmentation

The confusion matrix for the second datasets is given in Table 16. It can be shown that the accuracy attains 95.61%, 92.98%, 93.85%, and 98.24% respectively for grade I, II, III, and IV. The overall accuracy (OA) can reach about 90.35%. The results gained from the original dataset are not convincing enough and it cannot be trusted in the clinical environment due to the height error rate. The poor OA indicates the confused of some grades, which not acceptable.

7.4.2 Experimental Results After Data Augmentation

Due to the poor results obtained by the original dataset, we use various data augmentation to enhance the performance provided by the proposed model. The confusion matrix for the new dataset is presented in Table 17.

Based on the calculated performances, the OA is increased from 90.35 to 93.71%, which is much better than the OA gained without data augmentation. Similarly, accuracies for grades I, II, III, and IV are increased respectively from 95.61 to 96.32%, 92.98 to 95.31%, 93.85 to 96.18%, and 98.24 to 99.61%.

Various other performances can be calculated in order to have a clear indicator of the generalization of the model. Tables 16 and 17 summarizes the classification performance of each tumor grade.

The suggested model provides 90.79%, 95.83%, 90.84%, and 98.22% as sensitivity for Grade I, Grade II, Grade III, and Grade IV, respectively. For all the grades, the specificity values are high which indicates that the proposed model correctly recognizing the samples without a specific group. The F1-score is another factor of classification performance, which attains respectively 96.38%, 89.3%, 91.05%, and 99.1%.

According to the experimental results, it is evident that the use of data augmentation techniques enhances the classification performance.

7.4.3 Comparison with Previous Methods

A comparison with previous techniques is detailed in Table 18. Sachdeva et al. [48] used 428 MRI images in order to classify 5 classes. Various intensity and texture features are extracted from each image. The OA is about 94%. Pinaya et al. [50] classified MRI into Healthy control and Schizophrenia based on deep belief network and SVM. 231 images are exploited in this work. The OA attained is about 73.6%.

Recently, Sajjad et al. [47] propose a new deep CNN in order to classify 4 grades of brain tumors. The original dataset contains 121 images. The OA attains 87.38%. Due to the poor result and the reduced dataset used for Deep CNN model, the author uses data augmentation to provide several examples for the dataset. The new dataset contains 3630 images and the OA is increased to 90.67%.

In our work, two scenarios are realized, before and after data augmentation. The original data contain 121 images as used by Sajjad et al. [47]. It includes 4 grades: Grade I (Meni-giomas), Grade II (Gliomas), Grade III (Gliomas), and Grade IV (Glioblastomas). Each grade includes 36, 32, 25, and 28 images, respectively. The OA attained is 90.35%. Due to the poor result, we use several techniques for data augmentation, which are detailed in Table 4. The data set is augmented from 121 to 2178 images. The OA attained is 93.71%.

7.5 Experimental Results for REMBRANDT Dataset

7.5.1 Experimental Results Before Data Augmentation

The image size is reduced to 128×128 with the aim to reduce the time complexity. The confusion matrix for the third datasets is given in Tables 19, 20, 21, 22 and 23. From Table 19, it can be shown that the proposed model gained excellent accuracy, which attains 100%. This is due to the simplicity of this kind of classification and the accepted data volume exploited. According to Table 20, the normal class attains 100% accuracy, while LGG and HGG attain both 95%. The overall accuracy (OA) can reach about 95%. In C4 and based on Table 21, the normal class attains 98.6% accuracy. The accuracy for AST (G2+G3) can reach about 95.34%, and 99.06% and 95.81% for OLI (G2+G3) and GBM respectively. The OA is about 94.41%. Based on Table 22, the C5 accuracy attains 90.43% for AST (G2), 98.26% for AST (G3), 98.26% for OLI (G2), 96.52% for OLI (G2), and 88.69% for GBM. The OA reaches about 86.08%. According to Table 23, the last dataset proposed (C6) provide 97.67% as accuracy for normal class, 95.34% for AST (G2), 98.13% for AST (G3), 99.06% for OLI (G2), 99.06% for OLI (G3) and 94.88% for GBM. The OA attains 92.09%. The poor OA obtained indicates the confusion of some classes, which not acceptable in the clinical environment. We can surpass this issue using some data augmentation techniques.

Table 18 Comparison with previous works

Author	Brain tumor classe (number)	Method	Class accuracy (%)
Pinaya et al. [50]	Healthy control (83) Schizophrenia(148) Total = 231	DBN-SVM	Healthy control = 76 Schizophrenia = 70 OA = 73.6
Song et al. [52]	Cognitively normal (132) Early mild cognitive impairment (132) Late mild cognitive impairment (132) AD (132) Total = 528	GCNN	OA = 89%
Sachdeva et al. [48]	AS-118, GBM-59 MEN-97, MED-88, MET-66 Total = 428	Intensity and texture feature- SVM - ANN	AS-89.8 GBM-83.3 MED-95.6 MEN-91.8 MET-97.1 OA = 94.9
Ajjad et al. [47]	Before augmentation Grade I (36) Grade II (32) Grade III (25) Grade IV (28) Total = 121 After augmentation Grade I (1080) Grade II (960) Grade III (750) Grade IV (840) Total = 3630	Deep CNN (VGG-19)	Grade I = 90.03 Grade II = 89.91 Grade III = 84.11 Grade IV = 85.50 OA = 87.38 Grade I = 95.54 Grade II = 92.66 Grade III = 87.77 Grade IV = 86.71 OA = 90.67

Table 18 (continued)

Author	Brain tumor classe (number)	Method	Class accuracy (%)
Proposed method		New Deep CNN model	
	Before augmentation		
	Grade I (36)		Grade I = 95.61
	Grade II (32)		Grade II = 92.98
	Grade III (25)		Grade II = 93.85
	Grade IV (28)		Grade IV = 98.24
	Total = 121		OA = 90.35
	After augmentation		
	Grade I-648		Grade I = 96.32
	Grade II-576		Grade II = 95.31
	Grade III-450		Grade III = 96.18
	Grade IV-504		Grade IV = 99.61
	Total = 2178		OA = 93.71

Table 19 Confusion matrix for C2

Class	Accuracy	Sensitivity	Specificity	Precision	F1-score
Normal	100	100	100	100	100
With tumor	100	100	100	100	100

Table 20 Confusion matrix for C3

Class	Accuracy	Sensitivity	Specificity	Precision	F1-score
Normal	100	100	100	100	100
LGG	95	100	93.07	84.7	91.71
HGG	95	70	100	100	82.35

Table 21 Confusion matrix for C4

Class	Accuracy	Sensitivity	Specificity	Precision	F1-score
Normal	98.6	99	98.26	98.01	98.5
AST (G2+G3)	95.34	96.36	95	86.88	91.37
OLI (G2+G3)	99.06	92	100	100	95.83
GBM	95.81	80	98.88	93.33	85.87

Table 22 Confusion matrix for C5

Class	Accuracy	Sensitivity	Specificity	Precision	F1-score
AST (G2)	90.43	85.71	92.5	83.33	84.5
AST (G3)	98.26	90	100	100	94.73
OLI (G2)	98.26	86.66	100	100	92.85
OLI (G3)	96.52	80	98.09	80	80
GBM	88.69	85.71	90	78.94	82.18

Table 23 Confusion matrix for C6

Class	Accuracy	Sensitivity	Specificity	Precision	F1-score
Normal	97.67	100	95.65	95.23	97.55
AST (G2)	95.34	85.71	97.22	85.71	85.71
AST (G3)	98.13	90	98.97	90	90
OLI (G2)	99.06	86.66	100	100	92.85
OLI (G3)	99.06	80	100	100	88.88
GBM	94.88	82.85	97.22	85.29	84.05

7.5.2 Experimental Results After Data Augmentation

Various data augmentation techniques are exploited in the aim to enhance the performance provided by the proposed model. The confusion matrix for the new dataset is presented in Tables 24, 25, 26, 27 and 28.

Table 24 Confusion matrix for C2

Class	Accuracy	Sensitivity	Specificity	Precision	F1-score
Normal	100	100	100	100	100
With tumor	100	100	100	100	100

Table 25 Confusion matrix for C3

Class	Accuracy	Sensitivity	Specificity	Precision	F1-score
Normal	100	100	100	100	100
LGG	97.22	98.4	96.76	92.21	95.20
HGG	97.22	86	99.4	96.99	91.16

Table 26 Confusion matrix for C4

Class	Accuracy	Sensitivity	Specificity	Precision	F1-score
Normal	99.81	99.8	99.82	99.8	99.8
AST (G2 + G3)	97.58	97.09	97.75	93.68	95.35
OLI (G2 + G3)	98.69	90.4	99.78	98.26	94.16
GBM	97.95	93.71	98.77	93.71	93.71

Table 27 Confusion matrix for C5

Class	Accuracy	Sensitivity	Specificity	Precision	F1-score
AST (G2)	94.43	88.5	97	92.81	90.6
AST (G3)	96.69	94	97.26	87.85	90.82
OLI (G2)	97.91	96	98.2	88.88	92.3
OLI (G3)	96.69	62	100	100	76.54
GBM	92	90.85	92.5	84.12	87.35

Table 28 Confusion matrix for C6

Class	Accuracy	Sensitivity	Specificity	Precision	F1-score
Normal	99.81	100	99.65	99.6	99.8
AST (G2)	98.79	93.14	99.88	99.79	96.16
AST (G3)	97.48	88	98.46	85.43	86.69
OLI (G2)	99.25	98.66	99.3	91.35	94.86
OLI (G3)	98.79	76	99.9	97.43	85.39
GBM	97.3	94.85	88	89.24	87.67

According to the calculated performances, all the performances are increased which indicates that the suggested model works well with the augmented data. The OA for the augmented C2 stays fixed at 100%. The OA for the augmented C3 is increased from 95 to 97.22%. Similarly, OA for the augmented C4, C5 and C6, are increased respectively from 94.41 to 97.02%, 86.08% to 88.86%, and 92.09 to 95.72%. Several other performances can be calculated which are summarized in the previous tables in the aim to have a clear

Table 29 Comparison with previous works

Author	Classes	Method	Class accuracy (%)
Anaraki et al. [71]	3	CNN-1 (7-layer) CNN-2 (8-layer)	Acc I= 90.9 Acc II= 94.2
Yang et al. [4]	3	AlexNet: 8 layer GoogLeNet: 22 layer	Acc=93
Tandel et al. [72]	2, 3, 4, 5, 6	AlexNet: 8 layer	Acc C2= 100 Acc C3=95.97 Acc C4=96.65 Acc C5= 87.14 Acc C6=93.74
Proposed method	2, 3, 4, 5, 6	New Deep CNN model	Before augmentation: Acc C2= 100 Acc C3=95 Acc C4= 94.41 Acc C5= 86.08 Acc C6=92.09 After augmentation: Acc C2= 100 Acc C3=97.22 Acc C4=97.02 Acc C5= 88.86 Acc C6=95.72

indicator of the generalization of the model. Based on the experimental results, data augmentation techniques significantly ameliorate the classification performance.

7.5.3 Comparison with Previous Methods

A comparison with previous techniques used in the same dataset is detailed in Table 29. Anaraki et al. [71] evolved two CNN strategies based on GA. The first model exploits 7 layers in the aim to classify brain MRIs into three grades. The second model uses 8 layers to distinguish between 3 brain tumor classes. These two models can achieve 90.9% and 94.2% accuracy, respectively. The proposed approach is computationally expensive due to the complexity of parameter selection based on the GA method.

In a similar study, Yang et al. [4] exploit two well-known CNN models, which are AlexNet and GoogLeNet with transfer learning (TL) for glioma classification. The TL technique outperformed the “learning from scratch” technique for both models. The height accuracy provided is about 93%.

Recently, Tandel et al. [72] proposed a new technique based on the AlexNet model with the aim to classify 5 MRI brain tumor dataset. The used datasets include two, three, four, five, and six classes, respectively. The suggested technique outperforms several other machine learning techniques like SVM, Decision Tree, K-nearest neighbor, etc. The accuracy provided in about 100%, 95.97%, 96.65%, 87.14%, and 93.74% for the 5 used datasets, respectively.

In our work, we exploit two scenarios, before and after data augmentation. For the original dataset, the OA attained is about 100%, 95%, 94.41%, 86.08%, and 92.09% for the 5 proposed datasets, respectively. Due to the poor result, we use several techniques for data augmentation, which are mentioned above. The new OA achieved is about 100%, 97.22%, 97.02%, 88.86%, and 95.72% respectively

8 Discussion

In this paper, we have suggested a new deep CNN model for MRI brain tumor classification. Our model exploits various layers with different sizes and Softmax classifier. The experimental study of the proposed technique is carried out based on three public datasets, as discussed earlier.

The first brain tumor dataset is Figshare, which is a free and available dataset. It contains 3064 images correspond to meningioma, glioma, and pituitary tumors. The second exploited dataset is Radiopaedia, which is used as an extra validation to affirm that our suggested model can also classify the tumor grade with acceptable results. The original data contain 121 MRI images correspond to 4 grades. Due to the small data volume, we use several data augmentation techniques in order to increase the data set size which can enhance the model accuracy. We have employed rotation in different angles, Flipping, Gaussian Blur, and Sharpen with several parameters. The new dataset contains 2178 images. We adapted the proposed model for this dataset by modifying the last FC layer to have 4 neurons, which correspond to the number of grades. The classification results are analyzed, and various evaluation metrics are calculated as accuracy, sensitivity, specificity, precision, and F1-score.

In general, only the accuracy is exploited in the majority of the previous works to evaluate the proposed scheme. However, the use of the accuracy only for comparison can be misleading since it ignores the sensitivity to imbalanced data. In this case, some class performances can become better than others.

In this work, we use various measures in the aim to gain a clear indicator of the generalization of the proposed model. As shown in Table 14, the suggested Deep CNN provides good accuracy for the first data set, which reaches 95.23% for Meningioma, 95.43% for Glioma, and 98.43% for Pituitary tumor. The OA is 94.74%. Furthermore, the obtained results are compared with some new previous works. As detailed in Table 15, it is clear that our model achieved higher accuracy and outperformed the various previous works.

The second dataset is tested based on two scenarios: before and after augmentation. From Table 16, the original dataset attains 95.61%, 92.98%, 93.85%, and 98.24% as accuracy for grade I, grade II, grade III, and grade VI, respectively. The OA gained is about 90.35%. From Table 17, the augmented dataset achieves 96.32%, 95.31%, 96.18%, and 99.61% as accuracy for the 4 grades, respectively. The OA is about 93.71%. A comparison with some previous works is detailed in Table 18. It is clear that the proposed model achieved higher accuracy and outperformed the various previous works.

In the last dataset, we exploit two scenarios: before and after augmentation for the 5 sub-datasets. From the experiment results obtained, the OA gained before data augmentation is about 100%, 95%, 94.41%, 86.08%, and 92%, respectively. The augmented datasets achieve 100%, 97.22%, 97.02%, 88.86% and 95.72 as OA respectively. A comparison with some previous works that exploit the same datasets is detailed in Table 29. It is clear that

our model gained the highest accuracy and outperformed all the previous works. Despite the variety of grade, classes, position, and intensity for the used MRI, the suggested model generates promising results.

Based on the obtained results, the proposed model gained the highest accuracy in MRI brain tumor classification using the three datasets. It can generate promising results either for a small dataset. However, the MRI brain images classification represents a challenging problem owing to the variety in intensities, size, orientation, and shape. In addition to image contrast and the noise perturbations. Furthermore, the medical datasets exploited are often limited in size and hard to access. Several previous works [27, 29, 48] used the traditional machine learning technique, which needs manual feature extraction that is a tedious task and time consuming. Some other works that used CNN [33, 37] have exploited shallow networks with a few MRI data, which lead to lower accuracy value. In this study, the choice of architecture and the hyper-parameters for the brain tumor classification has proven to process the MRI images effectively and attain higher accuracy results.

In our opinion, the suggested scheme is effective to classify MRI brain tumors as type or grade, which help doctors to take the precise decision in a short time. We believe also that our model can be exploited to classify other tumor types as breast cancer, lung cancer, liver cancer... and other medical image types as ultrasound, X-ray.

9 Conclusion and Future Work

Our paper presents an innovative model for multi brain tumors classification based on CNN. It is an automatic system, which requires a minimum of pre-processing. The model was tested on three brain tumor datasets. Various performance metrics were studied to evaluate model accuracy and ascertain the robustness of the system. The suggested model recorded the best classification accuracies compared to previous related works on the same dataset. The experiment results show the effectiveness of this model despite the smaller amount of training data. The suggested approach can be used for other MRI classification since it requires minimum preprocessing and does not use handcrafted features. As future work, we attend to exploit images from different modalities as T1, T2, and Flair, which aim to augment the dataset size and added robustness to our scheme.

References

1. Razzak MI, Imran M, Xu G (2018) Efficient brain tumor segmentation with multiscale two-pathway-group conventional neural networks. *IEEE J Biomed Health Inform* 23(5):1911–1919
2. Siegel RL, Miller KD (2017) Jemal ACancer statistics, 2017. *CA Cancer J Clin* 67(1):7–30
3. Zhang Y, Li A, Peng C, Wang M (2016) Improve glioblastoma multiforme prognosis prediction by using feature selection and multiple kernel learning. *IEEE/ACM Trans Comput Biol Bioinform* 13(5):825–835
4. Yang Y, Yan LF, Zhang X, Han Y, Nan HY, Hu YC, Ge XW (2018) Glioma grading on conventional MR images: a deep learning study with transfer learning. *Front Neurosci* 12:804
5. Talo M, Baloglu UB, Yıldırım Ö, Acharya UR (2019) Application of deep transfer learning for automated brain abnormality classification using MR images. *Cogn Syst Res* 54:176–188
6. Kaur T, Saini BS, Gupta S (2017) Quantitative metric for MR brain tumour grade classification using sample space density measure of analytic intrinsic mode function representation. *IET Image Process* 11(8):620–632
7. Usman K, Rajpoot K (2017) Brain tumor classification from multi-modality MRI using wavelets and machine learning. *Pattern Anal Appl* 20(3):871–881

8. Anitha V, Murugavalli SJICV (2016) Brain tumour classification using two-tier classifier with adaptive segmentation technique. *IET Comput Vis* 10(1):9–17
9. El-Dahshan ESA, Mohsen HM, Revett K, Salem ABM (2014) Computer-aided diagnosis of human brain tumor through MRI: a survey and a new algorithm. *Expert Syst Appl* 41(11):5526–5545
10. Hemanth DJ, Anitha J, Naaji A, Geman O, Popescu DE (2018) A modified deep convolutional neural network for abnormal brain image classification. *IEEE Access* 7:4275–4283
11. Liu Y, Stojadinovic S, Hrycushko B, Wardak Z, Lau S, Lu W, Nedzi L (2017) A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery. *PLoS ONE* 12(10):e0185844
12. Qiu Y, Yan S, Gundreddy RR, Wang Y, Cheng S, Liu H, Zheng B (2017) A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology. *J Xray Sci Technol* 25(5):751–763
13. Sutskever I, Martens J, Hinton G (2011) Generating text with recurrent neural networks. In: *Proceedings of the 28th international conference on machine learning*, Bellevue, pp 1017–1024
14. Taigman Y, Yang M, Ranzato MA, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1701–1708
15. Karpathy A, Fei-Fei L (2015) Deep visual-semantic alignments for generating image descriptions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 3128–3137
16. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, Dieleman S (2016) Mastering the game of Go with deep neural networks and tree search. *Nature* 529(7587):484–489
17. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Berg AC (2015) Imagenet large scale visual recognition challenge. *IJCV* 115(3):211–252
18. Yousefi M, Krzyżak A, Suen CY (2018) Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning. *Comput Biol Med* 96:283–293
19. Gu Y, Lu X, Yang L, Zhang B, Yu D, Zhao Y, Zhou T (2018) Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs. *Comput Biol Med* 103:220–231
20. Charron O, Lallement A, Jarnet D, Noblet V, Clavier JB, Meyer P (2018) Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network. *Comput Biol Med* 95:43–54
21. Shao L, Zhu F, Li X (2014) Transfer learning for visual categorization: a survey. *IEEE Trans Neural Netw Learn Syst* 26(5):1019–1034
22. Kleesiek J, Urban G, Hubert A, Schwarz D, Maier-Hein K, Bendszus M, Biller A (2016) Deep MRI brain extraction: a 3D convolutional neural network for skull stripping. *NeuroImage* 129:460–469
23. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2016) Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging* 35(5):1299–1312
24. Chen H, Qi X, Cheng JZ, Heng PA (2016) Deep contextual networks for neuronal structure segmentation. In: *Proceedings of the 13th AAAI conference on artificial intelligence*, pp 1167–1173
25. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, Sánchez CI (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88
26. Deniz E, Şengür A, Kadiroğlu Z, Guo Y, Bajaj V, Budak Ü (2018) Transfer learning based histopathologic image classification for breast cancer detection. *Health Inf Syst Syst* 6(1):18
27. Cheng J, Huang W, Cao S, Yang R, Yang W, Yun Z, Feng Q (2015) Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PLoS ONE* 10(10):e0140381
28. Ismael MR, Abdel-Qader I (2018) Brain tumor classification via statistical features and back-propagation neural network. In: *2018 IEEE international conference on electro/information technology (EIT)*, pp 0252–0257
29. Tahir B, Iqbal S, Usman Ghani Khan M, Saba T, Mehmood Z, Anjum A, Mahmood T (2019) Feature enhancement framework for brain tumor segmentation and classification. *Microsc Res Tech* 82(6):803–811
30. Yu J, Tan M, Zhang H, Tao D, Rui Y (2019) Hierarchical deep click feature prediction for fine-grained image recognition. *IEEE Trans Pattern Anal Mach Intell*
31. Yu J, Li J, Yu Z, Huang Q (2019) Multimodal transformer with multi-view visual representation for image captioning. *IEEE Trans Circuits Syst Video Technol* 30(12):4467–4480
32. Yu J, Yao J, Zhang J, Yu Z, Tao D (2020) SPRNet: single-pixel reconstruction for one-stage instance segmentation. *IEEE Trans Cybern.* <https://doi.org/10.1109/TCYB.2020.2969046>
33. Paul JS, Plassard AJ, Landman BA, Fabbri D (2017) Deep learning for brain tumor classification. In: *Proc. SPIE*, vol 10137. pp 1–16

34. Afshar P, Mohammadi A, Plataniotis KN (2018) Brain tumor type classification via capsule networks. In: 2018 25th IEEE international conference on image processing (ICIP), pp 3129–3133
35. Zhou Y, Li Z, Zhu H, Chen C, Gao M, Xu K, Xu J (2018) Holistic brain tumor screening and classification based on densenet and recurrent neural network. In: International MICCAI Brainlesion workshop. Springer, pp 208–217
36. Pashaei A, Sajedi H, Jazayeri N (2018) Brain tumor classification via convolutional neural network and extreme learning machines. In: 2018 8th International conference on computer and knowledge engineering (ICCKE), pp 314–319
37. Abiwinanda N, Hanif M, Hesaputra ST, Handayani A, Mengko TR (2019) Brain tumor classification using convolutional neural network. In: World congress on medical physics and biomedical engineering 2018. Springer, Singapore, pp 183–189
38. Ghassemi N, Shoeibi A, Rouhani M (2020) Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images. *Biomed Signal Process Control* 57:101678
39. Wu B, Liu Y, Lang B, Huang L (2018) Dgcnn: disordered graph convolutional neural network based on the gaussian mixture model. *Neurocomputing* 321:346–356
40. Shi H, Zhang Y, Zhang Z, Ma N, Zhao X, Gao Y, Sun J (2018) Hypergraph-induced convolutional networks for visual classification. *IEEE Trans Neural Netw Learn Syst* 30(10):2963–2972
41. Fu S, Liu W, Tao D, Zhou Y, Nie L (2020) HesGCN: Hessian graph convolutional networks for semi-supervised classification. *Inf Sci* 514:484–498
42. Fu S, Liu W, Zhou Y, Nie L (2019) HpLapGCN: hypergraph p-Laplacian graph convolutional networks. *Neurocomputing* 362:166–174
43. Khan N, Chaudhuri U, Banerjee B, Chaudhuri S (2019) Graph convolutional network for multi-label VHR remote sensing scene recognition. *Neurocomputing* 357:36–46
44. Sichao F, Weifeng L, Shuying L, Yicong Z (2019) Two-order graph convolutional networks for semi-supervised classification. *IET Image Process* 13(14):2763–2771
45. Hemanth DJ, Vijila CKS, Selvakumar AI, Anitha J (2014) Performance improved iteration-free artificial neural networks for abnormal magnetic resonance brain image classification. *Neurocomputing* 130:98–107
46. Liu YH, Muftah M, Das T, Bai L, Robson K, Auer D (2012) Classification of MR tumor images based on Gabor wavelet analysis. *J Med Biol Eng* 32(1):22–28
47. Sajjad M, Khan S, Muhammad K, Wu W, Ullah A, Baik SW (2019) Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *J Comput Sci* 30:174–182
48. Sachdeva J, Kumar V, Gupta I, Khandelwal N, Ahuja CK (2016) A package-SFERCB-“Segmentation, feature extraction, reduction and classification analysis by both SVM and ANN for brain tumors”. *Appl Soft Comput* 47:151–167
49. Jun Cheng (2017) Brain Tumor Dataset (Version 5). Retrieved from <https://doi.org/10.6084/m9.figshare.1512427.v5>
50. Pinaya WH, Gadelha A, Doyle OM, Noto C, Zugman A, Cordeiro Q, Sato JR (2016) Using deep belief network modelling to characterize differences in brain morphometry in schizophrenia. *Sci Rep* 6:38897
51. Padole H, Joshi SD, Gandhi TK (2020) Graph wavelet-based multilevel graph coarsening and its application in graph-CNN for alzheimer’s disease detection. *IEEE Access* 8:60906–60917
52. Song TA, Chowdhury SR, Yang F, Jacobs H, El Fakhri G, Li Q, Dutta J (2019) Graph convolutional neural networks for alzheimer’s disease classification. In: Proceedings IEEE 16th Int. Symp. Biomed. Imag (ISBI), pp 414–417
53. Guo J, Qiu W, Li X, Zhao X, Guo N, Li Q (2019) Predicting alzheimer’s disease by hierarchical graph convolution from positron emission tomography imaging. In: 2019 IEEE international conference on big data (big data), pp 5359–5363
54. O’Shea K, Nash R (2015) An introduction to convolutional neural networks. CoRR [arXiv:1511.08458](https://arxiv.org/abs/1511.08458)
55. Ayadi W, Elhamzi W, Charfi I, Atri M (2019) A hybrid feature extraction approach for brain MRI classification based on Bag-of-words. *Biomed Signal Process Control* 48:144–152
56. Shang R, He J, Wang J, Xu K, Jiao L, Stolk R (2020) Dense connection and depthwise separable convolution based CNN for polarimetric SAR image classification. *Knowl. Based Syst* 105542
57. Yang C, Hou B, Ren B, Hu Y, Jiao L (2019) CNN-based polarimetric decomposition feature selection for PolSAR image classification. *IEEE Trans Geosci Remote Sens* 57(11):8796–8812
58. Li B, Zhang H, Luo H, Tan S (2019) Detecting double JPEG compression and its related anti-forensic operations with CNN. *Multimed Tools Appl* 78(7):8577–8601
59. Lei X, Pan H, Huang X (2019) A dilated CNN model for image classification. *IEEE Access* 7:124087–124095

60. Li G, Li N (2019) Customs classification for cross-border e-commerce based on text-image adaptive convolutional neural network. *Electron Commer Res* 19(4):779–800
61. Sundararajan SK, Sankaragomathi B, Priya DS (2019) Deep belief CNN feature representation based content based image retrieval for medical images. *Med Syst* 43(6):174
62. Wu XY (2019) A hand gesture recognition algorithm based on DC-CNN. *Multimed Tools Appl* 79(13–14):9193–9205
63. Swati ZNK, Zhao Q, Kabir M, Ali F, Ali Z, Ahmed S, Lu J (2019) Content-based brain tumor retrieval for MR images using transfer learning. *IEEE Access* 7:17809–17822
64. Huang M, Yang W, Wu Y, Jiang J, Gao Y, Chen Y, Lu Z (2014) Content-based image retrieval using spatial layout information in brain tumor T1-weighted contrast-enhanced MR images. *PLoS ONE* 9(7):e102754
65. Radiopaedia. <https://radiopaedia.org/>
66. Zhang YD, Dong Z, Chen X, Jia W, Du S, Muhammad K, Wang SH (2019) Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation. *Multimed Tools Appl* 78(3):3613–3632
67. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Tarbox L (2013) The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging* 26(6):1045–1057
68. Goodfellow I, Bengio Y, Courville A (2016) *Deep learning*. MIT Press, Cambridge
69. Hossin M, Sulaiman MN (2015) A review on evaluation metrics for data classification evaluations. *Int J Data Min Knowl Manag Process* 5(2):01–11
70. Afshar P, Plataniotis KN, Mohammadi A (2019) Capsule networks for brain tumor classification based on MRI images and coarse tumor boundaries. In: *ICASSP 2019–2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp 1368–1372
71. Anaraki AK, Ayati M, Kazemi F (2019) Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *Biocybern Biomed Eng* 39(1):63–74
72. Tandel GS, Balestrieri A, Ujaray T, Khanna NN, Saba L, Suri JS (2020) Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm. *Comput Biol Med* 122:103804

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.