

Automatic Brain Cancer Classification in MRI Images Based On Ensemble Light Weight Deep Learning Network

Jayachandran A (✉ ajaya1675@gmail.com)

Presidency University <https://orcid.org/0000-0002-8185-3237>

N. Zafar Ali Khan

Presidency University

Sudarson Rama Perumal

Rohini College of Engineering and Technology

Research Article

Keywords: Brain tumor, classification, Convolutional Neural Network, Two stage ensemble, Magnetic Resonance Imaging

Posted Date: February 28th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1352649/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

The brain tumor is one of the deadliest cancerous diseases and its severity has turned it to the leading cause of cancer related mortality. Automatic detection and classification of severity-level for a brain tumor using MRI is a complex process in multilevel classification and needs an improved learning method without computational complexity. In this research article, we propose a new encoder decoder model, which combines nested connection and efficient attention mechanism. The structure of the model consists of encoder, decoder and skip connection. By introducing the DropBlock regularization strategy, the local semantic information can be discarded more effectively to motivate the network to learn more robust and effective features. Efficient attention module uses appropriate cross channel interaction to capture richer global information. In the skip connection part, the nested connection strategy is adopted to effectively fuse the feature maps gathered from the intermediate decoder and the original feature maps from the encoder, which makes up for the semantic gap caused by direct simple connection. Finally, data augmentation is performed on the original image to improve the robustness and prevent the over-fitting problem caused by insufficient data. The proposed method is evaluated on different evaluation metrics such as sensitivity, specificity and accuracy. The experimental results show that this method shows the most competitive performance.

1. Introduction

Magnetic resonance imaging (MRI) is an imperative image utilized in medical practice as it provides images with elevated contrast. MRI is a well-known medical modality, which is utilized for examination and diagnosis of several neurological diseases that involves epilepsy, sclerosis and brain tumor (Gopal et al., 2020). The system fully managed by computer assist in automating the process for receiving precise and rapid outcomes. In MRI images analysis, the detection and classification of the pretentious region is lightheaded and time consuming. The structure of brain is distinguished and illuminated using MRI and has several imaging methods for examining and observing the interior structure of brain. Imaging tests carried out on MRI use expertise for provoking comprehensive brain images. CT or MRI is utilized for examining structure of brain. MRI scan is more effectual when contrasted using CT scan and does not utilize radiation (Muhammad et al., 2019). For cancer diagnosis, it is imperative for segmenting tissues of brain using MRI (Jayachandran et al., 2013a). MRI is a decisive unit in treatment and for planning treatment, which give considerably maximized facts regarding structures. Due to soft tissues, and highly spatial decree, it failed to generate destructive radiation and is considered as non-invasive method. MRI is effective in contrast to other imaging methods as it offers effective identification of brain tumour. Generally, the radiologist utilizes MRI for segmenting the tumorous regions (Jayachandran et al., 2013b and Deepak et al., 2019).

Brain is the significant part of our central nervous system that controls all our functionalities through a huge number of connected neurons(Gumaei et al., 2019). Any malfunction or abnormality in the brain cells affects the brain lesions connected to the corresponding part of the brain, consequently damaging the functionalities of that organ. Cancer originating in the brain and other nervous system is considered to be the 10th leading cause of death. The 5-year survival rate of the patients having cancerous brain is only 36% (Jayachandran et al., 2017). As brain tumor is caused by the unnatural and uncontrolled growth of brain

cells, its severe consequences can be life-threatening. Around 400,000 people are affected by brain tumor and 120,000 people have died in the past years all over the world, as reported by World Health organization (WHO) (Kavitha A and Chellamuthu C 2018). Early and proper detection can play an indispensable role in increasing the survival rate by accelerating the treatment process(Tonmoy et al., 2019).

Many radiologists state that diagnosing brain-related problems is challenging, especially when the tumour infiltrates into complex tissue structures. It is also tough to distinguish tumour affected region and the edema region present around it due to the intricacies of the brain. Oncologists and radiologists require an advanced computerized diagnosing system to enhance images' visualization for accurate anomaly detection and infectious tissue segmentation (Narmada et al., 2017). Mostly, Magnetic Resonance Image (MRI) is preferred by clinical experts for the diagnosis purpose as it offers more information compared to other imaging modalities like Computed Tomography (CT), Positron Emission Tomography (PET), and Ultrasound. However, they find it difficult to handle large numbers of MR image slices during diagnosis and treatment procedures; due to this, manual segmentation processing time goes long and is subject to inter and intra variations of brain structures, which may lead to inaccurate segmentation(Ali Pashaei et al., 2018 and Emrah Irmak 2021).

Difficulty in categorizing brain lesions and other tissue regions into several isolation parts without human intervention can be avoided by using automated segmentation methodology with specialized graphical tools. Manual detection of brain tumors can be tedious, time consuming and erroneous due to the variations on types and sizes. Proper and precise detection needs expertise and it's even harder for complicated cases. Hence, besides human inspection, we can't avoid the necessity of an automated process for precise detection and classification of brain tumors. Deep learning and convolutional neural network can significantly accelerate the whole diagnosis process making the classification task automated and conscientious. Our Light Weight Deep Convolutional Neural Network work (UNet++) takes full advantage of two networks with different dimensions, which can balance the complete semantic information and high-resolution detail information of a large-volume MRI image. The main contributions of the study are as follows:

- The proposed encoder-decoder model discards the local semantic information by introducing DropBlock (Ghiasi et al., 2018) regularization method, and encourages the network to learn more robust features. At the same time, an efficient attention module is used to capture more abundant global information.
- For the first time, the nested connected method is applied to vessel segmentation, which can effectively fuse the feature maps collected from the intermediate decoder and the original feature maps collected from the former encoder.
- In order to improve the robustness of the system, image augmentation strategies such as rotation, mirror image, scaling and contrast adjustment are adopted for the original data, and the problems of model convergence and over fitting caused by insufficient data are solved.
- In this paper, an improved lightweight model (UNet++) is proposed, and its different variants are trained and tested. Finally, the best performance model (UNet++) is obtained by adding incremental strategy and applied to MRI image classification.

- Our method obtains satisfactory segmentation precision for all types of tumor images, especially for Glioma, Meningioma and Pituitary. Moreover, it produces an accurate and smooth boundary without post processing.

The main structure of this paper is as follows: In Section 2, the details of the existing work related to brain tumor. In Section 3, the details of the proposed network are provided. The materials and experimental designs are described in Section 4. The results and discussion are illustrated in Sections 5. Finally, a brief conclusion is given in Section 6.

2. Related Work

The emerging technologies of machine learning and deep learning have profoundly developed different fields of applications. Particularly, a huge scope has been created in medical image processing, and numerous researches are ongoing for enhancing this research area. Automating the process of brain tumor segmentation and classification is a significant part of this research field.

Recently, deep learning technology has become a very popular choice in brain tumor classification from brain MRI images. Gumaiei et al(2019) introduced a hybrid method for feature extraction called PCA-NGIST. Then, from the extracted features, brain images are classified into three categories (meningioma, glioma, pituitary) using Regularized Extreme Learning Machine classifier and achieved an accuracy of 94.233%. Tandel et al(2020) proposed a deep learning model based on CNN for classification of brain tumor from five individual multi-class datasets. The highest achieved accuracy was 96.65%. Sajjad et al(2019) proposed an architecture, where they used Input Cascade CNN for tumor segmentation and fine-tuned VGG-19 for three types of tumor classification, where the achieved accuracy was 94.58% for precise detection and classification of brain tumors. Deep learning and convolutional neural network can significantly accelerate the whole diagnosis process making the classification task automated and conscientious.

Das et al (2019) presented a comparatively shallow CNN based model for the classification of three types of brain tumors, which could attain an accuracy of only 94.39%. Afshar et al (2019) proposed Capsule Network architecture for three tumor types classification namely glioma, meningioma and pituitary, where they got an accuracy of 90.89%. Hemanth et al (2018) introduced a modified deep CNN architecture to address the computational complexity of deep CNN model, and classified brain tumors into four classes. In this way, they achieved an average accuracy of 96.4%. Badža et al (2020) presented a new CNN architecture for classifying three types of brain tumors. After 10- fold cross validation, they achieved an accuracy of 96.56%. Ayadi et al (2021) suggested a new CNN model for multi- classification of brain tumors and attained average accuracy of 94.74%.

Sultan et al (2019) proposed a deep CNN based model for classifying tumors into three labels, and also different grades of glioma were differentiated. For two studies, they achieved best accuracy of 96.13% and 98.7%, respectively. Researchers in (Anaraki et al., 2019, Abiwinanda et al., 2019 and Kumar et al., 2021) applied different approaches of CNN architectures for classifying the tumors in three labels (meningioma, glioma, and pituitary). They achieved an overall accuracy of 94.2%, 97.08% and 84.19%, respectively. Deepak et al(2021) presented a CNN-SVM based classification model for three types of brain tumors and attained

95.82% accuracy. Swati et al (2019) used VGG- 19 pre-trained model and during training they applied block wise fine tuning, which resulted in a classification accuracy of 94.82%. Above literature survey suggested that there is a great scope of experimenting with multi stage ensemble technique. One of the common limitations found from the analysis of the existing models is the lack of generalization capability and robustness. To the best of our knowledge, no prior studies have conducted a multi- or two-stage ensemble approach for classification, which motivated us to devise a novel two-stage ensemble approach of deep CNN models for classification of brain tumors from brain MRI.

3. Proposed Methodology

In this section, the framework of our multi-dimensional cascaded net and the structures of each part are described in detail. This study is committed to building a more efficient neural network structure to complete the automatic classification of brain tumor in MRI images. U-net is found to be a more effective approach as compared to other network architectures. U-net, evolved from the traditional CNN, was first designed and applied in 2015 to process biomedical images. As a general convolutional neural network, it focuses on image classification, where input is an image and output is a one label, but in biomedical imaging, it requires us not only to distinguish whether there is a disease or not, but also to localize the area of abnormality. U-net provides fast and precise image segmentation and is dedicated in solving this problem as it is able to localize and distinguish borders by doing classification on every pixel, so that the input and output share the same size. This work emphasizes on the classification module of the CAD system, which presents a deep learning-based UNet for automated detection of cancer in MRI images. We propose an efficient attention network with nested connections, as shown in Fig. 1. This architecture is based on the popular UNet, which is designed to work well with a small number of training samples. The network is an encoder-decoder structure.

3.1. NESTED CONNECTION

The success of UNet depends largely on skip connections. The connection combines the coarse-grained features of the decoder with the fine-grained features of the encoder by adding elements. Zhou et al(2020) redesigned the skip connection of UNet and proposed a new model Light weight UNet (LUNet). It uses nested connection instead of the original simple connection, which can capture more abundant features of multi-level, and then integrate the multi-level features by adding elements. In the case of nested connection, the encoder is not directly connected with the decoder after getting the final aggregation feature mapping. Instead, the nested connection is accessed first, so that the rich features captured earlier can be better preserved.

Let's take a closer look at nested connection. Let $x^{i,j}$ denote the output value of node $X^{i,j}$, index i denote the maximum pooling operation from the encoder, and index j denote the transpose convolution operation from the decoder. The formula of the concatenated feature maps is expressed as

$$x^{i,j} = \begin{cases} F(x^{input}) & i = 0, j = 0 \\ F(M(x^{i-1,j})) & i > 0, j = 0 \\ F([\![x^{i,k}]_{k=0}^{j-1}, T(x^{i+1,j-1})\!\!]) & i \geq 0, j > 0 \end{cases}$$

Where function $F(\cdot)$ consists of four operations, two convolution layers $M(\cdot)$ and $T(\cdot)$ two DropBlock [14] layers arranged alternately, and $[\]$ represent the Max-pooling layer and transpose convolution layer respectively, and \oplus denotes the concatenation layer. When node $i = 0$ and $j = 0$, the function $F(\cdot)$ is executed on the input original image. When nodes $i > 0$ and $j = 0$ indicate that there is only one input in the front layer. When nodes $i \geq 0$ and $j > 0$, they will receive $j + 1$ inputs. Take decoder node $X^{0,2}$ as an example, it has three inputs, which are the previous decoder node $X^{1,1}$, the adjacent decoder nested node $X^{0,1}$ and the skip connection node $X^{0,0}$. The collection of multiple feature maps will provide richer semantic information for the final segmentation.

3.2. EFFICIENT ATTENTION

Since Vaswani et al (2017) proposed a new structure that connects encoders and decoders through Self-Attention (SA), which has achieved great success in machine translation tasks. Subsequently, many works have applied SA to computer vision tasks and achieved good results. The Squeeze Excitation (SE) (Hu et al., 2020) module starts from the perspective of channels and constructs information features by fusing the channel information of the local receptive fields of each layer. The SE module significantly improves the performance of the current most advanced Convolutional Neural Network (CNN) while slightly increasing the computational cost. Fu et al(2019) proposed a dual attention (DA) network that combines spatial attention and channel attention for adaptive fusion of local and global features. The best performance is achieved in scene segmentation tasks.

Wang et al.(2020) found through in-depth research on channel attention that appropriate cross-channel interaction is very important for learning more effective feature mapping. Local cross-channel interaction can be achieved through one-dimensional convolution, and nonlinear mapping is used to adaptively determine the size of the convolution kernel. The Efficient Attention (EA) module can be flexibly put into the existing CNN, and its structural details are shown in Fig. 3.

The function of efficient attention module is to obtain more abundant channel information. Group convolution is a method to manually adjust the size of convolution kernel according to the number of channels. However, manual adjustment consumes too much resources, so an adaptive adjustment strategy is designed. There is a mapping $\theta(\cdot)$ between the size k of convolution kernel and the channel dimension C :

$$C = \theta(k)$$

The channel dimension C is usually set to a power of 2. Therefore, a nonlinear function $\theta(\cdot)$ is designed:

$$\theta(k) = 2(\alpha * k - \beta)$$

When the channel dimension C is given, the convolution kernel size k can be calculated adaptively by

$$k = \frac{\log(C)}{\alpha} + \frac{\beta}{\alpha}$$

$$k = O(k)$$

Where function $O(\cdot)$ takes the nearest odd number. The later experiment, we set α and β to 2 and 1 respectively. In this way, higher dimensional channel will get larger convolution kernel size, while low dimensional channel can get smaller convolution kernel size through nonlinear mapping.

3.3. DropBlock regularization

In 2012, the Hinton team proposed an effective regularization method called Dropout (Srivastava et al., 2014) which can be used to prevent overfitting. Dropout is widely used in fully connected layers, but it is usually not effective for convolutional layers. The reason is probably because the adjacent elements in the feature map of the convolutional layer share semantic information in space, so although a unit is discarded, the adjacent elements can still retain the semantic information of the position, and the information can still be Circulate in convolutional networks. The mainstream network model that solves the segmentation problem just does not include a fully connected layer, so Google Brain proposed a DropBlock (Ghiasi et al., 2018) regularization method for this situation, which is a simple method similar to Dropout.

The main difference with Dropout is that it removes adjacent regions from the feature map of a layer instead of discarding independent random units. The first parameter μ of DropBlock is the size of the block to be dropped, and the second parameter γ controls the number of active units to be dropped. Similar to Dropout, we do not apply DropBlock in the inference process. In the experiment, we set a constant for all feature maps, regardless of the resolution of feature maps. When $\mu = 1$, DropBlock and Dropout are equivalent, when covering the entire feature map. For the setting of the value of γ , we assume that we want each activation unit to maintain a probability of p_{keep} . γ can be calculated as:

$$\gamma = \frac{\lambda^2(1 - p_{keep})}{\mu^2(\lambda - \mu + 1)^2}$$

In order to ensure that the μ^2 region is contained in the λ^2 region, we adjust the initial binary mask when sampling, so that the size of the effective seed region is $(\lambda - \mu + 1)^2$, and set p_{keep} as 0.9 and μ as 7.

3.4. FUSION LOSS FUNCTION

The most common loss function used for image semantic segmentation tasks is pixel-level cross-entropy loss. This loss function discriminates each pixel, and then compares the result with the one-hot encoding vector. Binary Cross Entropy (BCE) loss is for the case where there are only two categories, and its loss function formula is:

$$L = -\frac{1}{N} \sum_{i=1}^N [g_i \log(p_i) + (1 - g_i) \log(1 - p_i)]$$

Since the cross-entropy loss evaluates the category prediction of each pixel separately, and then averages the loss of all pixels, we essentially learn equally for each pixel in the image. If the distribution of multiple classes in the image is unbalanced, then this may lead to the training process being dominated by classes with a large number of pixels. The model will mainly learn the features of the large number of class samples, and the learned model will be more biased to predict the pixels for this category.

In order to overcome the problem that the foreground region is difficult to detect completely. Milletari et al(2016) proposed a new loss function based on dice coefficient (Dice). The expression of the dice coefficient D is

$$D = \frac{2 \sum_{i=1}^N p_i g_i}{\sum_{i=1}^N p_i^2 + \sum_{i=1}^N g_i^2}$$

Where $p_i = P$ represents the predicted binary vector and $g_i \in G$ represents the ground truth binary vector. The total number of pixels is N.

$$\frac{\partial D}{\partial p_j} = 2 \left[\frac{g_j \left(\sum_{i=1}^N p_i^2 + \sum_{i=1}^N g_i^2 \right) - 2p_j \left(\sum_{i=1}^N p_i g_i \right)}{\left(\sum_{i=1}^N p_i^2 + \sum_{i=1}^N g_i^2 \right)^2} \right]$$

Using this formula, we can establish an appropriate balance without assigning different weights to the parts of interest and background. In the fundus blood vessel segmentation task, we fuse the two loss functions of BCE loss and Dice loss by simple addition, which we call the fusion loss function.

4. Materials And Result Analysis

4.1 Material

In this article, three unique datasets have been used (Brain Tumor Dataset 1, Brain Tumor Dataset 2, and Brain Tumor Dataset 3). Dataset 1 contains total 3,064 T1-weighted contrast-enhanced images (where fat tissue is highlighted) from 233 patients with three kinds of brain tumor: meningioma (708 slices), glioma (1,426 slices), and pituitary tumor (930 slices). It includes axial, coronal and sagittal views of all the tumors. This dataset was publicly available (Jun Cheng 2021). Data set 2 contains total 3,264 MRI images of three types of brain tumor: meningioma (937 slices), glioma (926 slices), pituitary (901 slices) and normal brain tissue (500 slices). This images are the combination of T1 (type of MRI where fat tissue is highlighted and seems brighter), T2 (type of MRI where fat tissue and water are highlighted and seem brighter) and Flair types (same as T2 with free flowing water and fat seem dark), which is available(Sartaj Bhuvaji 2021). Data set 3 contains total 4,292 MRI images of three types of brain tumor: meningioma (1318 slices), glioma (1,038 slices), pituitary (1,255 slices) and normal brain tissue (681 slices). This is also publicly available (Mohamed 2021).

Minimum preprocessing is done on the datasets - image resizing and augmentation. All the images are resized to 256×256 pixels. Then six types of augmentation techniques (horizontal flipping, rotation, zoom, height shift, width shift, scaling) are applied on the dataset. Some samples from the datasets containing four types of brain MRI and the image preprocessing techniques are presented in Fig. 4(a) and 4(b), respectively.

4.2 Performance Metrics

For result analysis several performance metrics (precision, recall, f1-score, support, accuracy and ROC AUC score) are evaluated. Accuracy is the most common way to determine the performance of a classification model (Connor and Taghi 2019). It determines the ratio of the number of correct prediction to the total number of predictions. Precision is the ratio of true positive predictions and total positive predictions. Recall is the ratio of true positive and sum of true positive and false negative predictions. F1-score is the harmonic mean of precision and recall. ROC AUC score determines the model's capability to distinguish among different classes and higher value refers to a better performance. The formulas of determining precision, recall, f1 score and accuracy are shown in below Eqn.

Pr

In Eq. 6, 7, 8, 9, TP is the number of true positive predictions, TF is the number of true negative predictions, FP is the number of false positive predictions and FN is the number of false negative predictions. The result of our proposed ensemble CNN model is shown in Table 1 where the precision, recall, f1-score, support and accuracy for all individual datasets have been presented. Confusion matrix of each dataset is shown in Fig. 5. Here, 'M', 'G', 'P' and 'N' refer to Meningioma, Glioma, Pituitary and Normal brain images.

Table 1 Performance of the proposed two stage ensemble model considering different performance metrics

Dataset	Tumor labels	Precision	Recall	F1 Score	Validation	Validation	ROC AUC
					Accuracy %	Loss	
Dataset 1	Meningioma	1.00	0.99	1.00	99.67	0.024	0.994
	Glioma	0.99	1.00	0.99			
	Pituitary	1.00	1.00	1.00			
Dataset 2	Meningioma	0.99	0.96	0.97	98.16	0.063	0.982
	Glioma	0.97	0.99	0.98			
	Pituitary	0.99	1.00	0.99			
	No Tumor	0.98	0.98	0.98			
Dataset 3	Meningioma	1.00	0.99	1.00	99.76	0.019	0.998
	Glioma	0.99	1.00	1.00			
	Pituitary	1.00	1.00	1.00			
	No Tumor	1.00	1.00	1.00			

4.3. Validating model's robustness and generalization capability

Having a high generalization capability is a prerequisite for a robust classification model. Lack of generalization capability makes a model incompatible for real-world application. To validate the robustness

and generalization capability, three experiments have been done, i.e., training and validation on separate datasets, random splitting of dataset and k-fold cross validation, as discussed below.

4.3.1. Training and validation on Separate Datasets

To validate the generalization capability of the model, we adopted the leave-one-out approach, where out of three datasets, i.e., Dataset 1, Dataset 2, and Dataset 3, two of them have been used as the training set and the other as the validation set. In this way, we can find three combinations for validating our proposed ensemble model. The result of applying this approach is shown in Table 2. In Fig. 6 (a), the validation accuracy plot is also shown for the three combination of training/validation dataset. It is very clear that our model has shown quite satisfactory performance in each of the cases.

Table 2
Performance of the proposed model considering the leave-one-out approach for training and validation with the Dataset 1, Dataset 2, and Dataset 3.

Training Set	Validation Set	Validation Accuracy (%)
Dataset1 + Dataset2	Dataset3	98.96
Dataset2 + Dataset3	Dataset1	99.13
Dataset1 + Dataset3	Dataset2	98.84

4.3.2. Random Splitting of Datasets

Next, we split our datasets into different ratios of training and test sets as shown in Table 3. This is done to check the performance of our model when the dataset is randomly divided in any ratio. According to the result we get, the performance does not degrade so much due to this random splitting. We have used 3 ratios here. 90% in the training set, 80% in the training set and 70% in the training set. For all cases the accuracy remains quite competitive on all datasets.

Table 3 Performance of the proposed model considering different ratio of training and validation set.

Dataset	Training testing data splitting	Accuracy %	Average accuracy %
Dataset 1	90%-10%	99.67	
	80%-20%	99.28	99.2
	70%-30%	98.67	
Dataset 2	90%-10%	98.16	
	80%-20%	98.11	98.05
	70%-30%	97.89	
Dataset 3	90%-10%	99.76	
	80%-20%	98.67	98.62
	70%-30%	97.44	

4.3.3. K-Fold cross validation

Lastly, in the third experiment in proving model’s robust-ness, the proposed model is verified by K-fold cross validation. It is a more appropriate approach to validate a model as it uses every observation in test set and training set. In this process the whole dataset is divided into k equal parts, here eventually each part is considered as test set and the remaining as the training set. That’s how this ensures that no observation is remained without being tested by the model. The value of k depends on the size of the dataset. We have applied 8-fold cross validation for all the datasets and the result of each fold is presented in Table 4. Figure 6 (b) illustrates the model’s performance after applying 8-fold cross validation.

Table 4 Performance of the proposed model after 8 fold cross validation.

Validation Accuracy %								
Dataset	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8
Dataset 1	99.73	99.73	98.95	100	100	99.47	99.73	99.7
Dataset 2	99.75	100	98.15	98.29	100	100	99.5	99.87
Dataset 3	99.62	99.44	99.62	99.81	99.62	100	100	99.76

4.3.4 Data Augmentation

It is well known that training deep neural networks usually needs as much annotation data as possible. In the real world, many tasks do not have enough data for training. To solve this problem, the most intuitive and effective idea is to use the method of data augmentation to increase the amount of data. One of the important prerequisites to build a robust model is to use a dataset, which is large enough and diverse in quantity. It’s not always possible to collect a rich dataset due to its unavailability. New and different data can be formed by slightly modifying the existing data, which is called data augmentation. In this approach

model's robustness and performance can improve and also over-fitting can be avoided (Connor et al., 2019). In this work, six types of augmentation have been applied on all the datasets to acquire a robust model. We can clearly see from Fig. 7 (a), Fig. 7 (b), Fig. 7 (c), how data augmentation improves the validation accuracy during training on Dataset-1, Dataset- 2 and Dataset-3 respectively.. After augmentation the validation accuracy increases 1.35%, 1.04%, 1.24% for Dataset 1, Dataset 2 and Dataset 3 respectively.

4.4. Comparison with Other Existing Models

The proposed model is compared with other existing models in terms of different evaluation metrics. Here, the comparison is done based on two aspects. In the first case, individual dataset is taken under consideration for making comparison. In the second case, comparison is done by training and validating on separate datasets which means considering all the datasets.

4.4.1. Comparison on individual dataset

In this case, all three datasets (Dataset-1, Dataset-2, and Dataset-3) have been considered individually. The performance comparison of the proposed model with some existing recent models considering Dataset 1 is shown in Table 5. Here, we can see that the proposed model outperformed all the existing models with the highest precision, recall, f1 score and validation accuracy. For Dataset 2 and Dataset 3 we rebuilt some existing models and then trained and validated them using Dataset 2 and Dataset 3 individually. The result after validation of the existing models on Dataset 2 and Dataset 3 is shown in Table 6 and Table 7 respectively. From the performance comparison presented in these tables, it is evident that the proposed model performed quite better than the existing models in terms of all evaluation metrics considering all the datasets. The validation accuracy of all the existing models along with the proposed model is depicted in Fig. 8 using bar graphs.

Table 5

Performance comparison of the proposed model with other existing models considering Dataset 1.

Ref.	Year	Method used	Precision	Recall	F1 score	Validation Accuracy %	Validation Loss
Badža et al.	2020	CNN	0.97	0.98	0.974	97.28	–
Ayadi et al.	2021	CNN	0.94	0.94	0.94	94.74	–
Sultan et al.	2019	Deep neural network	0.96	0.94	–	96.13	–
Anaraki et al.	2019	CNN	0.943	0.942	–	94.2	–
Kumar et al.	2021	Residual network and global average pooling	0.972	0.972	0.972	97.08	–
Abiwinanda et al.	2019	CNN	–	–	–	84.19	–
Deepak et al.	2020	CNN and SVM	0.957	0.949	–	95.82	–
Swati et al.	2019	Fine tuned VGG19	0.895	0.942	0.917	94.82	–
Proposed model	–	Two stage of CNN model	0.996	0.996	0.996	99.67	0.024

Table 6

Performance comparison of the proposed model with other existing models considering Dataset 2.

Ref.	Year	Method used	Precision	Recall	F1 score	Validation Accuracy %	Validation Loss
Badža et al.	2020	CNN	0.9	0.88	0.885	89.45	0.414
Ayadi et al.	2021	CNN	0.957	0.957	0.955	95.71	0.226
Sultan et al.	2019	Deep neural network	0.9	0.91	0.905	90.21	0.614
Anaraki et al.	2019	CNN	0.912	0.897	0.905	90.51	0.785
Kumar et al.	2021	Residual network and global average pooling	0.952	0.952	0.955	95.1	0.176
Abiwinanda et al.	2019	CNN	0.892	0.88	0.885	88.68	0.67
Deepak et al.	2020	CNN and SVM	0.92	0.92	0.9	90.21	0.853
Proposed model	–	Two stage of CNN models	0.98	0.98	0.98	98.16	0.063

Table 7

Performance comparison of the proposed model with other existing models considering Dataset 3.

Ref.	Year	Method used	Precision	Recall	F1 score	Validation Accuracy %	VValidation Loss
Badža et al.	2020	CNN	0.935	0.935	0.935	93.48	0.26
Ayadi et al.	2021	CNN	0.985	0.98	0.98	98.13	0.118
Sultan et al.	2019	Deep neural network	0.972	0.96	0.965	96.64	0.137
Anaraki et al.	2019	CNN	0.9	0.892	0.897	89.76	0.323
Kumar et al.	2021	Residual network and global average pooling	0.992	0.992	0.992	99.3	0.031
Abiwinanda et al.	2019	CNN	0.93	0.917	0.925	92.55	0.471
Deepak et al.	2020	CNN and SVM	0.94	0.92	0.93	93.02	0.819
Proposed model	–	Two stage of CNN models	0.997	0.997	1	99.76	0.019

4.4.2. Comparison on separate training and validation data

All the existing models are trained and validated on separate datasets. As, we have three unique datasets, taking one of them as validation set we get three combinations. After training and validating in this way, the validation accuracy is compared is given in Fig. 9., where it shows that the proposed model obtained the highest accuracy. Figure 9 also depicts that the proposed two stage ensemble model outperformed all the existing models. This proves that the proposed model has better generalization capability than the state-of-the-art models. The consequences of brain tumors can be very dreadful and life-threatening as they can cause cancer in the long term. To avoid the heinous effect, this paper proposes an automated classification system for early and precise diagnosis. To the best of our knowledge, it is the first time a two-stage ensemble of deep CNN models is used for classifying three different types of tumors and normal brain cells. The whole process of two-stage ensemble and classification has been implemented by analyzing and scrutinizing the best CNN models and classifiers so that the final model becomes robust for precise classification. The impact of applying a two-stage ensemble, data augmentation shows the significant improvement in the model's performance in different aspects. Moreover, three experiments conducted on the datasets considering random splitting, 8-fold cross-validation, and different validation set also prove the robustness and generalization capability of the proposed model. Besides, the comparison with other state-of-the-art models shows the promising performance of the proposed model.

4.5 Explore Lightweight Unet (LUNet)

In this section, we will introduce how the proposed method LUNet is explored step by step. In the previous section, the UNet++ model shows better brain tumor classification accuracy under the condition of low computation and parameters. Therefore, we try to further reduce the parameters of the model based on the UNet++ model. As we all know, the main parameters of convolutional neural network are concentrated in the convolution layer and the fully connected layer. However, there is no fully connected layer in the UNet model, so we reduce the number of output channels of all the convolution layers to $\frac{1}{4}$ of the original, and keep the other settings unchanged. We call it Lightweight U-shaped Network (LUNet). On the MRI image dataset, experiments are carried out with different numbers of nodes for UNet++ and LUNet network. It can be seen that compared with UNet++, LUNet has a significant reduction in the amount of parameters, calculation and model size. At the same time, both UNet++ and LUNet achieve the best performance when the number of network nodes is 3. Surprisingly, the segmentation accuracy of the lightweight LUNet is not reduced, but slightly higher than that of UNet++. LUNet can greatly reduce the amount of parameter calculation while maintaining high classification accuracy, which also proves that LUNet is very effective in MRI image classification. In order to further improve the segmentation performance, we will continue to explore in the following sections on the basis of the 3 node LUNet.

4.6 Parameter Selection of Dropblock

The DropBlock regularization method mentioned in methodology section involves two parameters μ and p_{keep} . In this part, we evaluate their impact on classification results and verify the effectiveness of the selected parameters. Therefore, we take LUNet as the backbone model, and train them by setting μ to 3 to 9 and p_{keep} to 0.5 to 0.9, respectively. The results are shown in Fig. 9. Experimental results show that when the p_{keep} value is 0.5, no matter how the μ changes, the final segmentation index AUC is lower than the other two values of p_{keep} 0.7 and 0.9, and fluctuates around 0.9600. With the increase of the value of p_{keep} , AUC also increases. When p_{keep} is 0.9, it usually performs best and fluctuates around 0.9800. At this time, the value μ is 7, which has reached the maximum AUC value of 0.9821. These results show that it is very important to choose an appropriate μ to inactivate the pixel block. If the μ is too small, it cannot play the role of regularization, but too large will lead to the loss of useful feature information. The experimental performance is the best when the parameter p_{keep} is 0.9, which is mainly because the backbone model itself is a lightweight model with less parameters. The relatively low p_{keep} value deactivates too many nodes in the network, resulting in the reduction of feature extraction ability. Therefore, keeping a relatively high p_{keep} plays a regularization role and ensures the feature extraction ability of the network. Therefore, keeping a high p_{keep} can not only play a regularization effect, but also ensure the feature extraction ability of the network. The experimental result of proposed method LUNet is given in Table 8.

Table 8

COMPARE THE SEGMENTATION PERFORMANCE OF UNET ++ AND LUNET WITH DIFFERENT NUMBER OF NODES, AS WELL AS THE CORRESPONDING MODEL SIZE, NUMBER OF PARAMETERS AND AMOUNT OF CALCULATION

Models	Node	Se	Sp	F1	ACC	AUC	Params	Model size	FLOPs
UNet++	2	0.6577	0.9904	0.7450	0.9630	0.9742	0.44M	5.15MB	0.89M
	3	0.6972	0.9889	0.7691	0.9640	0.9769	2.07M	23.77MB	4.13M
	4	0.7095	0.9868	0.7641	0.9635	0.9763	8.56M	98.07MB	17.11M
	5	-	-	-	-	-	34.51M	395.15 MB	69.01M
LUNet	2	0.6757	0.9868	0.7413	0.9611	0.9686	0.03M	0.39MB	0.06M
	3	0.7553	0.9841	0.7820	0.9653	0.9780	0.13M	1.58 MB	0.26M
	4	0.7383	0.9852	0.7758	0.9648	0.9765	0.54M	6.27MB	1.07M
	5	0.7367	0.9836	0.7678	0.9633	0.9730	2.16M	24.88MB	4.31M

5. Conclusion

The consequences of brain tumors can be very dreadful and life-threatening as they can cause cancer in the long term. To avoid the heinous effect, this paper proposes an automated classification system for early and precise diagnosis. To the best of our knowledge, it is the first time a two-stage ensemble of deep CNN models is used for categorizing three different types of tumors and normal brain cells. The whole process of two-stage ensemble and classification has been implemented by analyzing and scrutinizing the best CNN models and classifiers so that the final model becomes robust for precise classification. Moreover, three experiments conducted on the datasets considering random splitting, 8-fold cross-validation, and different validation set also prove the robustness and generalization capability of the proposed model. Besides, the comparison with other state-of-the-art models shows the promising performance of the proposed model. In the future, we will expand the dataset and the number of annotated organs to improve the generalization performance of the algorithm. Semi-supervised or weakly-supervised methods will also be explored to reduce the demand for manual annotation. We will strive to push our method into practical application, providing guidance for doctors in disease diagnosis or surgical implementation.

Declarations

Funding Information

This work was not supported from any Funding Agencies

Conflict of interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

ETHICAL APPROVAL STATEMENT

This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Abiwinanda N, Muhammad Hanif ST, Hesaputra (2019) Astri Handayani, and Tati Rajab Mengko. Brain tumor classification using convolutional neural network. World congress on medical physics and biomedical engineering 2018. Springer, pp 183–189
2. Afser P, Konstantinos N (2019) Plataniotis, and Arash Mohammadi. Capsule networks for brain tumor classification based on mri images and coarse tumor boundaries. In ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1368–1372. IEEE,
3. Pashaei A (2018) Hedieh Sajedi, and Niloofar Jazayeri. Brain tumor classification via convolutional neural network and extreme learning machines. In 2018 8th International conference on computer and knowledge engineering (ICCKE), pages 314–319. IEEE,
4. Kabir AA, Ayati M, Kazemi F (2019) Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *biocybernetics and biomedical engineering* 39(1):63–74
5. Ayadi W, Wajdi, Elhamzi (2021) Imen Charfi, and Mohamed Atri. Deep cnn for brain tumor classification. *Neural Process Lett* 53(1):671–700
6. Badža Milica and Marko Barjaktarović (2020) Classification of brain tumors from mri images using a convolutional neural network. *Applied Sciences* 10(6):1999
7. Connor Shorten, Taghi M, Khoshgoftaar (2019) A survey on image data augmentation for deep learning. *J Big Data* 6(1):1–48
8. Connor Shorten, Taghi M, Khoshgoftaar (2019) A survey on image data augmentation for deep learning. *J Big Data* 6(1):1–48
9. Das S (2019) OFM Riaz Rahman Aranya, and Nishat Nayla Labiba. Brain tumor classification using convolutional neural network. In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), pages 1–5. IEEE,
10. Deepak S, Ameer PM (2021) Automated categorization of brain tumor from mri using cnn features and svm. *J Ambient Intell Humaniz Comput* 12(8):8357–8369
11. Deepak S, Ameer PM (2019) Brain tumor classification using deep cnn features via transfer learning. *Comput Biol Med* 111:103345
12. Emrah Irmak (2021) Multi-classification of brain tumor mri images using deep convolutional neural network with fully optimized framework. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, pages 1–22,
13. Fu J, Liu H, Tian Y, Li Y, Bao Z, Fang, Lu H (2019) “Dual Attention Network for Scene Segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3141–3149,
14. Ghiasi T-Y, Lin, Le QV (2018) “DropBlock: A regularization method for convolutional networks,” *arXiv:1810.12890*,

15. Gopal S, Tandel A, Balestrieri T, Jujaray NN, Khanna L, Saba, Jasjit S, Suri (2020) Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm. *Comput Biol Med* 122:103804
16. Gumaei A (2019) Mohammad Mehedi Hassan, Md Rafiul Hassan, Abdulhameed Alelaiwi, and Giancarlo Fortino. A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification. *IEEE Access* 7:36266–36273
17. Hemanth J, Anitha A, Naaji O, Geman DE, Popescu et al (2018) A modified deep convolutional neural network for abnormal brain image classification. *IEEE Access* 7:4275–4283
18. Hu J, Shen L, Albanie S, Sun G, Wu E “Squeeze-and-Excitation Networks,”*IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, pp. 2011– 2023, Aug. 2020.
19. Jayachandran A and R. Dhanasekaran, “Multi class brain tumor classification of MRI images using hybrid structure descriptor and fuzzy logic based RBF kernel SVM”,*Iranian Journal of Fuzzy Systems*14 (3),41–54,2017
20. Jayachandran A, Dhanasekaran R (2013a) Automatic detection of brain tumor in magnetic resonance images using multi-texton histogram and support vector machine||. *Int J Imaging Syst Technol* 23(2):97–103
21. Jayachandran A and R.Dhanasekaran (2013) ‘ Brain Tumor Detection using Fuzzy Support Vector Machine Classification based on a Texton Co-occurrence Matrix’,*Journal of imaging Science and Technology*, Vol 57, No 1, pp. 10507-1-10507-7(7),2013b.
22. Jun Cheng (2017) brain tumor dataset. https://figshare.com/articles/dataset/brain_tumor_dataset/1512427,
23. Kavitha A, Chellamuthu C (2018) Advanced brain tumour segmentation from mri images. *Basic Physical Principles and Clinical Applications. High-Resolution Neuroimaging*, pp 83–108
24. Kumar Lokesh R, Jagadeesh Kakarla BV, Isunuri, Singh M (2021) Multi-class brain tumor classification using residual network and global average pooling. *Multimedia Tools and Applications* 80(9):13429–13438
25. Milletari F, Navab N, Ahmadi S (2016) “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” in *Proc.- Int. Conf. 3D Vis.*, pp. 565–571,
26. Milletari F, Navab N, Ahmadi S (2016) “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” in *Proc.- Int. Conf. 3D Vis.*, pp. 565–571,
27. Mohamed Metwaly Sherif (2020) brain tumor dataset. <https://www.kaggle.com/mohamedmetwalysheerif/braintumordataset>,
28. Muhammad Sajjad S, Khan K, Muhammad W, Wu (2019) Amin Ullah, and Sung Wook Baik. Multi-grade brain tumor classification using deep cnn with extensive data augmentation. *J Comput Sci* 30:174–182
29. Narmada M, Balasooriya, Ruwan D, Nawarathna (2017) *IEEE International Conference on Industrial and Information Systems (ICIIS)*, pages 1–5. IEEE, 2017
30. Sajjad M, Salman Khan K, Muhammad W, Wu (2019) Amin Ullah, and Sung Wook Baik. Multi-grade brain tumor classification using deep cnn with extensive data augmentation. *J Comput Sci* 30:174–

31. Sartaj Bhuvaji (2020) Brain Tumor Classification (MRI). <https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri>,
32. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhut-dinov R (1958) "Dropout: a simple way to prevent neural networks from over-fitting," The Journal of Machine Learning Research, vol. 15, pp. 1929– Jan. 2014
33. Sultan H, Nancy M, Salem, Al-Atabany W (2019) Multi-classification of brain tumor images using deep neural network. IEEE Access 7:69215–69225
34. Khan SZarN, Zhao Q, Kabir M, Ali F, Ali Z, Ahmed S, Lu J (2019) Brain tumor classification for mr images using transfer learning and fine-tuning. Comput Med Imaging Graph 75:34–46
35. Tandel S, Gopal A, Balestrieri T, Jujaray NN, Khanna L, Saba, Jasjit S, Suri (2020) Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm. Comput Biol Med 122:103804
36. Tonmoy Hossain FS, Shishir M, Ashraf MD (2019) Abdullah Al Nasim, and Faisal Muhammad Shah. Brain tumor detection using convolutional neural network. In 2019 1st international conference on advances in science, engineering and robotics technology (ICASERT), pages 1–6. IEEE,
37. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (June 2020) UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. IEEE Trans Med Imaging 39:1856–1867

Figures

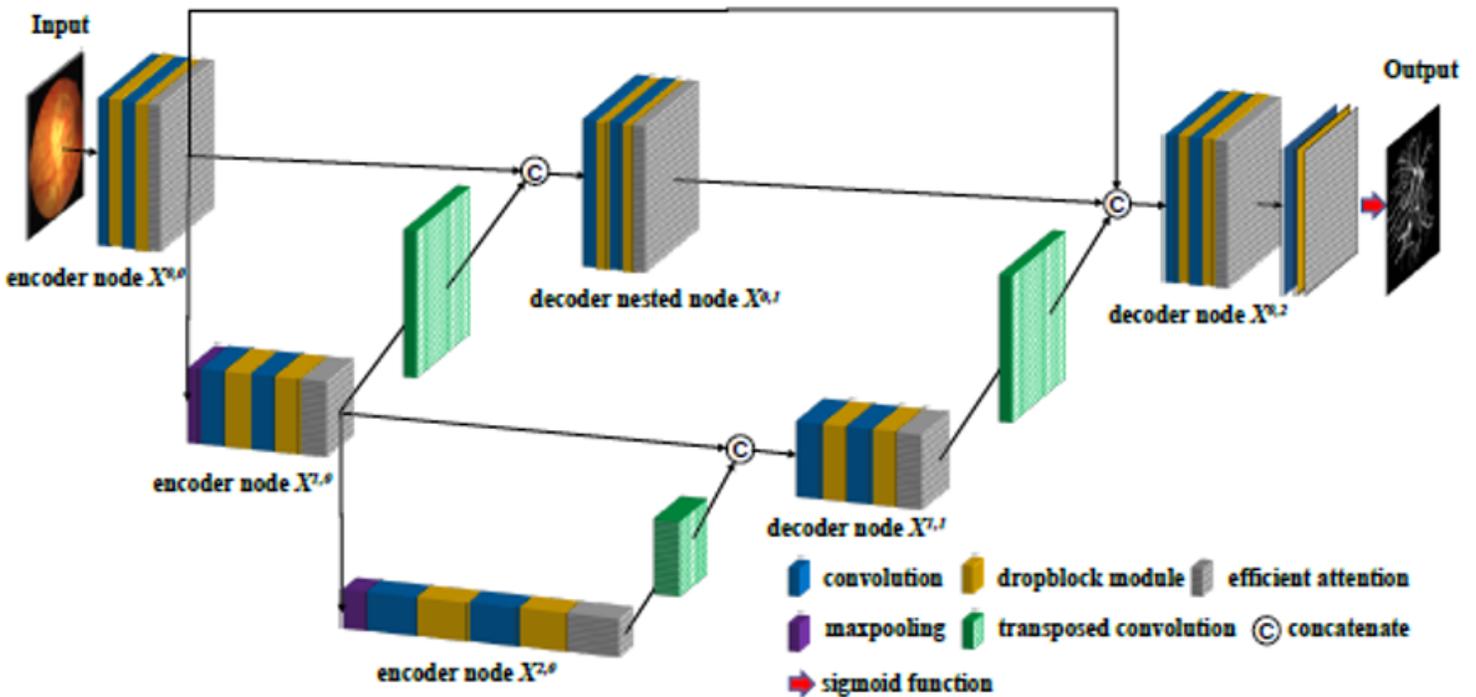


Figure 1

Structure of the proposed nested connection combined with efficient attention model (LUNet).

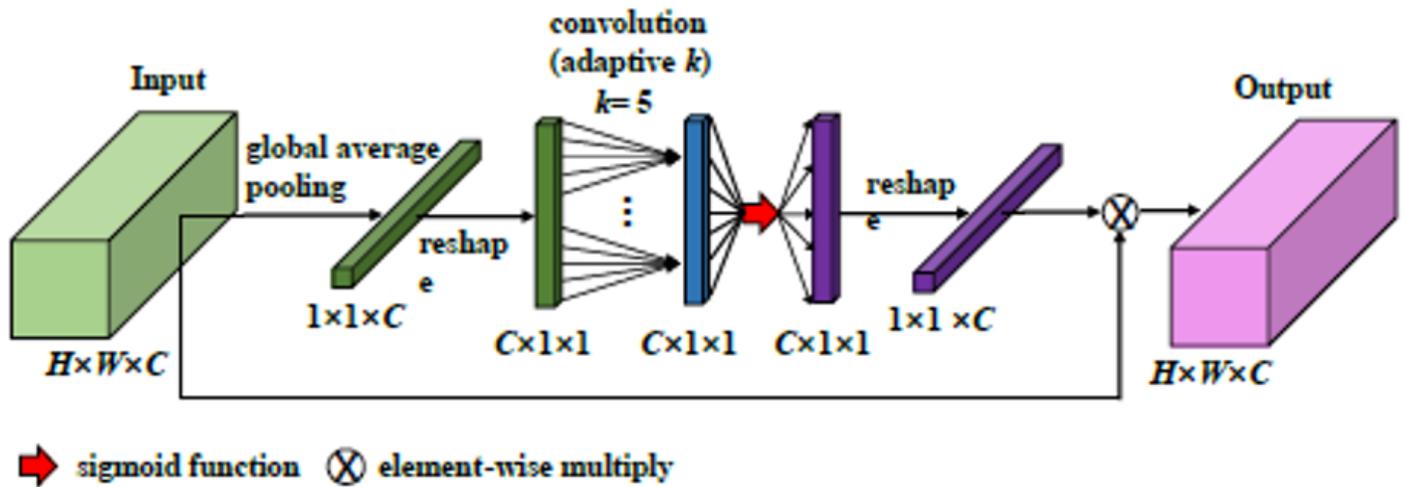


Figure 2

The structure of efficient attention module.

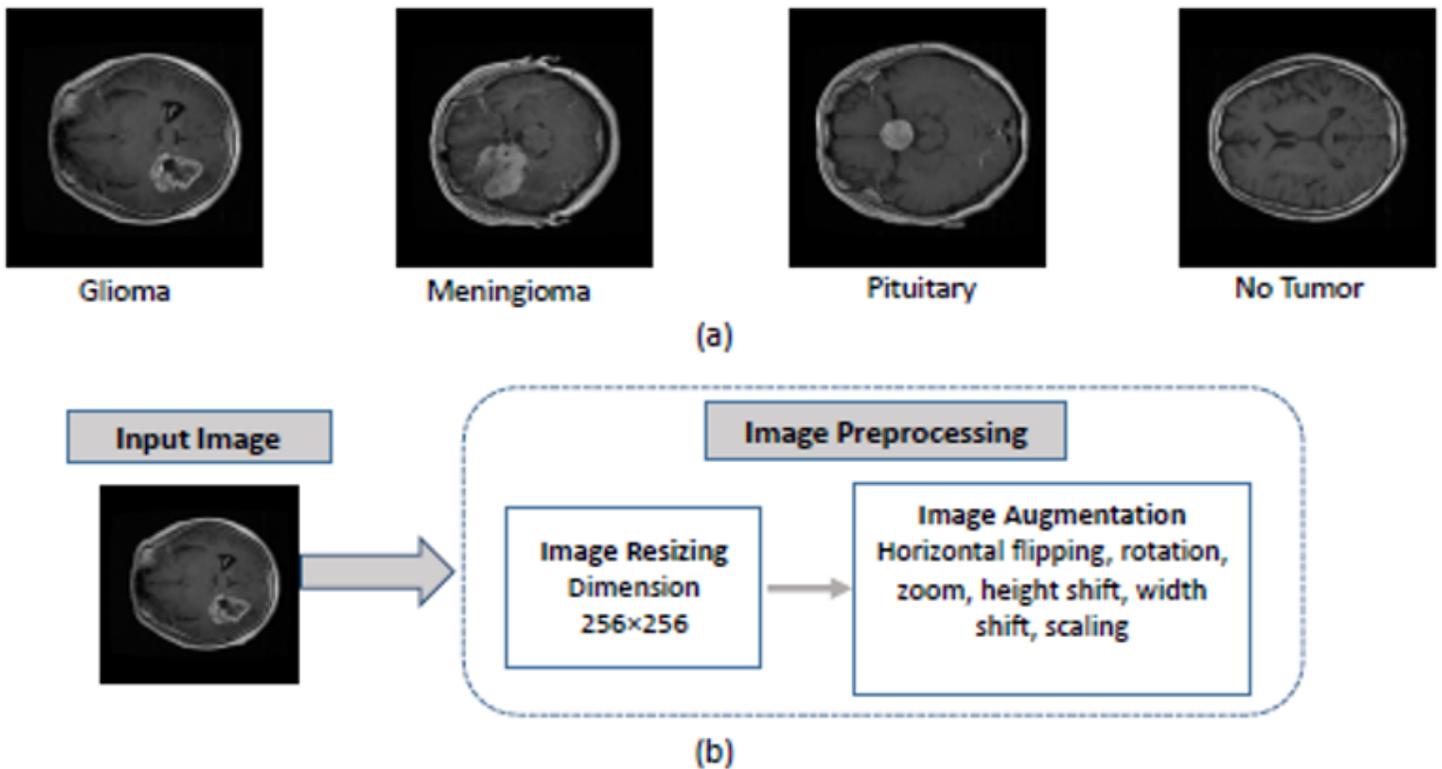


Figure 3

(a) Samples of the MRI datasets including glioma tumor, meningioma tumor, pituitary tumor and normal brain image. (b) Data preprocessing stages where image resizing and augmentation is done

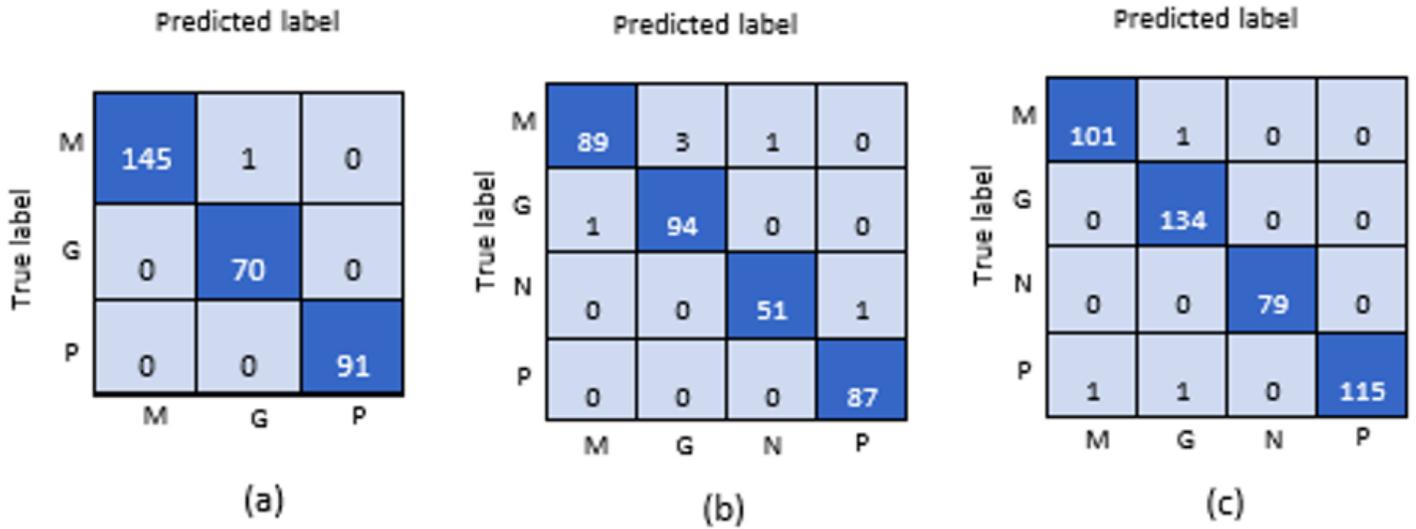


Figure 4

Confusion matrix after applying the proposed two stage ensemble model. Confusion matrices for (a) Dataset 1, (b) Dataset 2, (c) Dataset 3.

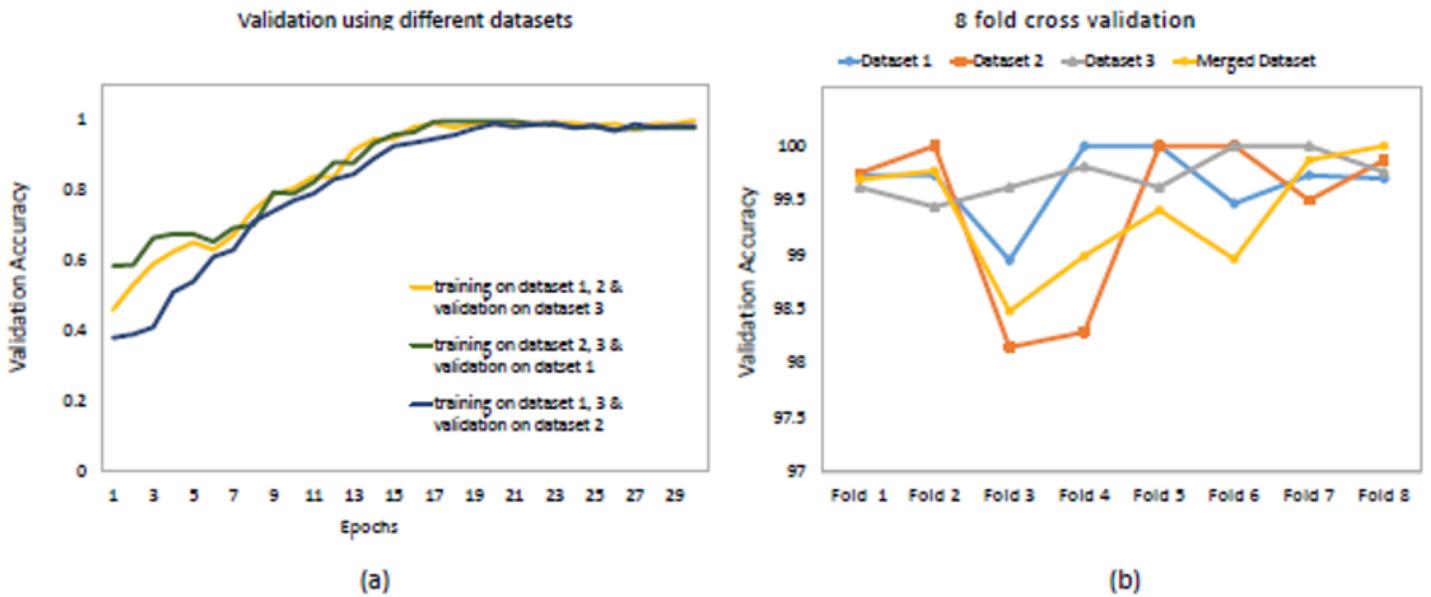


Figure 5

(a) Training and validation on separate datasets. Here, 3 cases are considered i) training with Dataset 1, 2 and validating with Dataset 3, ii) training with Dataset 2, 3, and validating with Dataset 1, and iii) training with Dataset 1, 3, and validating with Dataset 2. (b) Model accuracy after 8-fold cross validation for all datasets.

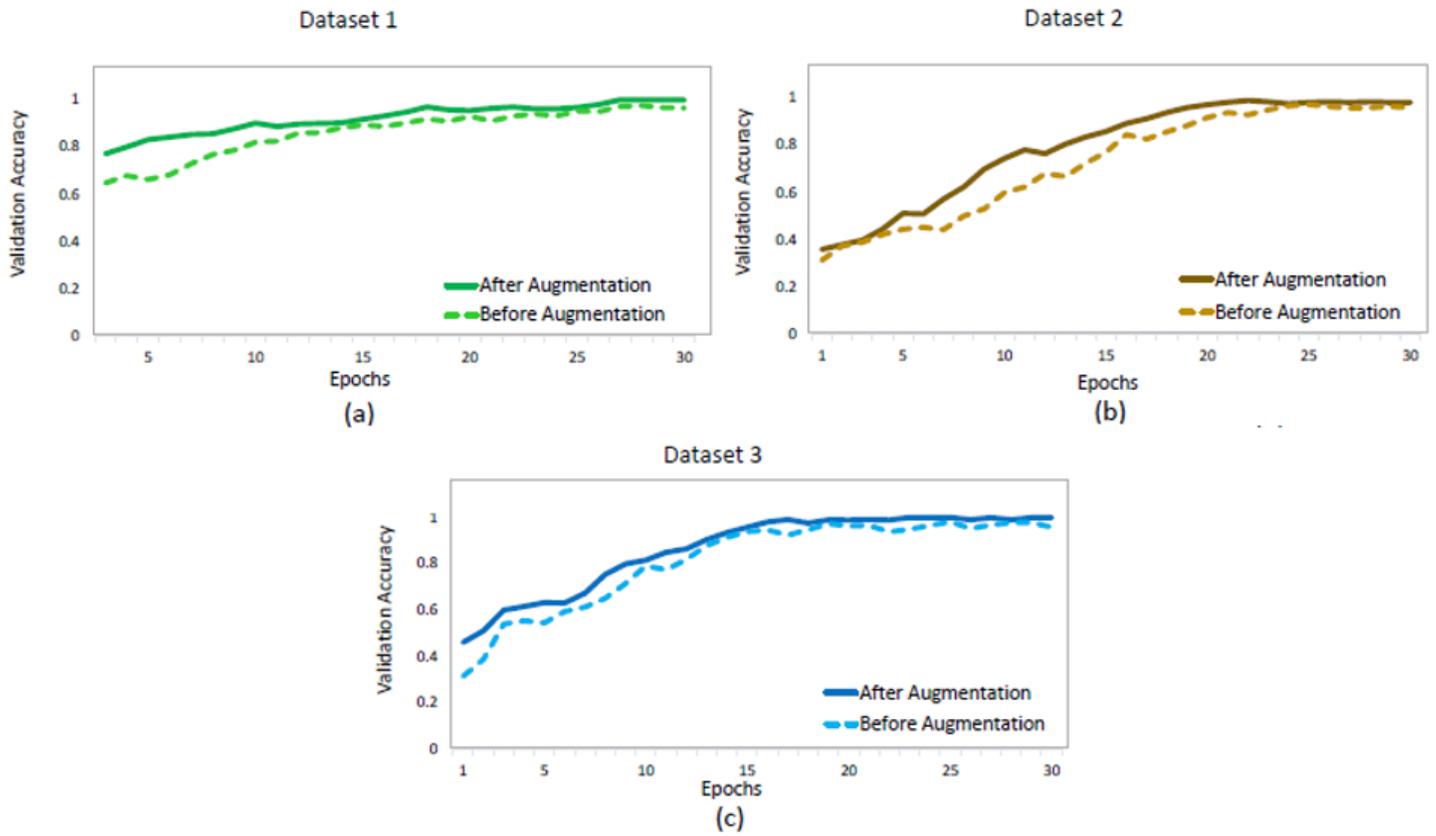


Figure 6

Impact of Data Augmentation. Epochs-vs-Validation accuracy before and after augmentation on (a) Dataset 1, (b) Dataset 2, (c) Dataset 3

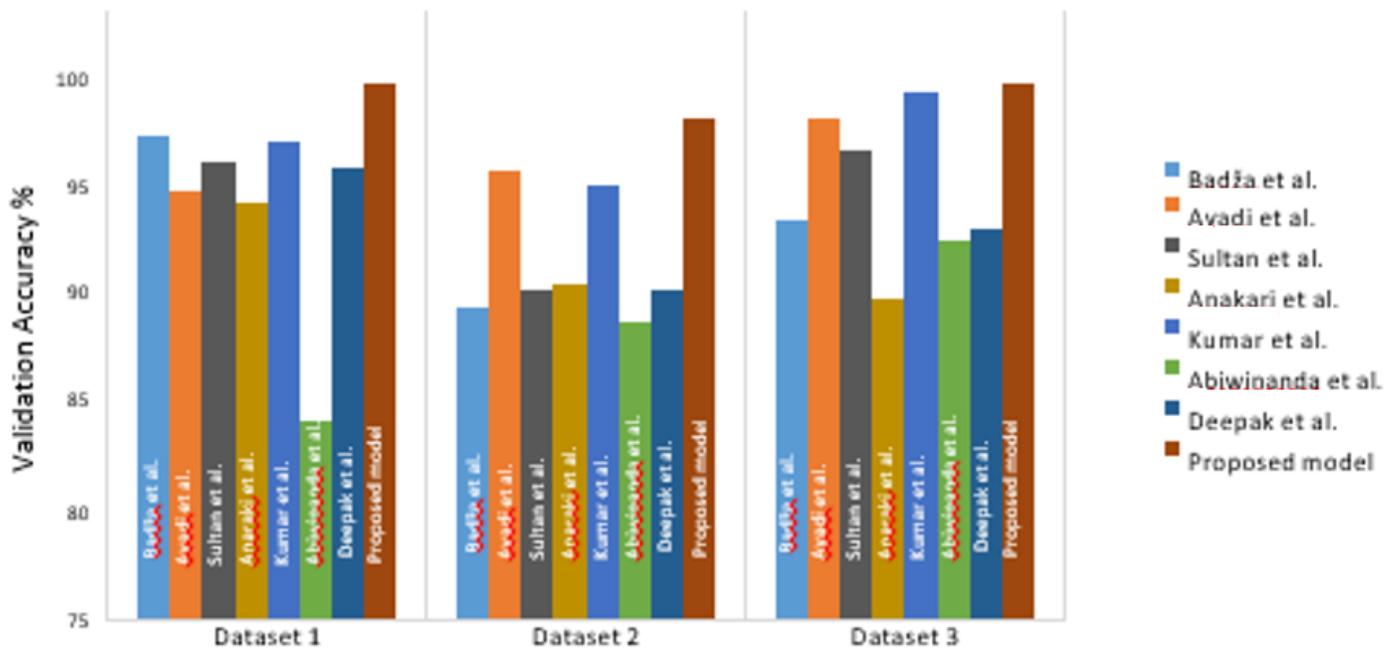


Figure 7

Performance comparison of the proposed two-stage ensemble model with some state-of-the-art models. The proposed model outperformed all the models considering all the datasets.

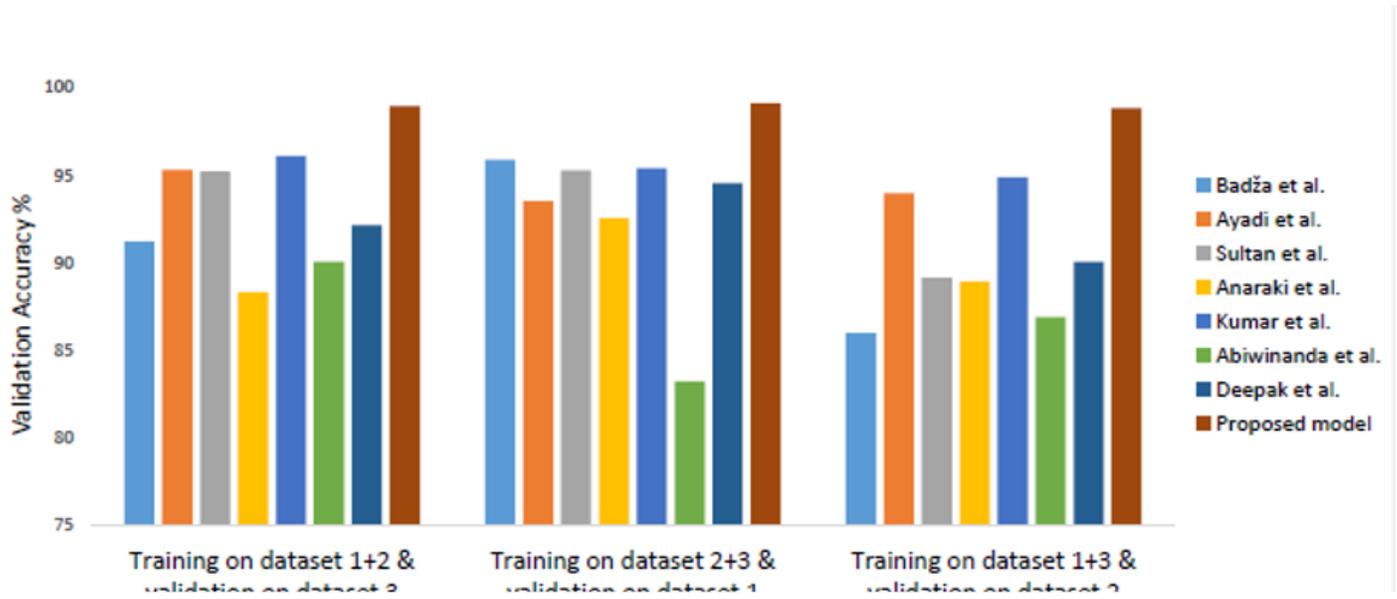


Figure 8

Performance comparison of the proposed model with other existing models considering separate training and validation data. The proposed model showed better performance in all three combinations.

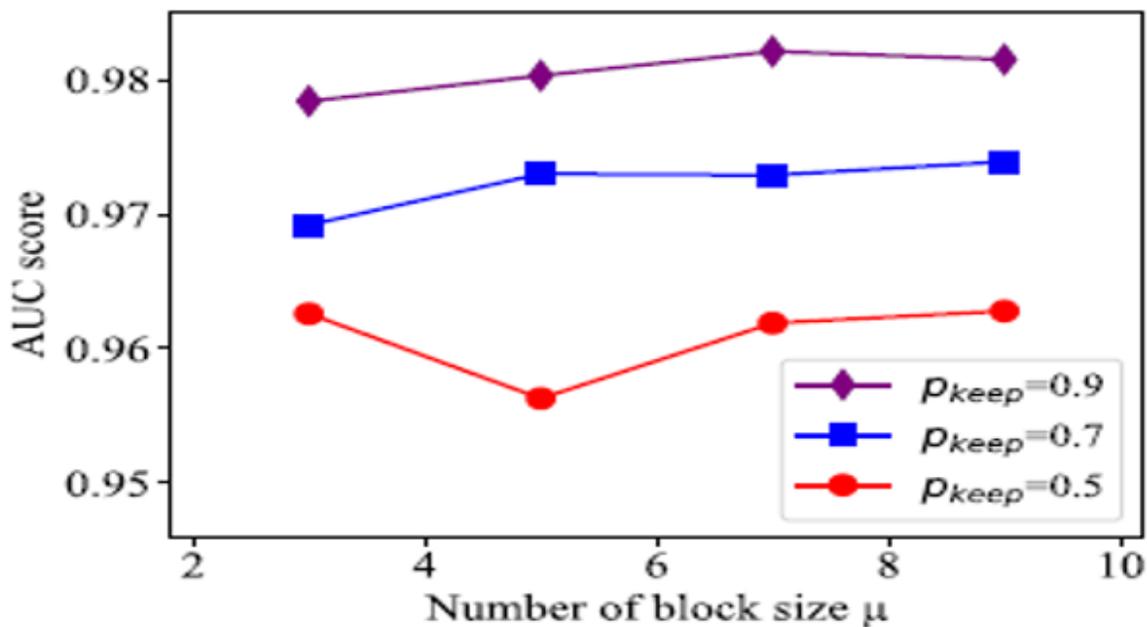


Figure 9

The results of different parameters p_{keep} and block size μ on MRI Data set 3.