RESEARCH PAPER

Multi-Classification of Brain Tumor MRI Images Using Deep Convolutional Neural Network with Fully Optimized Framework

Emrah Irmak[1] [ID]

Abstract

Brain tumor diagnosis and classification still rely on histopathological analysis of biopsy specimens today. The current method is invasive, time-consuming and prone to manual errors. These disadvantages show how essential it is to perform a fully automated method for multi-classification of brain tumors based on deep learning. This paper aims to make multi-classification of brain tumors for the early diagnosis purposes using convolutional neural network (CNN). Three different CNN models are proposed for three different classification tasks. Brain tumor detection is achieved with 99.33% accuracy using the first CNN model. The second CNN model can classify the brain tumor into five brain tumor types as normal, glioma, meningioma, pituitary and metastatic with an accuracy of 92.66%. The third CNN model can classify the brain tumors into three grades as Grade II, Grade III and Grade IV with an accuracy of 98.14%. All the important hyper-parameters of CNN models are automatically designated using the grid search optimization algorithm. To the best of author's knowledge, this is the first study for multi-classification of brain tumor MRI images using CNN whose almost all hyper-parameters are tuned by the grid search optimizer. The proposed CNN models are compared with other popular state-of-the-art CNN models such as AlexNet, Inceptionv3, ResNet-50, VGG-16 and GoogleNet. Satisfactory classification results are obtained using large and publicly available clinical datasets. The proposed CNN models can be employed to assist physicians and radiologists in validating their initial screening for brain tumor multi-classification purposes.

Keywords Brain tumor image classification · Convolutional neural network · Deep learning · Grid search · Hyper-parameter optimization · Tumor grading

1 Introduction

Brain tumors are known as the masses formed by the abnormal proliferation of brain cells by getting rid of the brain's control mechanisms. Tumors that may form in the skull can grow, put pressure on the brain and adversely affect body health. Early detection and classification of brain tumors is an important research domain in the field of medical imaging and accordingly helps in selecting the most convenient treatment method to save patients life. Brain tumors can be classified in several different ways. For instance, one of the popular classification types is to classify the brain tumors as benign and malignant tumors. Brain benign tumors are usually tumors that develop inside the skull but outside the brain tissue. Meningiomas form an important part of this group. Unlike benign tumors in other organs, brain benign tumors can sometimes cause life-threatening conditions. Some (for example, meningiomas) may rarely turn into malignant tumors. Since they usually do not spread to the surrounding brain tissue, they have a high chance of being removed by surgery. Tumors that start in pituitary glands which control hormones and regulate functions in the body are called pituitary tumors. Pituitary tumors are known as benign tumors and do not spread to other parts of the body. Although most of the pituitary tumors are benign, they rarely return to malignant tumors. The complications of pituitary tumors can cause permanent hormone deficiency and loss of vision. Cells in malignant tumors are abnormal cells that reproduce in an uncontrolled and irregular manner. These tumors can compress, infiltrate

✉ Emrah Irmak
emrah.irmak@alanya.edu.tr

1 Electrical-Electronics Engineering Department, Alanya Alaaddin Keykubat University, 07425 Alanya, Antalya, Turkey

🌸 ⚫ Springer

or destroy normal tissues. Metastatic brain tumors are known as brain tumors that emerge from another part of the body and spread to the brain. They mostly originate from the lung, breast, large intestine, stomach, skin or prostate. Gliomas are the most common brain malignant tumors. They are the cause of most of the brain cancers and contain cells with uncontrolled proliferation. Although they can very rarely spread to the spinal cord or even to other organs of the body, they grow rapidly and may extend into the surrounding healthy tissues.

Gliomas can further be classified according to their grades. Today, the most widely accepted classification of glioma tumors is The World Health Organization (WHO) (Banan and Hartmann 2017) grading system which classifies gliomas into four grades starting from grade I to grade IV (from benignant to malignant) (Kleihues, Paul, Burger and Scheithauer 1993). This classification is based on survival data as well as histopathological features. Grade I and Grade II are referred to as "low-grade" or "benign," while Grade III (anaplastic astrocytoma) and Grade IV (glioblastoma multiforme) are considered "high-grade" or "malignant." Grade I is the least aggressive tumor grade, which does not tend to infiltrate nearby tissues. They generally grow quite slowly and can be cured with surgical operations. Grade II tumors are another slowly growing brain tumor type except that they tend to invade nearby tissues and can become faster-growing tumors over some time. Grade III brain tumors have an abnormal appearance under the microscope. They need other medical interventions other than surgical intervention because their tendency to invade other brain tissues is strong. Lastly, Grade IV tumors are known as the fastest growing tumors, which typically need the most aggressive treatment (National Cancer Institute 2020).

Early diagnosis, true grading and classification of brain tumors are vital in cancer diagnosis, treatment planning and evaluation of treatment outcome. Looking at the current medical technological advances, the detection, classification and grading of brain tumors still rely on histopathological diagnosis of biopsy specimens. The final diagnosis is usually made after clinical examination and interpretation of imaging modalities such as magnetic resonance imaging (MRI) or computed tomography (CT) followed by pathological examinations. It is known that the biggest disadvantages of this diagnostic method are that they are invasive, time-consuming and open to sampling errors. With the help of computer-aided fully automated detection and diagnosis systems that aim to make fast and accurate decisions by experts, it is possible to increase the diagnostic abilities of clinicians and radiologists to shorten the time required for a correct diagnosis.

The objective of this paper is to designate three fully automatic CNN models for multi-classification of the brain tumors using publicly available datasets. To the best of author's knowledge, this is the first attempt of multi-classification of brain tumors from given MRI images, using CNN whose almost all hyper-parameters are automatically tuned by the grid search optimizer. The rest of this paper is organized as follows: Section 2 presents related studies and a detailed review of these studies. Section 3 introduces the proposed CNN models in detail. Experimental results are reported in Sect. 4. Section 5 includes the discussions of the experimental results and a detailed comparison of the proposed method with state-of-the-art methods. Section 6 is the last section and concludes the paper.

2 Related Work

Brain tumor classification using machine learning methods has previously been studied by researchers especially over the past years. The development of artificial intelligence and deep learning-based new technologies has made a great impact in the field of medical image analysis, especially in the field of disease diagnosis (Mehmood et al. 2020, 2021; Yaqub et al. 2020). Parallel to this, many studies have been conducted on brain tumor detection and brain tumor multi-classification using CNN. This section is devoted to literature review of brain tumor multi-classification using CNN. It is possible to examine the studies in the literature in several aspects. For example, there are researchers who have performed brain tumor classification with the CNN models, which they have designed by their own, as well as those researchers who have adopted the transfer learning approach for the same purpose. The following researchers have designed their own CNN models for brain tumor classification. For example, Badža and Barjaktarović 2020designed a 22-layered CNN architecture for brain tumor-type classification using 3064 T1-weighted contrast-enhanced MRI images. Their proposed model achieved to classify the brain tumor as meningioma, glioma and pituitary with 96.56% accuracy. In another study, Mzoughi et al. (2020) presented a deep multi-scale 3D CNN model for brain tumor grading from volumetric 3D MRI images. The proposed method achieved 96.49% accuracy in classifying the brain tumor images as low-grade glioma and high-grade glioma. Ayadi et al. (2021) suggested CNN-based computer-assisted diagnosis (CAD) system for brain tumor classification. Experiments performed on three different datasets using the 18-weighted layered CNN model have achieved 94.74% classification accuracy for brain tumor-type classification and 90.35% classification accuracy for tumor grading. In 2018, Pereira et al. (2018) used CNN for predicting tumor grade directly from imaging data by overcoming the need for expert annotations of regions of interest. They evaluated two prediction approaches:

from the whole brain and from an automatically defined tumor region. They achieved accuracy of 89.5% using the grade prediction from whole brain and accuracy of 92.98% using the grade prediction from the tumor ROI. Abiwinanda et al. (2019) implemented the simplest possible architecture of CNN to recognize three most common types of brain tumors, i.e., the glioma, meningioma and pituitary achieving a validation accuracy of 84.19% at best. In 2019, Hossam et al. (2019) proposed a CNN architecture to classify brain tumors into meningioma, glioma and pituitary and differentiated between the three glioma grades (Grade II, Grade III and Grade IV).

The following researchers have adopted pre-trained CNN models using transfer learning approach for brain tumor classification. For instance, Çinar and Yildirim (2020) used a modified form of pre-trained ResNet-50 CNN model by replacing its last 5 layers with 8 new layers for brain tumor detection. They achieved 97.2% accuracy using MRI images with this modified CNN model. In a similar manner, Khawaldeh et al. (2017) proposed a modified version of AlexNet CNN model to classify brain MRI images into healthy, low-grade glioma and high-grade glioma. An overall accuracy of 91.16% was obtained using 4069 brain MRI images. Talo et al. (2019) suggested the pre-trained ResNet-34 CNN model to detect brain tumor from MRI images. Although they achieved a detection accuracy of 100%, the number of images they used for the deep learning model was 613, which were not considered as a high number for machine learning studies. Rehman et al. (2020) proposed using three pre-trained CNN models known as AlexNet, GoogleNet and VGG16 to classify the brain tumors into glioma, meningioma and pituitary. The best classification accuracy of 98.69% was achieved by the VGG-16 during this transfer learning approach. They used 3064 brain MRI images collected from 233 patients. Mehrotra et al. (2020) made use of deep learning-based transfer learning technique to classify the brain tumor images as malignant and benign using 696 T1-weighted MRI images. The most popular CNN models such as ResNet-101, ResNet-50, GoogleNet, AlexNet and SqueezeNet have been used for the classification study and compared with each other. They achieved the highest accuracy of 99.04% with the help of transfer learning through pre-trained AlexNet CNN model. Deepak and Ameer (2019) used pre-trained GoogleNet CNN model to differentiate among glioma, meningioma and pituitary brain tumor types. A mean classification accuracy of 98% was obtained in this 3-class classification problem using MRI images. In 2018, Yang et al. (2018) investigated the effect of CNN trained with transfer learning and fine-tuning to noninvasively classify low-grade glioma (LGG) and high-grade glioma (HGG) by analyzing on conventional MRI images. They achieved the accuracy of 86.6% using

pre-trained GoogleNet and 87.4% using pre-trained AlexNet.

There are also researchers who perform brain tumor classification by combining the deep learning concept with other methods. For instance, Mohsen et al. (2018) used deep neural network (DNN) classifier combined with discrete wavelet transform (DWT) and principal component analysis (PCA) to classify brain MRI images into four classes as normal brain, glioblastoma, sarcoma and metastatic bronchogenic carcinoma tumors. The accuracy rate was found to be 96.97%. Khan et al. (2020) proposed a deep learning method for classification of brain tumors into cancerous and non-cancerous using 253 real brain MRI with data augmentation. They used edge detection to find the region of interest in MRI image prior to extracting the features by a simple CNN model. They obtained 89% classification accuracy. In 2019, Kabir Anaraki et al. (2019) proposed CNN and genetic algorithm (GA)-based method to noninvasively classify different grades of glioma using MRI images. They achieved an accuracy of 90.9% for classifying three glioma grade and accuracy of 94.2% for glioma, meningioma and pituitary tumor types. Ertosun and Rubin (2015) developed a deep learning pipeline with ensemble of CNN for the problem of classification and grading of glioma from pathology images. Their method was considered quite successful in cases of lack of data, which is a common problem in the domain of deep learning approaches. They achieved 96% accuracy for HGG vs. LGG classification task and 71% accuracy for LGG Grade I versus LGG Grade II classification task.

Researchers and readers who are interested in further papers on brain tumor classification using CNN can examine the following review articles (Litjens et al. 2017; Lotan et al. 2019; Muhammad et al. 2021; Shaver et al. 2019; Shirazi et al. 2020; Tandel et al. 2019; Tiwari et al. 2020), which are very rich resources on this topic.

3 Materials and Methods

3.1 Dataset

Four different datasets, which are available from publicly available databases, are used in this study. The first dataset is called the reference image database to evaluate therapy response (RIDER) (Barboriak 2015). RIDER dataset is a targeted data collection containing MRI-multi-sequence images from 19 patients with glioblastoma (Grade IV). The total number of images in this dataset is 70,220. The second dataset is called The Repository of Molecular Brain Neoplasia Data (REMBRANDT) (Lisa et al. 2015). REMBRANDT dataset contains MRI multi-sequence images from 130 patients with glioma of Grade II, Grade

III and Grade IV. The total number of images in this dataset is 110,020. The third dataset is called the cancer genome atlas low-grade glioma (TCGA-LGG) (Pedano et al. 2016). TCGA-LGG data collection contains 241,183 MRI images of 199 patients with low-grade glioma (grade I and grade II). These three datasets are from the cancer imaging archive (TCIA) project (Clark et al. 2013). Each case was multimodal with T1-contrast-enhanced and FLAIR images. Another dataset (Cheng et al. 2015) used in this study contains 3064 T1-weighted contrast-enhanced images from 233 patients with three kinds of brain tumor: glioma (1426 slices), meningioma (708 slices) and pituitary (930 slices). Figure 1 shows some of the samples from the datastore. For Classification-1 task, a total of 2990 images are collected, including 1640 tumor and 1350 no tumor images. For Classification-2 task, a total of 3950 images are collected, including 850 normal, 950 glioma, 700 meningioma, 700 pituitary and 750 metastatic images. For Classification-3 task, a total of 4570 images are collected, including 1676 grade II, 1218 grade III and 1676 grade III. All the details about the datasets can be seen in Table 1.

3.2 Convolutional Neural Networks

The most commonly used deep learning model among neural networks is CNN model. A typical CNN model
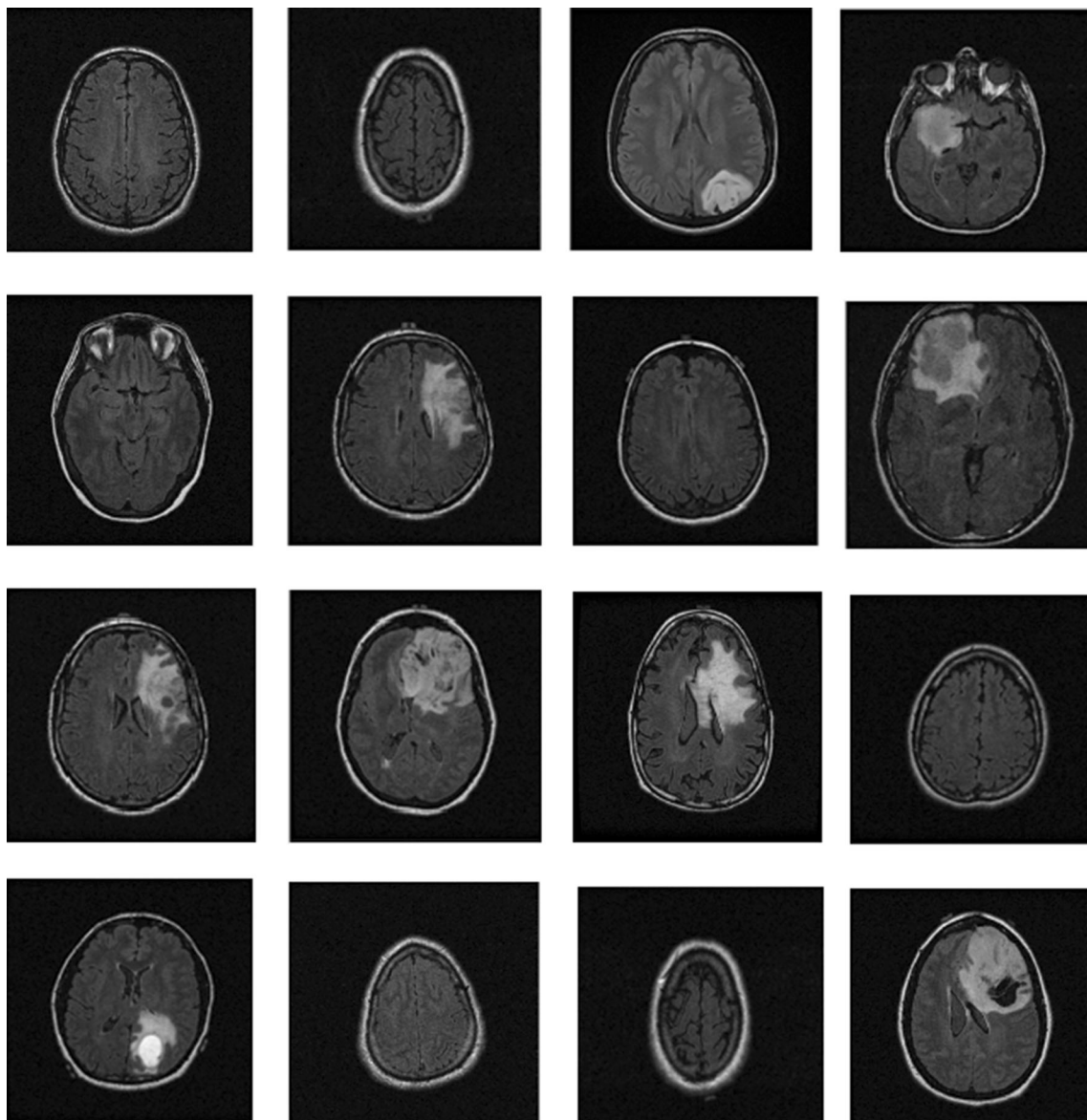

Fig. 1 Examples of brain tumor MRI images with different grades from datastore

Table 1 Number of MRI images in the dataset

Classification task	Classification groups	Number of images for each group	Total number of images
Classification-1	Tumor	1640	2990
	No tumor	1350	
Classification-2	Normal	850	3950
	Glioma	950	
	Meningioma	700	
	Pituitary	700	
	Metastatic	750	
Classification-3	Grade II	1676	4570
	Grade III	1218	
	Grade IV	1676	

consists of two parts: feature extraction and classification. CNN architecture generally includes five main layers: input layer, convolution layer, pooling layer, fully connected layer and classification layer. CNN performs feature extraction and classification through sequentially trainable layers placed one after the other. Feature extraction part of the CNN generally includes the convolutional and pooling layers, whereas the classification part includes the fully connected and classification layers. Although CNNs focus on image classification and accept images as input data in recent years, they have been also widely used in many other fields whose input data can be any signal such as audio and video (Doğantekin et al. 2019).

This paper proposes to create three fully automatic CNN models using MRI images for brain tumor multi-classification. Important hyper-parameters of the CNN models are automatically tuned by grid search optimization. The first of these CNN models is used to detect the brain tumor; hence, it decides whether a given MRI image of a patient has a tumor or not. This task is called Classification-1 throughout this paper. The proposed CNN model for Classification-1 has 13 weighted layers (1 input, 2 convolutions, 2 ReLU, 1 normalization, 2 max pooling, 2 fully connected, 1 dropout, 1 softmax and 1 classification layers) as shown in Fig. 2. Because the first CNN model is designed to classify a given image into 2 classes, the output layer has two neurons. The last fully connected layer, which is a two-dimensional feature vector, is given as an input to softmax classifier, which makes the final prediction

whether there is tumor or not. Refer to Table 2 for more information about the CNN architecture.

The second CNN model classifies the brain tumor into five brain tumor types, i.e., normal, glioma, meningioma, pituitary and metastatic. This task is called Classification-2 throughout this paper. The proposed CNN model for Classification-2 has 25 weighted layers (1 input, 6 convolutions, 6 ReLU, 1 normalization, 6 max pooling, 2 fully connected, 1 dropout, 1 softmax and 1 classification layers) as can be seen in Fig. 3. Because the second CNN model is designed to classify a given image into 5 classes, the output layer has five neurons. The last fully connected layer, which is a five-dimensional feature vector, is given as an input to softmax classifier, which makes the final prediction about the tumor type. Refer to Table 3 for more information about the CNN architecture.

The third CNN model classifies the glioma brain tumors into three grades as Grade II, Grade III and Grade IV. This task is called Classification-3 throughout this paper. The proposed CNN model for Classification-3 has 16 weighted layers (1 input, 3 convolutions, 3 ReLU, 1 normalization, 3 max pooling, 2 fully connected, 1 dropout, 1 softmax and 1 classification layers) as shown in Fig. 4. Because the last CNN model is designed to classify a given image into 3 classes, the output layer has three neurons. The last fully connected layer, which is a three-dimensional feature vector, is given as an input to softmax classifier, which makes the final prediction about the tumor grade. Refer to Table 4 for more information about the CNN architecture.
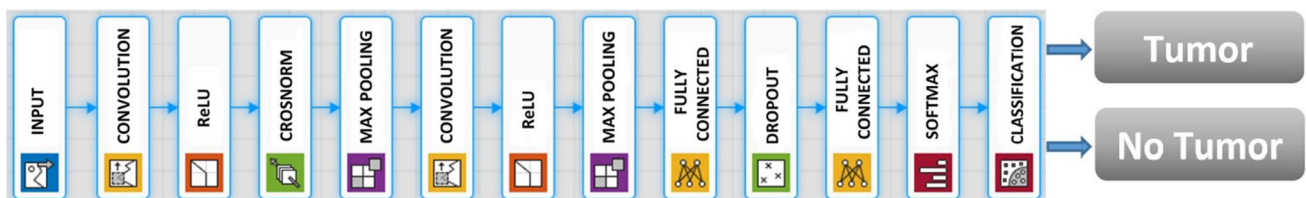

Fig. 2 Architecture of the proposed CNN model for Classification-1 task

Table 2 Details of CNN architecture used for Classification-1 task

	CNN layer	Layer type	Layer activations	Learnables parameters	Total learnables
1	227 × 227 × 3 input layer	Input	227 × 227 × 3	–	0
2	128 6 × 6 × 3 convolutions with stride [4 4] and padding [0 0 0 0]	Convolutional	56 × 56 × 128	Weights: 6 × 6 × 3 × 128	13,952
				Bias: 1 × 1 × 128	
3	ReLU-1	ReLU	56 × 56 × 128	–	0
4	Cross-channel normalization	Normalization	56 × 56 × 128	–	0
5	2 × 2 max pooling with stride [2 2] and padding [0 0 0 0]	Max pooling	28 × 28 × 128	–	0
6	96 2 × 2 × 128 convolutions with stride [1 1] and padding [2 2 2 2]	Convolutional	31 × 31 × 96	Weights: 2 × 2 × 128 × 96	49,248
				Bias: 1 × 1 × 96	
7	ReLU-2	ReLU	31 × 31 × 96	–	0
8	2 × 2 max pooling with stride [2 2] and padding [0 0 0 0]	Max pooling	15 × 15 × 96	–	0
9	512 fully connected layer	Fully connected	1 × 1 × 512	Weights: 512 × 21,600	11,059,712
				Bias: 512 × 1	
10	30% dropout	Dropout	1 × 1 × 512	–	0
11	2 fully connected layer	Fully connected	1 × 1 × 2	Weights: 2 × 512	1026
				Bias: 2 × 1	
12	Softmax	Softmax	1 × 1 × 2	–	0
13	**Output with ' No tumor' and 'tumor'**	**Classification**	–	–	**0**

Fig. 3 Architecture of the proposed CNN model for Classification-2 task

3.3 Performance Evaluation

It is very important to evaluate the classification performance in image classification studies to scientifically support the results of the study. Otherwise, the classification study would remain incomplete and academically weak. There are various performance evaluation metrics that have been used for a long time in image classification studies and have become standard performance evaluation metrics in similar studies. These are accuracy, specificity,

Table 3 Details of CNN architecture used for Classification-2 task

|	CNN layer	Layer type	Layer activations	Learnables parameters	Total learnables
1	$227 \times 227 \times 3$ input layer	Input	$227 \times 227 \times 3$	–	0
2	$128\ 6 \times 6 \times 3$ convolutions with stride [4 4] and padding [0 0 0 0]	Convolutional	$56 \times 56 \times 128$	Weights: $6 \times 6x3 \times 128$ Bias: $1 \times 1 \times 128$	13,952
3	ReLU-1	ReLU	$56 \times 56 \times 128$	–	0
4	Cross-channel normalization	Normalization	$56 \times 56 \times 128$	–	0
5	2×2 max pooling with stride [2 2] and padding [0 0 0 0]	Max pooling	$28 \times 28 \times 128$	–	0
6	$96\ 6 \times 6 \times 128$ convolutions with stride [1 1] and padding [2 2 2 2]	Convolutional	$27 \times 27 \times 96$	Weights: $6 \times 6 \times 128 \times 96$ Bias: $1 \times 1 \times 96$	442,464
7	ReLU-2	ReLU	$27 \times 27 \times 96$	–	0
8	2×2 max pooling with stride [2 2] and padding [0 0 0 0]	Max pooling	$13 \times 13 \times 96$	–	0
9	$96\ 2 \times 2 \times 96$ convolutions with stride [1 1] and padding [2 2 2 2]	Convolutional	$16 \times 16 \times 96$	Weights: $2 \times 2 \times 96 \times 96$ Bias: $1 \times 1 \times 96$	36,960
10	ReLU-3	ReLU	$16 \times 16 \times 96$	–	0
11	2×2 max pooling with stride [2 2] and padding [0 0 0 0]	Max pooling	$8 \times 8 \times 96$	–	0
12	$24\ 6 \times 6x96$ convolutions with stride [1 1] and padding [2 2 2 2]	Convolutional	$7 \times 7 \times 24$	Weights: $6 \times 6 \times 96 \times 24$ Bias: $1 \times 1 \times 24$	82,968
13	ReLU-4	ReLU	$7 \times 7 \times 24$	–	0
14	2×2 max pooling with stride [2 2] and padding [0 0 0 0]	Max pooling	$3 \times 3 \times 24$	–	0
15	$24\ 6 \times 6 \times 24$ convolutions with stride [1 1] and padding [2 2 2 2]	Convolutional	$2 \times 2 \times 24$	Weights: $6 \times 6 \times 24 \times 24$ Bias: $1 \times 1 \times 24$	20,760
16	ReLU-5	ReLU	$2 \times 2 \times 24$	–	0
17	2×2 max pooling with stride [2 2] and padding [0 0 0 0]	Max pooling	$1 \times 1 \times 24$	–	0
18	$32\ 4 \times 4 \times 4$ convolutions with stride [1 1] and padding [2 2 2 2]	Convolutional	$2 \times 2 \times 32$	Weights: $4 \times 4 \times 24 \times 32$ Bias: $1 \times 1 \times 32$	12,320
19	ReLU-6	ReLU	$2 \times 2 \times 32$	–	0
20	2×2 max pooling with stride [2 2] and padding [0 0 0 0]	Max pooling	$1 \times 1 \times 32$	–	0
21	512 fully connected layer	Fully connected	$1 \times 1 \times 512$	Weights: 512×32 Bias: 512×1	16,896
22	30% dropout layer	Dropout	$1 \times 1 \times 512$	–	0
23	5 fully connected layer	Fully connected	$1 \times 1 \times 5$	Weights: 5×512 Bias: 5×1	2565
24	Softmax	Softmax	$1 \times 1 \times 5$	–	0
25	**Output with** normal, glioma, meningioma, pituitary, metastatic	**Classification**	–	–	**0**
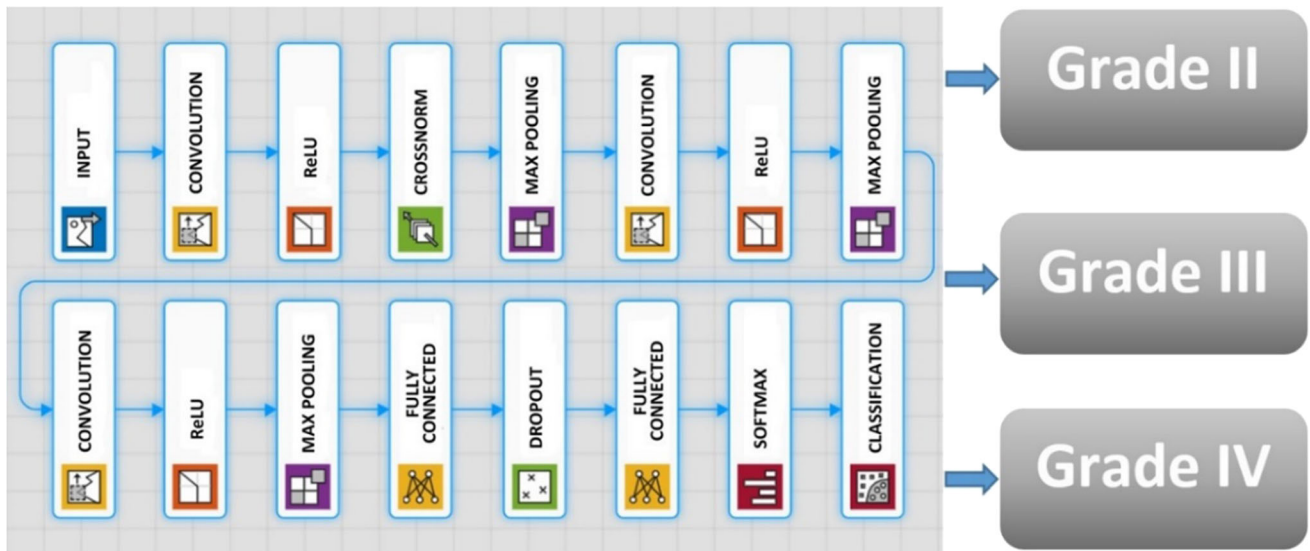
Fig. 4 Architecture of the proposed CNN model for Classification-3 task

sensitivity and precision. These metrics that are accepted as standard performance evaluation metrics in image classification studies are also used to measure the accuracy and reliability of the classification process in this paper. Moreover, the performance of the models is evaluated using the area of the receiver operation characteristic curve (ROC) known as AUC of ROC curve. Corresponding formulas regarding each of these metrics can be seen in Eq. 1. TP, TN, FP and FN are true positive, true negative, false positive and false negative, respectively.

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

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \qquad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

4 Experimental Results

4.1 Hyper-Parameter Optimization

With the increasing use of CNNs in the field of medical image processing, some difficulties have emerged in the use of CNN. As the architectures, which are developed to achieve more successful results, are getting deeper and the input images are getting higher quality, more computational costs arise. Both the reduction of these computation costs and the achievement of more successful results highly depend on the use of powerful hardware and optimizing the hyper-parameters of the established network. Therefore, almost all the important hyper-parameters of the proposed CNN models are automatically tuned using grid search optimization method. Grid search optimization method is an efficient alternative for hyper-parameter optimizations of CNN's when value range is a small search space. The grid search aims to select the best combination of which the network is trained in all the specified range combinations.

CNN models are quite complicated architectures, which include many hyper-parameters. Typically these hyper-parameters can be classified as architectural hyper-parameters and fine adjustment hyper-parameters. Number of convolutional pooling layers, number of fully connected layers, number of filters, filter sizes and activation function are known as architectural hyper-parameters. On the other hand, l_2 regularization, momentum, mini-batch size and learning rate are known as fine adjustment hyper-parameters. In this study, architectural hyper-parameters are tuned first using Algorithm 1. The fine adjustment hyper-parameters are tuned using Algorithm 2 after the architectural hyper-parameters are determined.

Algorithm 1: Grid Search algorithm to optimize the architectural hyper-parameters	
Step 1: Set 5-dimensional grid for five hyper-parameters to be optimized	: *Number of convolutional and max pooling layers* : *Number of FC layers* : *Number of filters* : *Filter sizes* : *Activation function*
Step 2: Set potential value intervals corresponding each dimension	: *Number of convolutional and max pooling layers =* *[1, 2, 3, 4]* : *Number of FC layers = [1, 2, 3, 4]* : *Number of filters = [16, 24, 32, 48, 64, 96, 128]* : *Filter sizes = [3, 4, 5, 6, 7]* : *Activation function = [ELU, SELU, ReLU, Leaky ReLU]*
Step 3: Look for all the candidate combinations and get the best one which optimizes the overall accuracy	: e.g. $C_1 = (2, 4, 16, 3, ELU) \rightarrow acc = 95\%,$ $C_2 = (4, 4, 96, 7, SELU) \rightarrow acc = 97\%,$ $C_3 = (2, 3, 64, 6, ReLU) \rightarrow acc = 99\%$

Algorithm 2: Grid Search algorithm to optimize the fine adjustment hyper-parameters	
Step 1: Set 4-dimensional grid for four hyper-parameters to be optimized	: ℓ_2 *regularization* : *Momentum* : *Mini-batch size* : *Learning rate*
Step 2: Set potential value intervals corresponding each dimension	: ℓ_2 regularization = [0.0001, 0.0005, 0.001, 0.005] : *Momentum* = [0.80, 0.85, 0.9, 0.95] : *Mini-batch size* = [4, 8, 16, 32, 64] : *Learning rate* = [0.0001, 0.0005, 0.001, 0.005]
Step 3: Look for all the candidate combinations and get the best one which optimizes the overall accuracy	: e.g. $C_1 = (0.0001, 8, 0.8, 0.0001) \rightarrow acc = 94\%,$ $C_2 = (0.0005, 32, 0.9, 0.001) \rightarrow acc = 98\%,$ $C_3 = (0.001, 64, 0.9, 0.001) \rightarrow acc = 99\%$

The grid search is performed on the training set with a fivefold cross-validation procedure in this proposed study.

The dataset is divided into fivefold out of which 4 sets are used for training and the remaining one is used for testing. There are 2990 images for Classification-1 task, 3950 images for Classification-2 task and 4570 images for Classification-3 task. For each classification task, the dataset is randomly separated as training, validation and test sets having the ratio of 60:20:20. The grid search algorithm basically tries all possible combinations of parameter values and returns the combination with the highest accuracy. In Algorithm 1, there are five parameters needed to be optimized to obtain the best accuracy. These parameters have various numbers of combinations such as 4, 4, 7, 5 and 4, respectively. Therefore, the total number of combinations to be checked is $4 \times 4 \times 7 \times 5 \times 4 = 2240$. The grid search algorithm designed to optimize the architectural hyper-parameters of the CNN model is executed a total of 11,200 times because there are 2240 combinations to be checked with fivefold cross-validation procedure. Similarly, there are four parameters needed to be optimized to obtain the best accuracy in Algorithm 2. These parameters also have various numbers of combinations such as 4, 4, 5 and 4, respectively. Therefore, the total number of combinations to be checked is $4 \times 4 \times 5 \times 4 = 320$. The grid search algorithm designed to optimize the fine adjustment hyper-parameters of the CNN model is executed a total of 1600 times because there are 320 combinations to be checked with fivefold cross-validation procedure. Tables 5, 6 and 7 demonstrate optimum hyper-parameters achieved for Classification 1, Classification 2

Table 4 Details of CNN architecture used for Classification-3 task

	CNN layer	Layer type	Layer activations	Learnable parameters	Total learnables
1	227 × 227 × 3 input layer	Input	227 × 227 × 3	–	0
2	128 6 × 6 × 3 convolutions with stride [4 4] and padding [0 0 0 0]	Convolutional	56 × 56 × 128	Weights: 6 × 6 × 3 × 128 Bias: 1 × 1 × 128	13,952
3	ReLU-1	ReLU	56 × 56 × 128	–	0
4	Cross-channel normalization	Normalization	56 × 56 × 128	–	0
5	2 × 2 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	28 × 28 × 128	–	0
6	96 6 × 6 × 28 convolutions with stride [1 1] and padding [2 2 2 2]	Convolutional	27 × 27 × 96	Weights: 6 × 6 × 128 × 96 Bias: 1 × 1 × 96	46,752
7	ReLU-2	ReLU	27 × 27 × 96	–	0
8	2 × 2 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	13 × 13 × 96	–	0
9	96 2 × 2 × 6 convolutions with stride [1 1] and padding [2 2 2 2]	Convolutional	16 × 16 × 96	Weights: 2 × 2 × 96 × 96 Bias: 1 × 1 × 96	36,864
10	ReLU-3	ReLU	8 × 8 × 96	–	0
11	2 × 2 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	6 × 6 × 256	–	0
12	512 fully connected layer	Fully Connected	1 × 1 × 512	Weights: 512 × 6144 Bias: 512 × 1	3,146,240
13	30% dropout	Dropout	1 × 1 × 512	–	0
14	3 fully connected layer	Fully Connected	1 × 1 × 3	Weights: 3 × 512 Bias: 3 × 1	1539
15	Softmax	Softmax	1 × 1 × 2	–	0
16	**Output with 'Grade II, 'Grade III' and 'Grade IV'**	**Classification**	–	–	**0**

and Classification 3 tasks, respectively, by grid search optimization algorithm.

4.2 Results obtained by optimized CNN models

The performance of the proposed model is evaluated using fivefold cross-validation procedure for Classification 1 task. The dataset is divided into fivefold out of which 4 sets are used for training and the remaining one is used for testing. The experiments are repeated five times. Classification performance for the task is evaluated for each fold, and the average classification performance of the model is calculated. High accuracies from training and validation phases are not meaningful without testing the trained and hyper-parameter-tuned CNN on predicting unseen samples.

Therefore, a test dataset is randomly assigned and separated along with training and validation datasets to test the performance of trained CNN on predicting samples; otherwise, the high accuracy may be due to biased dataset assignment (e.g., obvious images with strong characteristics from severe tumor patients). For Classification-1 task, as the study has 2990 samples, there are enough images to be randomly separated as training, validation and test sets having the ratio of 60:20:20 as shown in Table 8. Two hundred and ninety-nine images are randomly excluded from the dataset of each class, and they are used for test purposes.

Displaying the activations of convolution layers of the CNN is a great way to see the features learned by CNN upon training. This visualization is of great benefit to the

Table 5 Optimum hyper-parameters results achieved by grid search for Classification-1 task

Parameters	Range of parameters	Optimum value
Number of convolution and max pooling layers	[1, 2, 3, 4]	2
Number of FC layers	[1, 2, 3, 4]	2
Number of filters	[16, 24, 32, 48, 64, 96, 128]	128, 96
Filter size	[3, 4, 5, 6, 7]	6, 6
Activation function	[ELU, SELU, ReLU, Leaky ReLU]	ReLU
Mini-Batch Size	[4, 8, 16, 32, 64]	32
Momentum	[0.80, 0.85, 0.9, 0.95]	0.9
Learning Rate	[0.0001, 0.0005, 0.001, 0.005]	0.0001
l_2 Regularization	[0.0001, 0.0005, 0.001, 0.005]	0.0001

Table 6 Optimum hyper-parameters results achieved by grid search for Classification-2 task

Parameters	Range of parameters	Optimum value
Number of convolution and max pooling layers	[1, 2, 3, 4]	6
Number of FC layers	[1, 2, 3, 4]	2
Number of filters	[16, 24, 32, 48, 64, 96, 128]	128, 96, 96, 24, 24, 32
Filter size	[3, 4, 5, 6, 7]	6, 6, 2, 6, 6, 4
Activation function	[ELU, SELU, ReLU, Leaky ReLU]	ReLU
Mini-batch size	[4, 8, 16, 32, 64]	64
Momentum	[0.80, 0.85, 0.9, 0.95]	0.9
Learning rate	[0.0001, 0.0005, 0.001, 0.005]	0.0001
l_2 Regularization	[0.0001, 0.0005, 0.001, 0.005]	0.001

Table 7 Optimum hyper-parameter results achieved by grid search for Classification-3 task

Parameters	Range of parameters	Optimum value
Number of convolution and max pooling layers	[1, 2, 3, 4]	3
Number of FC layers	[1, 2, 3, 4]	2
Number of filters	[16, 24, 32, 48, 64, 96, 128]	128, 96, 96
Filter size	[3, 4, 5, 6, 7]	6, 6, 2
Activation function	[ELU, SELU, ReLU, Leaky ReLU]	ReLU
Mini-batch size	[4, 8, 16, 32, 64]	32
Momentum	[0.80, 0.85, 0.9, 0.95]	0.9
Learning rate	[0.0001, 0.0005, 0.001, 0.005]	0.005
l_2 Regularization	[0.0001, 0.0005, 0.001, 0.005]	0.001

researcher to see what the network has learned. The activations of first and second convolutional layers are shown in Fig. 5a and b, respectively. The first convolutional layer of the CNN is used to learn features like color and edges, whereas the second convolutional layer is used to learn more complicated features like brain tumor borders. The subsequent (deeper) convolutional layers build up their features by combining features learned by the earlier convolutional layers. There are 128 channels in the first convolutional layer of CNN for Classification-1 task, and 96 of these channels are shown in Fig. 5a. In the second convolutional layer, there are 96 channels and these channels are shown in Fig. 5b. Channels are 2-D arrays which form every layer of CNN.

Each image in the grid of Fig. 5a is the output of each channel in the first convolutional layer. White pixels in these images show strong positive activations, and black ones show strong negative activations. Likewise, gray

pixels show not-strongly activated channels on the input image. Activations of a specific channel and the strongest activation channel in the first convolutional layer are shown in Fig. 6b and c, respectively. White pixels in the channel of Fig. 6c show that this channel is strongly activated at tumor position. It can be concluded that the CNN has learned that tumors are characteristic features to distinguish between classes of images although it has never been told to learn about tumors. Unlike previous artificial neural network approaches which are often manually designed specific to the problem, these convolutional neural networks can learn useful features for themselves by their own. In this paper, learning to identify tumors helps to distinguish between a tumorous image and non-tumorous image.

After the classification process has been carried out, the performance of CNN models should be tested by various reliable methods. The performance evaluation of the models in this paper is made using accuracy, specificity, sensitivity, precision metrics and AUC of ROC curve. A thorough review of the test results is made, and the results are added in this paper. Figure 7 is the accuracy and loss plot of the proposed CNN for Classification-1 task. Classification accuracy of 99.33% is achieved after 444 iterations using the proposed model for Classification-1 task. It is quite obvious from Fig. 7 that after about 200 iterations, almost 100 percent accuracy is achieved. The AUC value of the ROC curve is 0.9995 as shown in Fig. 9. These results proof the ability of the proposed CNN model for brain tumor detection. Please see Fig. 8 for confusion matrix, Fig. 9 for ROC curve and Table 9 for accuracy metrics in terms of TP, TN, FP, FN, accuracy, specificity, sensitivity and precision. Figure 10 shows classification

results and the predicted probabilities of four test images for Classification-1 task.

The performance of the proposed model is evaluated using the fivefold cross-validation procedure for Classification-2 task. The dataset is divided into fivefold out of which 4 sets are used for training and the remaining one is used for testing. The experiments are repeated five times. Classification performance for the task is evaluated for each fold, and the average classification performance of the model is calculated. For Classification-2 task, as the study has 3950 samples, there are enough images to be randomly separated as training, validation and test sets having the ratio of 60:20:20 as shown in Table 8. One hundred and fifty-eight images are randomly excluded from the dataset of each class to be used for testing the model. Figure 11 is the accuracy and loss plot of the proposed CNN model for Classification-2 task. Classification accuracy of 92.66% is achieved after 294 iterations using the proposed CNN model for Classification-2 task. The AUC value of the ROC curve is 0.9981 as shown in Fig. 13. These results show the ability of the proposed CNN model for brain tumor-type classification. Please see Fig. 12 for confusing matrix and Table 9 for accuracy metrics in terms of TP, TN, FP, FN, accuracy, specificity, sensitivity and precision. As shown in Table 9, accuracy of 97.85% is achieved to classify glioma, 97.60% for meningioma, 97.47% for metastatic, 95.44 for healthy brain (normal) and 96.96% for pituitary tumor type for Classification-2 task (Fig. 13).

The performance of the proposed model is evaluated using the fivefold cross-validation procedure for Classification-3 task. The dataset is divided into fivefold out of which 4 sets are used for training and the remaining one is used for testing. The experiments are repeated five times. Classification performance for the task is evaluated for

Table 8 Learning scheme of the CNN models

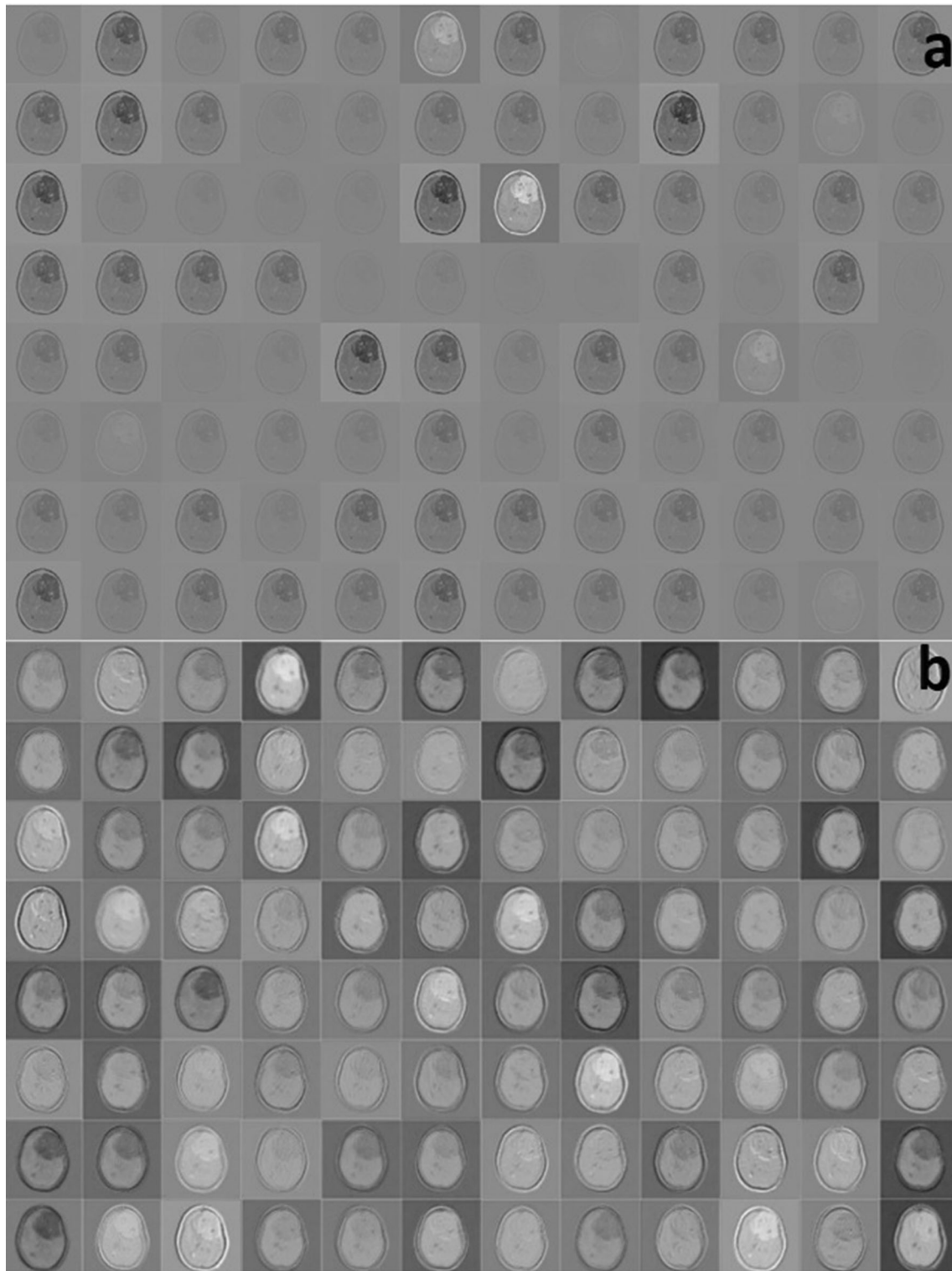| Classification task	Classification groups	Number of images				
		Each group	Total	Training set (60%)	Validation set (20%)	Test set (20%)
Classification-1	Tumor	1640	2990	1794	598	598
	No tumor	1350				
Classification-2	Normal	850	3950	2370	790	790
	Glioma	950				
	Meningioma	700				
	Pituitary	700				
	Metastatic	750				
Classification-3	Grade II	1676	4570	2742	914	914
	Grade III	1218				
	Grade IV	1676				

Fig. 5 First **a** and second **b** convolutional layer activations for Classification-1 task. Each image in the grid is the output of each channel. White regions show strong positive activations, whereas gray sections show not-strongly activated channels

each fold, and the average classification performance of the model is calculated. For Classification-3 task, as the study has 4570 samples, there are enough images to be randomly separated as training, validation and test sets having the ratio of 60:20:20 as shown in Table 8. Three hundred and five images are randomly excluded from the dataset of each

Fig. 6 a Input image,
b activations in a specific
channel and **c** the strongest
activation channel of the first
convolutional layer for
Classification-1 task. White
pixels in **c** show strong
activations which shows that
this channel is strongly
activated at tumor position
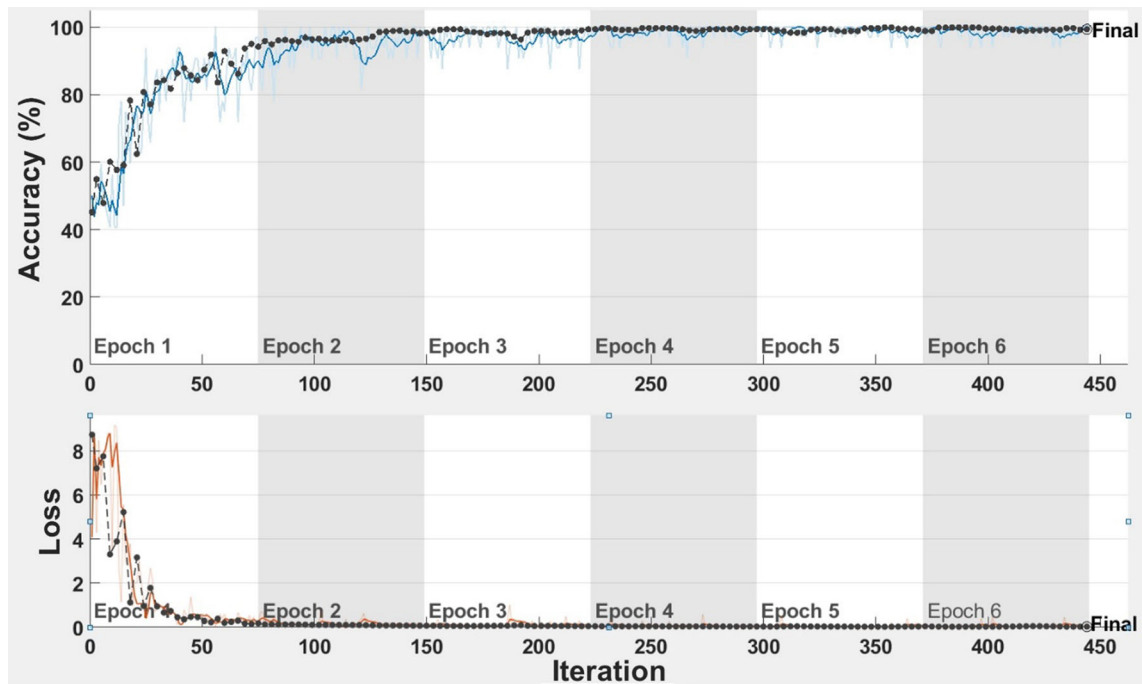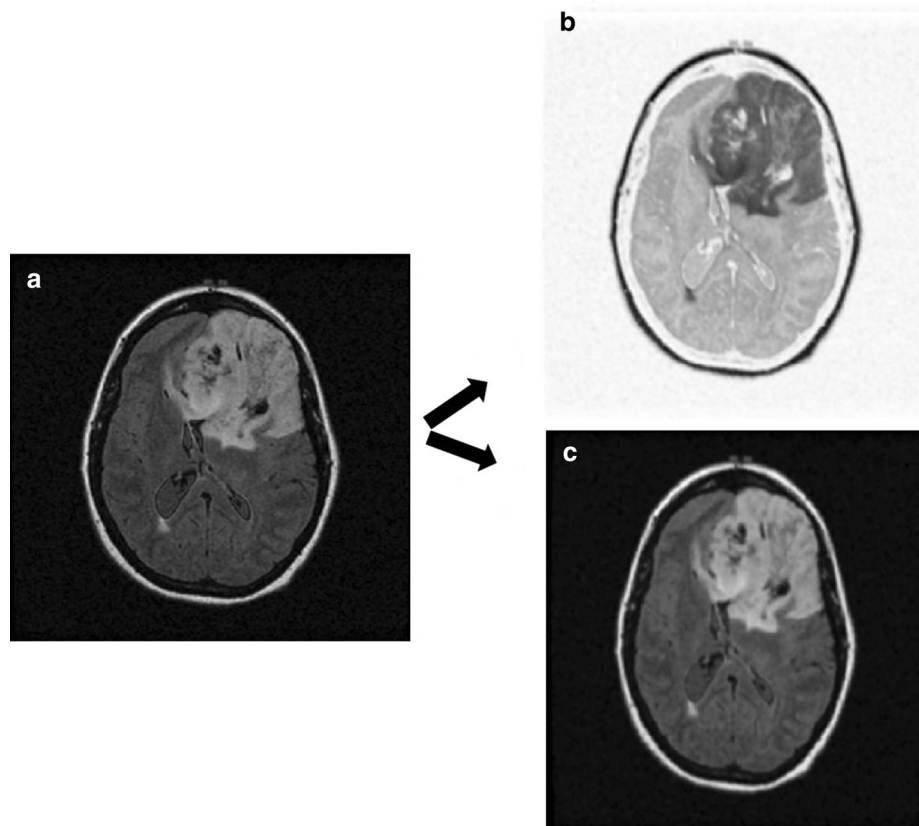
Fig. 7 Accuracy and loss curves for Classification-1 task

class to be used for testing the model. Figure 14 is the accuracy and loss plot of the proposed CNN for Classification-3 task. Classification accuracy of 98.14% is

achieved after 342 iterations using the proposed model for Classification-3 task. The AUC value of the ROC curve is 0.9994 as shown in Fig. 16. These results proofs the ability

Target Class

Fig. 8 Confusion matrix for Classification-1 task

of the proposed CNN model for brain tumor grading. Please see Fig. 15 for confusion matrix and Table 9 for accuracy metrics in terms of TP, TN, FP, FN, accuracy, specificity, sensitivity and precision. As shown in Table 9, accuracy of 98.14% is achieved to classify grade II, 100% for grade III and 98.14% for grade IV for brain tumor grade classification in Classification-3 task (Fig. 16).
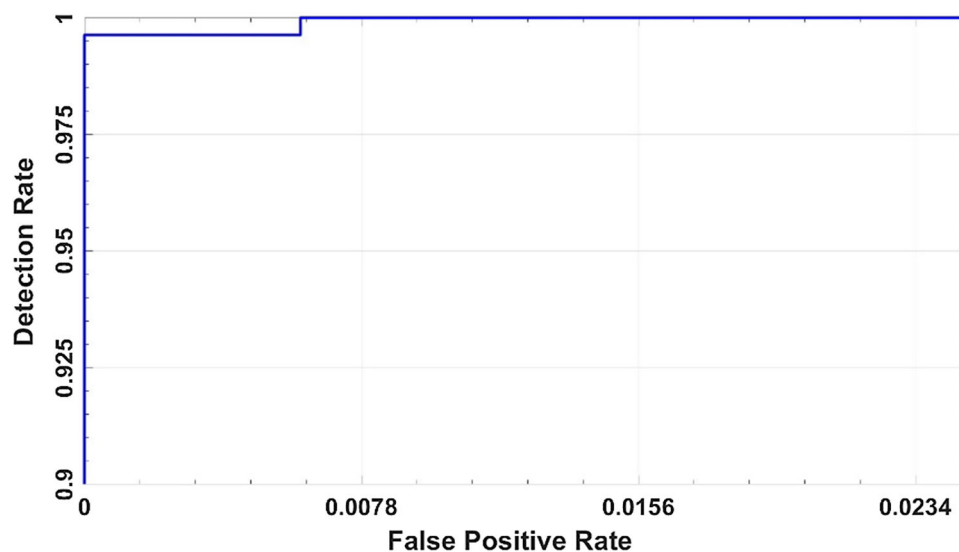
5 Discussions

Image classification using convolutional neural network is frequently used in the diagnosis of medical diseases recently. It is not possible and realistic to design an efficient CNN model that can be used jointly for all classification problems and can give good results. For this reason, a unique CNN model is designed for each problem type. The structure and complexity of the CNN model vary according to the type of problem, inputs and expected outputs. In this study, three different CNN models are designed for three classification purposes. The first model is designed to detect brain tumor from input MRI images. The second model is designed to find brain tumor type, and lastly, the third model is designed to predict the brain tumor grade. One of the difficulties encountered in convolutional neural networks is choosing the most successful network model for the specific problem. Obtaining successful results especially in convolutional neural networks is highly dependent on choosing the right hyper-parameters. In this study, grid search optimizer is used in order to design the most successful CNN model and to optimize the hyper-parameters of the CNN model. Satisfactory classification results are obtained using large and publicly available clinical datasets. For example, brain tumor detection is achieved with a highly satisfactory accuracy as 99.33% using the first designed CNN model. In addition, the brain tumor type classification is performed with an accuracy of 92.66%. Lastly, the brain tumor grading is succeeded with a high accuracy as 98.14%. The results of the proposed models are validated using performance evaluation metrics such as AUC value of ROC curve, accuracy, specificity, sensitivity and precision.

Fig. 9 ROC curve for Classification-1 task

Table 9 Accuracy metrics in terms of TP, TN, FP, FN, accuracy, specificity, sensitivity and precision

Metrics	Classes	TP	TN	FP	FN	Accuracy (%)	Specificity	Sensitivity	Precision	Total
Architectures										
Proposed CNN model for										
Task 1	No tumor	270	324	4	0	**99.33**	0.988	**1**	0.985	270
	Tumor	324	270	0	4	**99.33**	**1**	0.988	**1**	328
Task 2	Glioma	181	592	8	9	**97.85**	0.978	**0.944**	0.934	190
	Meningioma	130	641	9	10	97.60	0.984	0.942	0.923	140
	Metastatic	136	634	6	14	97.47	**0.990**	0.900	**0.955**	150
	Normal	159	595	25	11	95.44	0.965	0.921	0.880	170
	Pituitary	126	640	10	14	96.96	0.9846	0.880	0.909	140
Task 3	Grade II	328	569	10	7	98.14	0.9827	0.9791	0.9704	335
	Grade III	244	670	0	0	**100**	**1**	**1**	**1**	244
	Grade IV	325	572	7	10	98.14	0.9879	0.9701	0.9789	335
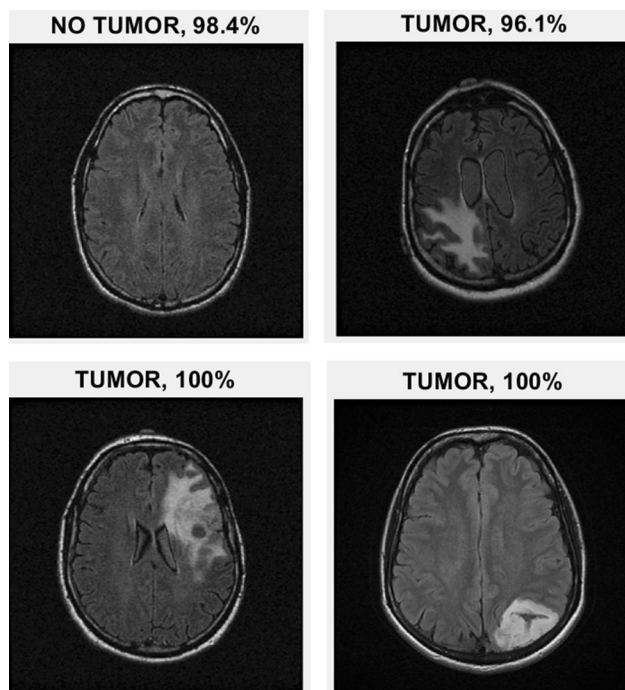
Fig. 10 Classification results and the predicted probabilities of four test images for Classification-1 task

It is worth comparing the results obtained by the proposed CNN models with the results of existing popular state-of-the-art CNN models. For this purpose, the same experiments with the same dataset are conducted using the popular well-known pre-trained CNN models such as AlexNet, Inceptionv3, ResNet-50, VGG-16 and Google-Net. The results obtained with these networks are shown in Table 10. The proposed CNN models and these popular networks are compared in terms of accuracy and AUC obtained during the experiments. Table 10 shows that the

proposed CNN models outperform other networks in each classification task. In brain tumor detection task (Classification-1 task), the pre-trained ResNet-50 model which achieves 92.79% classification accuracy is the closest model to the proposed model. On the other hand, the pre-trained VGG-16 model achieves 88.87% classification accuracy in brain tumor-type classification task (Classification-2 task) and is the closest model to the proposed CNN model. A classification accuracy of 94.13% is obtained with pre-trained GoogleNet model which becomes the best network after the proposed CNN model in tumor grading (Classification-3 task). One possible reason about the superiority of the proposed CNN models to pre-trained networks is obviously that those pre-trained deep learning models are designed and trained on general datasets for general image classification problems. On the contrary, the proposed CNN models are designated for more specific problems such as brain tumor detection, tumor type and grade classification. In addition, the proposed models are trained and tested on brain tumor MRI images. Another possible reason why the proposed CNN models give better results than the pre-trained models is that the proposed CNN architectures have been optimized for the specific purposes and used the hyper-parameters that give the best results for the specific problems in question. There are similar CNN-based image classification studies that use grid search optimizer to tune the hyperparameters of the CNN to obtain better accuracy results. For instance, the study (Irmak 2021), proposes a novel CNN model, which is also tuned using grid search for COVID-19 disease detection. Although the same optimization method is used in both that paper and the proposed paper, they differ in type of disease that they diagnose. In addition, the CNN architectures are different
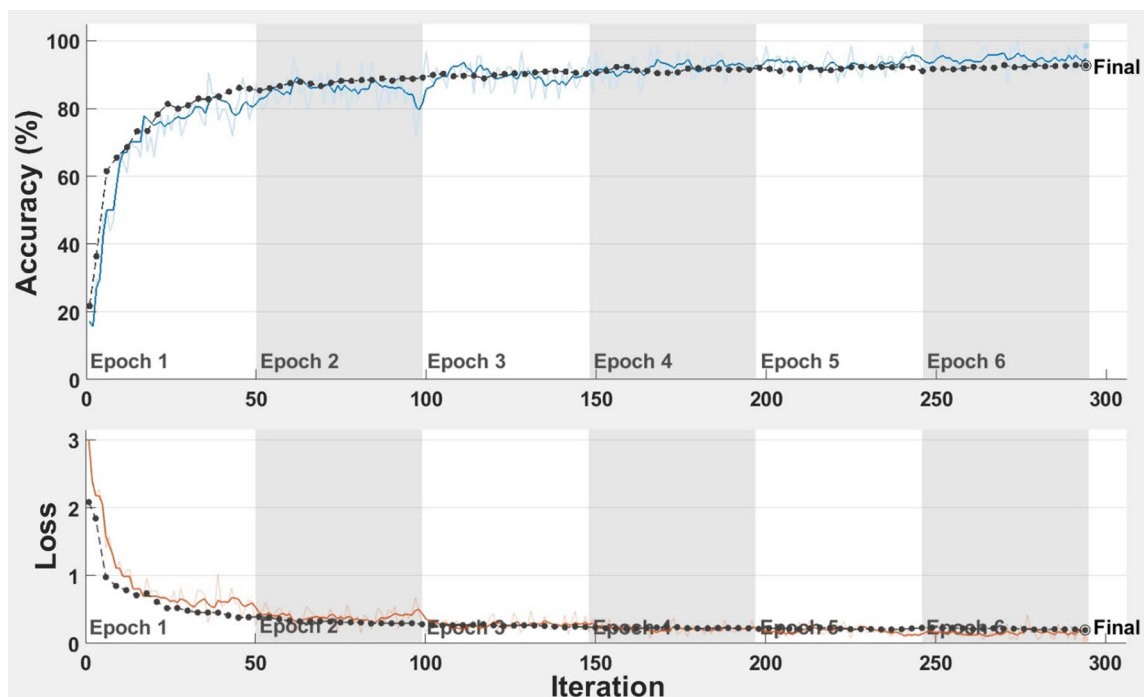
Fig. 11 Accuracy and loss curves for Classification-2 task

Fig. 12 Confusion matrix for Classification-2 task

from each other. The study, (Irmak 2020), is another successful application of CNN model whose hyper-parameters are tuned by grid search optimizer.

Looking at the literature, one can see that some researchers have studied to classify the images into its grades, while other researchers have made brain tumor type classification. Moreover, there are other researchers who

have classified the MRI images into brain tumor and brain non-tumor images. Since all these three tasks are achieved in the proposed method, comparison of these tasks is made with individual studies in the literature. Researchers in (Sultan et al. 2019) have presented a CAD system to classify the brain tumor MR images into three types (glioma, meningioma and pituitary) which can be considered as Classification-2 task and further classify gliomas into different grades (grade I, grade II, grade III and grade IV), which can be considered as Classification-3 task. These researchers achieved a classification accuracy of 96.13% for Classification-2 task and 98.7% for Classification-3 task. Another researcher in (Kabir Anaraki et al. 2019) achieved a classification accuracy of 94.2% for Classification-2 task and 90.9% for Classification-3 task using CNN with genetic algorithms. Sajjad et al. (2019) used convolutional networks with extensive data augmentation to perform Classification-2 task with an overall accuracy 90.81% and Classification-3 task of 90.67%. El-Dahshan et al. (2010) got an overall accuracy of 98% for Classification-3 task using hybrid intelligent techniques to classify MR images into tumor and non-tumor images. Seetha et al. (2018) had an overall accuracy of 97.5% for Classification-1 task. The proposed individual models in this paper are able to perform three classification tasks. The proposed CNN models in this paper achieved an overall accuracy of 99.33% for Classification-1 task, 92.66% for Classification-2 task and 98.14% for Classification-3 task. Although the proposed model for Classification-2 task

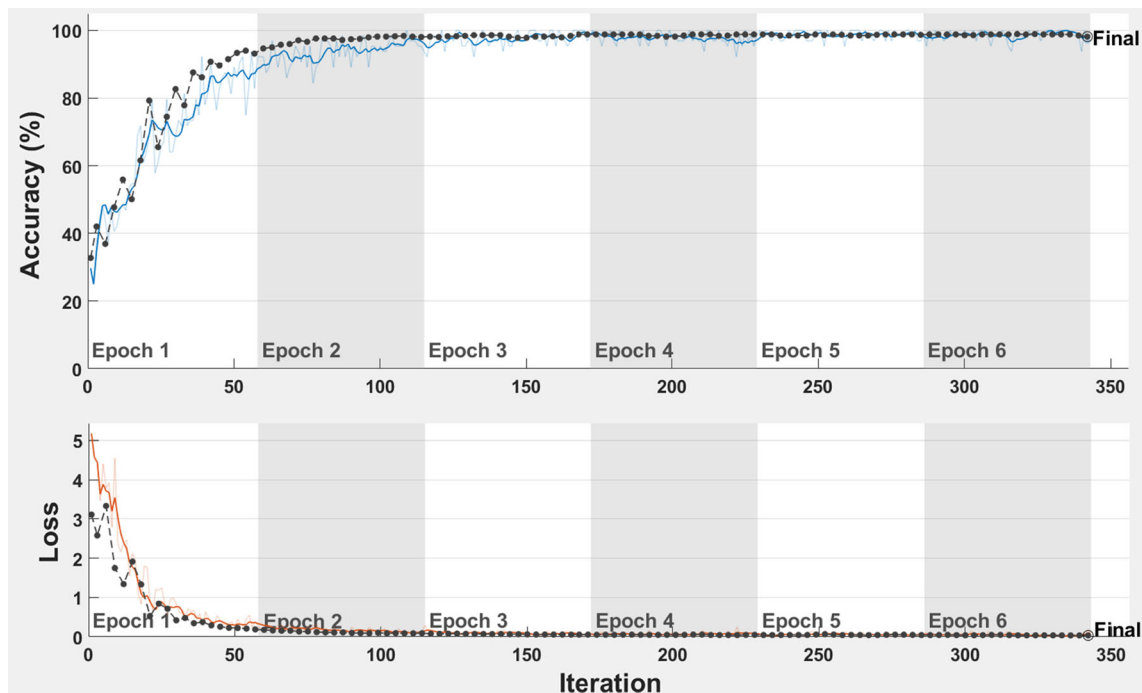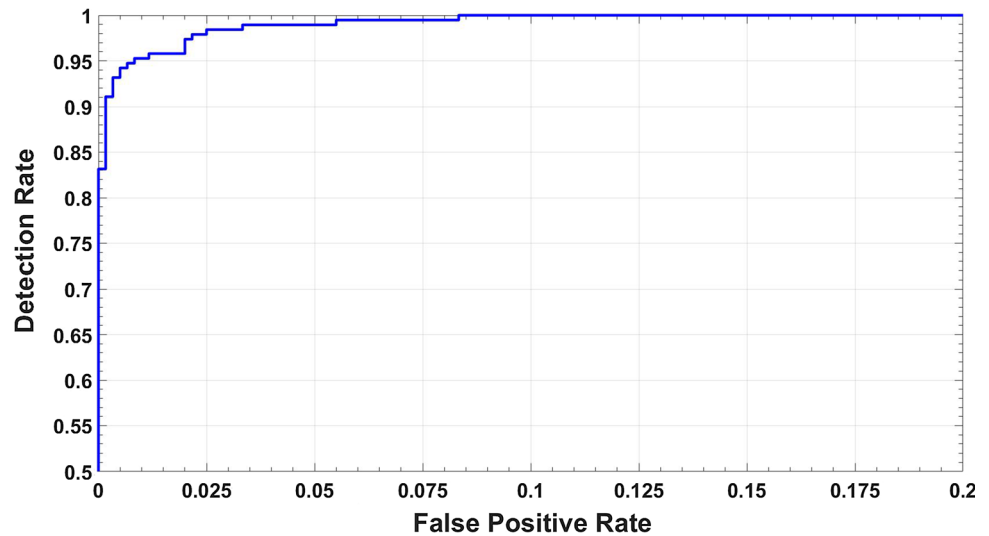**Fig. 13** ROC curve for Classification-2 task

Fig. 14 Accuracy and loss curves for Classification-3 task

classifies the tumor images into five types (glioma, meningioma, pituitary, normal brain and metastatic), it still gives a high accuracy. It is clear that the CNN models proposed in this paper are superior to existing methods for multi-classification purposes of brain tumor MRI images. Table 11 presents the comparison of the proposed models with the state-of-the-art methods in the literature in terms of performance evaluation, datasets used and classification type. Considering the literature carefully, to the best of author's knowledge the main advantages and contributions of the proposed approach in this paper are as follows:

- This study is the first study for multi-classification of brain tumor MRI images, using CNN whose almost all hyper-parameters are automatically determined by the grid search optimizer.
- Grid search optimization algorithm can be used to select the best CNN architecture and hyper-parameters of the selected CNN model.
- Thanks to the proposed novel CNN model for Classifications-1 task, detection of brain tumor can be achieved with high classification accuracy such as 99.33%.

Target Class

Fig. 15 Confusion matrix for Classification-3 task

- Thanks to the proposed novel CNN model for Classifications-2 task, glioma tumor, meningioma tumor, pituitary tumor, metastatic tumor types and healthy MR images can be classified with high classification accuracy such as 92.66%.
- Thanks to the proposed novel CNN model for Classifications-3 task, Grade II, Grade III and Grade IV brain tumor images can be classified with high classification accuracy such as 98.14%.

6 Conclusion

The state-of-the-art advances in deep learning lead the studies and researches in machine learning to evolve from feature engineering into architectural engineering. This paper presents the multi-classification of brain tumors for the early diagnosis purposes using CNN models whose almost all hyper-parameters are automatically tuned using grid search. Three robust CNN models for three different brain tumor classification tasks by means of publicly medical image datasets are designated. Detection of brain tumor is achieved with a high accuracy such as 99.33%.
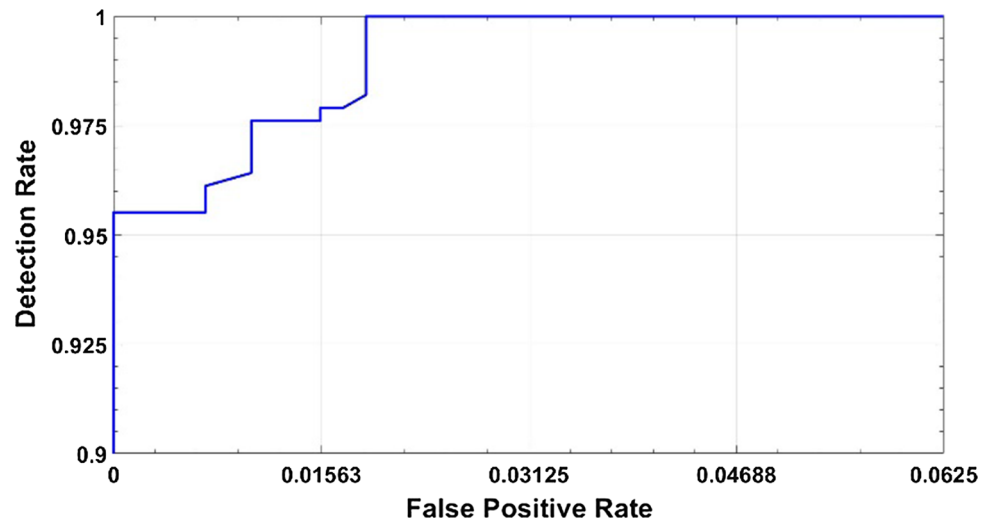
Fig. 16 ROC curve for Classification-3 task

Table 10 Performance comparison of the proposed model with existing popular state-of-the-art CNN networks

	Classification-1 task		Classification-2 task		Classification-3 task	
	Accuracy (%)	AUC	Accuracy (%)	AUC	Accuracy (%)	AUC
AlexNet	88.12	0.8976	83.12	0.8421	90.43	0.9565
Inceptionv3	86.09	0.8734	82.38	0.8319	85.01	0.8671
ResNet-50	92.79	0.9317	75.93	0.8022	85.79	0.8790
VGG-16	87.89	0.9105	88.87	0.8998	83.97	0.8567
GoogleNet	72.32	0.8009	78.87	0.8117	94.13	0.9579
Proposed CNN Model	99.33	0.9995	92.66	0.9981	98.14	0.9994

Table 11 Comparison of the proposed study with related studies

Method	Accuracy result for Classification-1	Accuracy result for Classification-2	Accuracy result for Classification-3	Classification type	Datasets used
(Sultan et al. 2019)	–	96.13%	98.7%	Glioma, meningioma and pituitary classification Glioma grade classification	Tianjing Medical University REMBRANDT
(Kabir Anaraki et al. 2019)	–	94.2%	90.9%	Glioma, meningioma and pituitary classification Glioma grade classification	Brain tumor dataset TCGA-GBM TCGA-LGG REMBRANDT Hazrat-e Rasool General Hospital
(Sajjad et al. 2019)	–	90.81%	90.67%	Glioma, meningioma and pituitary classification Glioma grade classification	Radiopaedia dataset Brain tumor dataset
Zacharaki et al.	–	–	88%	Glioma grade classification	The patients were imaged using a 3.0-T MRI scanner system
(Paul et al. 2012)	–	91.43%	–	Glioma, meningioma and pituitary classification	Publicly available dataset
(Cheng et al. 2015)	–	91.28%	–	Glioma, meningioma and pituitary classification	Publicly available dataset
(Papageorgiou et al. 2008)	–	–	92%	Glioma grade classification	Department of Pathology of the University Hospital of Patras
(Ertosun and Rubin 2015)	–	–	96%	Glioma grade classification	TCGA-GBM TCGA-LGG
(El-Dahshan et al. 2010)	–	–	98%	Glioma grade classification	Harvard Medical School website
(Özyurt et al. 2019)	95.62%	–	–	Brain tumor and brain non-tumor classification	TCGA-GBM
(Seetha et al. 2018)	97.5%	–	–	Brain tumor and brain non-tumor classification	IMAGENET
Proposed method	99.33%	92.66%	98.14%	Brain tumor and brain non-tumor classification Glioma, meningioma, pituitary, normal brain and metastatic classification Glioma grade classification	RIDER REMBRANDT TCGA-LGG

Moreover, classification of brain MR into glioma, meningioma, pituitary, normal brain and metastatic is obtained with satisfying accuracy of 92.66%. Finally, classification of glioma brain tumors into grade II, grade III and grade IV is performed with an accuracy of 98.14%. The proposed CNN models are trained and tested using sufficiently large number of medical images. Results obtained using the proposed CNN models and the comparisons with state-of-the-art methods show the effectiveness of the CNN models created with the proposed optimization framework. The CNN models established in this paper can be employed to assist physicians and radiologists in validating their initial screening for brain tumor multi-classification purposes.

Declarations

Conflict of interest The author declares that there is no conflict of interest regarding the publication of this article.

References

Abiwinanda N, Hanif M, Hesaputra ST, Handayani A, Mengko TR (2019) Brain tumor classification using convolutional neural network. IFMBE Proc 68(1):183–189. https://doi.org/10.1007/978-981-10-9035-6_33

Ayadi W, Elhamzi W, Charfi I, Atri M (2021) Deep CNN for brain tumor classification. Neural Process Lett 53(1):671–700. https://doi.org/10.1007/s11063-020-10398-2

Badža MM, Barjaktarović MC (2020) Classification of brain tumors from MRI images using a convolutional neural network. Appl Sci 10(6):1–13. https://doi.org/10.3390/app10061999

Banan R, Hartmann C (2017) The new WHO 2016 classification of brain tumors—what neurosurgeons need to know. Acta Neurochir 159(3):403–418. https://doi.org/10.1007/s00701-016-3062-3

Barboriak D (2015) Data from RIDER_NEURO_MRI. Cancer Imag Arch. https://doi.org/10.7937/K9/TCIA.2015.VOSN3HN1

Cheng J, Huang W, Cao S, Yang R, Yang W, Yun Z, Wang Z, Feng Q (2015) Enhanced performance of brain tumor classification via tumor region augmentation and partition. PLoS ONE 10(10):1–13. https://doi.org/10.1371/journal.pone.0140381

Çinar A, Yildirim M (2020) Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture. Med Hypotheses 139:109684. https://doi.org/10.1016/j.mehy.2020.109684

Clark K, Vendt B, Smith K, Freymann J, Kirb J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M (2013) The cancer imaging archive (TCIA): Maintaining and operating a public information repository. J Digit Imag 26(6):1045–1057. https://doi.org/10.1007/s10278-013-9622-7

Deepak S, Ameer P (2019) Brain tumor classification using deep CNN features via transfer learning. Comput Biol Med 111:103345. https://doi.org/10.1016/j.compbiomed.2019.103345

Doğantekin A, Özyurt F, Avcı E, Koç M (2019) A novel approach for liver image classification PH-C-ELM. Measurement 137:332–338. https://doi.org/10.1016/j.measurement.2019.01.060

El-Dahshan ESA, Hosny T, Salem ABM (2010) Hybrid intelligent techniques for MRI brain images classification. Digital Signal Process 20(2):433–441. https://doi.org/10.1016/j.dsp.2009.07.002

Ertosun MG, Rubin DL (2015) Automated grading of gliomas using deep learning in digital pathology images: a modular approach with ensemble of convolutional neural networks. Annu Symp Proc AMIA Symp 2015:1899–1908

Irmak E (2020) Implementation of convolutional neural network approach for COVID-19 disease detection. Physiol Genom 52(12):590–601. https://doi.org/10.1152/physiolgenomics.00084.2020

Irmak E (2021) COVID-19 disease severity assessment using CNN model. IET Image Process. https://doi.org/10.1049/ipr2.12153

Kabir Anaraki A, 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. https://doi.org/10.1016/j.bbe.2018.10.004

Khan HA, Jue W, Mushtaq M, Mushtaq MU (2020) Brain tumor classification in MRI image using convolutional neural network.

Math Biosci Eng 17(5):6203–6216. https://doi.org/10.3934/MBE.2020328

Khawaldeh S, Pervaiz U, Rafiq A, Alkhawaldeh RS (2017) Noninvasive grading of glioma tumor using magnetic resonance imaging with convolutional neural networks. Appl Sci 8(1):1–17. https://doi.org/10.3390/app8010027

Kleihues P, Burger PC, Scheithauer BW (2012) Histological typing of tumours of the central nervous system, 2nd edn. Springer, Berlin

Lisa S, Flanders Adam E, Mikkelsen JR, Tom Andrews DW (2015) Data From REMBRANDT. Cancer Imag Arch. https://doi.org/10.7937/K9/TCIA.2015.588OZUZB

Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, Vander Laak JAWM, Van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. Med Image Anal 42:60–88. https://doi.org/10.1016/j.media.2017.07.005

Lotan E, Jain R, Razavian N, Fatterpekar GM, Lui YW (2019) State of the art: Machine learning applications in glioma imaging. Am J Roentgenol 212(1):26–37

Mehmood A, Maqsood M, Bashir M, Shuyuan Y (2020) A deep siamese convolution neural network for multi-class classification of alzheimer disease. Brain Sci 10(2):1–15. https://doi.org/10.3390/brainsci10020084

Mehmood A, Yang S, Feng Z, Wang M, Ahmad ALS, Khan R, Maqsood M, Yaqub M (2021) A transfer learning approach for early diagnosis of alzheimer's disease on MRI images. Neuroscience 15(460):43–52. https://doi.org/10.1016/j.neuroscience.2021.01.002

Mehrotra R, Ansari MA, Agrawal R, Anand RS (2020) A Transfer learning approach for AI-based classification of brain tumors. Mach Learn Appl 2(9):1–12. https://doi.org/10.1016/j.mlwa.2020.100003

Mohsen H, El-Dahshan ESA, El-Horbaty ESM, Salem ABM (2018) Classification using deep learning neural networks for brain tumors. Future Comput Informat J 3(1):68–71. https://doi.org/10.1016/j.fcij.2017.12.001

Muhammad K, Khan S, Ser JD, Albuquerque VHC (2021) Deep learning for multigrade brain tumor classification in smart healthcare systems: a prospective survey. IEEE Trans Neural Netw Learn Syst 32(2):507–522. https://doi.org/10.1109/TNNLS.2020.2995800

Mzoughi H, Njeh I, Wali A, Slima M, Ben BenHamida A, Mhiri C, Mahfoudhe K (2020) Deep Multi-Scale 3D convolutional neural network (CNN) for MRI gliomas brain tumor classification. J Digit Imag 33(4):903–915. https://doi.org/10.1007/s10278-020-00347-9

National Cancer Institute. (2020). cancer.org

Özyurt F, Sert E, Avci E, Dogantekin E (2019) Brain tumor detection based on Convolutional Neural Network with neutrosophic expert maximum fuzzy sure entropy. Measurement 147(106803):1–7. https://doi.org/10.1016/j.measurement.2019.07.058

Papageorgiou EI, Spyridonos PP, Glotsos DT, Stylios CD, Ravazoula P, Nikiforidis GN, Groumpos PP (2008) Brain tumor characterization using the soft computing technique of fuzzy cognitive maps. Appl Soft Comput J 8(1):820–828. https://doi.org/10.1016/j.asoc.2007.06.006

Pedano N, Flanders AE, Scarpace L, Mikkelsen T, Eschbacher JM, Hermes B, Ostrom Q (2016) Radiology data from the cancer genome atlas low grade glioma [TCGA-LGG] collection. Cancer Imag Arch. https://doi.org/10.7937/K9/TCIA.2016.L4LTD3TK

Pereira S, Meier R, Alves V, Reyes M, Silva CA (2018) Automatic brain tumor grading from MRI data using convolutional neural networks and quality assessment. Understanding and interpreting machine learning in medical image computing applications. Springer, Cham, pp 106–114

Rehman A, Naz S, Razzak MI, Akram F, Imran M (2020) A Deep Learning-based framework for automatic brain tumors classification using transfer learning. Circuits, Syst, Signal Process 39(2):757–775. https://doi.org/10.1007/s00034-019-01246-3

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. https://doi.org/10.1016/j.jocs.2018.12.003

Seetha J, Raja SS (2018) Brain tumor classification using convolutional neural networks. Biomed Pharmacol J 11(3):1457–1461

Shaver MM, Kohanteb PA, Chiou C, Bardis MD, Chantaduly C, Bota D, Filippi CG, Weinberg B, Grinband J, Chow DS, Chang PD (2019) Optimizing neuro-oncology imaging: a review of deep learning approaches for glioma imaging. Cancers 11(6):1–14. https://doi.org/10.3390/cancers11060829

Shirazi AZ, Fornaciari E, McDonnell MD, Yaghoobi M, Cevallos Y, Tello-Oquendo L, Inca D, Gomez GA (2020) The application of deep convolutional neural networks to brain cancer images: A survey. J Personal Med 10(4):1–27. https://doi.org/10.3390/jpm10040224

Sultan HH, Salem NM, Al-Atabany W (2019) Multi-classification of brain tumor images using deep neural network. IEEE Access 7:69215–69225. https://doi.org/10.1109/ACCESS.2019.2919122

Talo M, Baloglu UB, Yıldırım Ö, Rajendra Acharya U (2019) Application of deep transfer learning for automated brain abnormality classification using MR images. Cogn Syst Res 54(12):176–188. https://doi.org/10.1016/j.cogsys.2018.12.007

Tandel GS, Biswas M, Kakde OG, Tiwari A, Suri HS, Turk M, Laird JR, Asare CK, Ankrah AA, Khanna NN, Madhusudhan BK, Saba L, Suri JS (2019) A review on a deep learning perspective in brain cancer classification. Cancers 11(1):1–32. https://doi.org/10.3390/cancers11010111

Tiwari A, Srivastava S, Pant M (2020) Brain tumor segmentation and classification from magnetic resonance images: review of selected methods from 2014 to 2019. Pattern Recogn Lett 131:244–260. https://doi.org/10.1016/j.patrec.2019.11.020

Yaqub M, Jinchao F, Zia MS, Arshid K, Jia K, Rehman ZU, Mehmood A (2020) State-of-the-art CNN optimizer for brain tumor segmentation in magnetic resonance images. Brain Sci 10(7):1–19. https://doi.org/10.3390/brainsci10070427

Yang Y, Yan LF, Zhang X, Han Y, Nan HY, Hu YC, Hu B, Yan SL, Zhang J, Cheng DL, Ge XW (2018) Glioma grading on conventional MR images: a deep learning study with transfer learning. Front Neurosci 12:804. https://doi.org/10.3389/fnins.2018.00804