



Impact of the data augmentation on the detection of brain tumor from MRI images based on CNN and pretrained models

Samir Benbakreti¹ · Mohamed Benouis² · Ahmed Roumane¹ · Soumia Benbakreti³

Received: 10 November 2022 / Revised: 11 August 2023 / Accepted: 18 September 2023 /

Published online: 2 October 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Deep Learning has significantly push forward the research on cancer magnetic resonance imaging (MRI) images. These images are widely used to diagnose the presence of a deformed tissue within the brain in which the cells replicate indefinitely without control, i.e. a brain tumor. Radiologist have to deeply examine a set of MRI images for each patient in order to decide whether the tumor is benign (noncancerous) or malignant (cancerous). The latest have very severe consequences and have a very high mortality rate, but this could be significantly reduced if the cancer is diagnosed at an earlier stage. The classification task is very complicated due to neurological and radiological similarities of different tumors. In order to assist the radiologists, our objective in this paper is to achieve a correct classification of the MRI images. The studied deep classification models have been trained over three types of tumors: meningioma, glioma and pituitary tumor, on sagittal, coronal and axial views in addition to MRI of normal patients. The proposed model consists of combining a set of several classifiers that uses the features extracted by a convolutional neural network (CNN). We will, also, explore the impact of data augmentation and image resolution on the classification performance with the goal of obtaining the best possible accuracy. We used a CNN architecture and several pre-trained models. The model ResNet 18 gave the highest accuracy of 95.7%.

Keywords Deep learning · Image classification · Brain tumor · MRI images · CNN · Pretrained models

1 Introduction

According to the World Health Organization (WHO), Cancer is a leading cause of death worldwide, accounting for nearly 10 million deaths in 2020, or nearly one in six deaths [1]. But an effective treatment is possible to many cancers if detected early. Early detection has to parts: early diagnosis and screening. The diagnostics of a cancer is a very complicated task that requires precise differentiation of symptoms, infrastructure and coordination between medical services. Brain tumor results from cells that grow uncontrollably in the brain to create

an abnormal mass of tissue (tumor). There are two different types: benign and malignant. Benign tumors, which have a slow growth rate, are easily separated from brain tissue. Malignant brain tumors grow faster and can spread, invade and damage nearby brain tissue [2].

Being the most important organ in the human body, the brain, if affected by tumor is very deadly, this will significantly impact the quality of the whole life, and will eventually change everything for a patient and its surrounding family and friends. This tumor interferes with brain functions such as memory, muscle control and other bodily operations [3, 4]. Depending on the type of tumor and its location in the brain, symptoms may include seizures, memory problems, unusual behavior, confusion, vision issues, and difficulty with balance [5, 6]. Brain tumors are the main cause of deadly cancers among children under 14 years, it is estimated that about 4,170 of them will also be diagnosed with a brain tumor this year in the United States [7].

In practice, several brain tumors have been documented, but the two main groups are: primary (tumors originating from brain tissue) and metastatic (formed by the splash of cancerous cells from other part of the body) also called secondary.

Despite being primary or metastatic, brain tumors have several types. Meningioma is the most prevalent (almost benign) tumor, making up 36.8% of all tumors, while Glioma [8] is the most prevalent cancer (malignant), making up 75% of central nervous system cancers, with an annual incidence of six cases per 100,000 persons. Glioma is categorized in stages (I to IV): Pilocytic Astrocytoma is the least aggressive (stage I), low-grade Astrocytoma in (the stage II), Anaplastic Astrocytoma is the stage III and Glioblastoma is the most dangerous form of stage IV tumor [9, 10].

Understanding brain tumor types is crucial to prevent and carry out the correct healing procedure. For this purpose, MRI is currently the preferred means for early detection of human brain tumors because it is non-invasive, painless [11] and generates more detailed images than computed tomography (CT) [12]. However, the interpretation of MRI is mostly centered on the opinion of radiologists.

Early diagnosis of brain tumors can change the course of the disease and save lives. Using computers in this task may produce automatic, fast and accurate results. For this reason, computer-aided diagnostic (CAD) systems are frequently used today. CAD systems analyze complex relationships in medical data and assist the physician in decision making [13–17].

For about twenty years, machine learning techniques such as random forests and support vector machines (SVMs) were also used for brain tumor classification. Recently, deep neural networks have been developed and have shown great promise in accurately classifying brain tumors and discovering new image features. We have taken the CNN model to try to achieve the best possible results, and on the other hand, to see the impact of reducing the resolution of the training images and data augmentation on the classification results. Later, we will compare the CNN architecture with pre-trained models, which have proven to be much more effective than CNN. The details of the different experiments performed are mentioned in Sections 4 and 5.

2 Related works

In [18], by comparing several segmentation strategies, the authors of this work aim to enhance the performance of tumor identification. Additionally, genetic algorithms are applied for automatic tumor stage categorization to increase classification accuracy. The extraction of the pertinent features and the area calculation serve as the foundation for

the categorization step's choice. Based on the segmentation score, precision, sensitivity, and specificity, the experimental findings of the proposed technique are assessed and validated for the examination of performance and quality on brain magnetic resonance images (MRI). The main limitation of this work is the very limited number of images in the dataset. In addition, all MRI scans are axial, which significantly limits the diagnosis of cancer on other scans (sagittal or coronal).

In this article, Rahul Singh et al. [19] proposed an automated computer-assisted network for the diagnosis of the HGG and LGG types of MR brain tumors. For classifying brain tumors, the classifier is based on a Gabor modulated convolutional filter. The suggested architecture can learn comparatively smaller feature maps, which reduces the need for network parameters, thanks to the incorporation of Gabor's filter dynamics. To measure performance, the authors added hop connections to networks that had already been trained. Unfortunately, the authors compare the proposed method with works using pre-trained models on non-medical image recognition tasks, which makes the comparison inconsistent.

In [20], the research focuses the segmentation of brain anatomy, including tumors, in medical imaging to assist neurologists in making an accurate diagnosis. Because of the complexity of these structures, automatic segmentation techniques—often created by MRI—must be used. By combining three different techniques, the goal of this study is to create and implement an automatic system for the identification and localization of tumor sites. The Random Forest algorithm is used to first identify the region of interest, or the pixels that belong to the tumor, while the remaining pixels in the image are assumed to belong to the background. As a result, the Graph Cut approach automatically obtains tumor and background seeds for segmentation. The starting contour for the level set (LVS) segmentation can be obtained thanks to this segmentation. The proposed method for brain tumor segmentation presented in this study combines three different techniques to mitigate the disadvantages of one, while utilizing the advantages of the other. It would be very interesting to combine this method with a classifier as suggested by the authors in the end.

Radiologists' diagnostic abilities are improved by computer-aided diagnostic (CAD) systems, which also speed up the process of making a correct diagnosis. The study of [21] makes a suggestion for an intelligent computer-assisted method for automatically identifying brain cancers from MRI images. The suggested technique employs the following computation strategies: Grayscale co-occurrence matrices (GLCM) and the Support Vector Machine (SVM) are used to extract features from this segmented portion of the brain tumor and to classify the input MRI images into normal and abnormal. K-means clustering is used to separate the brain tumor from other parts of the brain. The complete project is completed using 64 images, of which 42 show a brain tumor (benign and malignant) and 22 are normal images. Applied to the Brain Tumor Segmentation (BraTS) 2015 datasets [22], this method's overall classification accuracy is 99.28%.

In the paper [23], a radial basis function neural network (RBFNN)-based intelligent diagnostic technique for the early detection of brain cancers is proposed. The developed method consists of four primary modules: segmentation, feature extraction, classification and learning modules. The tumor region is segmented using the Grab cut method in the segmentation module. A convolutional neural network (ConvNet) is employed in the feature extraction module to extract fresh deep features from segmented images. The classification module feeds RBFNN with the extracted deep features. The RBFNN's learning algorithm significantly affects how well the network performs. As a result, the learning module used a novel learning algorithm based on the bee's algorithm (BA) [24]. The proposed method was applied to the (BraTS) 2015 datasets [22], and the acquired findings demonstrated its efficacy and suitability for use in computer-aided brain tumor detection

systems. That being said, RBF networks are known to be sensitive to noisy data, which can be fatal for classification. It would have been wise to perform pre-processing on the input images to better generalize.

The challenge of having a limited number of samples in a dataset is of utmost importance when working with deep learning. To tackle this problem, numerous techniques have been proposed and discussed in the literature. Below, we highlight some noteworthy examples that have effectively addressed this issue. Although the areas of application may vary, the works listed below can provide inspiration for addressing complex problems in the field of medical imaging.

Authors in [25] treated the few-shot learning problem in the recognition of flare soot density with the meta-learning technics which is used to extract the intrinsic features from a bunch of similar tasks (e.g., dark clouds, cigarette smoke, and vapor). Then the obtained weights are used as initial values to train the network in the main task i.e. soot density recognition. Using this approach with ensemble learning lead to a strong generalization capability of the proposed model.

In [26] Authors observed that a rise of PM_{2.5} concentration has a consequence of lowering the information amount in a photographic image. Thus, information abundance “IA” measure has been used as an indicator to monitor the concentration of PM_{2.5} in the atmosphere in megacities. The extracted IA features serve as input to the Wide and Deep Learning “WD” networks that are trained separately. Finally, the whole IAWD model is finetuned to find the best mapping between the IA features and the PM_{2.5} concentration.

Ke Gu et al. [27] discuss the difficulties in detecting smoke from images and propose a novel deep neural network with a dual-channel (DCNN) for this purpose. The DCNN consists of two sub-networks that complement each other: one sub-network is proficient at extracting detailed smoke features such as texture, while the other sub-network can capture basic smoke features like contours. The DCNN also incorporates two critical components, a skip connection to enhance feature propagation and global average pooling to mitigate the overfitting problem. By using augmented data obtained through rotating training images, the proposed DCNN can rapidly and steadily converge towards perfect performance.

3 Description of the dataset and data augmentation

The dataset consist of 2D slices collected from 233 patients between 2005 and 2010 at Tianjing Medical University and Nanfang Hospital in Guangzhou, China [28]. The 3264 slices are labeled using four classes, namely: Meningiomas (937 slices), Gliomas (926 images), Pituitary tumors (901 slices) and No-tumor. (See Table 1)

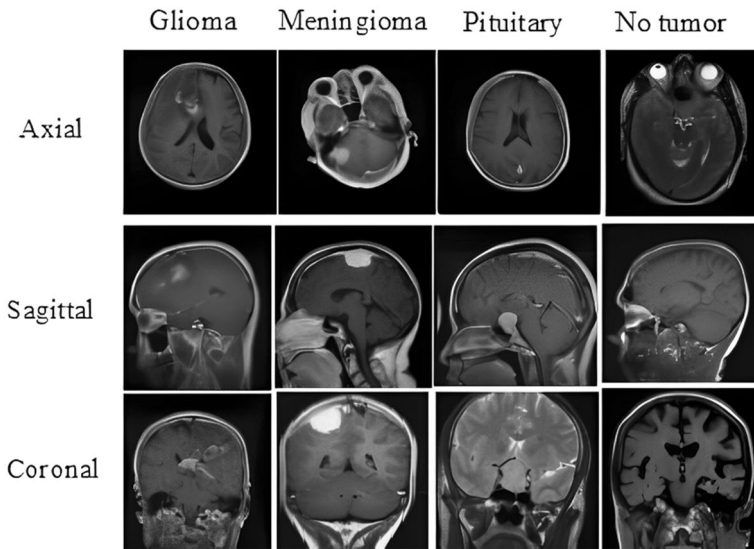
Image are taken from different perspectives (sagittal, coronal, and axial). A sample of these three categories of tumors are shown in Fig. 1.

The MRI slice thickness is 6 mm with a slice gap of 1 mm. The image resolution is 512 * 512 pixels. A team of three expert radiologists manually marked the tumor boundaries. Every slice in the dataset is associated to a data structure that includes the patient’s PID, the tumor type label: 1 for Meningioma, 2 for Glioma, and 3 for Pituitary tumor, as well as the coordinates vector of the points that make up the tumor border and the tumor mask.

The training was performed using Matlab 2021 installed on a work station with 64-bit operating system, windows 10 Pro, 24 GB of Random Access Memory (RAM), with an Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40 GHz and Graphical Processing Unit

Table 1 Description of the dataset

Class	Number of images
Glioma	926
Meningioma	937
Pituitary Tumor	901
No tumor	500

**Fig. 1** Four categories of tumors

(GPU). 80% of the datasets are used for training and 20% for testing (evaluating the model performance).

Additionally, we employed a reflection-based data augmentation technique (Mirror effect) for the sake of preventing overfitting. In fact, by using the reflection effect and keeping the labels, we will increase the quantity of images intended for training. The two procedures are as follows:

- Vertical mirror image: Each image is flipped vertically, where x and y are the original positions of each pixel in the image. Then x' and y' are the new positions of each pixel after reflection around the X axis, as shown in Eq. 1:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (1)$$

- Horizontal mirror image: Each image is flipped horizontally, where x and y are the original positions of each pixel in the image. Then x' and y' are the new positions of each pixel after reflection around the Y axis, as shown in Eq. 2:

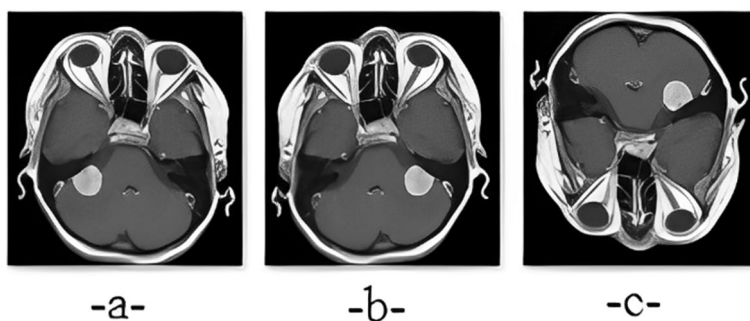


Fig. 2 Data augmentation reflection (DAR) results: **a** Original image, **b** horizontal mirror image, **c** Vertical mirror image

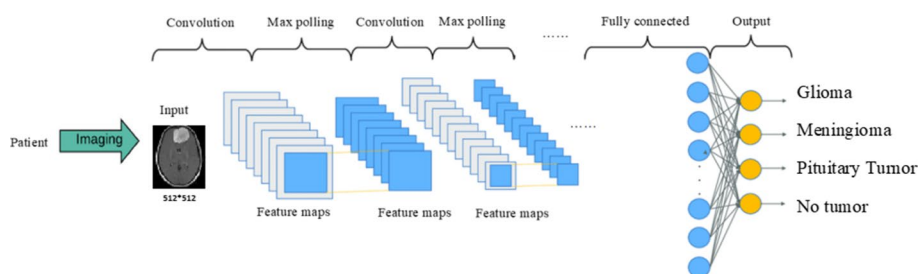


Fig. 3 The proposed model with different algorithms (CNN Vs Pretrained models)

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (2)$$

Figure 2 highlights the result of using the mirror effect (horizontal and vertical) on an original image of a meningioma.

By using the data augmentation technique, the database is passed from 2612 to 7836 images, i.e. we triple its size. The model becomes more robust as the training set grows.

4 Proposed model

In this study, we analyzed the different techniques for detection of brain tumor from MRI images (as mentioned in Fig. 3), then examined CNN's architecture that is based on research on the visual cortex of the cat by Hubel and Wiesel [29], and different pre-trained models: AlexNet, ResNet18, GoogleNet, Mobilenet, Densenet, VGG19 and YOLOv3 in order to see the variation of answers in our work.

The pre-trained model has a specific architecture and provided a simpler approach to quickly retrain neural networks on selected datasets with high accuracy (all the pretrained models architectures are detailed in Section 5.2). The choice of a pre-trained model mainly depends on the specific task to be performed and the nature of the data to be processed. In this study, it is clear that the goal is to perform classification of medical images; therefore

it is more appropriate to use pre-trained models on a similar dataset, such as ImageNet, which is the case for all the models used. The exception is YOLOv3, which was trained on several datasets, including COCO (Common Objects in Context) [30] and PASCAL VOC (Visual Object Classes) [31]. Because although it was originally designed for object detection (through segmentation), this method, which is also based on CNN, can be used for classification purposes.

5 Experiements, results and discussion model

5.1 Experiment 1: the proposed CNN

The structure of our CNN model as shows in Table 2 includes several layers. The input is MRI image of size 512×512 pixels and is used to feed six convolution layers : with Conv 1 consist of 32 feature map, C2 consist of 64 feature map, C3 consist of 128 feature map. ., As shown in Table 2 six pooling layers are used, Layer S1, S2. . and S6 are subsampling layers, whose number of feature maps is equal to the maps number of their previous convolution. After the pre-processing layers, comes to the fully connected layers FC with 4 neurons, each one connects to all the maps of last pooling layer followed by the output layer is also a full connected layer.

The trained parameters used in this model are in the options side where all the hyperparameters used were defined including the number of epochs used (1 or 5), the mini batch size (64), the learning rate is 0.001 and frequency validation is 20. The given CNN was trained using different parameters to test the accuracy for this model. The training parameters of the CNN model are mentioned in the Table 3.

When:

- SGDM: the stochastic gradient descent with momentum Solver [32].
- Momentum is the contribution to the current iteration of stochastic gradient descent with momentum of the parameter update stage of the previous iteration.

Table 2 The architecture of the CNN model

Name	Type	Size
Input layer	Input data	512*512
Conv1	Convolution + Relu	32*32*8
S1	Max pooling	3,2
Conv2	Convolution + Relu	64*64*3
S2	Max pooling	3,2
Conv3	Convolution + Relu	128*128*5
S3	Max pooling	3,2
Conv4	Convolution + Relu	256*256*5
S4	Max pooling	3,2
Conv5	Convolution + Relu	512*512*5
S5	Max pooling	3,2
Conv6	Convolution + Relu	1024*1024*5
S6	Max pooling	3,2
Fc	Fully connected	1 Fc(4)

- Mini Batch size: at each iteration the stochastic gradient descent algorithm evaluates the gradient and updates the parameters using a subset of the training data. A different subset, called a mini-batch, is use at each iteration to evaluate the gradient of loss function and update the weights.
- The Validation Frequency value is the number of iterations between evaluations of validation metrics.
- Learn Rate Drop factor: To avoid oscillations and divergence in the optimization process, a technique of gradually reducing the learning rate by a given factor is used.

To evaluate the performance of the trained models, we used the accuracy parameter. Accuracy is termed as the ratio of correctly classified images over the total number of images. The following formula 3 is utilized:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Where:

TP or True Positive	is the number of pathology images perfectly identified.
True Negatives or TN	is the number of non-pathological images perfectly detected.
False Positive or FP	is the number of images detected wrongly as positive, which are actually non-pathological images.
False Negative or FN	is the number of images detected wrongly as non-pathological, which are actually pathology images.

The results are mentioned in Table 4.

In order to evaluate the impact of data augmentation with reflection (DAR), we compared the accuracy with and without DAR. The first top accuracy after training the model using 11 epochs was 68.9% (Table 4) for our 4 classes classification. As we train the model using only five epochs did not give a good result, we had to increase the number of epochs and see the performance of our model and the best results for our model gave us 89.9% without DAR and 9 epochs and 79.75% with DAR and 7 epochs. In Tables 4 and 5, the result mentioned in bold corresponds to the best result obtained.

Additionally, we observe that the data augmentation using reflection does not gives a better results, on the contrary it deteriorates the results. We would rather wait before making a firm conclusion.

Before performing the same experiment, we done a pre-processing that involved reduction the image resolution (from 512*512 to 299*299). The results are shown in Table 5.

Table 3 The training parameters of the CNN model

Parameter	Value
Initial learning rate	0.001
Optimizer	SGDM
Momentum	0.9
Max epoch	5 to 11
Mini batch size	64
Validation frequency	20
Learn Rate Drop Factor	0.2

Table 4 The results of the CNN model (Image input 512*512)

The CNN model		5 epochs	7 epochs	9 epochs	11 epochs
Without DAR technique	Accuracy	68.9%	79.8%	89.9%	82.18%
	Time GPU execution	211 min 27 s	291 min 21 s	374 min	455 min 12 s
With DAR technique	Accuracy	59.2%	79.75%	66.8%	76.53%
	Time GPU execution	211 min 01s	310 min 10s	395 min 37s	460 min 22s

The best results were obtained after 9 epochs, with 89.3% for the DAR technique and 90.2% without it. We note once again that this data augmentation technique does not improve the recognition rate (see the Figs. 4 and 5). However, reducing the size of the images (299*299 instead of 512*512) significantly improves performance (see Tables 4 and 5). In addition, a huge gain is noted in terms of processing time. That is why we will set the input image size to 299*299 from now on.

We have observed that some data augmentation technic can decrease the accuracy. This can happen if the augmentation removes relationships (even if false) in the data. Another explanation is that a reflected image can be invalid in a given context and can corrupt the training process by providing the network with invalid inputs that can lead to incorrect models being learned.

Secondly, the background of the image can play a role in improving classification when compressing the image. Images with a lot of background information can make the classification task more difficult for the neural network. By reducing the amount of background information, the network can focus on the important features of the image, which can improve performance. However, this also depends on the amount of important information preserved during compression and the nature of the image being processed.

We believe that decreasing the resolution can have varying effects on the image classification, depending on the image's characteristics. If the area of interest is concentrated in a specific location (such as a tumor), then the impact of reducing the resolution may be positive. However, if the area of interest is scattered throughout the image (example of recognition of several objects of different sizes in the same image), then the classification accuracy may decline.

So, we observe that while increasing the number of epochs the execution time rises up, without enhancement in the accuracy. Reason why we did not go beyond 11 epochs.

The use of accuracy alone may be insufficient to properly evaluate the performance of a model, especially when the classes are imbalanced. For the selected resolution which is

Table 5 The results of the CNN model (Image input 299*299)

The CNN model		5 epochs	7 epochs	9 epochs	11 epochs
Without DAR technique	Accuracy	89.3%	90%	90.2%	89.2%
	Time GPU execution	64 min 13s	88 min 36s	113 min 31s	138 min 11s
With DAR technique	Accuracy	84.8%	87.8%	89.3%	88.4%
	Time GPU execution	90.3 min 31 s	90 min 9s	111 min 10 s	138 min 34s

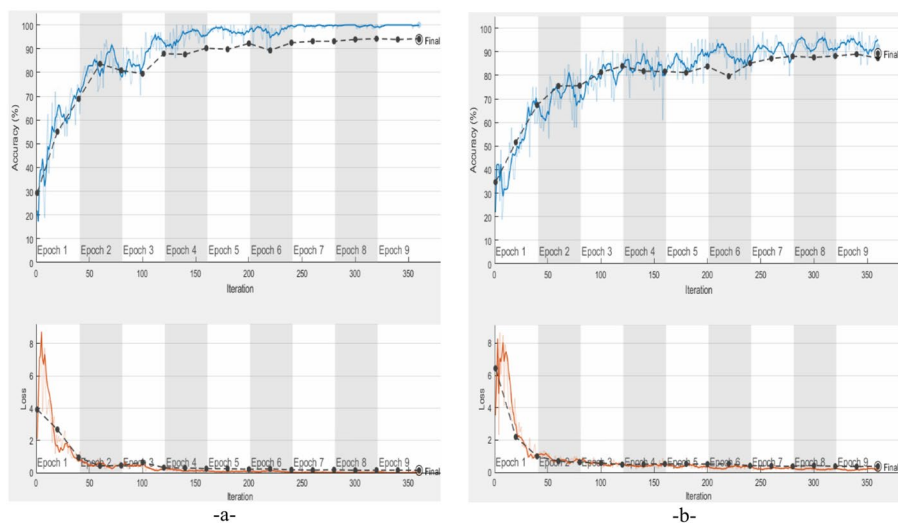


Fig. 4 The results of CNN model after 9 epochs: **a** Without DAR, **b** With DAR

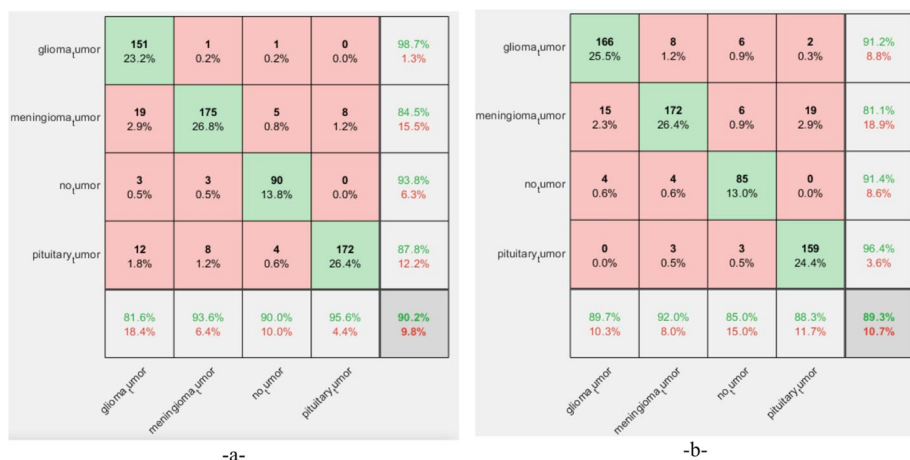


Fig. 5 Confusion matrix for CNN model with 9 epochs: **a** With DAR, **b** Without DAR

299*299 (best results), additional measures such as precision, recall and F1score can provide more detailed information on the model's performance for each class. Equations (4), (5) and (6) show how to calculate precision, recall and F1score respectively:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

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

$$F1score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (6)$$

We will apply these last metrics on the best results obtained previously (see Table 6). The goal of this is to have a clear indicator of the generalization of the model on the unbalanced brain tumors MRI dataset. Since accuracy is influenced by class imbalance, precision, recall, and F1 score help indicate the overall performance of the model, regardless of the number of individual classes.

5.2 Experiment 2: the proposed pretrained models

5.2.1 AlexNet

AlexNet is an 8-layer convolutional neural network [33], consisting of 5 convolutional layers, with convolutional filter size 3×3 and 2×2 window size of max pooling operation. The last 3 layers is fully connection layers. Different from previous neural networks, AlexNet uses ReLU as the activation function instead of the traditional sigmoid and tanh functions. ReLU is a no saturated activation function, which not only effectively improves the training speed of the model, but also better controls the problem of gradient disappearance and gradient explosion, it is easy to train a deeper network. The standard input size for AlexNet model is $227 \times 227 \times 3$.

5.2.2 GoogleNet (Inception v3)

GoogleNet is a convolutional neural network which has a depth of 50 layers [34]. It was built and trained by Google titled “Going deeper with convolutions”. The main feature of GoogleNet/Inception architecture is the innovation of the inception module, which is a series of 1-by-1 convolutional layers/blocks used for dimensionality reduction and feature aggregation. This model comprised a total 22 layers with 9 inception modules. The pre-trained version of Inceptionv3 with the ImageNet dataset [35] weights can classify up to 1000 objects. The image input size of this network was 299×299 pixels.

Table 6 The best results of the CNN model with Accuracy, Precision, Recall, F1 score and Time GPU metrics

	The image resolution	The metric	CNN with 9 epochs
Without DAR technique	299*299	Accuracy	90.2%
		Precision	91.18%
		Recall	90.19%
		F1 score	90.68%
		Time GPU execution	113 min 31s
With DAR technique	299*299	Accuracy	89.3%
		Precision	89.8%
		Recall	89.6%
		F1 score	89.53%
		Time GPU execution	111 min 10s

5.2.3 ResNet18

ResNet18 is a convolutional neural network which has a depth of 18 layers. It was built and trained by Microsoft in 2015, titled Deep Residual Learning for Image Recognition [36]. ResNet architectures introduced the use of residual layers and skip connection to solve the problem of vanishing gradient that may impact the weightage change in neural network and to ease the training and become much deeper with better performance. This model is trained on more than 1 million images from the ImageNet database. It can classify up to 1000 objects and the network was trained on 224×224 pixels colored images.

5.2.4 VGG 19

VGG is a convolutional neural network which has a depth of 19 layers. It was built and trained by Karen Simonyan and Andrew Zisserman at the University of Oxford University's Visual Geometry Group in 2014 [37] where they demonstrated that using small filters of size 3×3 in each convolutional layer throughout the network may result in better performance. The main idea behind VGG architecture is that multiple small filters can make design simpler and reproduce similar results compared to that of larger filters. The VGG-19 network is trained using more than 1 million images from the ImageNet database. Naturally, you can import the model with the ImageNet trained weights. The network was trained on 224×224 pixels colored images.

5.2.5 MobileNet

Howard et al. created the lightweight architecture known as MobileNet [38]. The MobileNet model from Google uses deep separable convolutions and is intended for usage in mobile applications. In comparison to a network with ordinary convolutions of the same depth in the nets, it greatly reduces the number of parameters.

Two straightforward global hyperparameters are presented by Howard et al. to effectively trade of latency and accuracy. These hyperparameters enable the model builder to select the appropriate model size for their application to the problem's constraints. Except for the first layer, which is a full convolution, the MobileNet structure is composed of depthwise separable convolutions. We are able to quickly investigate network topologies to identify a good network by defining the network in such easy terms. With the exception of the final fully connected layer, which has no nonlinearity and feeds into a softmax layer for classification, all layers are followed by a batchnorm [39] and ReLU nonlinearity. Prior to the completely connected layer, a final average pooling brings the spatial resolution down to 1. MobileNet features 28 layers if depthwise and pointwise convolutions are considered independent layers.

5.2.6 Densenet

In the dense convolutional network (DenseNet) proposed by Huang et al. [40], each layer is beforehand connected directly to every other layer (in each dense block). Dense connectivity is the name given to this kind of connection. Each layer treats the feature maps from all preceding layers as distinct inputs, while passing its own feature maps as inputs to all succeeding layers. The DenseNet is separated into Dense Blocks, each of which has the

same dimensions but differs in a number of filters. It is a crucial step in CNN that Transition Layer applies batch normalization via down-sampling. The same architecture as [30] is used in this paper.

5.2.7 YOLOv3

Inspired by ResNet, YOLOv3 uses a new network to extract features, the Darknet-53. The architectures of YOLOv3 has 53 convolutions, contains skip connections, like ResNet and 3 prediction heads. Each processing the image at a different spatial compression. And a total of nine prior bounding boxes with different sizes can be obtained by clustering. The same architecture of YOLOv3 is used like mentioned in [41]. The original MRI images were converted into RGB images and resized to fit into the input image size of each pre-trained model. For the training options, the learning rate remained constant throughout training; the training data were shuffled before each training epoch, and the validation data were shuffled before each network validation.

Basic identical parameters of the seven networks AlexNet, GoogleNet, ResNet18 ... and YOLOv3 are mentioned in the Table 7.

The parameters that have been introduced are:

Weight Learn Rate Factor : can be used to adjust the learning rate for each weight in the network.

Bias Learn Rate Factor: is a hyperparameter used to adjust the learning rate for biases in a deep neural network.

Accordingly, our contributions have been summed up as follows.

- We use the CNN architecture to evaluate the performance.
- We have evaluated the impact of the size of the input image as well as the data augmentation on the classification results.
- We suggested the use of pretrained models like AlexNet, ResNet 18, Vgg and YOLOv3 to detect glioma, meningioma, pituitary and no tumor using MRI images and improve the recognition.
- We trained the models separately to discriminate between the four categories.

In addition to the accuracy parameters, we estimated the precision, the recall, the F1 score and time GPU execution for each model. The results obtained are shown in Table 8 and the two best results are highlighted in bold.

Table 7 The training parameters of the pretrained models

Parameter	Value
Initial learning rate	0.001
Weight Learn Rate Factor	10
Bias Learn Rate Factor	10
Optimizer	SGDM
Max epoch	5
Mini batch size	20
Activation function	Softmax

The Figs. 6 and 7 represent the confusion matrix of the AlexNet, GoogleNet, ResNet and YOLOv3 without and with data augmentation respectively, which gave the best results in terms of accuracy.

5.3 Discussion

The experiments that were done on classic CNN and pre trained models gave us the accuracy on how we can correctly classify the images found in our dataset. For CNN, the accuracy is 90.2% without DAR and 89.3% with DAR. When the same dataset was used using the same hyper parameters the best accuracy found was 93.6%, 94.1%, 95.7% and 87% for AlexNet, GoogleNet, ResNet and YOLOv3 respectively in the case of without DAR. But when using DAR, all these results deteriorate except for YOLO v3 which improves accuracy reaching 91.25%. According to our experiments, we found that data augmentation using image reflection has resulted in a unique improvement for the YOLO v3 model. Overall, we observed that using pre-trained architectures can enhance results compared to using CNN alone. For instance, ResNet18 achieved an accuracy of 95.7%.

We notice that the DAR can decrease the accuracy, if you are over fitting. This could happen if the model found (probably spurious) relationships in the data that the augmentation removed. Example, class 1 often coincides with some pixel pattern on the left, class 2 has the pattern on the right. The models based on CNN with deep learning discovered this. Augmentation removed that relationship, so accuracy decreased.

Other explication, a reflected image may turn out to be an invalid image in this domain. This will falsify the training process because we are providing a network with invalid inputs which can learn false patterns.

The increased classification rate attained by ResNet 18 can be attributed to the network's use of novel techniques to lessen over-fitting in its model. The first way was to augment the data, where they artificially enlarged the dataset using a label-preserving transformation. This included generating image translations and horizontal reflections by extracting random patches (224×224 for ResNet 18) and training the network on these patches, while changing the intensities of the RGB channels in the training images. The second method to reduce over-fitting was "dropout", where neurons that do not contribute to the forward pass and do not participate in back-propagation are dropped. This reduces the complex co-adaptations of neurons and forces the model to learn more robust features.

Another very important remark, YOLOv3 which is generally used for object detection in images gives an excellent result, has several advantages over algorithms based on the use of classifiers. YOLOv3 examines the entire image at test time so that its predictions are influenced by the overall context of the image. It also performs predictions with a single network evaluation, unlike other two-stage systems such as region-based convolutional neural networks, which require thousands of images to have predictions for a single image. This makes the YOLOv3 algorithm extremely fast.

The GPU execution time is optimal in the case of AlexNet but what requires our attention is that the pretrained models converge much faster than the CNN. The time saving is very considerable; almost 2 h for CNN and less than a half hour for AlexNet, GoogleNet and ResNet 18 for a better results. We can also note that these results can be improved but for a coherent comparison, we have preserved the same training parameters for all the pretrained models.

Table 8 The results of the pretrained models

The CNN model Vs Pretrained models										
Without DAR technique	The metric	AlexNet	GoogleNet	ResNet 18	YOLO v3	VGG 19	Dense Net	MobileNet	CNN	
	Accuracy (%)	93.6	94.1	95.7	87	75.31	66	72.2	90.2	90.2
With DAR technique	Precision (%)	93.73	94.56	95.64	87	74.29	68.33	73.5	91.18	91.18
	Recall (%)	93.52	94.64	95.65	87.29	75.91	70.33	76.9	90.19	90.19
	F1score (%)	93.62	94.6	95.64	87.14	75.09	69.32	74.2	90.68	90.68
	Time GPU execution	11 min 52 s	31 min 39s	26 min 37s	221 min 17s	121 min 17s	119min52s	118min34s	113 min 31s	113 min 31s
	Accuracy (%)	85.6	93.4	94.5	91.3	81.75	61.35	59	89.3	89.3
With DAR technique	Precision (%)	86.19	93.13	94.6	91.25	82.76	62.32	60.07	89.8	89.8
	Recall (%)	85.57	93.99	94.77	91.77	83.66	59.54	62.52	89.6	89.6
	F1score (%)	85.87	93.56	94.68	91.55	83.21	61.01	61.27	89.53	89.53
	Time GPU execution	12 min 56 s	31 min 20s	26 min 54s	221 min 21s	121 min 21s	119min57s	118min58s	111 min 10s	111 min 10s

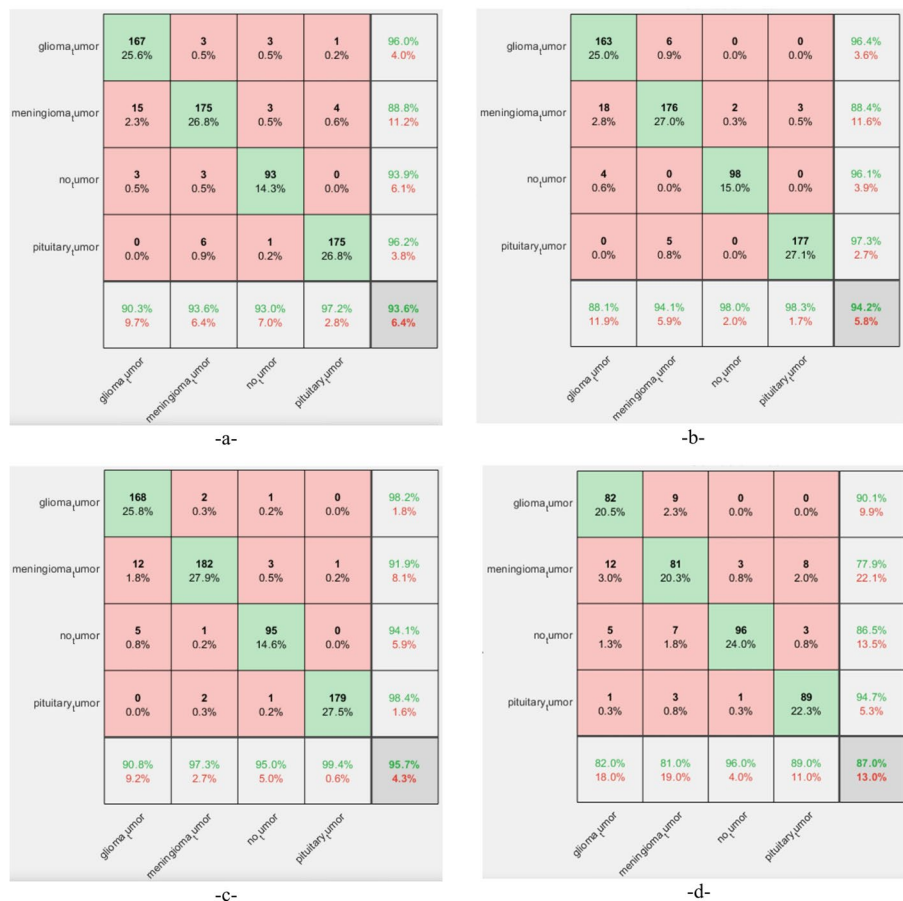


Fig. 6 Confusion matrix for the best pretrained models without DAR: **a** AlexNet, **b** GoogleNet, **c** ResNet18, **d** YOLOv3

6 Conclusion

This work aimed to develop a convolutional neural network (CNN) model that will help classify a three types of brain tumors and no-tumor using MRI images. The model used in this work was CNN as well as pre-trained models including AlexNet, GoogleNet, ResNet18 and YOLOv3. With 6 convolution layers and 5 epochs, CNN gave a result with 90.2% accuracy for classifying the four classes after training 80% of the dataset and testing on 20%. Because of the unsatisfactory performance, we decided to try the pre-trained models. These were used on the same dataset gave the better results were 93.6%, 94.1%, 95.7%, 91.25%, 81.75%, 66% and 72.2% for AlexNet, GoogleNet, ResNet18, YOLOv3, VGG19, DenseNet and Mobilenet respectively. Same remark for the use of data augmentation i.e. the results of the pretrained models are better than CNN but the DAR degraded the results for the reasons mentioned in the above discussion. The best result was obtained with ResNet18 reaching 95.7%. We also conclude

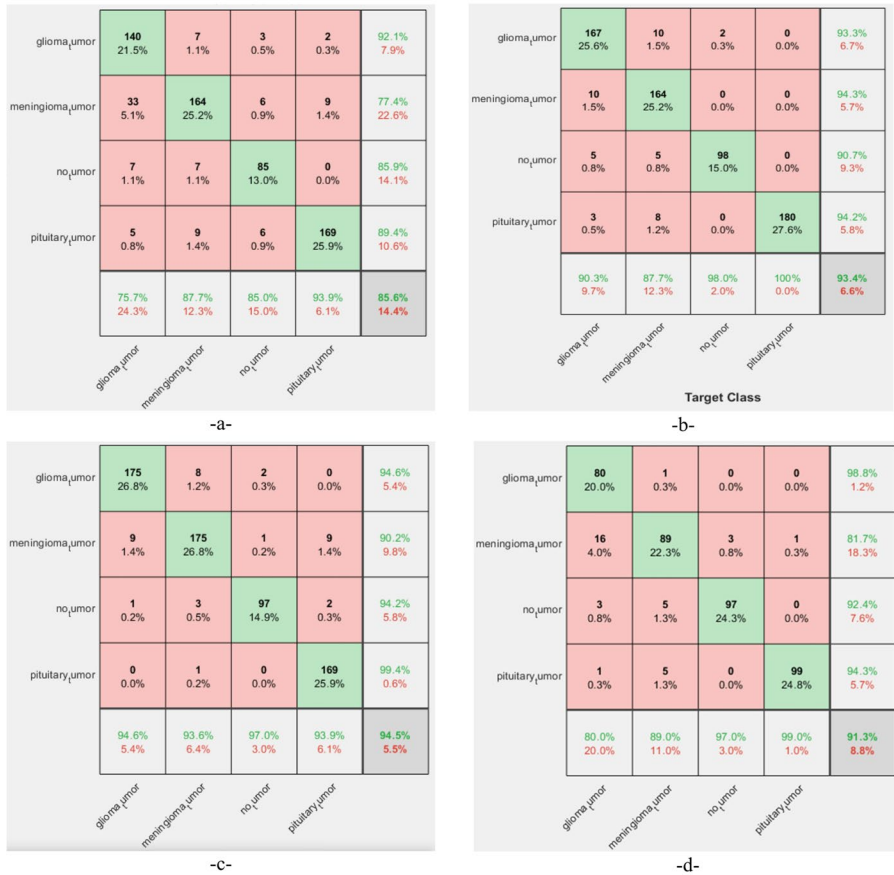


Fig. 7 Confusion matrix for the best pretrained models with DAR: -a- AlexNet, -b- GoogleNet, -c- ResNet18, -d- YOLOv3

that the increase in data is not necessarily a guarantee of performance, because it can degrade the results (except ResNet18 in our work). The choice of the data augmentation technique is crucial for a better convergence of the model. The size of the input image has a significant impact on deep learning and classification results. A larger input image size generally allows for capturing more details and complex structures, but may require more computing resources for training (as is the case for 512*512). A smaller input image size may lead to a loss of details, but can also speed up training time by requiring less computing resources. Consequently, it is important to find a balance between the input image size and the desired classification performance. In the future, we plan to acquire additional datasets and train the images using deep neural networks, including other pre-trained models. We also intend to explore hybrid approaches that enable us to leverage multiple networks simultaneously.

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

All authors designed the study. The author performed the simulations and wrote the manuscript. All authors read, edited, and approved the manuscript.

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J Clin* 71(3):209–249. <https://doi.org/10.3322/caac.21660>
2. Buckner JC, Brown PD, O'Neill BP, Meyer FB, Wetmore CJ, Uhm JH (2007), October Central nervous system tumors. In: Mayo Clinic Proceedings (Vol. 82, No. 10). Elsevier, Amsterdam, pp 1271–1286. <https://doi.org/10.4065/82.10.1271>
3. Thangudu S, Cheng FY, Su CH (2020) Advancements in the blood–brain barrier penetrating nano-platforms for brain related disease diagnostics and therapeutic applications. *Polymers* 12(12):3055
4. Komori T (2020) Updating the grading criteria for adult diffuse gliomas: beyond the WHO2016CNS classification. *Brain Tumor Pathol* 37(1):1–4. <https://doi.org/10.1007/s10014-020-00358-y>
5. Koriyama S, Nitta M, Kobayashi T, Muragaki Y, Suzuki A, Maruyama T, Kawamata T (2018) A surgical strategy for lower grade gliomas using intraoperative molecular diagnosis. *Brain Tumor Pathol* 35:159–167. <https://doi.org/10.1007/s10014-018-0324-1>
6. Asano K, Kurose A, Kamataki A, Kato N, Ogawa K, Katayama K, Ohkuma H (2018) Importance and accuracy of intraoperative frozen section diagnosis of the resection margin for effective carmustine wafer implantation. *Brain Tumor Pathol* 35:131–140. <https://doi.org/10.1007/s10014-018-0320-5>
7. Noone AM, Howlader N, Krapcho M, Miller D, Brest A, Yu M (1975) Cronin KA. SEER cancer statistics review, 2015
8. Ohgaki H, Kleihues P (2005) Epidemiology and etiology of gliomas. *Acta Neuropathol* 109:93–108. <https://doi.org/10.1007/s00401-005-0991-y>
9. Akagi Y, Yoshimoto K, Hata N, Kuga D, Hatae R, Amemiya T, Iihara K (2018) Reclassification of 400 consecutive glioma cases based on the revised 2016WHO classification. *Brain Tumor Pathol* 35:81–89. <https://doi.org/10.1007/s10014-018-0313-4>
10. Kuwahara K, Ohba S, Nakae S, Hattori N, Pareira ES, Yamada S, Hirose Y (2019) Clinical, histopathological, and molecular analyses of IDH-wild-type WHO grade II–III gliomas to establish genetic predictors of poor prognosis. *Brain Tumor Pathol* 36:135–143. <https://doi.org/10.1007/s10014-019-00348-9>
11. Spatharou A, Hieronimus S, Jenkins J (2020) Transforming healthcare with AI: The impact on the workforce and organizations. McKinsey & Company: Sydney, NSW, Australia, vol 10
12. Jerban S, Chang EY, Du J (2020) Magnetic resonance imaging (MRI) studies of knee joint under mechanical loading. *Magn Reson Imaging* 65:27–36. <https://doi.org/10.1016/j.mri.2019.09.007>
13. Spadaccini M, Iannone A, Maselli R, Badalamenti M, Desai M, Chandrasekar VT, Repici A (2021) Computer-aided detection versus advanced imaging for detection of colorectal neoplasia: a systematic review and network meta-analysis. *Lancet Gastroenterol Hepatol* 6(10):793–802. [https://doi.org/10.1016/S2468-1253\(21\)00215-6](https://doi.org/10.1016/S2468-1253(21)00215-6)
14. Saygılı A (2021) A new approach for computer-aided detection of coronavirus (COVID-19) from CT and X-ray images using machine learning methods. *Appl Soft Comput* 105:107323. <https://doi.org/10.1016/j.asoc.2021.107323>
15. Jungblut L, Blüthgen C, Polacin M, Messerli M, Schmidt B, Euler A, Martini K (2022) First performance evaluation of an artificial intelligence-based computer-aided detection system for pulmonary nodule evaluation in dual-source photon-counting detector CT at different low-dose levels. *Invest Radiol* 57(2):108–114. <https://doi.org/10.1097/RLI.0000000000000814>
16. Jarnalo CM, Linsen PVM, Blazis SP, van der Valk PHM, Dieckens DBM (2021) Clinical evaluation of a deep-learning-based computer-aided detection system for the detection of pulmonary nodules in a large teaching hospital. *Clin Radiol* 76(11):838–845. <https://doi.org/10.1016/j.crad.2021.07.012>


17. Alshayeji MH, Ellethy H, Gupta R (2022) Computer-aided detection of breast cancer on the Wisconsin dataset: an artificial neural networks approach. *Biomed Signal Process Control* 71:103141. <https://doi.org/10.1016/j.bspc.2021.103141>
18. Bahadure NB, Ray AK, Thethi HP (2018) Comparative approach of MRI-based brain tumor segmentation and classification using genetic algorithm. *J Digit Imaging* 31:477–489. <https://doi.org/10.1007/s10278-018-0050-6>
19. Singh R, Goel A, Raghuvanshi DK (2021) Computer-aided diagnostic network for brain tumor classification employing modulated Gabor filter banks. *Visual Comput* 37(8):2157–2171. <https://doi.org/10.1007/s00371-020-01977-4>
20. Abdelaziz M, Cherfa Y, Cherfa A, Alim-Ferhat F (2021) Automatic brain tumor segmentation for a computer-aided diagnosis system. *Int J Imaging Syst Technol* 31(4):2226–2236. <https://doi.org/10.1002/ima.22594>
21. Samanta AK, Khan AA (2018) Computer aided diagnostic system for automatic detection of brain tumor through MRI using clustering based segmentation technique and SVM classifier. In: *The International Conference on Advanced Machine Learning Technologies and Applications (AMLT2018)*. Springer International Publishing, pp 343–351. https://doi.org/10.1007/978-3-319-74690-6_34
22. Kleesiek J, Biller A, Urban G, Kothe U, Bendszus M, Hamprecht F (2014) Ilastik for multi-modal brain tumor segmentation. *Proceedings MICCAI BraTS (brain tumor segmentation challenge)*, pp 12–17
23. Addeh A, Iri M (2021) Brain tumor type classification using deep features of MRI images and optimized RBFNN. *ENG Trans* 2:1–7
24. Cruz DPF, Maia RD, da Silva LA, de Castro LN (2016) BeeRBF: a bee-inspired data clustering approach to design RBF neural network classifiers. *Neurocomputing* 172:427–437
25. Gu K, Zhang Y, Qiao J (2020) Ensemble meta-learning for few-shot soot density recognition. *IEEE Trans Industr Inf* 17(3):2261–2270. <https://doi.org/10.1109/TII.2020.2991208>
26. Gu K, Liu H, Xia Z, Qiao J, Lin W, Thalmann D (2021) PM2.5 monitoring: use information abundance measurement and wide and deep learning. *IEEE Trans Neural Networks Learn Syst* 32(10):4278–4290. <https://doi.org/10.1109/TNNLS.2021.3105394>
27. Gu K, Xia Z, Qiao J, Lin W (2019) Deep dual-channel neural network for image-based smoke detection. *IEEE Trans Multimedia* 22(2):311–323. <https://doi.org/10.1109/TMM.2019.2929009>
28. Cheng J, Huang W, Cao S, Yang R, Yang W, Yun Z, ... Feng Q (2015) Correction: enhanced performance of brain tumor classification via tumor region augmentation and partition. *PLoS One* 10(12):e0144479. <https://doi.org/10.1371/journal.pone.0144479>
29. Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195(1):215–243. <https://doi.org/10.1113/jphysiol.1968.sp008455>
30. Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, *Proceedings, Part V* 13. Springer International Publishing, pp 740–755
31. Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A (2010) The pascal visual object classes (voc) challenge. *Int J Comput Vision* 88:303–338. <https://doi.org/10.1007/s11263-009-0275-4>
32. Ruder S (2016) An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*. <https://doi.org/10.48550/arXiv.1609.04747>
33. Li S, Wang L, Li J, Yao Y (2021) Image classification algorithm based on improved AlexNet. In: *Journal of Physics: Conference Series* (Vol.1813, No. 1). IOP Publishing, p 012051. <https://doi.org/10.1088/1742-6596/1813/1/012051>
34. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, ... Rabinovich A (2015) Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1–9
35. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition*. IEEE, pp 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
36. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 770–778
37. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. <https://doi.org/10.48550/arXiv.1409.1556>
38. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, ... Adam H (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. <https://doi.org/10.48550/arXiv.1704.04861>

39. Ioffe S, Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning, PMLR, pp 448–456
40. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
41. Redmon J, Farhadi A (2018) Yolo3: An incremental improvement. arXiv preprint arXiv:1804.02767. <https://doi.org/10.48550/arXiv.1804.02767>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Samir Benbakreti¹  · Mohamed Benouis² · Ahmed Roumane¹ · Soumia Benbakreti³

✉ Samir Benbakreti
Samir.benbakreti@ensttic.dz

Mohamed Benouis
Mohamed.benouis@uni-a.de

Ahmed Roumane
Ahmed.roumane@ensttic.dz

Soumia Benbakreti
souben2223@gmail.com

¹ Department of Speciality, Ecole Nationale Supérieure des Telecommunications et des Technologies de l'Information et de la Communication (ENSTTIC), Street of Senia, Oran 31000, Algeria

² Chair of Human-Centered Artificial Intelligence, University of Augsburg, Augsburg, Germany

³ Laboratory of Matematic, University of Djillali Liabes, Sidi Bel Abbes 22000, Algeria