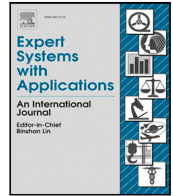




Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Robust clinical applicable CNN and U-Net based algorithm for MRI classification and segmentation for brain tumor

Atika Akter^a, Nazeela Nosheen^a, Sabbir Ahmed^a, Mariom Hossain^a, Mohammad Abu Yousuf^{a,*},
 Mohammad Ali Abdullah Almoyad^b, Khondokar Fida Hasan^c, Mohammad Ali Moni^{d,e,**}

^a Institute of Information Technology, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh

^b Department of Basic Medical Sciences, College of Applied Medical Sciences in Khamis Mushyt, King Khalid University, 47 Abha, Mushait, PO Box. 4536, ZIP. 61412, Saudi Arabia

^c School of Professional Studies, University of New South Wales (UNSW), 37 Constitution Avenue, Canberra 2606, Australia

^d Artificial Intelligence & Data Science, School of Health and Rehabilitation Sciences, Faculty of Health and Behavioural Sciences, The University of Queensland St Lucia, QLD 4072, Australia

^e Artificial Intelligence and Cyber Futures Institute, Charles Sturt University, Bathurst, NSW, 2795, Australia

ARTICLE INFO

Keywords:

Brain tumor
 Machine learning
 Deep convolutional neural network
 Classification of brain MRI
 Segmentation

ABSTRACT

Early diagnosis of brain tumors is critical for enhancing patient prognosis and treatment options, while accurate classification and segmentation of brain tumors are vital for developing personalized treatment strategies. Despite the widespread use of Magnetic Resonance Imaging (MRI) for brain examination and advances in AI-based detection methods, building an accurate and efficient model for detecting and categorizing tumors from MRI images remains a challenge. To address this problem, we proposed a deep Convolutional Neural Network (CNN)-based architecture for automatic brain image classification into four classes and a U-Net-based segmentation model. Using six benchmarked datasets, we tested the classification model and trained the segmentation model, enabling side-by-side comparison of the impact of segmentation on tumor classification in brain MRI images. We also evaluated two classification methods based on accuracy, recall, precision, and AUC. Our developed novel deep learning-based model for brain tumor classification and segmentation outperforms existing pre-trained models across all six datasets. The results demonstrate that our classification model achieved the highest accuracy of 98.7% in a merged dataset and 98.8% with the segmentation approach, with the highest classification accuracy reaching 97.7% among the four individual datasets. Thus, this novel framework could be applicable in clinics for the automatic identification and segmentation of brain tumors utilizing MRI scan input images.

1. Introduction

The most prevalent form of brain disease is brain tumors, which are also the cause of brain cancer. In this kind of cancer, which is deadly, and prompt, the diagnosis of brain tumors is critical (Arokia Jesu Prabhu & Jayachandran, 2018). It is caused by an abnormal and unregulated increase in the number of brain cells (Razzak, Imran, & Xu, 2018). Brain tumors are classified in a variety of ways. One of the most common types of classification is the division of brain tumors into benign and malignant tumor types. Benign-type brain tumors form on the inner side of the skull, but not in the brain tissue. Benign tumors in the brain can occasionally be life-threatening. Meningiomas are a

common kind of benign tumor (almost 30%). Meningiomas are slow-growing tumors that make up about 85% of all cases. Meningiomas are more frequently diagnosed in women than in men. They have a good probability of being surgically removed as they rarely spread to adjacent brain tissue. Although meningiomas can occasionally develop into malignant tumors. Pituitary tumors originate in the pituitary glands, and these glands govern hormones and physiological processes. Pituitary tumors are noncancerous tumors that do not spread to other organs. Although pituitary tumors seldom progress to cancer, the problems can lead to long-term hormone insufficiency and loss of vision. Malignant tumor cells are abnormal cells that increase uncontrolled and

* Corresponding author.

** Corresponding author.

E-mail addresses: atika.iit.ju@gmail.com (A. Akter), nazeela.stu2016@juniv.edu (N. Nosheen), sabbir.iit.ju@gmail.com (S. Ahmed), mariom.stu2016@juniv.edu (M. Hossain), yousuf@juniv.edu (M.A. Yousuf), maabdullah@kku.edu.sa (M.A.A. Almoyad), fida.hasan@qut.edu.au (K.F. Hasan), m.moni@uq.edu.au (M.A. Moni).

<https://doi.org/10.1016/j.eswa.2023.122347>

Received 30 December 2022; Received in revised form 24 October 2023; Accepted 24 October 2023

Available online 28 October 2023

0957-4174/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

irregularly. Normal tissues can be compressed, infiltrated, or destroyed by these tumors. The most common variety of malignant brain tumors is noted to be gliomas. Gliomas account for approximately 33% of all brain tumors. They rarely develop in the spine or other organs of the body, but they can develop quickly and can invade healthy tissues in the surrounding area (Irmak, 2021a). Since brain tumors are a life-threatening disease, initial detection of brain tumors can be fruitful in saving a patient's life by delivering optimal treatment. However, it can take some time because a patient may need to undergo various brain tumor diagnosis measures.

Machine learning can be advantageous in determining the presence and type of brain tumors; however, additional human intervention is necessary, as machine learning models are based on predictions. In contrast, deep learning models can learn and detect features on their own through neural networks, which is the ultimate goal of incorporating the entire detection process in an automated way. Some of the research works incorporated machine learning approaches into their work. Mohamed Shakeel, Tobely, Al-Feel, Manogaran, and Baskar (2019) employed a Near Infrared Imaging (IR) mechanism for detecting brain tumors with a size of less than 3 mm (arbitrary random value) that could not be detected using Computed Tomography (CT) or MRI scans. They sent the thermal information via WSN. The feature vector, gray level co-occurrence matrix, support vector machine (SVM), statistical features, and backpropagation neural network are shown here. Sharif, Li, Khan, and Saleem (2020) used a pre-trained Inception architecture 3 to extract features, and then the features were incorporated with dominate Rotate LBP for superior texture analysis. The feature vectors were optimized employing a particle swarm optimization algorithm, and the softmax classifier performed the classification.

The majority of research papers utilized a segmentation strategy with a binary classification focus on determining whether or not an MRI contained a tumor. Segmentation methods are used before classification, but the classifier and segmentation datasets differ from each other. These approaches add more error to classifier training since the first model provides a probabilistic segmented tumor image, which itself cannot have 100 percent accuracy. Again, none of the more recent studies has demonstrated the effects of segmentation before classification and the result of whole MRI because, biologically speaking, the classification of tumors depends on their position in the human brain.

Deep learning networks have gained significant popularity in medical diagnosis due to their ability to efficiently process large volumes of complex data, such as medical images, and accurately identify patterns indicative of diseases (Islam et al., 2023; Suzuki, 2017; Talukder et al., 2022). In recent years, numerous deep learning-based models have been examined for brain tumor detection. Of these models, the Convolutional Neural Network (CNN) architecture has been predominantly utilized, owing to its exceptional performance and flexibility. The CNN model executes feature extraction and classification through various training layers, which can be increased or decreased with respect to the resultant performance (Islam, Hossain, Akhtar, Moni, & Hasan, 2022; Yamashita, Nishio, Do, & Togashi, 2018). Regarding these aspects, in this research, the CNN architecture is used. MRI images of brain tumors are used as datasets because of their high-quality and non-ionizing radiation properties. The conventional practice for segmenting tumor areas is manual segmentation. However, it is exorbitant, exhausting, and prone to inter-observer variability (Menze et al., 2014). As a result, automated tumor segmentation is preferred, especially when dealing with massive amounts of data and the need for constant tumor observation and pliable treatment planning. Due to the wide range of tumor sites, forms, and architectures, effective automated tumor segmentation is generally difficult (Naser & Deen, 2020). A U-Net architecture-based model is created to segment the tumor portions and the model offers segmented tumor areas for MRIs.

The following are the main contributions of the article:

- A classification model for tumor MRI multiclassification has been proposed and evaluated on four separate individual datasets and two merged datasets that are the composite of these individual datasets. The classification model was tested against 5 pre-trained models.
- A segmentation model that accepts input MRIs and creates masked pictures has been suggested. Although the model was trained using a manually constructed mask from the Merged dataset 1, it can also produce masks for other datasets.
- The typical strategy of doing segmentation first and then classification has been evaluated, and the results show that the classification approach with segmentation performs better by a very small margin. Consequently, the classification may be performed without even performing segmentation, which undoubtedly cuts down on the time it takes to train two models (segmentation and classification models).
- Merged dataset 1 has been manually masked with expert assistance and consists of a large number of images divided into four groups (glioma, meningioma, pituitary, and no tumor)

The rest of the paper is structured as follows: Section 2 includes a comprehensive review of all datasets and techniques, including information on single, merged, augmented, and segmented datasets, data pre-processing, and classification and segmentation methods. Section 3 details the results as well as the additional training and validation methods. Finally, the conclusion is appended in Section 4.

2. Literature review

Due to the significant contributions that deep learning models can make to this topic, some pertinent articles have been extensively covered in this section.

Mohsen, El-Dahshan, El-Horbaty, and Salem (2018) have created a Deep Neural Network (DNN) classifier integrated with principal component analysis and discrete wavelet transform to classify MRIs into 4 classes. Images were segmented using fuzzy C-means clustering. Although they performed classification with a very small dataset of 66 brain MRIs. Özyurt, Sert, Avci, and Dogantekin (2019) implemented a method to categorize brain tumors as malignant or benign using a hybrid neutrosophic set-expert maximum fuzzy-sure entropy convolutional neural network classification technique. A comparison carried out for two classifiers, and Support Vector Machine (SVM) gave an accuracy of 95.62%. However, they only classified tumors as benign or malignant.

Sajid, Hussain, and Sarwar (2019) devised a method that implemented a hybrid CNN with 2 path CNN and 3 path CNN using a patch-based technique considering local and contextual information. Glioma tumor regions are segmented using a patch-based approach. They segmented tumors into Higher Grade Glioma (HGG) and Lower Grade Glioma (LGG), but no classification was performed. Irmak (2021b) put forward a model for classifying brain MRIs with Deep CNN and grid search optimizer-tuned hyper-parameters with the motivation to build three different models for the classification of tumors. Although their second model, which classifies tumors into five classes, achieved less accuracy 92.66% compared to the other two models.

Sultan, Salem, and Al-Atabany (2019) proposed an approach using CNN architecture to classify brain MRIs into three distinct groups and distinguish into different glioma grades. However, the dataset used to distinguish glioma grades is smaller and should be evaluated on a larger group of datasets. Havaei et al. (2017) proposed a two-pathway CNN architecture that is capable of extracting local features along with global features of the brain tumors concurrently. In the two-training phase, the CNN base output is fed again into the CNN subsequently. However, the architecture is required to be observed in large datasets for efficiency in handling huge magnetic resonance imaging. Badža and Barjaktarović (2020) introduced a CNN approach to classifying tumors

into three classes: meningioma, glioma, and pituitary. Four evaluation approaches have been used by aggregating two databases and two 10 cross-validation approaches. Both the 10 fold record and the subject-wise cross-validation approach have been used on the original and augmented images for evaluation. Amin, Sharif, Yasmin, and Fernandes (2018) divided the input image into a number of patches and, using a pre-trained CNN model, the core pixel of each patch is calculated then combining all the predicted results. They classified the tumor into HGG and LGG with different tumor regions. They claimed that the average processing time is only 5.02 s.

An automated deep multi-scale 3D CNN is recommended by Mzoughi et al. (2020). Their proposed architecture may combine lower-weight global and local contextual information. However, they classified the tumor into two subclasses of gliomas: HGG and LGG. Siar and Teshnehlab (2019) proposed a strategy that used the ImageNet feature extraction model. Their model showed good performance on a first-order clustering algorithm and CNN softmax fully connected layer has been devised to distinguish fat tissue and tumor tissue, which increased accuracy. Although no specific tumor type categorization has been performed; only benign or malignant tumors have been classified. Irsheidat and Duwairi (2020) implemented a CNN model which took MRI input in grayscale mode and increased the dataset that made it fourteen times larger than the initial size. The model predicted the presence or absence of a tumor with an accuracy of 96.7% and 88.25% in validation and test data, respectively. Although they simply did a binary classification and used a smaller dataset to test their model. Ayadi, Elhamzi, Charfi, and Atri (2021) came up with a CNN model that included multiple layers to classify brain tumor MRIs as meningioma, glioma, and pituitary using three publicly available datasets. Figshare dataset, Radiopaedia dataset, and REMBRANDT dataset for comparison with the previous work.

Naser and Deen (2020) incorporated a deep learning strategy in which CNN relied on U-Net to segment regions affected by tumors. A transfer learning VGG-16 model and classifiers has been developed to grade cancerous tumors. No independent dataset was available to compare with for testing and a moderately small LGG data has been used for validation. Nevertheless, only glioma tumors have been categorized as LGG and HGG. Shahzadi, Tang, Meriadeau, and Quyyum (2018) used a CNN based 3-D medical image classification cascading with Long Short Term Memory (LSTM) to identify LGG and HGG. They utilized VGG-16 for feature extraction and LSTM for glioma classification. Because they employed a small number of samples and did not apply any data augmentation techniques to expand the number of samples, their model has lesser accuracy than other research.

Noreen et al. (2020) implemented two separate multi-level architectures using Inception-v3 and DenseNet201 architectures with softmax classifier. In these pre-trained architectures, features are extricated from various modules and percolated through the softmax layer for classification after the concatenation of the extracted features. The super-resolution fuzzy c-means technique has been proposed by Özyurt, Sert, and Avcı (2020) in tumor detection. The features were extracted from the SqueezeNet architecture and then the extreme learning machine classification technique was performed. Anyway, tumors are mainly divided into benign and malignant categories in their work (see Tables 1 and 2).

A method based on the Resnet-50 was proposed by Öinar and Yildirim (2020). The endmost five layers of the model are eliminated, and eight additional layers are inserted as a substitute. However, the type of tumor was not characterized and only a binary classification has been implemented. A CNN of 22 layers is proposed to classify lung cancer into five categories (Faruqui et al., 2021). Ghassemi, Shoeibi, and Rouhani (2020) proposed a Generative Adversarial Network (GAN) based model where a deep CNN is used as the discriminator for the detection of fake images that are generated by the generative model and a pre-trained CNN network is fine-tuned to perform as a classifier. The input size was limited to 64×64 on account of some GAN restrictions.

A new method in deep learning features, as well as a fusion of hand-crafted features for brain tumor detection, was established by Saba, Mohamed, El-Affendi, Amin, and Sharif (2020) where input images went through segmentation employing the GrabCut algorithm. The incorporated VGG-19 with Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). In the end, multiple classifiers are used for the classification. Between glioma and healthy brain images, classification is limited. Ahamed et al. (2021) developed a deep learning approach for detecting covid-19 cases using chest CT scans and X-ray pictures. Saouli et al. (2018) proposed an automatic incremental CNN end-to-end for tumor segmentation in which they implemented deep learning models entitled 2Cnet, 3Cnet, and EnsembleNet. EnsembleNet is an integrated model of 2Cnet and 3Cnet. The model is only suitable for the segmentation of MRIs. Aurna, Yousuf, Taher, Azad, and Moni (2022) also used the ensemble technique to classify brain MRIs.

Chelghoum, Ikhlef, Hameurlaine, and Jacquir (2020) involved the use of nine pre-trained CNNs namely ResNet-50, AlexNet, VGG-16, ResNet-101, GoogleNet (Szegedy et al., 2015), VGG-19, ResNet-18, SENet, and ResNet-Inception-v2 for comparative analysis. However, all models have been implemented on a single data set. Swati et al. (2019) used the blockwise VGG-16 network centered on transfer learning and fine-tuning to classify MRIs as: glioma, meningioma, and pituitary. The model went through a minimal pre-processing procedure and did not use any of the hand-made features.

A segmentation approach employing a CNN was implemented followed by an extensive data augmentation strategy to classify multi-graded brain tumors by Sajjad et al. (2019). Due to the small-scale datasets, the segmented data have been augmented using eight different types strategies, including various geometric transformations and noise invariances. Elazab et al. (2020) devised a model named GP-GAN, which predicts the growth of glioma tumors at an early stage using stacked 3D GANs. Using a 3D U-Net architecture, this model's generator was created. GP-GAN was used in this study which presupposes that tumors would always grow. Furthermore, the shrinkage of tumors as a result of treatment is not taken into account in the method.

Hamghalam, Wang, and Lei (2020) used a multistage attention-GAN to increase the contrast of the tumor image. In their work, tumors are segmented into whole tumors, core tumors, and enhancing tumor portions. The high tissue contrast synthetic image takes a significant amount of time to generate 27 ms. Due to the multistage architecture of their models, the computational complexity and number of parameters rise as the number of regions of Interest (ROI) increases. Rezaei et al. (2017) gave out a conditional GAN approach for segmentation into three distinctive sub-regions namely the whole tumor, the core tumor, and the enhancing tumor region with different labels that are used further to evaluate the survival days of patients after tumor diagnosis. Han et al. (2019) designed a model to detect brain metastases at desired position using progressive growing of GAN and a highly rough bounding box. They divided the image into smaller regions and tried to predict the bounding box and achieved great sensitivity in tumor detection by working on random shapes rather than appropriate segmentation, spontaneously at desirable positions and sizes. But they failed to achieve high sensitivity by adding more synthetic images.

3. Datasets and methodology

The datasets have been utilized in accordance with the classification and segmentation methods. Four individual datasets and two merged datasets are taken into consideration for classification. These individual datasets are combined to form the merged datasets. One of these two merged datasets is manually segmented to produce masks with the assistance of a specialist. The proposed segmentation model is trained using this masked dataset.

Data have undergone preprocessing steps. While data augmentation is performed, certain pre-processing techniques like resizing, zooming, mode filter, sobel filter, unsharp masking, and grayscale are used. Both

Table 1
Recent research using deep learning for the classification and segmentation of brain tumor MRIs.

Ref.	Contribution	Limitations
Mohsen et al. (2018)	Accomplished a DNN classifier integrated with PCA and DWT for categorizing tumors into 4 classes. MRIs were segmented using the Fuzzy C-Means (FCM) clustering.	Classified a very small dataset of 66 brain MRIs.
Özyurt et al. (2019)	Implemented a method to categorize brain tumors using a hybrid NS-EMFSE-CNN classification technique. They acquired an accuracy of 95.62% by SVM.	No specific tumor type classification has been carried out.
Sajid et al. (2019)	Devised a method that implemented a hybrid CNN with 2 path CNN and 3 path CNN using a patch-based technique considering local and contextual information.	Segmented MRIs of HGG and LGG and no classification was performed.
Irmak (2021b)	Put forward a model for classifying brain MRI images with Deep CNN and grid search optimizer-tuned hyper-parameters with the motivation to build three different models for the classification of tumors.	The second model achieved less accuracy compared to the other two models.
Sultan et al. (2019)	Offered an approach of using the CNN architecture for classifying brain MRIs into three distinct groups and distinguishing in different glioma grades	The used dataset is smaller and must be evaluated on a bigger group of datasets
Havaei et al. (2017)	Proposed a CNN which is a two-path architecture that is capable of extracting local features along with global features of the brain tumors concurrently.	Required to be observed on large datasets for efficiency of handling huge MRIs.
Irsheidat and Duwairi (2020)	A generic CNN was implemented which predicted the presence of a tumor or not with an accuracy of 96.7% and 88.25% in validation and test data respectively.	A binary classification is performed and used a smaller dataset to test their model.
Naser and Deen (2020)	Incorporated a deep learning strategy in which CNN hinged on U-Net for segmenting tumor affected regions. A transfer learning VGG-16 model and classifiers has been developed to grade cancerous tumors.	No independent dataset was available to compare with for testing and a moderately small LGG data has been used for validation.
Öinar and Yildirim (2020)	A method based on the Resnet-50 from CNN was proposed, The endmost five layers of the model are eliminated, and eight additional layers are inserted as a substitute.	Only binary classification has been implemented.
Han et al. (2019)	Designed CPGANs to detect brain metastases at desired position using PGGANs and a highly rough bounding box.	Failed to achieve high sensitivity by adding more synthetic images.
Saouli, Akil, and Kachouri (2018)	An end-to-end automatic incremental CNN was proposed for the segmentation of tumors in which they implemented deep learning models entitled 2Cnet, 3Cnet, and EnsembleNet using their proposed training strategy.	The model is solely suitable for the segmentation of MRI.
Elazab et al. (2020)	Devised the GPGAN, which predicts the growth of glioma tumors at an early stage using stacked 3D GANs.	Presupposes that tumors would always grow and the shrinking tumors as a result of treatment are not taken into account.

Table 2
The descriptions for each layer in the proposed classification model have two blocks and a total of 39 layers.

Block No.	Layer name	Layer count	Filter count	Filter size
	Convolution	1	32	3×3
	Maxpooling	1		2×2
	Batch Normalization	1		
	Dropout	1		
Block 1	Convolution	4	64	$1 \times 1, 3 \times 3$
	Maxpooling	3		2×2
	Batch Normalization	3		
	Dropout	1		
Block 2	Convolution	4	128	$1 \times 1, 3 \times 3$
	Maxpooling	3		2×2
	Batch Normalization	3		
	Dropout	1		
	Convolution	2	128	$5 \times 5, 3 \times 3$
	Maxpooling	1	128	2×2
	Batch Normalization	1		
	Dropout	3		
	Flatten	1		
	Dense	3		

the suggested model and pre-trained architectures have been employed over the classification datasets separately and the models classify MRI images into four classes individually. In Fig. 2, the complete working model for classifying tumors is displayed. Fig. 2 shows that MRI images are augmented, then the MRIs are sent out for preprocessing purposes. After that, the MRIs are sent to the proposed model and transfer learning models separately to classify the images as glioma, meningioma, pituitary, or no tumor. Two scenarios are taken into consideration, the first in which the classification is done without the segmentation being done first. And in the second scenario, segmentation is followed by sending the MRIs to the classification models. Sending the segmented MRI to the classification model minimizes computational complexity and processing time since segmentation only extracts the tumored

portion of an MRI. Fig. 1 illustrates the workflow of the proposed methodology. A variety of metrics is used to compare the classification performance of pre-trained and proposed models as well as the performance of the segmentation model. Finally, the two methods of classifying tumors while segmenting them and classifying them without segmentation are compared.

3.1. Classification datasets

A total of four individual datasets have been outlined, and two combinations of these datasets have been integrated to create two new merged datasets. As a result, our work has been applied to a total of six datasets (four individual and two merged datasets). The four datasets are referred to as Dataset a, Dataset b, Dataset c, and Dataset d throughout the paper. Merged dataset 1 and Merged dataset 2 are used to refer to the two merged datasets respectively.

3.1.1. Dataset a

There are 3064 T1-weighted Contrast Enhanced-MRI images in this dataset. The images are taken from 233 patients who have three types of tumors: meningioma (708 slices), glioma (1426 slices), and pituitary tumor (930 slices). This dataset has been widely used in most research papers (Ayadi et al., 2021; Badža & Barjaktarović, 2020; Chelghoum et al., 2020; Noreen et al., 2020; Sultan et al., 2019; Swati et al., 2019). The data was gathered from Nanfang Hospital in Guangzhou, China, and General Hospital, Tianjin Medical University in China between 2005 and 2010. The dataset is made publicly available by Cheng (2017).

3.1.2. Dataset b

This dataset comprises 3264 T1, T2, and fluid-attenuated inversion recovery MRI images. The images are split up into two directories: training and testing. Glioma, meningioma, pituitary, and normal brain MRI are the four subclasses in each directory. Glioma (100 slices), meningioma (115 slices), pituitary (74 slices), and no tumor (105 slices) are all included in the testing directory. Glioma (826 slices), meningioma (822 slices), pituitary (827 slices), and no tumor (395 slices) are all listed in the training directory. The dataset is publicly available by Bhuvaji (2020).



Fig. 1. The suggested methodology's workflow entails the selection of the dataset, MRI preprocessing, MRI segmentation, models shown for classification, MRI classification classes, and ultimately results analysis over the findings.

3.1.3. Dataset c

This dataset includes a total of 10000 images of 3 classes of tumor MRIs and MRIs with no tumor. There is glioma (2500 slices), meningioma (2500 slices), pituitary (2500 slices), and MRIs with no tumor (2500 slices). This dataset is also publicly available by Pradeep (2021).

3.1.4. Dataset d

There are a total of 4292 images in this dataset. The dataset is made publicly available by Sherif (2020). The dataset is divided into two

directories: training and testing. Each of these two directories contains four subclasses of glioma, meningioma, pituitary, and no tumor. There is glioma (1038 slices), meningioma (1318 slices), pituitary (1255 slices), and normal brain MRIs (681 slices).

3.1.5. Merged dataset 1

This dataset is a combination of Dataset a and Dataset b. Glioma, meningioma, pituitary, and normal brain MRIs are all included in

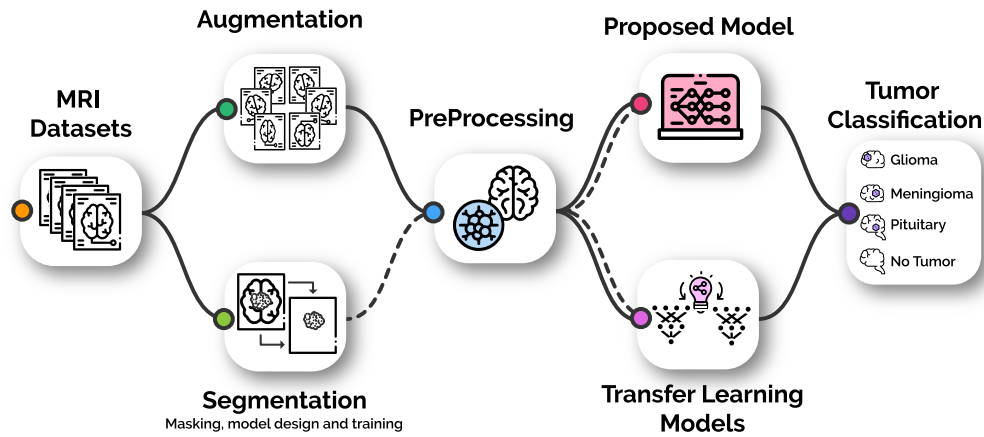


Fig. 2. The classification technique is depicted schematically, with MRIs segmented, and preprocessed before passing them to the proposed classification model and transfer learning models.

the combined dataset. Because the number of normal brain MRIs is low compared to the other three classes, normal brain MRIs have been taken from the (Hamada, 2020) dataset. Hamada (2020) dataset includes MRIs of normal and tumor-affected brain tissue. As a result, the Merged dataset 1 contains a total of 7023 images. Glioma (1621 slices), meningioma (1645 slices), pituitary (1757 slices), and normal brain MRIs (2000 slices) are included in the Merged dataset 1. The sample images from this dataset are shown in Fig. 4(a). The dataset is publicly available by Nickparvar (2021).

3.1.6. Merged dataset 2

The Merged dataset 2 has been created by combining the Merged dataset 1 with Dataset c and Dataset d. As a result, this merged dataset 2 combines all four datasets used in our work (Dataset a, Dataset b, Dataset c, and Dataset d) along with the dataset by Hamada (2020) having normal tumor images. Glioma (5159 slices), meningioma (5465 slices), normal brain MRI (5181 slices), and pituitary (5512 slices) are included in this dataset.

3.2. Augmented merged dataset

MRI images for 4 classes, glioma, meningioma, no tumor, and pituitary, are contained in the Merged dataset 1, which is split into two folders for training and testing. The number of images in every class's testing and training folder is not equal. In the event of training, each tumor class has 5000 images, but in the event of testing, each class contains 1000 images. Therefore, the merged augmented dataset contains 24,000 images. Because the dataset comprises images of varied sizes, they have been reduced in the augmentation stage to a constant size of 256 by 256. Rotation, zooming, height and width shifting, shearing, horizontal flipping and mode filling are the preprocessing methods employed while augmenting the MRI images. To keep maximal features in the images, the rotation, zooming, shifting, and shearing settings have been modified for different classes of tumors in the training and testing folders. It also includes augmenting the dataset such that it has 80% of images for training the model whereas 20% for testing and validation (see Fig. 3).

3.3. Segmentation dataset

The merged dataset 1 has been manually segmented because the Merged dataset 1 contains MRIs from four different classes. As after the MRIs have been segmented, they will be forwarded to a classification model where they will be divided into four categories: pituitary, glioma, and meningioma and no tumor. Four individuals manually segmented the MRIs of the four classes. Before segmentation, an expert with a medical background examines the MRIs. For each MRI,

four individuals create four mask images. Then the masked MRIs are provided back to the expert. Out of the four masks, the expert selects the one that is closest and then makes the required corrections. There is a corresponding generated mask that exists for every MRI. Example images of this dataset are shown in Fig. 4(b).

3.4. Data preprocessing

Image resizing, RGB to grayscale conversion, unsharp masking, and sobel filtering (Vincent & Folorunso, 2009) are some of the preprocessing techniques that have been used for MRI images. The MRIs are downsized to the size of $160 \times 160 \times 3$ in the preprocessing stage. They are transformed to grayscale using the weighted average approach, which reduces its dimension to $160 \times 160 \times 1$. Unsharp masking with a radius of 100 pixels is applied to the grayscale MRI. Then the MRIs are subjected to the sobel filter.

3.4.1. Weighted average method

Eq. (1) is representing the grayscale weighted average, Y

$$Y = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (1)$$

In Eq. (1), R, G, and B are integers with values ranging from 0 to 255 that indicate red (R), green (G), and blue (B).

3.4.2. Unsharp masking

Unsharp masking is a versatile and strong technique for improving image sharpness. A lower radius adds smaller-scale detail by affecting the size of the edges to be enhanced or the width of the edge rims. The threshold determines how much of a brightness shift will be sharpened. A blurred image is subtracted from the original MRI, which is referred to as an unsharp masking algorithm. An unsharp or blurred image is created by spatial filtering the original image using a Gaussian low-pass filter.

3.4.3. Sobel filter

A basic 3×3 convolution is used in sobel filter. Separation of the sobel kernels is an additional optimization option. It is designed to operate on first-order derivatives. It computes the MRI's first derivatives individually for the X and Y axes.

3.5. Classification models

A pre-trained model is a saved model that has already been trained on datasets to perform a certain job. Since ImageNet includes 1000 classes, pre-trained models have been taught to handle a wide range of tasks. It takes less time and effort to create the architecture of a model

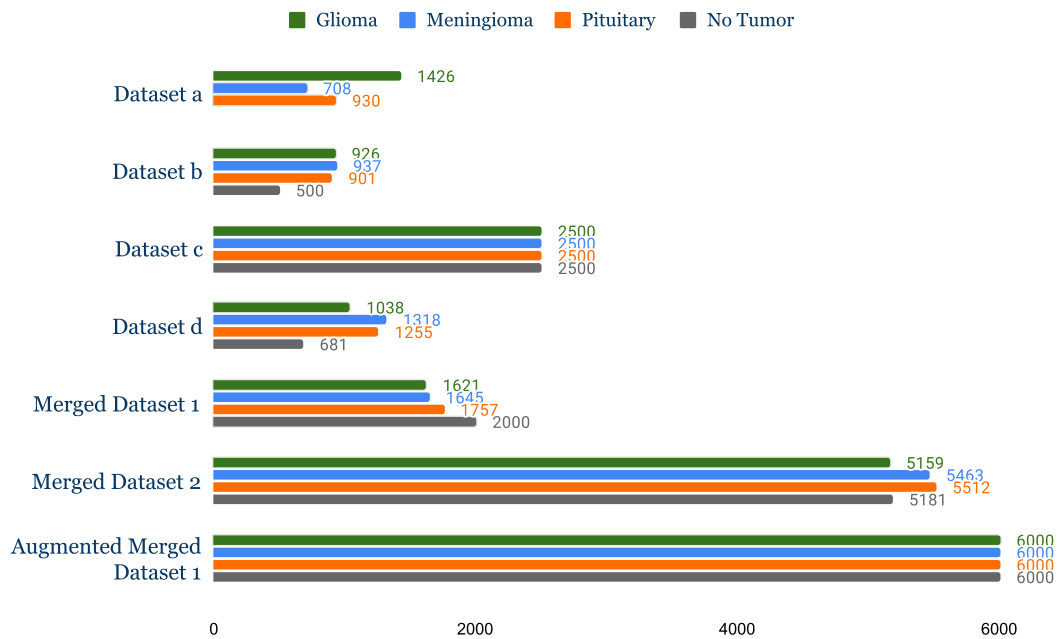
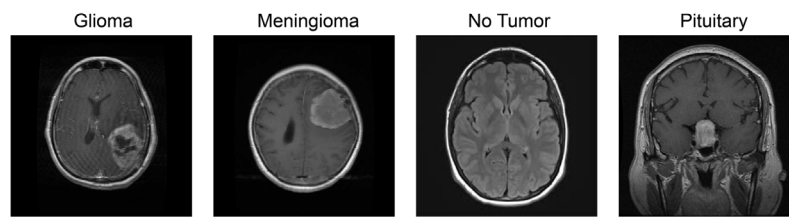
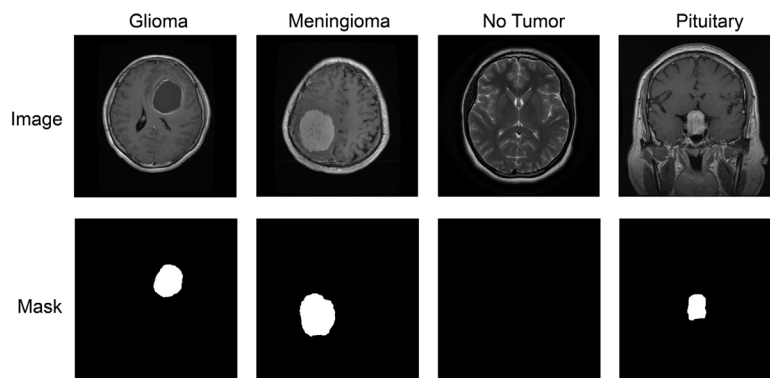


Fig. 3. Dataset distribution for classification includes four individual datasets (Datasets a, b, c, and d), two merged datasets (Merged Dataset 1 combines three datasets, and Merged Dataset 2 combines Merged Dataset 1 and Datasets c and d).



(a)



(b)

Fig. 4. (a) Sample images from Merged dataset 1 where MRIs are from four classes named: Glioma, Meningioma, No Tumor, and Pituitary. (b) Sample images from the segmentation dataset where images with the corresponding masks are generated manually.

that has already been trained. As the classification models, five pre-trained architectures have been used. Pre-trained architectures include VGG16 (Simonyan & Zisserman, 2015), VGG19 (Simonyan & Zisserman, 2015), EfficientNet B0 (Tan & Le, 2019), and EfficientNet B7 (Tan & Le, 2019), ResNet152V2 (He, Zhang, Ren, & Sun, 2016). The pre-trained models are shown in Fig. 6. Along with the pre-trained models, a proposed classification methodology has also been presented. Fig. 7 shows the proposed classification model.

3.5.1. VGG16

VGG16 is an extensively applied convolutional neural network. This has achieved a lot of appeal in the research because of its straightforward methodology and the fact that pre-trained weights were publicly available online, allowing new tasks to be fine-tuned in the simplest way possible. VGG16 has 16 convolutional layers and is well-liked for its relatively consistent architecture. 5 maxpooling layers, 2 fully connected layers, and 1 softmax layer make up the architecture in addition to the convolutional layers. For classification, a flatten layer, 2

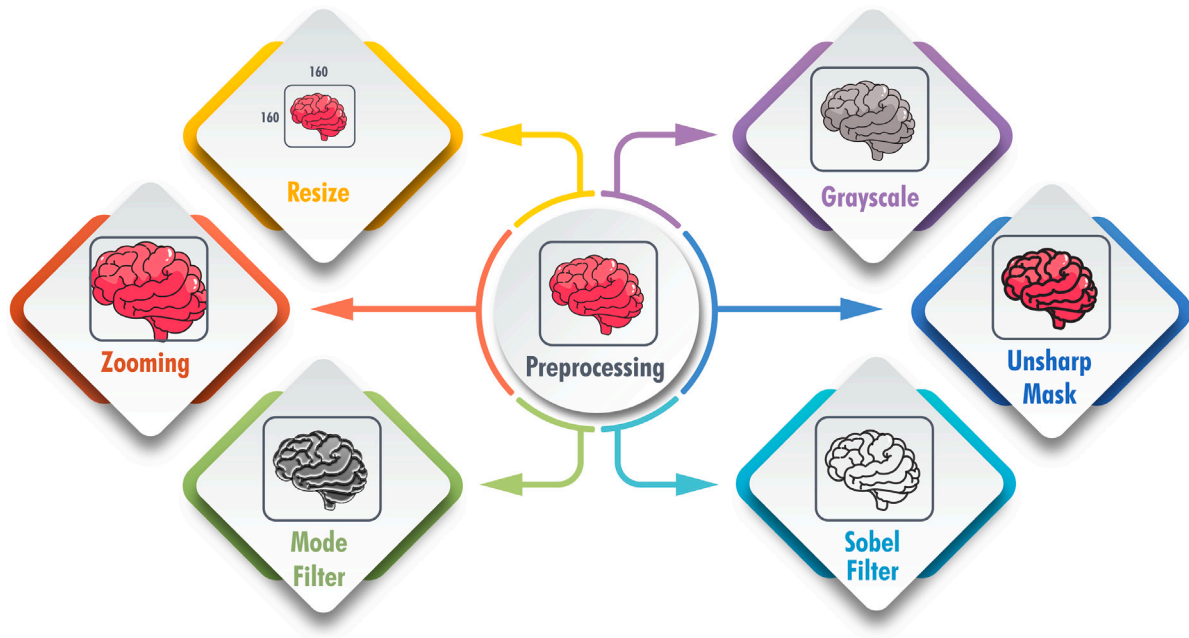


Fig. 5. Input MRIs are preprocessed before being sent into the classification model and preprocessing methods are used during augmentation too.

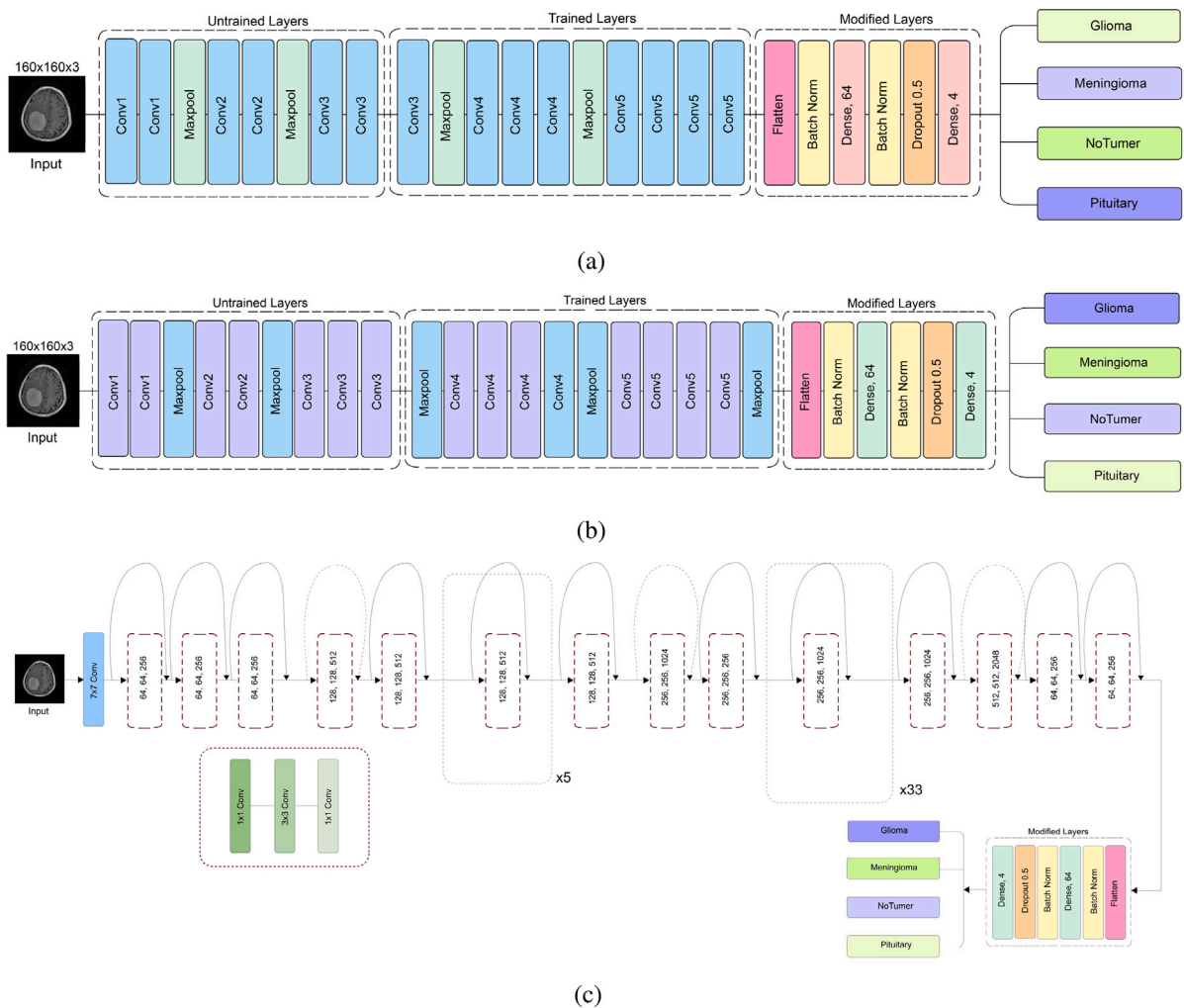


Fig. 6. Pre-trained architectures incorporated in this work, (a) Modified VGG16, (b) Modified VGG19, (c) Modified ResNet152V2.

batch normalization layers, 1 dropout with the dropout rate of 25%, and 2 dense layers are substituted for fully connected layers and softmax layer in the modified VGG16 shown in Fig. 6(a), which accepts input of size 160×160 .

3.5.2. VGG19

One of the VGG (Simonyan & Zisserman, 2015) architectures, VGG19, has 16 convolutional layers, 5 maxpool layers, 3 fully connected layers, and 1 softmax layer. The modified VGG19 architecture is made up of these convolutional and max pooling layers. And the fully connected layers and softmax layer are replaced with 6 layers that are organized as flatten, batch normalization, dense, dropout layer with a dropout rate of 25%, and finally a dense layer to classify MRIs into four classes. Fig. 6(b) depicts the modified architecture that has been tailored for MRI classification.

3.5.3. EfficientNet B0

In order to maximize accuracy and floating-point calculations, the researchers used a multi-objective neural network search to create EfficientNet B0 (Tan & Le, 2019). EfficientNet B0 to EfficientNet B7 has been developed using B0 as a baseline model, and they have achieved the highest level of accuracy on ImageNet while being significantly more efficient than its rivals. There are 237 layers in EfficientNet B0. To classify MRIs using modified EfficientNet B0 architecture, the flatten, batch normalization, dense, batch normalization, dropout layer with a dropout rate of 25%, and lastly, a dense layer have been added.

3.5.4. EfficientNet B7

EfficientNet B7 (Tan & Le, 2019) is a convolutional neural network that uses compound coefficients for the uniform scaling of width, height, and resolution. The scaling technique of EfficientNet B7 invariably and systematically enlarges the dimension of the network and resolution with a series of predefined scaling coefficients. Although in standard practice, these factors are arbitrarily adjusted. There are 813 layers in EfficientNet B7. In the modified EfficientNet B7 architecture, the flatten, batch normalization, dense, batch normalization, dropout layer, and finally a dense layer have been integrated into the base EfficientNet B7, and the input size is also taken into consideration to be 160×160 in this case.

3.5.5. Resnet152V2

ResNet (He et al., 2016) has become a game changer because it efficiently and considerably trains deep neural networks. ResNet operates on building deeper networks than conventional simple networks while determining the optimal amount of layers. These residual networks may get accuracy from far more depth and are simpler to tune. Before Resnet152V2, training operations for a deep neural network was challenging for vanishing gradient problem. The Resnet152V2 has 564 layers in all. The modified Resnet152V2 design adds 6 extra layers to the basic Resnet152V2 architecture shown in Fig. 6(c).

3.5.6. Proposed classification model

Fig. 7 shows the classification model that has been proposed. The model comprises 39 layers with two blocks between the input and output layers. The output layer predicts MRIs as glioma, meningioma, pituitary, or no tumor. The input image is processed via convolution layers, activation functions followed by the convolutional layers used to select the features, max-pooling, batch normalization, and dropout layers. Dropout layers are employed to prevent overfitting. To forecast the output, the dense layers and softmax layers are utilized.

The first layer provides information about the input size. The input MRI size to the model is 160×160 . Following that is the convolution layer. A convolutional layer is operated by performing a convolution operation on the original image and the kernel. A 2D convolutional layer employs a number of filters, with the kernel moving horizontally or vertically over the image in a given number of steps known as strides.

The filter size for the convolution layer in between input and the first block is 3×3 with 32 filters. After the convolution layer, a maxpooling layer is added with a pool size of 2×2 . The maxpooling layer is used to downsample the image. Maxpooling is used to discard less significant data while simultaneously addressing the issue of overfitting. After the pooling layer, the batch normalization and dropout layer (dropout rate 25%) are added.

The architecture of the two blocks in the model is identical, with the exception that the first block uses 64 filters while the second block uses 128 kernels. Each of the blocks has 3 paths. Each path has a distinct kernel size for the convolution layers, which is $1 \times 1, 3 \times 3$. For the maxpooling operation, a 2×2 kernel is utilized, and one of the three pathways has a dropout layer added with a 25% dropout rate. Paths guarantee that layers are applied at the same level. To make the model wider rather than deeper and to reduce the computing complexity of the model, layers with varying kernel sizes are used at the same level. The three paths each extract a different set of characteristics that are then combined at the end of each block.

Two convolution layers with filter sizes 5×5 and 3×3 filter sizes are integrated between the second block and the output layer. An additional maxpool layer with a 2×2 filter size is added after the convolution layer. Then dense layers, dropout, flatten, and batch normalization are included. The dropout rate for the latter two layers is 50 percent. Dropout layers are introduced to speed up the training process. The dense layer is followed by softmax. The dense layer's output is delivered to the softmax activation function. The softmax activation function is used to calculate the relative probabilities of having a certain type of tumor. The proposed model enables 2,051,872 worth of parameters to be trained.

3.6. Hyper parameters for proposed classification model

A lot of trial and error is required when tuning the hyper-parameters. While the model is being trained, these hyper-parameters act as controls that may be changed. It is determined what these hyperparameters should be set at in order to achieve the optimum results. The number of layers, neurons, optimizer, input and output activation functions, batch size, number of epochs, and the loss function have been taken into account as model hyperparameters.

Adaptive Moment Estimation (Adam) (Kingma & Ba, 2017) is chosen as optimizer. When dealing with complex problems requiring a lot of data or factors, this strategy is incredibly effective. And this strategy also uses minimal memory. In Adam, for a specific iteration (t), moving averages are dependent on the parameters: exponential decay rates for the first-moment and second-moment estimates which are denoted by β_1 and β_2 respectively, and gradient (gt). Eq. (2) represents the bias correction formula for moving averages.

$$\hat{p}_t = \frac{p_t}{1 - \beta_1^t}, \hat{q}_t = \frac{q_t}{1 - \beta_2^t} \quad (2)$$

In Eq. (2), \hat{p}_t and \hat{q}_t represents first and second moment vector respectively. The weights and biases are updated accordingly. In order to find the best parameter configuration, the suggested model uses the Grid Search technique. All conceivable parameter combinations are tested using grid search.

The output layer uses softmax for multi-classification while the hidden layers employ the Rectified Linear Unit (ReLU) activation function. Eq. (3) and Eq. (4) represent ReLU and Softmax activation functions, respectively.

$$f(x) = \max(0, x) \quad (3)$$

$$\text{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad (4)$$

In Eq. (3), any negative input causes the function to return 0, while any positive value x causes it to return that value. As a result, it generates an

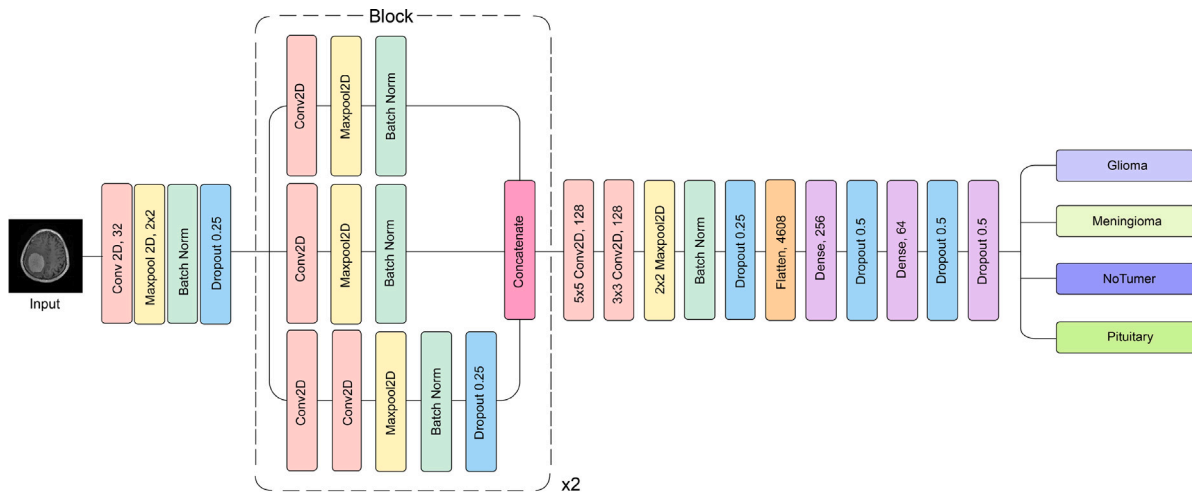


Fig. 7. The proposed classification model, which consists of 39 layers and includes two blocks, classifies MRIs into four classes: Glioma, Meningioma, No Tumor, and Pituitary.

Table 3
Hyperparameters of the proposed classification model.

Hyperparameters	Values
Number of trainable parameters	2,051,872
Optimizer	Adam
Searching strategy	Grid search
Output layer activation function	Softmax
Hidden layer activation function	ReLU
Batch size	64
Number of epochs	100
Loss function	Categorical cross entropy

output whose range is from 0 to infinite. In Eq. (4), the values from the output layer’s neurons are represented by the x . The non-linear function is represented by the exponential. After normalizing, these values are divided by the total of exponential values and then transformed into probabilities.

For the purpose of multi-classification in the proposed model, the categorical cross-entropy loss function has been employed. The probability for each class is provided by this loss function and is an excellent indicator of how easily two discrete probability distributions can be distinguished from one another.

Eq. (5) denotes the loss formula of the cross-entropy.

$$CELoss = - \sum_i^C g_{t_i} * \log(s_i) \tag{5}$$

In Eq. (4), the score and ground truth are denoted by s_i and g_{t_i} respectively for each class i in C .

The batch size is fixed at 64 and the epoch number is 100. Table 3 represents the hyperparameters utilized in the proposed model.

3.7. Segmentation model

Ronneberger, Fischer, and Brox (2015) established the U-Net structure. U-Net has been widely exploited for image segmentation in the medical field and has also shown competitive performance. The proposed segmentation model is based on U-Net. The U-Net architecture is divided into two paths: downsampling and upsampling. This model is made up of two networks: encoder and decoder. The encoder network has 2D convolutional layers in a total of five convolution blocks. For 2D convolutional layers, the number of filters is 32, 64, 128, 256, and 512. Following these convolutional layers, the batch normalization layer, the activation function (ReLU), and the dropout layer are employed to reduce overfitting. For a total of four upsampling blocks, the decoder network contains transposed convolutional layers. Using a stride size

of 2,2 for the transposed convolutional layers, the spatial dimension is widened. Each of the upsampling blocks has corresponding encoder block output concatenated, convolutional layers, ReLu activation function, batch normalization layer, and dropout layer. Fig. 8 illustrates the architecture. The image’s context is captured via the downsampling route. The standard stack of convolutional and maxpooling layers is all that is used for the downsampling procedure. The second method also called the upsampling path, is symmetric expanding and allows for exact localization using transposed convolutions.

The model’s first convolution layer of the model accepts the input of $160 \times 160 \times 3$. This convolution layer produces an output with a volume of $160 \times 160 \times 32$ by using 32 kernels of size 3×3 . Because the padding is equal to 1 and is maintained, the size of the output feature maps matches that of the input feature maps. Initializing the weights of the kernels uses a uniform distribution in the proposed segmentation. The area in the input volume that a certain feature extractor or kernel encompasses at a given time is known as the receptive field or context. The pooling process takes place between two layers in order to make the feature map smaller and transmit fewer parameters through the model. In the suggested model, the first max pooling layer uses a feature map with a size of $160 \times 160 \times 32$ and produces a feature map with a size of $80 \times 80 \times 32$. In this instance, the pool size is 2×2 . This procedure is carried out to keep the features that best capture the MRI’s context. In this manner, the pooling procedure in the downsampling route reduces the MRI size. The size of the MRI decreases as the network deepens, while the receptive field expands at the same time. The number of kernels is increasing at each stage of the downsampling process to extract increasingly intricate features from the input MRI. The image is upsampled in the upsampling layer using the transposed convolution approach. The feature map is routed via a 2×2 deconvolutional layer with strides equal to 2. It is transmitted through two 3×3 convolutional layers, much like in the downsampling path, and is sequentially concatenated with the prior feature map. The aforementioned procedure is continued until an image of size $160 \times 160 \times 32$ is obtained, at which point a 1×1 convolution layer is applied to produce an output of size $160 \times 160 \times 3$. The detailed description for each layer of the proposed segmentation model is given in Table 4.

The corresponding MRI and mask are used by this model from the manually masked segmentation dataset. For each of the four MRI classes to which the data set corresponds, the U-Net model will be trained.

3.8. Hyper parameters for segmentation model

The optimizer, input and output activation functions, loss function, batch size, number of epochs, training-validation splitting, and

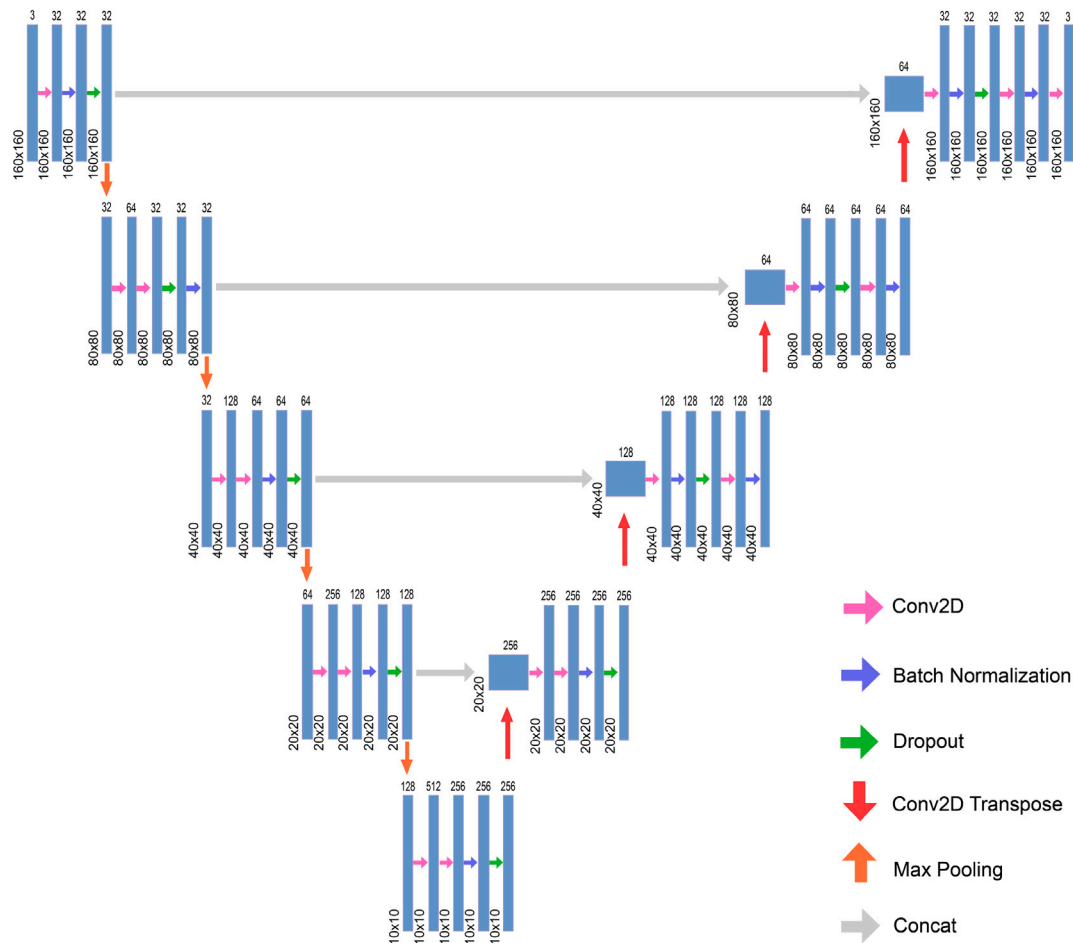


Fig. 8. The suggested segmentation model is based on the U-Net architecture and contains a downsampling and an upsampling route that accepts input MRIs and outputs a segmented mask image.

early-stopping patience have all been taken into account as model hyperparameters for the proposed segmentation model.

The optimizer used is Adam (Kingma & Ba, 2017). ReLU activation functions are used with convolution in the hidden layers of the model. The last convolution layer that generates the output makes use of the sigmoid activation function.

The segmentation model’s last layer has a binary cross-entropy loss function and a sigmoid activation function. The proposed segmentation model makes use of the binary cross-entropy loss function, which measures the divergence between the ground truth MRI masks and the predicted masks. The ground truth masks are the manually segmented mask of the merged dataset 1 in this case. The training approach used a grid search technique to identify the best parameter values for each combination of hyperparameters.

The batch size in the segmentation model is 32, and 200 epochs are taken into account. The proportion between training and validation is 80% and 20%. Early stopping patience is maintained at 100. The hyperparameter values are represented in Table 5.

3.9. Classification approaches

Two categorization methods were suggested by the methodology used in this research.

The first approach extracts the MRI from the dataset and sends it for preprocessing shown in Fig. 5. Pre-processing in the augmentation stage also applies to MRIs. The classification of MRIs into the four groups of glioma, meningioma, no tumor, and pituitary is then carried out by both the proposed model and the pre-trained networks. In this method, segmentation is not done before the classification goal.

The conventional approach is to first segment the image before sending it to be classified. As segmentation only pulls features from the MRI of the areas with the tumor. Time and computational cost are reduced if only the tumored area of the MRI is supplied to the classification model, rather than the entire image. Consequently, in the second method, the MRIs are delivered to the classification model after being segmented into tumor sections.

4. Results

The experiments are divided into a number of sections, including comparisons and individual outcomes. Each DL model’s effects have been examined independently, taking performance into account for each epoch. For training and testing, a total of 6 classification datasets were used. Among those, 4 are individual datasets (Dataset a (Cheng, 2017), Dataset b (Bhuvaji, 2020), Dataset c (Pradeep, 2021), Dataset d (Sherif, 2020)), and 2 are merged datasets (Merged dataset 1 and Merged dataset 2). One segmentation dataset has been employed, which includes a custom mask for the Merged dataset 1 mentioned before. Five transfer learning models, including the neural network that has been proposed, are trained and tested on all of the classification datasets. And a different segmentation model, has been used throughout the studies. The results of data augmentation and comparison between the results of segmentation in brain tumor classification are also presented separately.

4.1. Experimental setup

The experiment has been run on both cloud services like google colab and using personal computing devices. Since the scope of the

Table 4

The description for each layer of the proposed segmentation model based on the U-Net architecture having downsampling and upsampling paths.

Layer No.	Layer name	Filter count	Filter size	Dropout rate
Layer 1	Convolution+ReLU	32	3 × 3	10%
	Batch Normalization			
	Dropout			
	Maxpooling			
Layer 2	Convolution+ReLU	64	3 × 3	20%
	Convolution+ReLU			
	Dropout			
	Batch Normalization			
Layer 3	Maxpooling	128	3 × 3	20%
	Convolution+ReLU			
	Convolution+ReLU			
	Batch Normalization			
Layer 4	Dropout	256	3 × 3	20%
	Maxpooling			
	Convolution+ReLU			
	Convolution+ReLU			
Layer 5	Batch Normalization	128	3 × 3	30%
	Dropout			
	Conv2DTranspose			
	Convolution+ReLU			
Layer 4	Convolution+ReLU	256	3 × 3	20%
	Batch Normalization			
	Dropout			
	Conv2DTranspose			
Layer 3	Convolution+ReLU	128	3 × 3	20%
	Batch Normalization			
	Dropout			
	Conv2DTranspose			
Layer 2	Convolution+ReLU	64	3 × 3	20%
	Batch Normalization			
	Dropout			
	Conv2DTranspose			
Layer 1	Convolution+ReLU	32	3 × 3	10%
	Batch Normalization			
	Dropout			
	Convolution+ReLU			
	Convolution+ReLU	3	1 × 1	

Table 5

Hyperparameters of the proposed segmentation model.

Hyperparameters	Values
Number of trainable parameters	4,162,499
Batch size	32
Number of epochs	200
Optimizer	Adam
Hidden layer activation function	ReLU
Output layer activation function	Sigmoid
Early stopping patience	100
Training/validation split	80/20%
Loss function	Binary cross entropy

experiment is quite huge, which includes executing multiple transfer learning and proposed neural networks on a variety of classification and segmentation datasets, simultaneously running them was crucial. Neural networks are built using python 3.9 with several packages such as Tensorflow, Sklearn, Numpy, and more.

4.2. Performance metrics

Accuracy is considered as a significant evaluation metric for classification purposes. Precision indicates how much a classifier can be trusted when it indicates that an instance belongs to the positive class. A high precision rating indicates that there are very few false positives, and the classifier is very strict in the criteria for classifying something as positive. The ratio of all instances properly classified in the positive class to the total number of real members of the positive class is defined as recall. In other words, it indicates how many of the total number of positive instances are classified properly. A metric for sensitivity and specificity is called ROC AUC. The ROC is a graph that depicts the relationship between True Positive Rate (TPR) and False Positive Rate (FPR). The resultant score, known as the ROC AUC score, is the area under this ROC curve and broadly illustrates how well the model can predict classes.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{6}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{7}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{8}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{9}$$

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{10}$$

$$TPR = \frac{TP}{(TP + FN)} \tag{11}$$

$$FPR = \frac{FP}{(FP + TN)} \tag{12}$$

$$FNR = 1 - TPR \tag{13}$$

$$TNR = 1 - FPR \tag{14}$$

The terms TP, TN, FP, and FN in equations refer to the number of true positive predictions, true negative predictions, false positive predictions, and false negative predictions, respectively.

The F1 score, also known as the F-measure, is specified as the harmonic mean of recall and precision and has been used as a classification evaluation metric. It is a statistical measure of a model's accuracy. It is mathematically represented as follows:

$$F1Score = \frac{2 * Precision * Recall}{(Precision + Recall)} \tag{15}$$

Dice coefficient, dice loss, binary cross entropy dice loss and Intersection Over Union (IOU), recall, and precision has been considered as the evaluation metrics for the segmentation performance metrics. IOU also referred to as the Jaccard Index, is the area of overlap between the ground truth and the predicted segmentation and is divided by the area of union between the ground truth and the predicted segmentation.

$$JaccardIndex = \frac{AreaOfOverlap}{AreaOfUnion} \tag{16}$$

4.3. Effect of augmentation

Data augmentation plays an important role in the case of an imbalanced dataset or a lower quantity of data. Since brain tumor MRI data are hard to find and labeling these data requires an expert opinion, the amount of data along with the number of datasets are pretty rare. To observe the effect of augmentation, Merged dataset 1, which was created by concatenating three individual datasets and had imbalance, ie: with different amounts of class-wise MRI, were passed through several operations to create and augmented datasets. These operations

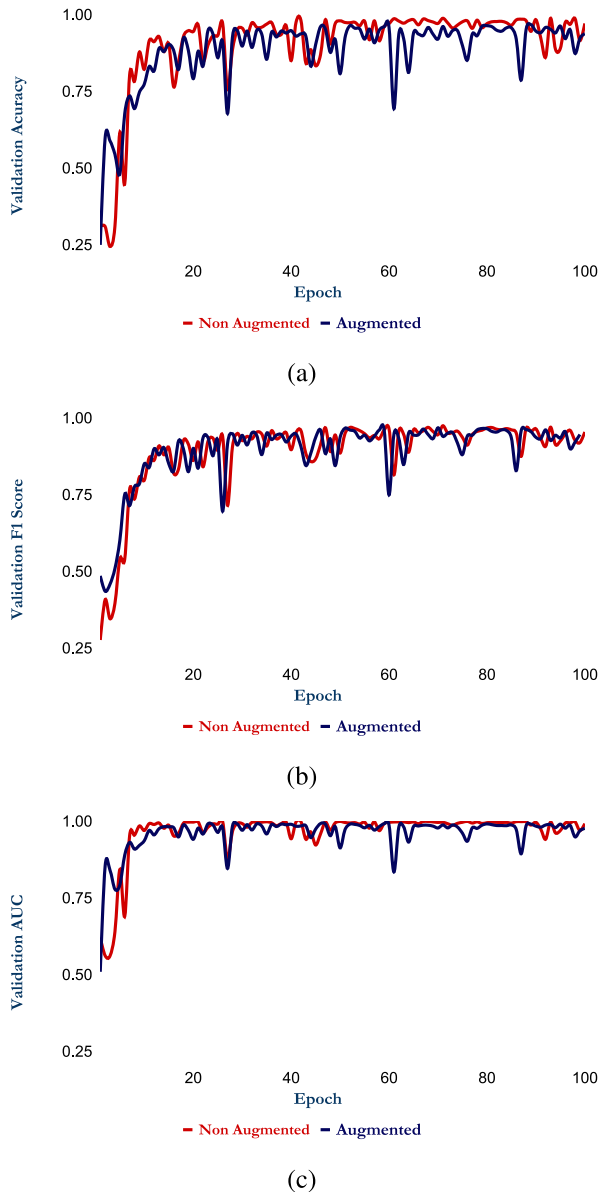


Fig. 9. Line charts showing suggested classification model validation accuracy, F1 score, and AUC vs epochs using Merged dataset 1 and Augmented Merged dataset 1. (a) The chart showing the validation accuracy comparison for these two datasets (b) the validation F1 score comparison (c) the validation AUC comparison.

include rotation, zooming, height and width shifting, shearing, horizontal flipping, and mode filtering. To show the effect of augmentation, the accuracy of the epoch-wise validation, the F1 score, and the AUC of the augmented and non-augmented data set are depicted in Fig. 9(a), Fig. 9(b) and Fig. 9(c) respectively. All of these figures represent similar trends and values while having more oscillations in the case of augmented data. Especially for validation accuracy Fig. 9(a) augmented results are a little below the non-augmented ones, with additional fluctuations or drops. The validation F1 score shown in Fig. 9(b) and AUC in Fig. 9(c) were almost identical epoch-wise with slight changes for AUC values in some epochs, and to our surprise, there is not much distinction between the augmented datasets and non-augmented ones, even in the last epochs the results of augmentations have deteriorated. Since the 5 merge augmentation has almost the same results that are not included in the section.

4.4. Impact of datasets in our proposed classification model

The proposed model has been tested and validated on all the individual datasets and merged datasets. Throughout the study, test sets had been kept constant to ensure clear comparisons. This part of the experiment is concerned with the effect of different datasets in our proposed models. To represent that effect, a handful of line charts and a confusion matrix for each dataset's validation part are given.

The proposed model has been trained and tested for 100 epochs. The line graph in Fig. 10(a) illustrates the training and validation accuracy and AUC over the epochs for Dataset a. Since the proposed model is deployed for the classification task, AUC is given since it represents the separability of tumor classes. The highest validation accuracy is 96.73 percent with a training accuracy of 99.83 percent. On the other hand, both the training and validation AUC are quite high at 99.99 and 99.56 percent respectively. The training validation trends for both of these metrics are quite similar, with some minor oscillations. The next Fig. 10(b) showcases the confusion matrix for the testing data where the actual vs the predicted value is given. Only a few glioma and meningioma were misclassified, although these numbers are very low compared to the correct classification.

Similarly, for Dataset b, epoch-wise accuracy, and AUC curve have been illustrated in Fig. 11(a), which has a slight difference between the training and validation results. Since this dataset was imbalanced, without the augmentation process, there was some deviation in each epoch, although the fluctuations were minimal. The maximum training accuracy is 99.93 percent and the AUC is 99.99 percent whereas the validation AUC fell down to 88.85 percent. In Fig. 11(b) the corresponding validation confusion matrix is given.

Again in Fig. 12(a), the same as previously, the accuracy and AUC curve is illustrated, while Fig. 12(b) delineated the confusion matrix for Dataset c. Similar to other single datasets there were some ups and downs in the validation curves, as training progressed, and the maximum results were quite high, as the training and validation AUC are 99.9% and 96.88% respectively. Again in the confusion matrix, the correct classifications are way higher than the false ones. Dataset d also follows a similar pattern to Dataset c which is shown in Fig. 13(a) and Fig. 13(b). Fig. 13(a) shows similar oscillations, although they stabilize at the end, and the validation AUC has the value of 0.9688. On the other hand, Fig. 13(b) depicts the confusing matrix for Dataset d.

4.5. Classification comparison on two merged datasets

To better understand the impact on performance datasets, two merged data sets were also validated in the proposed model. Fig. 14(a) and Fig. 14(b) demonstrate the validation confusion matrix for the Merged dataset 1 and Merged dataset 2. Since the merged dataset is slightly imbalanced, there were differences in the count of separate classes, however, most of the predicted classes are as can be seen in Fig. 14(a). Comparably, in Fig. 14(b), there are only a few misclassifications, in total 103, compared to the correct classification which is around 4214.

The comparative analysis of various evaluation metrics for different transfer learning models and the proposed model is illustrated in Figs. 9 to 19. The first comparison is done on the Merged dataset 1, where training and testing accuracy of the proposed model along with VGG16, VGG19, Resnet152v2, EfficientNet B0, EfficientNet B7 is given. Although a lot more transfer learning models have been tested for this purpose, the models mentioned above predicted surprisingly well, sometimes on par with the proposed model, compared to others. Fig. 15(a) represents the comparisons of training metrics, where the proposed model achieved an astonishing 99.8 percent accuracy and F1 scores. The results of VGG16, Resnet152v2, and EfficientNet B0 are also quite similar, reaching the 99 percent mark. In contrast, in the case of the accuracy of the tests in Fig. 15(b) and the F1 score, EfficientNet B0 comes down to 97 percent. Again, for testing metrics, the proposed

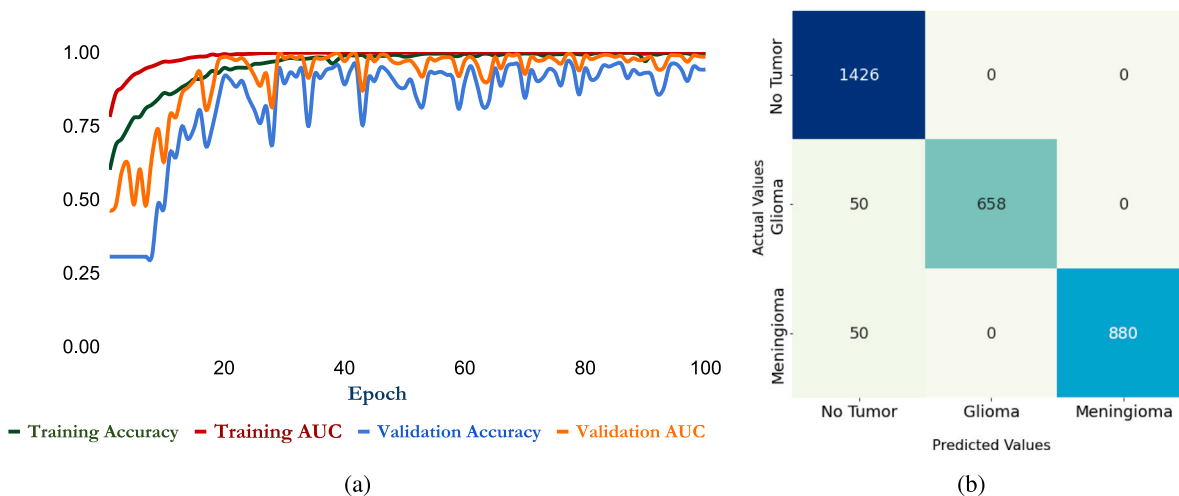


Fig. 10. Graph showing training and validation accuracy and AUC comparison for Dataset a of the proposed classification model (b) The confusion matrix for Dataset a showing the number of actual and predicted images for each class of the proposed classification model.

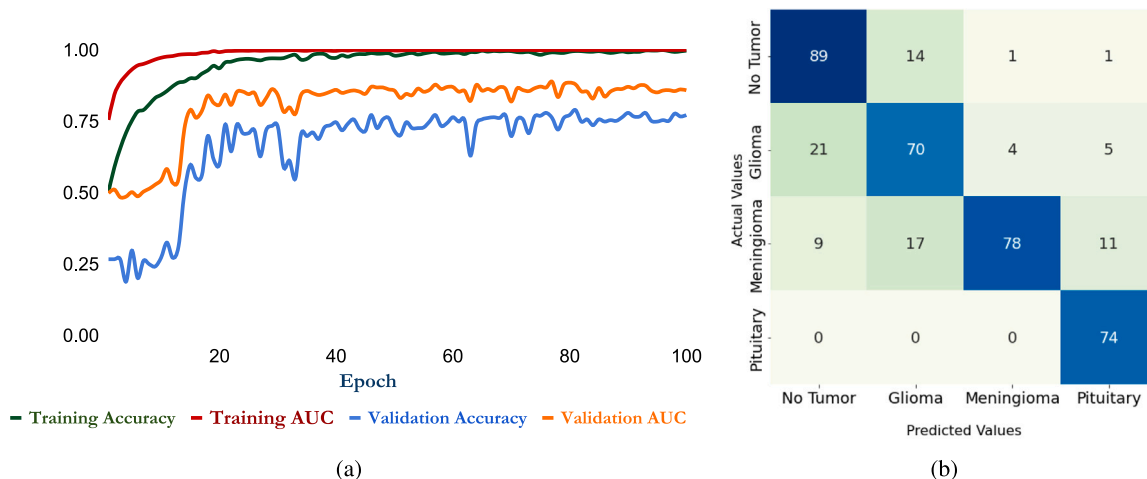


Fig. 11. (a) Line chart comparing training and validation accuracy and AUC for the proposed classification model on Dataset b. (b) Dataset b's confusion matrix displays the number of actual and predicted images for each class of the proposed model.

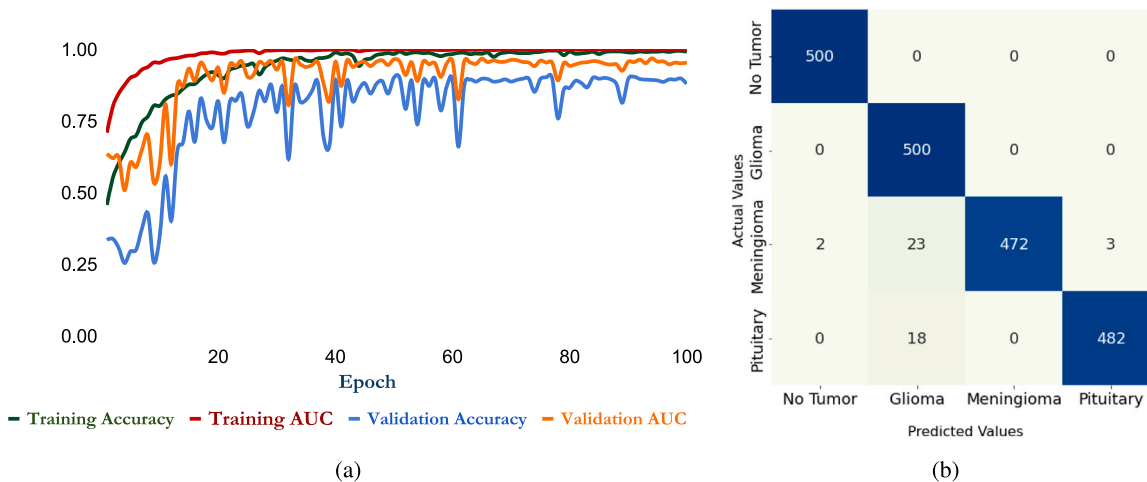


Fig. 12. (a) Graph comparing training and validation accuracy and AUC for Dataset c of the proposed classification model. (b) The confusion matrix for Dataset c shows the number of actual and predicted images for each class.

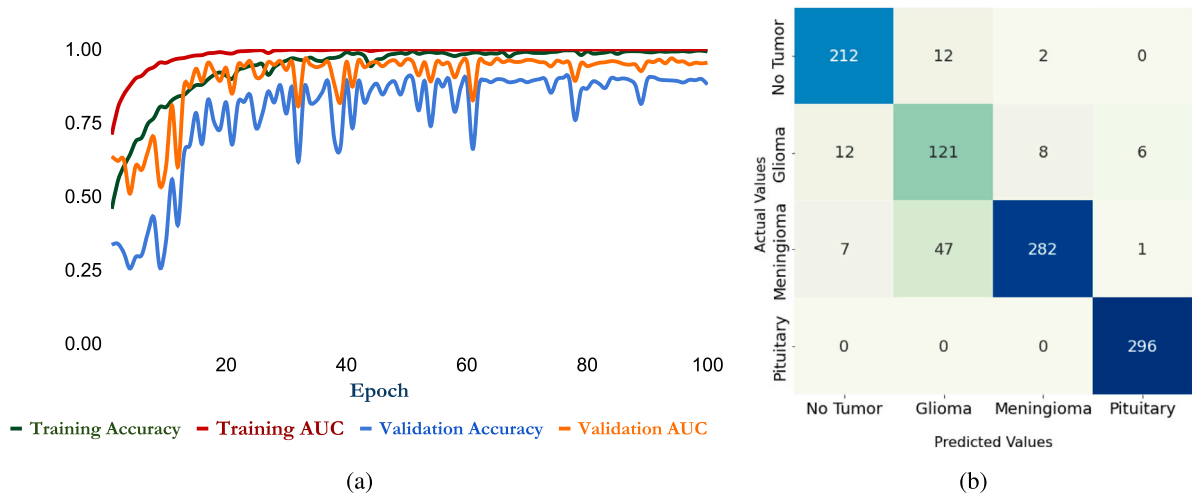


Fig. 13. (a) Graph showing training and validation accuracy and AUC comparison for Dataset d of the proposed classification model (b) The confusion matrix for Dataset d showing the number of actual and predicted images for each class of the proposed classification model.

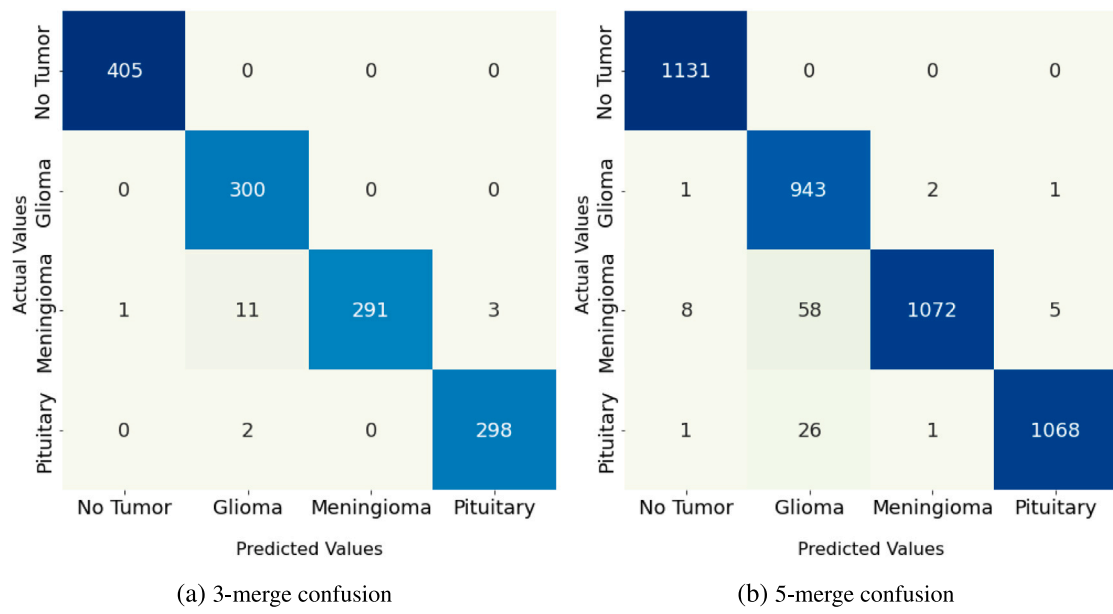


Fig. 14. Confusion matrix illustrating the number of actual and predicted MRIs following application of the proposed classification model. (a) Merged dataset 1, (b) Merged dataset 2.

model achieved a higher value than any other model, at a value of 98.7%, which is also very close to train accuracy, provided the fact that the model was well trained. The recall, precision, and F1 score were also quite similar to accuracy, ranging from 98.7 to 98.8 percent.

Fig. 16 shows the accuracy of training and validation and AUC for the proposed model and the best transfer learning model from the previous comparison-VGG16, which indicates that the proposed model suffers from less oscillation than the VGG16 model. Also, the training and validation accuracy are almost close to each other, so there might not be any overfitting or underfitting issues.

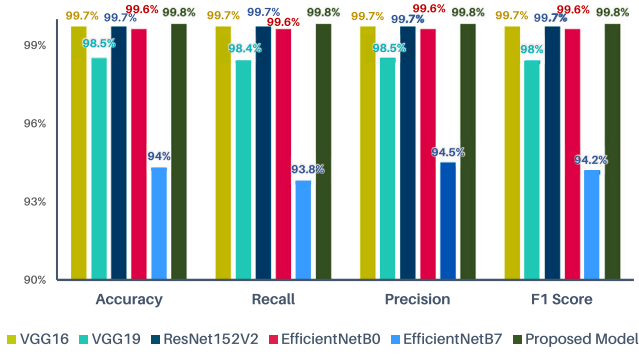
Since we have also created a separate augmented dataset from the Merged dataset 1, another bar chart is given in Fig. 17 to showcase validation performance comparisons, where our proposed model performs better than the other models in terms of results. The proposed model achieved validation accuracy and F1 score of 96.7 percent, and 99.1 percent AUC.

Again in Fig. 18(a) and Fig. 18(b) training and testing evaluation are given to create comparisons among models for Merged dataset 2. Similarly in the comparison of the Merged dataset 1, the models achieve

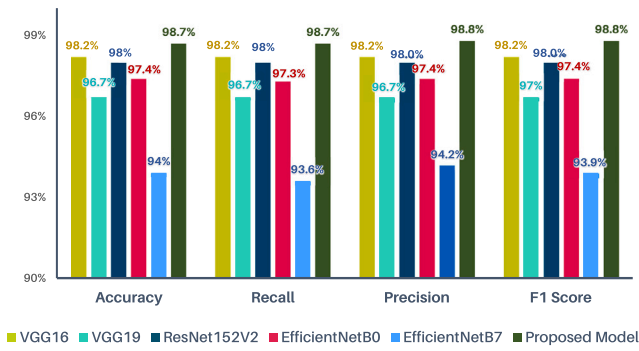
almost similar results in training; however, VGG19 and EfficientNet B7 perform significantly worse than the proposed model which resulted in an accuracy of 99.9 percent. Although in validation performance, the proposed model outclasses the other models, with accuracy and a F1 score of 97.6 percent, where all transfer learning models have these values less than or equal to 95 percent. Fig. 19 represents another epoch-wise line chart comparing the VGG16 model performance with the proposed model. Similar to Fig. 16, there were only minor fluctuations in the learning process of the proposed model. Although the proposed model has a lot fewer parameters than the transfer learning models, it performs better in terms of both training and validation.

4.6. Overall classification performance analysis

As the proposed model has been trained and tested overall 6 datasets, the performance metrics value on all these datasets are examined thoroughly in this section. Table 6 shows the proposed model's performance metrics on all these datasets. All of these parameters, including the values of accuracy, recall, precision, F1 score, FPR, TPR,



(a)



(b)

Fig. 15. Performance metrics for 5 transfer learning techniques vs the proposed approach (a) For the situation of training Merged dataset 1, where the metrics values are relatively near (b) For validation of Merged dataset 1, where the proposed model outperforms.

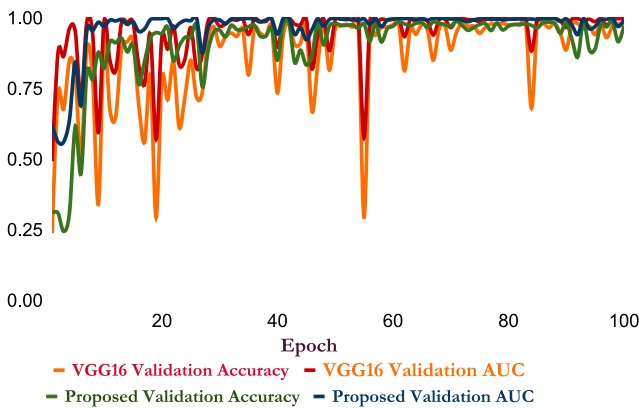


Fig. 16. Graph comparing training and validation accuracy and AUC for the proposed classification model to VGG16 for Merged dataset 1.

False Negative Rate (FNR), True Negative Rate (TNR), sensitivity, and specificity, have been taken into account. Table 6 represents ten statistical metrics on different datasets for our proposed models. On Dataset b and Dataset d there are a few false negative predictions which result in lower recall and sensitivity in terms of performance. This might be due to fewer images in these datasets since the false-positive rate is lower. However, in the case of merging these four datasets, the number of increased images and balancing each class balance the FPR and FNR problem.

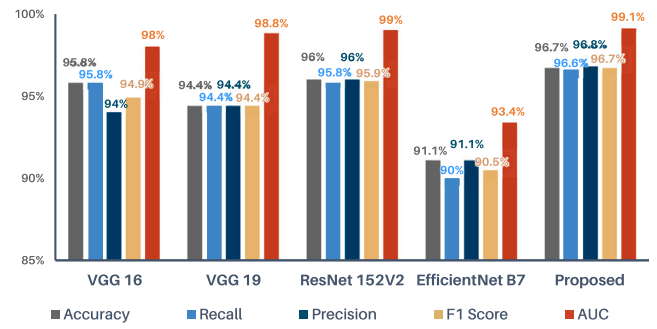
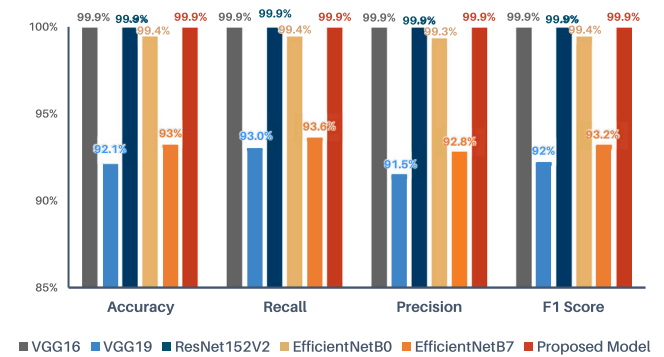
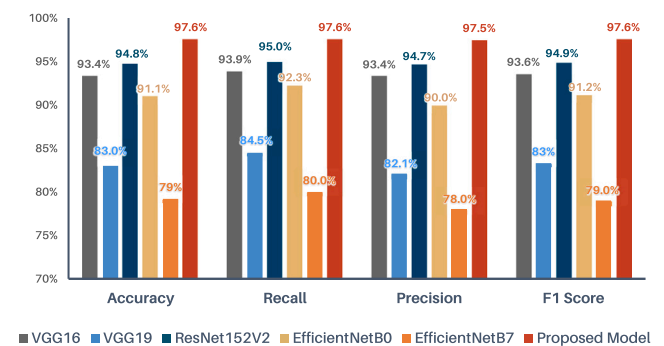


Fig. 17. Comparison of the proposed model with 5 transfer learning models on Augmented Merged dataset 1, where proposed models surpass.



(a)



(b)

Fig. 18. Metrics of performance for 5 transfer learning approaches against the suggested methodology (a) For the training Merged dataset 2 condition, where the metrics values are pretty close (b) For the validation Merged dataset 2 situation, where the proposed model outperforms.

4.7. Segmentation result analysis

To separate the tumor portion from brain MRI images, a custom-made U-Net was also trained on the segmentation dataset. Fig. 20 illustrates the output of the proposed segmentation model for different classes in the dataset with the ground truth image.

The input MRI and the predicted mask generated for Dataset d using the proposed segmentation model are shown in Fig. 21. The graphic demonstrates how the model conveniently segments the tumor sections from the input MRI.

The segmentation model was also tested using a variety of statistical methods and image similarity measurements. Fig. 22(a) and Fig. 22(b) represent training and validation precision and cross-entropy loss over

Table 6
Statistical Analysis of the Performance Metrics for the Proposed Model Considering All Datasets.

Performance metrics	Dataset a	Dataset b	Dataset c	Dataset d	Merged dataset 1	Merged dataset 2
FPR	0	7.02	0.79	3.14	0.43	0.71
FNR	6.32	21.06	2.39	9.44	1.29	2.61
TPR	93.67	78.9	97.6	90.5	98.7	97.3
TNR	100	92.9	99.2	96.8	99.5	99.2
Accuracy	96.7	89.4	97.7	95.2	98.7	97.6
Recall	93.6	78.9	97.6	90.5	98.7	97.6
Precision	100	78.9	97.7	90.5	98.8	97.6
F1 Score	96.7	78.9	97.7	90.5	98.7	97.6
Sensitivity	93.6	78.9	97.6	90.5	98.7	97.3
Specificity	100	92.9	99.2	96.8	99.5	99.2

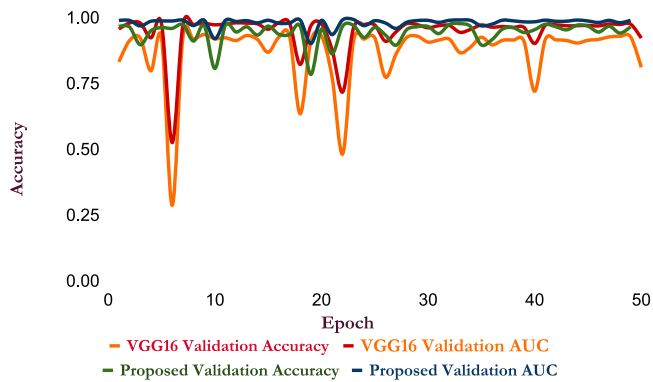


Fig. 19. Graph comparing training and validation accuracy and AUC for the proposed classification model to VGG16 for Merged dataset 2.

Table 7
Performance Metrics for the proposed segmentation model.

Performance metrics	Values
Dice Coefficient	0.89
Dice Loss	0.10
Binary cross entropy Dice loss	0.79
Intersection Over Union (IOU)	0.81

200 epochs for the segmentation model. Both loss values are quite low, close to 0.1 and the precision is above 90 percent. There were almost no fluctuations in the training process.

Fig. 23 represents the statistical and similarity measure for the validation data, the validation cross-entropy loss is 6.6%, with 98.4 percent precision, and the recall is 71%.

From Table 7, the values of performance metrics are noted as the dice coefficient value is 0.89, the Jaccard index (IOU) is 0.81, dice loss and binary cross entropy dice losses are 0.10 and 0.79 respectively.

4.8. Effect of segmentation

Another comparison is made between segmented and non-segmented classification. Since segmentation significantly reduces the size of the data, the overall time of classification models is reduced. Fig. 24 illustrates the validation results of different classification models in a segmented tumor image created from a Merged dataset 1. The proposed model achieved 99 percent accuracy, 98.7 percent F1 score and 99.6 percent AUC for segmented tumor classification. Compared to the proposed model, the best transfer learning model is Resnet152v2 with a validation accuracy of 96 percent, which is lower than the proposed model. In terms of the effect of segmentation, Fig. 25 describes the difference between segmented and non-segmented tumor classification. The statistical metrics are quite similar, varying only by 1 percent to the utmost. The proposed models achieved 98.8 percent validation accuracy with the segmentation process, while without segmenting the

tumor portion from the MRIs, they achieved 98.7 percent validation accuracy. The time needed to train the classification model without segmentation was 21 min 66 seconds, compared to the segmentation process, which requires a total time of 58 minutes 33 seconds. Since in the second approach two models, one U-Net for segmentation, and one classification model are needed to train, it had a higher training time. However, in case of just classification after segmentation, the time reduces to only 13 minutes, which is 40% lower than the previous approach.

4.9. Comparison with state-of-the-art papers

A number of State-of-the-Art studies have been compared to the proposed work. Glioblastoma and glioma classes are binary classified by Rajinikanth, Kadry, and Nam (2021). This study seeks to categorize tumors using a segmentation approach. Whereas our segmentation classification accuracy is 98.8%, their classification accuracy is 98%. With our suggested model's segmentation strategy, the accuracy is greater. In addition, our suggested model has an accuracy of 98.7% when classifying tumors into four unique classes without segmenting the MRI data. Using a combination of CNN and AlexNet models, Badjie and Ülker (2022) conducted binary classification between normal MRIs and tumor MRIs. Although their model classified the MRIs with greater accuracy, they did not segment the MRIs. This was a concern of our research because it compares how the MRIs are classified with and without segmentation. Maqsood, Damaševičius, and Maskeliūnas (2022) proposed a 17-layer deep neural network design for segmentation. They developed a Multiclass SVM to divide MRIs into three categories (glioma, meningioma, and pituitary). In their instance, the feature selection procedure for classifying tumor MRI takes a while. Woźniak, Siłka, and Wieczorek (2021) proposed a correlation learning mechanism that combines CNN with classic architecture. This model achieves an accuracy of 95.09% on Dataset a, which is lower than in our instance because the accuracy of the proposed model is 97.7% on an individual dataset. And both with and without conducting the segmentation process, the merged datasets offer superior accuracy. The authors of paper Khan et al. (2022) created a technique for detecting and classifying brain tumors. The approach uses a saliency map and is based on a deep learning feature optimization strategy. Although the suggested model has a few stages, its greatest accuracy is 95.94%.

The proposed model is compared to several existing classification models in relation to Dataset a. As a result, all of the papers in Table 8 conducted MRI image multiclassification and classified the images as glioma, meningioma, pituitary, or no tumor. Proposed model and other state-of-the-art papers comparison on Dataset a is shown in Fig. 26.

Ayadi et al. (2021) also implemented a CNN approach and achieved an accuracy of 94.74%. Ghassemi et al. (2020) proposed a GAN model where a deep convolutional neural network is used as a discriminator for the detection of fake images that are generated by the generative model and a pre-trained CNN network is fine-tuned to perform as the classifier for brain tumor classifications and the model achieved an accuracy of 95.6%. Ismael and Abdel-Qader (2018) combined the

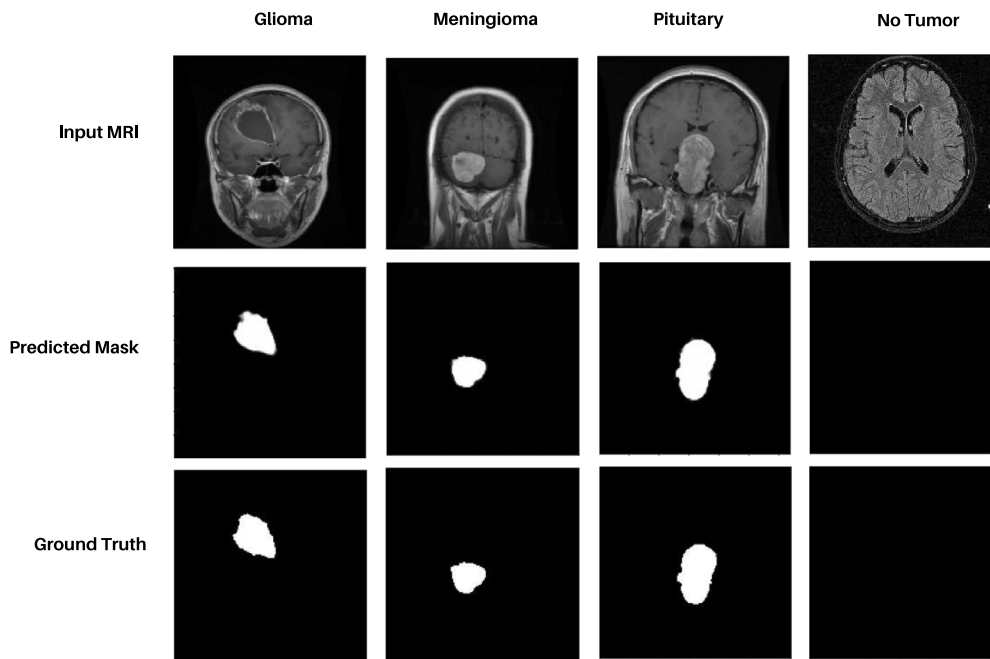


Fig. 20. The proposed segmentation model’s projected mask and growth truth comparison, which is based on the U-Net architecture and accepts input MRIs and outputs segmented mask image.

a

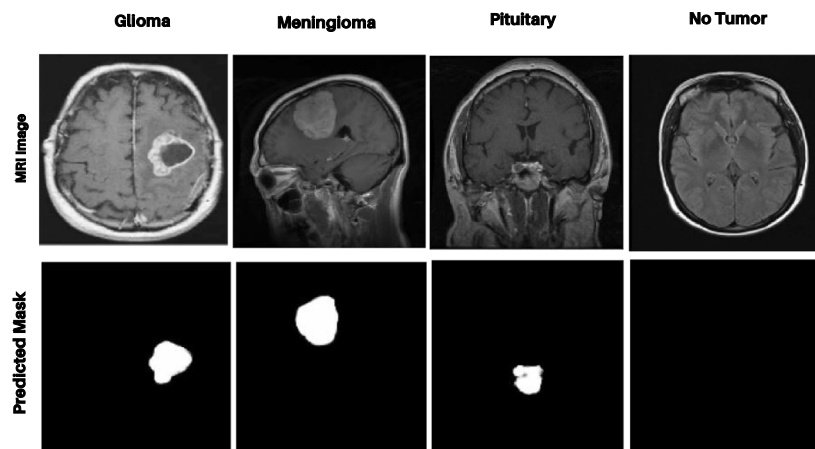


Fig. 21. Predicted mask generated from input MRI using proposed segmentation model for classification Dataset d.

Table 8

Comparison among proposed model and State of The Art papers considering the Dataset a.

Ref.	Year	Method	Accuracy (Percentage)
Wozniak et al. (2021)	2021	CNN with classic architecture	95.09
Ayadi et al. (2021)	2021	CNN	94.74
Ghassemi et al. (2020)	2020	GAN+CNN	95.6
Badža and Barjaktarović (2020)	2020	CNN	96.56
Swati et al. (2019)	2019	AlexNet, VGG16, VGG19	94.82
Sultan et al. (2019)	2019	CNN	96.13
Pashaei, Sajedi, and Jazayeri (2018)	2018	CNN	93.68
Ismael and Abdel-Qader (2018)	2018	Neural Network	91.9
Afshar, Mohammadi, and Plataniotis (2018)	2018	Capsule Network	86.56
Proposed Model		Deep CNN	96.7

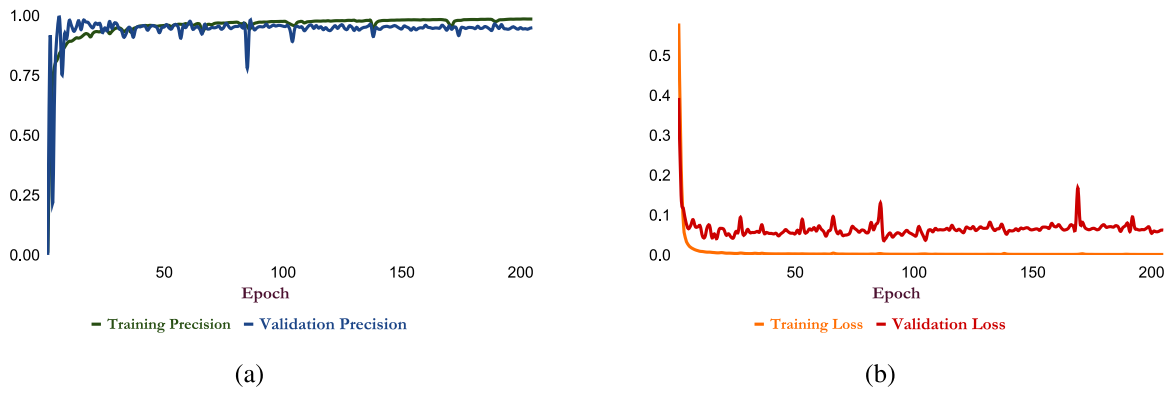


Fig. 22. The proposed segmentation model’s training and validation (a) precision vs. epoch (b) loss vs. epoch comparison, where the validation precision and loss both nearly match their respective training curves.

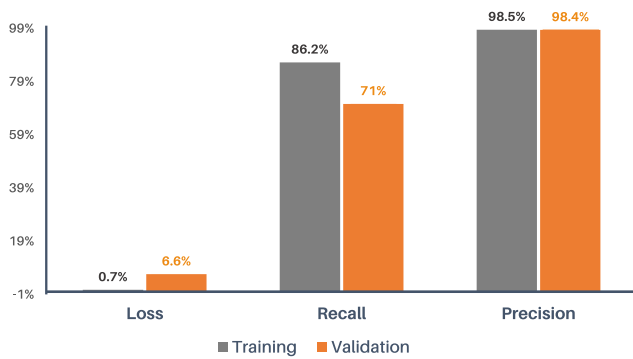


Fig. 23. Comparison of training and validation performance measures (loss, recall, and accuracy) for the proposed segmentation model.

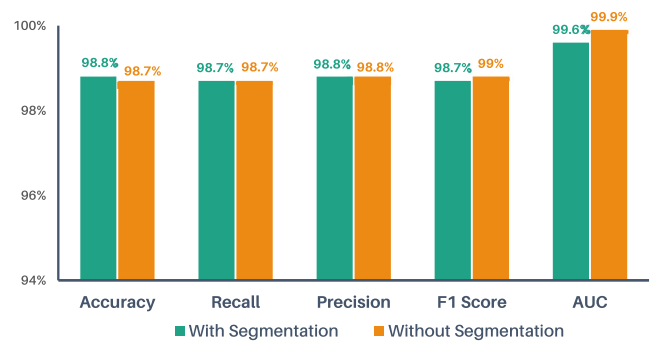


Fig. 25. Proposed classification model validation performance metrics comparison for Merged dataset 1 when tumor MRIs are segmented using the proposed segmentation model and when they are not segmented.

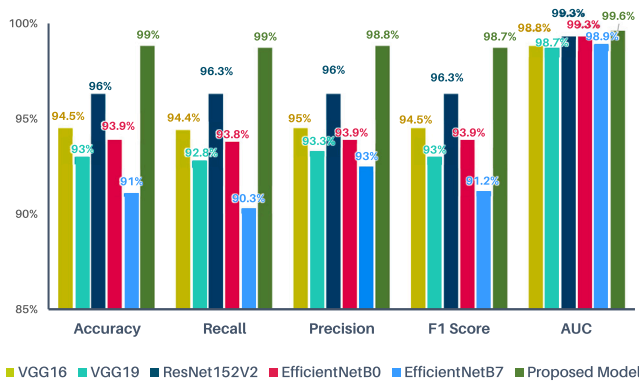


Fig. 24. Validation performance comparison of proposed classification model with 5 transfer learning methods after segmenting tumor MRIs using the proposed segmentation model and transferring them to the classification model on Merged dataset 1.

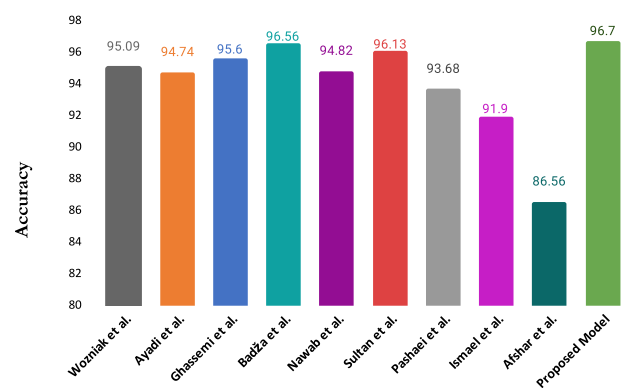


Fig. 26. Comparison of classification accuracy with proposed model and other State of The Art researches on Dataset a.

2D DWT and 2D gabor extraction methods for feature extraction, the features were fed to a traditional neural network and achieved an accuracy of 91.9%. Afshar et al. (2018) employed a capsule network on segmented tumor regions to address CNN’s shortcomings related to the loss of active features at the specific location in the subsampling layers and poor training results regarding small datasets. Pashaei et al. (2018) proposed a method that extracts features of brain tumors using CNN and further classifies the obtained features into three classes of tumors such as meningioma, glioma, and pituitary tumors using kernel extreme learning machines. With an accuracy of 96.7%, the proposed classification model beats the results of the research described above using Dataset a.

5. Discussions

This research provides a comprehensive comparison of several tumor classification algorithms. First, the influence of datasets was evaluated on different accessible datasets to determine whether it had any effect on classification. Because the size and output classes of these datasets (Dataset a, Dataset b, Dataset c, and Dataset d) vary, the findings were fairly disparate. However, the proposed strategy produced some promising results where dataset c showed the highest accuracy with 97.7%. The suggested model was then trained and validated in two data sets (Merged dataset 1 and Merged dataset 2) developed by combining existing ones. There are three types of tumor MRI in these datasets, with normal MRIs. And Merged dataset 1 showed an

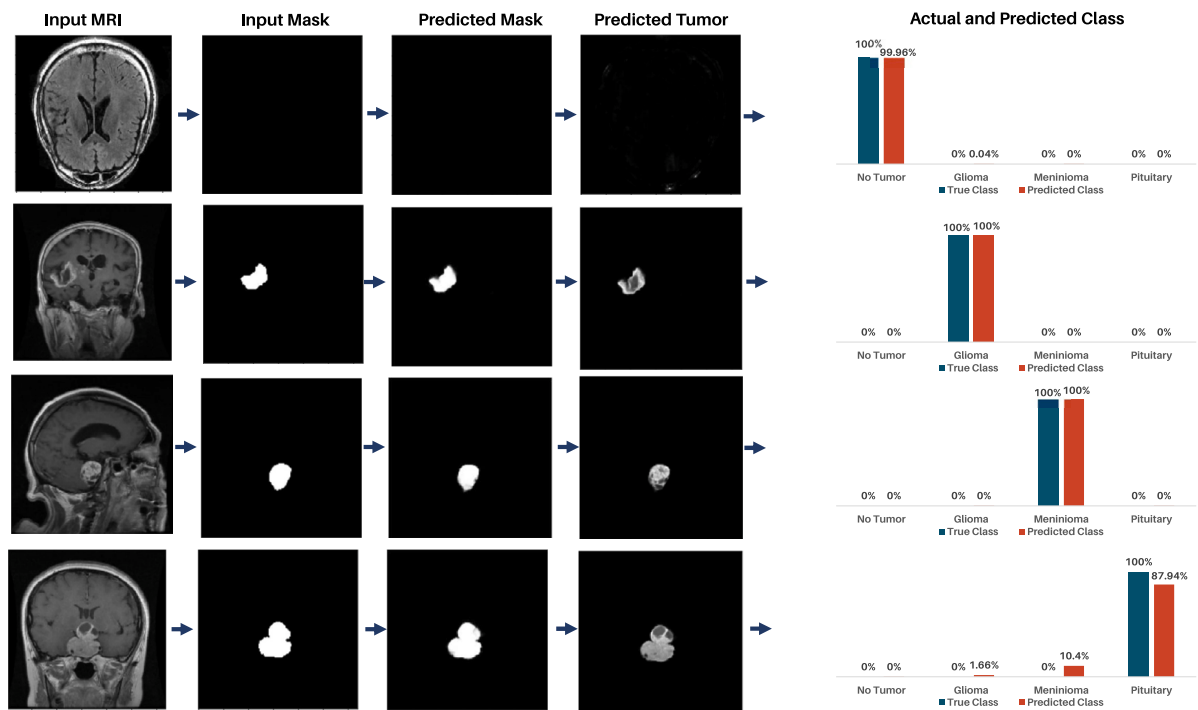


Fig. 27. The whole classification process including the segmentation technique is shown, including the input MRI, input mask, generated mask, tumor area, actual class, and predicted class.

accuracy of 98.7% which is higher than both of the merged datasets. The model was then compared with several transfer learning models to demonstrate that, even without a huge number of parameters, the proposed model matches or outperforms the transfer learning models in most circumstances. The effect of segmentation was then demonstrated by training and validating both segmented and non-segmented MRIs. The findings clearly illustrate that, whereas segmentation requires two steps of model training, classification obtained equivalent or even superior results in less time. The segmentation classification approach showed an accuracy of 98.8%. So, if training time has no influence on the application, segmentation can improve classification performance. However, if a minor variation in performance is not a problem, the suggested model may predict fairly well without any segmentation with an accuracy of 98.7%. The input MRIs, input mask, predicted mask from the segmentation model, predicted tumor area, and finally the percentage of prediction of the tumor MRIs identified from the classification model, which uses segmented MRIs before completing the classification, are presented in Fig. 27. As shown in the Figure, the classification model with segmentation detects normal MRIs with 99.96% accuracy, glioma and meningioma classes with 100% accuracy, and pituitary tumors with 87.94% accuracy.

6. Conclusion

Early identification of brain tumors can be crucial in saving lives as they can be extremely harmful and even deadly. This study proposes an automated classification strategy for a rapid, early, and accurate diagnosis to prevent disastrous consequences. A deep CNN model was used to classify brain MRIs into four groups: glioma, meningioma, no tumor, and pituitary. An accurate and automatic segmentation model was applied to segment brain MRIs from the original input MRIs. Effective automated tumor segmentation can be challenging due to the wide variety of tumor locations, shapes, and structures. For segmenting the tumor sections, a U-Net architecture-based model was developed. The classification techniques were found to be not significantly different whether segmentation was used or not. However, non-segmentation classification reduced the time required by the classification model.

The impact of dataset augmentation was also studied. The classification model was evaluated across multiple datasets and with five pre-trained models. Furthermore, the proposed model outperformed the pre-trained models as well as state-of-the-art articles. The segmentation approach also generated a more accurate segmented mask, allowing for the successful segmentation of every MRI image from any dataset. However, our proposed model could be more useful if we have more real-life MRI data for the segmentation and training of the model.

CRedit authorship contribution statement

Atika Akter: Conceptualization, Formal analysis, Methodology, Writing – original draft. **Nazeela Nosheen:** Software, Validation, Formal analysis. **Sabbir Ahmed:** Formal analysis, Investigation, Software. **Mariom Hossain:** Resources, Data curation, Visualization. **Mohammad Abu Yousuf:** Supervision, Data curation. **Mohammad Ali Abdullah Almoayad:** Project administration. **Khondokar Fida Hasan:** Writing – original draft, Formal analysis, Investigation. **Mohammad Ali Moni:** Visualization, Project administration, Supervision.

Declaration of competing interest

We do not have any conflict of interest.

Data availability

Data will be made available on request.

Acknowledgment

This research was supported by the Deanship of Scientific Research Large Groups at King Khalid University, Kingdom of Saudi Arabia (RGP.2/219/43).

References

- Afshar, P., Mohammadi, A., & Plataniotis, K. N. (2018). Brain tumor type classification via capsule networks. In *2018 25th IEEE international conference on image processing* (pp. 3129–3133). IEEE.
- Ahamed, K., Islam, M., Uddin, M. A., Akhter, A., Paul, B. K., Yousuf, M., et al. (2021). A deep learning approach using effective preprocessing techniques to detect COVID-19 from chest CT-scan and X-ray images. *Computers in Biology and Medicine*, *139*, Article 105014. <http://dx.doi.org/10.1016/j.compbiomed.2021.105014>.
- Amin, J., Sharif, M., Yasmin, M., & Fernandes, S. L. (2018). Big data analysis for brain tumor detection: Deep convolutional neural networks. *Future Generation Computer Systems*, *87*, 290–297, Publisher: Elsevier.
- Arokia Jesu Prabhu, L., & Jayachandran, A. (2018). Mixture model segmentation system for parasagittal meningioma brain tumor classification based on hybrid feature vector. *Journal of Medical Systems*, *42*(12), 251. <http://dx.doi.org/10.1007/s10916-018-1094-3>.
- Aurna, N. F., Yousuf, M. A., Taher, K. A., Azad, A. K. M., & Moni, M. A. (2022). A classification of MRI brain tumor based on two stage feature level ensemble of deep CNN models. *Computers in Biology and Medicine*, *146*, Article 105539. <http://dx.doi.org/10.1016/j.compbiomed.2022.105539>, URL <https://www.sciencedirect.com/science/article/pii/S0010482522003316>.
- Ayadi, W., Elhamzi, W., Charfi, I., & Atri, M. (2021). Deep CNN for brain tumor classification. *Neural Processing Letters*, *53*(1), 671–700, Publisher: Springer.
- Badjie, B., & Ülker, E. D. (2022). A deep transfer learning based architecture for brain tumor classification using MR images. *Information Technology and Control*, *51*(2), 332–344.
- Badža, M. M., & Barjaktarović, M. (2020). Classification of brain tumors from MRI images using a convolutional neural network. *Applied Sciences*, *10*(6), 1999, Publisher: Multidisciplinary Digital Publishing Institute.
- Bhuvaji, S. (2020). Brain tumor classification (MRI). <https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri>. (Accessed 12 April 2021).
- Chelghoum, R., Ikhlef, A., Hameurlaine, A., & Jacquir, S. (2020). Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images. In *IFIP international conference on artificial intelligence applications and innovations* (pp. 189–200). Springer.
- Cheng, J. (2017). Brain tumor dataset. https://figshare.com/articles/dataset/brain_tumor_dataset/1512427/5. (Accessed 9 April 2021).
- Elazab, A., Wang, C., Gardezi, S. J. S., Bai, H., Hu, Q., Wang, T., et al. (2020). GP-GAN: Brain tumor growth prediction using stacked 3D generative adversarial networks from longitudinal MR images. *Neural Networks*, *132*, 321–332, Publisher: Elsevier.
- Faruqui, N., Yousu, M., Whaiduzzaman, M., Azad, A. K. M., Barrosean, A., & Moni, M. A. (2021). LungNet: A hybrid deep-CNN model for lung cancer diagnosis using CT and wearable sensor-based medical IoT data. *Computers in Biology and Medicine*, *139*, Article 104961. <http://dx.doi.org/10.1016/j.compbiomed.2021.104961>.
- Ghassemi, N., Shoeibi, A., & Rouhani, M. (2020). Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images. *Biomedical Signal Processing and Control*, *57*, Article 101678, Publisher: Elsevier.
- Hamada, A. (2020). Br35H :: Brain tumor detection. <https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection>. (Accessed 14 April 2021).
- Hamghalam, M., Wang, T., & Lei, B. (2020). High tissue contrast image synthesis via multistage attention-GAN: application to segmenting brain MR scans. *Neural Networks*, *132*, 43–52, Publisher: Elsevier.
- Han, C., Murao, K., Noguchi, T., Kawata, Y., Uchiyama, F., Rundo, L., et al. (2019). Learning more with less: Conditional PGGAN-based data augmentation for brain metastases detection using highly-rough annotation on MR images. In *Proceedings of the 28th ACM international conference on information and knowledge management* (pp. 119–127).
- Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., et al. (2017). Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, *35*, 18–31, Publisher: Elsevier.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE conference on computer vision and pattern recognition* (pp. 770–778). Las Vegas, NV, USA: IEEE. <http://dx.doi.org/10.1109/CVPR.2016.90>, URL <http://ieeexplore.ieee.org/document/7780459/>.
- Irmak, E. (2021a). 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*, *45*(3), 1015–1036, Publisher: Springer.
- Irmak, E. (2021b). 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*, 1–22, Publisher: Springer.
- Irsheidat, S., & Duwairi, R. (2020). Brain tumor detection using artificial convolutional neural networks. In *2020 11th international conference on information and communication systems* (pp. 197–203). <http://dx.doi.org/10.1109/ICICS49469.2020.239522>.
- Islam, M. S., Hasan, K. F., Sultana, S., Uddin, S., Quinn, J. M., Moni, M. A., et al. (2023). HARD: A novel ECG-based heartbeat classification method to detect arrhythmia using hierarchical attention based dual structured RNN with dilated CNN. *Neural Networks*.
- Islam, M. M., Hossain, M. B., Akhtar, M. N., Moni, M. A., & Hasan, K. F. (2022). CNN based on transfer learning models using data augmentation and transformation for detection of concrete crack. *Algorithms*, *15*(8), 287.
- Ismael, M. R., & Abdel-Qader, I. (2018). Brain tumor classification via statistical features and back-propagation neural network. In *2018 IEEE international conference on electro/information technology* (pp. 0252–0257). IEEE.
- Khan, M. A., Khan, A., Alhaisoni, M., Alqahtani, A., Alsubai, S., Alharbi, M., et al. (2022). Multimodal brain tumor detection and classification using deep saliency map and improved dragonfly optimization algorithm. *International Journal of Imaging Systems and Technology*, Publisher: Wiley Online Library.
- Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. arXiv URL <http://arxiv.org/abs/1412.6980>[cs].
- Maqsood, S., Damaševičius, R., & Maskeliūnas, R. (2022). Multi-modal brain tumor detection using deep neural network and multiclass SVM. *Medicina*, *58*(8), 1090, Publisher: MDPI.
- Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., et al. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, *34*(10), 1993–2024, Publisher: IEEE.
- Mohamed Shakeel, P., Tobely, T. E. E., Al-Feel, H., Manogaran, G., & Baskar, S. (2019). Neural network based brain tumor detection using wireless infrared imaging sensor. *IEEE Access*, *7*, 5577–5588. <http://dx.doi.org/10.1109/ACCESS.2018.2883957>, Conference Name: IEEE Access.
- Mohsen, H., El-Dahshan, E.-S. A., El-Horbaty, E.-S. M., & Salem, A.-B. M. (2018). Classification using deep learning neural networks for brain tumors. *Future Computing and Informatics Journal*, *3*(1), 68–71, Publisher: Elsevier.
- Mzoughi, H., Njeh, I., Wali, A., Slima, M. B., BenHamida, A., Mhiri, C., et al. (2020). Deep multi-scale 3D convolutional neural network (CNN) for MRI gliomas brain tumor classification. *Journal of Digital Imaging*, *33*, 903–915, Publisher: Springer.
- Naser, M. A., & Deen, M. J. (2020). Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images. *Computers in Biology and Medicine*, *121*, Article 103758, Publisher: Elsevier.
- Nickparvar, M. (2021). Brain tumor MRI dataset. <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>. (Accessed 16 April 2021).
- Noreen, N., Palaniappan, S., Qayyum, A., Ahmad, I., Imran, M., & Shoaib, M. (2020). A deep learning model based on concatenation approach for the diagnosis of brain tumor. *IEEE Access*, *8*, 55135–55144, Publisher: IEEE.
- Öinar, A., & Yildirim, M. (2020). Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture. *Medical Hypotheses*, *139*, Article 109684, Publisher: Elsevier.
- Özyurt, F., Sert, E., & Avci, D. (2020). An expert system for brain tumor detection: Fuzzy C-means with super resolution and convolutional neural network with extreme learning machine. *Medical Hypotheses*, *134*, Article 109433, Publisher: Elsevier.
- Özyurt, F., Sert, E., Avci, E., & Dogantekin, E. (2019). Brain tumor detection based on Convolutional Neural Network with neurosophic expert maximum fuzzy sure entropy. *Measurement*, *147*, Article 106830, Publisher: Elsevier.
- Pashaei, A., Sajedi, H., & Jazayeri, N. (2018). Brain tumor classification via convolutional neural network and extreme learning machines. In *2018 8th international conference on computer and knowledge engineering* (pp. 314–319). IEEE.
- Pradeep (2021). Brain MRI. <https://www.kaggle.com/datasets/pradeep2665/brain-mri>. (Accessed 12 April 2021).
- Rajinikanth, V., Kadry, S., & Nam, Y. (2021). Convolutional-neural-network assisted segmentation and SVM classification of brain tumor in clinical MRI slices. *Information Technology and Control*, *50*(2), 342–356.
- Razzak, M. I., Imran, M., & Xu, G. (2018). Efficient brain tumor segmentation with multiscale two-pathway-group convolutional neural networks. *IEEE Journal of Biomedical and Health Informatics*, *23*(5), 1911–1919, Publisher: IEEE.
- Rezaei, M., Harmuth, K., Gierke, W., Kellermeier, T., Fischer, M., Yang, H., et al. (2017). A conditional adversarial network for semantic segmentation of brain tumor. In *International MICCAI brainlesion workshop* (pp. 241–252). Springer.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 234–241). Springer.
- Saba, T., Mohamed, A. S., El-Affendi, M., Amin, J., & Sharif, M. (2020). Brain tumor detection using fusion of hand crafted and deep learning features. *Cognitive Systems Research*, *59*, 221–230, Publisher: Elsevier.
- Sajid, S., Hussain, S., & Sarwar, A. (2019). Brain tumor detection and segmentation in MR images using deep learning. *Arabian Journal for Science and Engineering*, *44*(11), 9249–9261, Publisher: Springer.
- Sajjad, M., Khan, S., Muhammad, K., Wu, W., Ullah, A., & Baik, S. W. (2019). Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *Journal of Computational Science*, *30*, 174–182, Publisher: Elsevier.
- Saouli, R., Akil, M., & Kachouri, R. (2018). Fully automatic brain tumor segmentation using end-to-end incremental deep neural networks in MRI images. *Computer Methods and Programs in Biomedicine*, *166*, 39–49, Publisher: Elsevier.
- Shahzadi, I., Tang, T. B., Meriadeau, F., & Quyyum, A. (2018). CNN-LSTM: Cascaded framework for brain Tumour classification. In *2018 IEEE-EMBS conference on biomedical engineering and sciences IECBES*, (pp. 633–637). IEEE.
- Sharif, M. I., Li, J. P., Khan, M. A., & Saleem, M. A. (2020). Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images. *Pattern Recognition Letters*, *129*, 181–189, Publisher: Elsevier.

- Sherif, M. M. (2020). Brain Tumor Dataset. <https://www.kaggle.com/datasets/mohamedmetwallysherif/braintumordataset>. (Accessed 14 April 2021).
- Siar, M., & Teshnehlab, M. (2019). Brain tumor detection using deep neural network and machine learning algorithm. In *2019 9th international conference on computer and knowledge engineering ICCKE*, (pp. 363–368). IEEE.
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. arXiv URL <http://arxiv.org/abs/1409.1556>[cs].
- Sultan, H. H., Salem, N. M., & Al-Atabany, W. (2019). Multi-classification of brain tumor images using deep neural network. *IEEE Access*, *7*, 69215–69225, Publisher: IEEE.
- Suzuki, K. (2017). Overview of deep learning in medical imaging. *Radiological Physics and Technology*, *10*(3), 257–273.
- Swati, Z. N. K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., et al. (2019). Brain tumor classification for MR images using transfer learning and fine-tuning. *Computerized Medical Imaging and Graphics*, *75*, 34–46, Publisher: Elsevier.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). Going deeper with convolutions. In *2015 IEEE conference on computer vision and pattern recognition* (pp. 1–9). Boston, MA, USA: IEEE, <http://dx.doi.org/10.1109/CVPR.2015.7298594>, URL <http://ieeexplore.ieee.org/document/7298594/>.
- Talukder, M. A., Islam, M. M., Uddin, M. A., Akhter, A., Hasan, K. F., & Moni, M. A. (2022). Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning. *Expert Systems with Applications*, *205*, Article 117695.
- Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th international conference on machine learning* (pp. 6105–6114). PMLR, URL <https://proceedings.mlr.press/v97/tan19a.html>.
- Vincent, O. R., & Folorunso, O. (2009). A descriptive algorithm for sobel image edge detection. In *Proceedings of informing science & IT education conference*, vol. 40 InSITE, (pp. 97–107).
- Woźniak, M., Siłka, J., & Wiczorek, M. (2021). Deep neural network correlation learning mechanism for CT brain tumor detection. *Neural Computing and Applications*, 1–16, Publisher: Springer.
- Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, *9*, 611–629.