An Enhanced Deep Learning Model for MRI Image Classifications

Hanaa Torkey, Amira M.Abdelnabi[*], and Nirmeen A. El-Bahnasawy

Abstract— The correct classification of the type of brain tumor is critical in the early detection of the tumor, which can mean the difference between life and death. Implementing automated computer-aided approaches can help improve tumor diagnosis. We proposed a method for brain tumor classification via EfficientNetB3, a pre-trained model based on the transfer learning strategy. First, preprocessing images utilizing various methods, followed by classification of the preprocessed images using the fine-tuned EfficientNetB3 model. The suggested technique of fine-tuning pre-trained EfficientNetB3 is executed by first loading ImageNet weights to the EfficientNeB3 model, then adding several layers for the classification of brain tumor classes. A global average pooling (GAP) layer is used in our design to avoid over-fitting and Batch normalization layer to reduce losses. The proposed model was evaluated on 5712 images divided into four classes: glioma, meningioma, pituitary tumors, and normal which are shared publicly on Kaggle website. In addition, Multiple tests were run to assess the reliability of the proposed fine-tuned model in comparison to other traditional pre-trained models as well as other studies in the literature. The proposed framework achieved an accuracy of 97.7% with a minimum loss of 0.17. Also, the proposed method scored 95.6% for precision and F1-score respectively with only 20 epochs with Exponential Linear Unit (ELU) activation function at a threshold of 0.2 and Adam optimizer. We also evaluated the proposed model on two additional datasets to enhance generalizability. This model will certainly minimize detection complications and aid radiologists without requiring invasive procedures.

Keywords—*Magnetic Resonance Imaging (MRI), Convolutional Neural Network (CNN), Transfer Learning (TL), Artificial Intelligence (AI), Deep Learning (DL)*

Hanaa Torkey is with Computer Science & Engineering Department, Faculty of Electronic Engineering, Menoufia University, PO Box 32952, Menouf, Egypt (e-mail: htorkey@el-eng.menofia.edu.eg).

Amira M.Abdelnabi, is with Communication and Electronics dept. Delta Higher Institute of Engineering and Technology Mansoura, Egypt (e-mail: amira.awad@el-eng.menofia.edu.eg).

Nirmeen A. El-Bahnasawy is with Computer Science & Engineering Department , Faculty of Electronic Engineering, Menoufia University, PO Box 32952, Menouf, Egypt (e-mail: nirmeena.el-bahnasawy@el-eng.menofia.edu.eg).

I. INTRODUCTION

Brain tumors are the most common and deadly problem today. Every day, several individuals die as a consequence of a tumor late discovering and these lives could have been preserved if the tumor had been diagnosed previously. Brain tumors are frequently categorized as benign or low grade (grade I and II) or malignant (grade III and IV). Benign tumors are non-progressive (non-cancerous) and therefore considered less aggressive; they originate in the brain and grow gradually; and they are unable to spread to other parts of the body. The American Cancer Society (ASC) forecasts that 24, 810 humans will be diagnosed with malignant tumors by 2023, with 18, 990 dying as a consequence [1]. The most common procedure for classifying brain tumors is biopsy, but biopsy is carried out only after a surgery in which a small tissue sample is removed from the brain and examined under a microscope to determine whether or not it is a tumor.

Although a surgical nature with the biopsy method, as well as the risk of bleeding or even functional loss, constitute hazardous method for a diagnosis. Because of the variability and size of the lesion, diagnosing this disease is difficult [2]. Health care providers have begun to use medical imaging techniques more frequently in order to save time and give better results. MRI (Magnetic Resonance Imaging) is preferred by physicians over the other two techniques, and researchers are heavily focused on it. Anomalies in brain tissue can be discovered using MRIs, and detailed information about the brain's structure can be obtained [3]. Because previous methods were trained on a single data set, the system could not identify brain tumors in a variety of environments using MRI scans [4]. (Computed Tomography) CT and MRI scans are more effective imaging technologies for brain tumors. Manual brain tumor definition is time-consuming for many MRI scans. Despite this, MRI scans are of greater value than CT images in terms of tumor texture and shape [5]. Due to a malignant condition generally grows to neighboring tissues, detecting and treating it early increases the chances of effective treatment and thus staying alive. As a result, accurate diagnosis and classification of brain tumors are critical for effective treatment [6]. Quick identification of tumors is critical, and experts use a variety of techniques to assess disease attributes. Glioma and meningioma are both major tumors which may be fatal unless diagnosed early [7]. Gliomas are the most common type of brain tumors that begin in the brain's glial cells. Gliomas account for 30% of all brain tumors and 80% of all malignant brain tumors [8]. Meningioma is a benign tumor that grows on the membrane that covers the brain and spinal cord inside the human skull

[9]. Even so, pituitary tumors begin in the pituitary glands, which control hormones and regulate body functions. Pituitary tumors are primarily caused by adenohypophysis cells, which are members of the neuroendocrine epithelial cell family that secrete hormones [10]. Pituitary tumor issues can result in ongoing hormone absence and sight loss [11].

In light of the details presented above, prompt identification and categorization of brain tumors have become critical tasks in case evaluation, aiding in the selection of the best treatment method to save the lives of patients. In addition, in some complicated cases, the classification stage may be a confusing and time-consuming task for physicians or radiologists. These cases require experts to work on them in order to localize the tumor, compare tumor tissues with adjacent regions, apply filters to the image if important to make it clearer for human vision, and finally determine whether it is a tumor, as well as its type and grade if on hand. This task takes time, which is why a Computer Aided Diagnosis (CAD) system is needed to recognize brain tumors in much less time without any human involvement.

Many artificial intelligence strategies have been used in recent years to automate the process. Deep learning (DL) is a subdomain of machine learning (ML) which classifies images using the convolutional neural network (CNN) scheme [12]. CNN is a different well-known deep learning technique that performs well on both 2D and 3D medical images [13]. One of the primary advantages of CNNs over common ML and vanilla neural networks are feature learning and limitless accuracy, that can be achieved through increasing the number of samples used for training, which produces a more robust and accurate model [14]. In a similar way the transfer learning (TL) technique is commonly used to save time when data and computational resources are scarce. It means that the model uses the pre-trained model's convent weights when training only the final dense layer [15]. This method applies knowledge gained from one task to solve related ones.

To assist healthcare professionals in determining the effective course of action for treating and preventing premature deaths that result from brain tumors, it's necessary to establish a powerful DL model that can reliably forecast brain tumors in a short amount of time. As a result, this study emphasizes on developing a useful as well as structured framework for classifying brain tumors into four harmful forms of brain tumor: normal, glioma, meningioma, and pituitary tumor. The model's implementation extracts number of features from MRI images using the EfficientNetB3 pretrained model. We employ a variety of preprocessing procedures for preparing our dataset, adjust the TL architecture, alongside add additional layers. In this study, adjusting is the procedure of adding layers with the modified design which allows us to develop a new DL construction for identifying brain tumors successfully with high performance and minimal loss. Additionally, we assess the impact of our proposed DL model with a variety of performance metrics such as Accuracy, Precision, Recall, F1-score, and others. A comprehensive assessment of the suggested method is then performed. Results from this study show that our approach can classify brain tumors by achieving an outstanding accuracy rate of 97.7%.

The main contributions to this work are outlined as follows:
- Develop a framework for automated brain tumor classification which is based on finetuning EfficientNetB3 model.
- The proposed framework is based on combining EfficientNetB3 model with GAP and BN layers and ELU activation function for reducing loss and providing overfitting.
- The proposed framework achieved high performance with minimum loss despite the few training dataset.
- The suggested fine-tuned EfficientNetB3 is computationally inexpensive, lightweight, and performs well on previously unseen samples.

The other parts of this paper are arranged as follows. Section II provides related work also; section III gives complete description for material and method of the proposed work. Section IV displays performance metrics. Section V introduces results and discussion in addition, section VI presents conclusion and future work.

II. RELATED WORK

Several techniques to recognize brain tumors in MRI images were invented over the years. These approaches range from traditional image processing to neural network-based machine learning. Qinghua and Kabir [16] suggested a block-wise fine-tuning method based on the TL and VGG19 pre-trained model. The suggested approach was tested on Figshare's brain tumor dataset. The approach they employed was adaptable as it was free of a feature extractor, needed minimal preprocessing, and obtained an average accuracy of 94.82% after five-fold cross validation tests. Afshar et al. [17] established the BayesCap, a Bayesian CapsNet structure that is capable of giving not only mean estimates but also entropy as an indicator of anticipating uncertainties, through leveraging capsule networks' ability to regulate small datasets and control uncertainty. They used Figshare brain tumor dataset to test their model and gained an accuracy rate of 78% for recognizing brain tumors. Five popular DL architectures were used by Sohaib and Zhao [18] to create a system for diagnosing brain tumors. They applied Xception, DenseNet201, DenseNet121, ResNet152V2, and InceptionResNetV2 models for multi-classification tasks. The findings indicated that the proposed model based on the Xception architecture was the best performance and scored an accuracy of 95.8% for the four classes classification.

To diagnose brain abnormalities, Lu et al. proposed a fine-tuned Alex-Net framework [19]. Only 291 images were used in this study. Sajjad et al. used a fine-tuned VGG19 model for multiclass brain tumor detection on 121 images in a similar study [20]. They had a prediction accuracy of 87.4% overall. Chato and Latifi [21] developed a model for cancer diagnosis that achieves 90% accuracy through a combination of trial and error and adjustments made manually. The answers, however, did not achieve their optimal outcome and accuracy using conventional methods. Alanazi et al. [22] proposed a CNN for categorizing brain Magnetic Resonance Imaging (MRI) scans into two groups: tumor and non-tumor. This study used a dataset of 3000 images from the Kaggle platform, which included glioma, meningioma, pituitary

tumor, and no tumor classes. The proposed CNN model performed admirably, with an accuracy of 96.89% in determining and classifying these images. Irmak et al. [23] proposed two distinct CNN models with 13 and 25 layers to categorize brain images into two and five classes, as well. The proposed model's accuracy decreased to 92.66% as the number of classes increased. The approach used two different models for brain tumor detection and identification was also a flaw.

We collected a number of brain tumor MRI images for the purpose of research. We also considered various machine learning models to see how well our proposed CNN model worked. In our study, CNN performed better than the TL models. Other authors, on the other hand, have demonstrated good results when working on TL models, with results better than 90%. Unlike most other researchers, we used a large dataset of 5712 MRI images, which is significantly larger than what is typically used in many studies. Despite the limitations of other research studies, we worked to improve our method, shorten the training period, and improve performance.

III. MATERIAL AND METHOD

The computerized technique for recognizing and evaluating for brain tumors is being tested through collecting MRI images from Kaggle website[1,2]. The used dataset contained 5712 MRI scans for training and 1311 MRI scans for testing. This dataset also includes four types of brain tumors: normal, pituitary, glioma, and meningioma as shown in Table 1. The images in this dataset are collected from Figshare, SARTAJ, and Br35H datasets. This dataset's scans have three different viewpoints: axial, coronal, and sagittal. Fig.1 describes a sample of each class in the used dataset.

Table 1: Dataset distribution

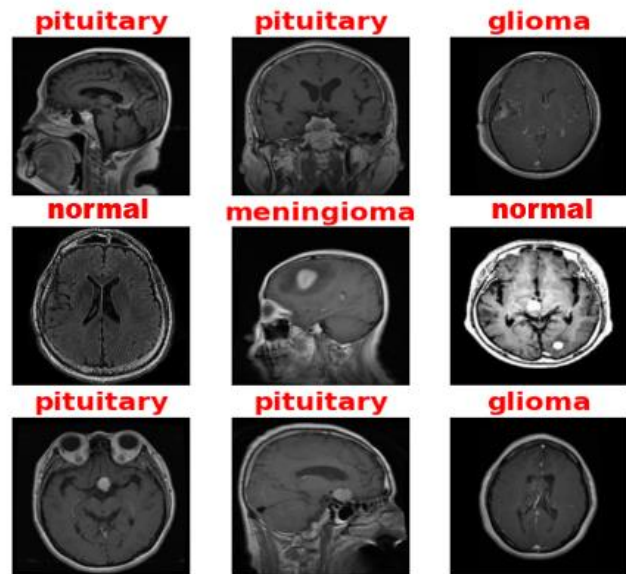|	Class	No. of images in each class
Train	Pituitary	1457
	Glioma	1321
	Meningioma	1339
	Normal	1595
	Total	**5712**
Test	Pituitary	300
	Glioma	300
	Meningioma	306
	Normal	405
	Total	**1311**

Fig. 1: Samples of the used dataset.

A. The proposed framework

This section explores the methodology used to classify brain tumor types from MRI images. The proposed method's overall workflow is depicted as a block diagram in Fig. 2. First, MRI images containing slices of various brain tumor types are preprocessed in order to make samples for training. The proposed approach is based on pre-trained EfficientNets and the variations of them that are transferred to the learning setting. Here, EfficientNetB3 is explicitly adjusted on the dataset's MRI sequences for extraction of features and categorization. This model was chosen because it is computationally inexpensive, requires fewer Floating-Point Operations per second (FLOPs) for inference, and outperforms different state-of-the-art pre-trained deep CNN architecture on ImageNet [24]. The FLOPs demonstrate the model's complexity by counting the number of transactions in a frozen CNN network. The

following sections provide an extensive description of every stage within the suggested approach.

B. Data Preprocessing

We apply a variety of processing methods before feeding the images into the classifier. Data preprocessing, including contrast enhancement and normalization, is required for medical image analysis. As a result, in order to read the image, we must increase its size. Following that, we convert all of the images into NumPy arrays (available in Python) so that our model takes up a smaller area. The size of the input images is first rescaled to 240 x 240 x 3 to have the pre-trained EfficientNetB3 model correspond with the dimension tensor being used with that of the needed input shape. Image resizing additionally assists avoid computational overload while training the network by preserving the image's content and features.
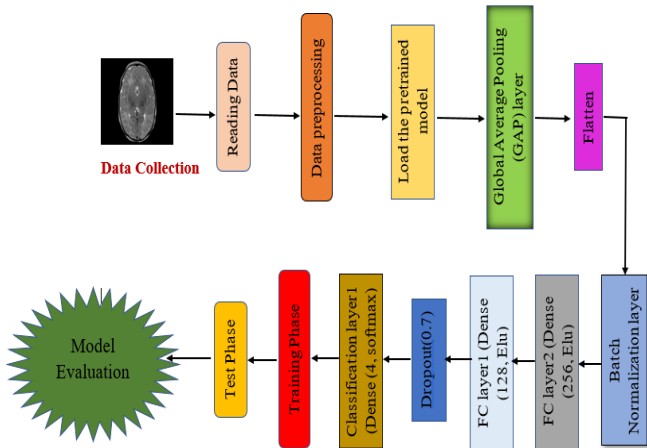

Fig. 2. The proposed model structure.

Fig. 3. Transfer learning idea.

The dataset was initially normalized to intensity values before being mapped to the 256 levels of grayscale using the following Equation:

$$I_{norm} = \frac{I - I_{min}}{I_{max} - I_{min}} \tag{1}$$

where I_{norm} describes the normalized data, while variables I_{max} and I_{min} are the maximum and minimum pixel intensity in the raw data, accordingly [18]. After the normalization step, we shuffled the data before splitting it so that our model could train on unorganized data. The samples of data are shuffled in order to stop the network from learning the dataset's narrow band and to allow the system to train on unsorted images. The next step is dividing the dataset into three phases: training, testing, and validation. About half of the data is used for training, 35% for validation, and 15% for testing. After this step, the results of class labels in the train, test, and validation sets have been assigned to 0,1,2, and 3 for glioma, meningioma, normal, and pituitary tumor, as well.

C. Transfer learning concept

Contrary to traditional ML techniques, CNN allows for the automatic acquisition of both low-level and high-level feature maps from the model's convolutional base. The generated feature maps have been transformed into a one-dimensional feature vector, which is initially classified using an arrangement of single or multiple fully connected layers. Regardless of its widespread popularity, one of CNN's limitations is that it requires a large number of data samples for effective model training and preventing overfitting. TL employs the understanding of structures that were previously developed on a larger benchmark dataset, such as ImageNet, to problems with fewer data points, such as the classification of brain tumors from MRI. Fig. 3 illustrates the general concept of TL. Fine-tuning is a method of upgrading the weights taken from some top layers of a deep CNN structure, which were initially developed on a massive dataset to address additional certain issues. Pre-trained mechanisms may be fine-tuned by unfreezing all or some layers in the convolutional base [25], [26], or by employing pre-trained frameworks as fixed feature extractors and then feeding them to other algorithms for classification, like SVM [27].
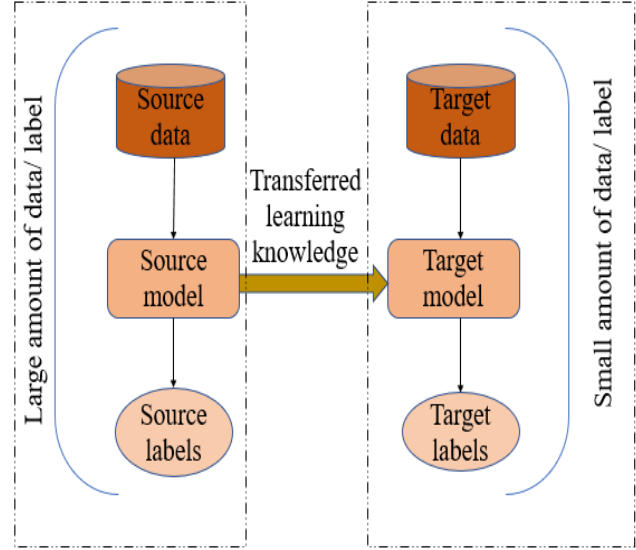
TL is performed on the pre-trained EfficientNets, namely EfficientNetB3, that was originally trained on the ImageNet benchmark dataset. Fig. 4 depicts the network architecture of the modified EfficientNetB3. The pre-trained EfficientNetB3 is fine-tuned by first creating a base model utilizing the pre-trained ImageNet weights as its backbone. For lower dimensionality, the Global Average Pooling (GAP) layer is added on top of the EfficientNets backbone, while the convolutional base of all blocks is retained. The GAP layer also helps to simplify the system in terms of several parameters without compromising the model's accuracy. GAP layer reduces the spatial dimensions of the feature map, resulting in rapid training and less utilization of memory. Following the GAP layer, a flatten and batch normalization layers were added. The benefit of the batch normalization layer is to help the model converge faster for the learning rate. In addition, batch normalization improves training stability and lowers the risk of vanishing or exploding gradients. After this layer, two fully connected layers were added with an Exponential Linear Unit (ELU) activation function with a threshold of 0.2. This function is described by Eq (2).

$$ELU(I_r) = \begin{cases} I_r, & I_r \geq 0 \\ \sigma(e^{I_r} - 1), & I_r < 0 \end{cases} \tag{2}$$

where I_r describes the input data and σ is the threshold of the activation function. Fully connected layers were included as well to capture complicated structures, patterns, and nonlinear connections in data. After that, a dropout layer with a probability of 0.7 is added to the network. Dropout is a regularizer that prevents the model from becoming overfit. At last, for the classification of four types, the primary output layer of one thousand units was substituted by an output layer of 4 units and a softmax layer which is defined as:

$$SM = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \tag{3}$$

where i= 1, ..., k and z= $(z_1, ..., z_k) \in R^k$

For convenience, the exponential function is given to each element z_i of the input vector z, and the results are normalized by dividing by the sum of all exponentials. Normalization ensures that each of the components of the output vector SM sum to 1.

D. System requirements

Simulations are performed on Google Colab, a free GPU resource provided by Google to individuals for investigation and academic use. Google Colab tool also serves as a tool to execute the codes due to the speed of its processor and inbuilt libraries. The algorithms used are trained on a K-4 GPU with 12GB of RAM. The software is written in Python, with Tensorflow and Keras API as backend and frontend, respectively.

E. Hyper-parameter setting

By applying a try and error method, various hyper-parameters including batch size, optimizers, learning rate, epochs, and loss function have been directly adjusted until the most appropriate sets of hyper-parameters for model training and achieving the desired outcomes are obtained. Batch size is defined as the number of training instances used in one iteration. On the other hand, learning rate defines the step size at each iteration while moving towards the lowest value of a loss function. Because sparse categorical cross-entropy is used as a loss function for brain tumor classification, which is a multi-class classification task. This loss function is used in multi-class classification algorithms with multiple output labels. The function computes the variation between the expected probability distribution and the actual label according to the following equation:

$$L (y, f(x)) = -\log (f(x) - y) \qquad (4)$$

where y represents the correct label and f(x) is the expected probability distribution across the classes. Adam [29] is the optimizer used for the model, with an initial learning rate of 0.0001. The images from the train set are all loaded in 64-bit mini-batch mode. EfficientNetB3 that has been fine-tuned is trained for 20 epochs. Each variation of the EfficientNetB3, is trained and evaluated using the same experimental and hyper-parameter conditions. Table 2 provides the optimal values for all hyper-parameters utilized during the simulation.
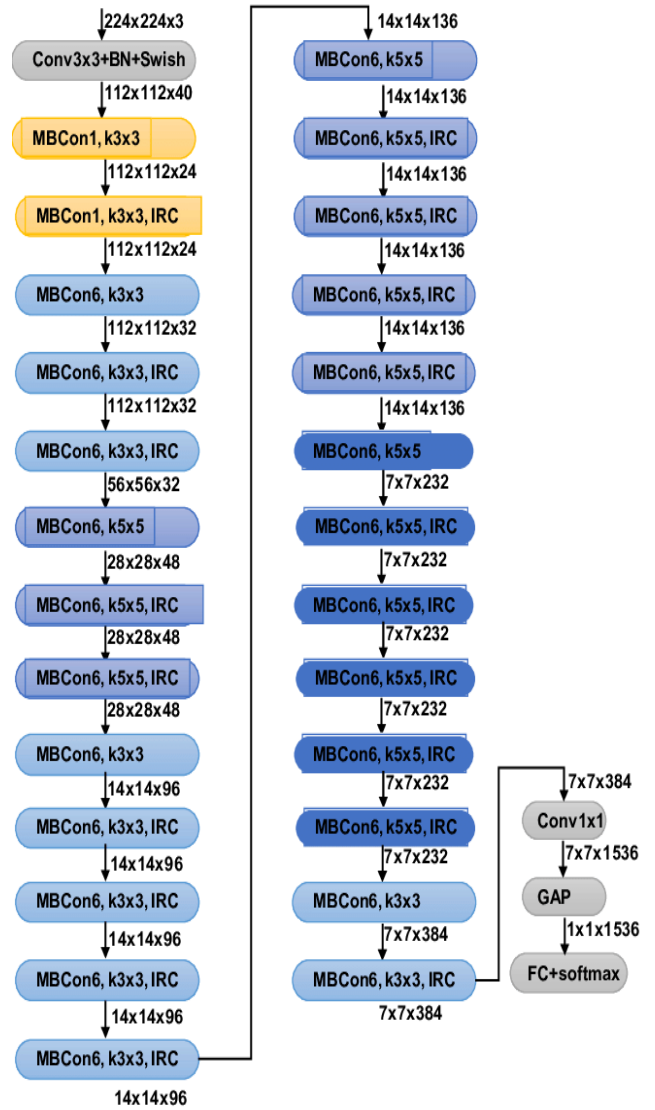

Fig. 4. The network architecture for the modified EfficientNetB3.


```
In [16]:
NB_CLASSES = 4

model = Sequential()
model.add(baseModel)
#resnet50.add(AvgPool2D())

model.add(GlobalAveragePooling2D())
model.add(layers.Flatten())
model.add(BatchNormalization())
model.add(layers.Dense(256, activation= tf.keras.layers.ELU(alpha=0.2)))
model.add(layers.Dense(128, activation= tf.keras.layers.ELU(alpha=0.2)))
#model.add(BatchNormalization())
model.add(layers.Dropout(0.2))
#model.add(LeakyReLU(alpha=0.2))

model.add(layers.Dropout(0.5))
model.add(layers.Dense(NB_CLASSES, activation='softmax'))
```

Fig. 5. Snap shot of the classification code.

Table 2 Hyper-parameter setting

Hyper parameter	Value
Input shape	240 x 240 x 3
Drop connect rate	0.7
Output layer activation function	Softmax
Dense layers activation function	ELU
Threshold of activation function	0.2
Batch Normalization parameters	$ɣ = 0.1, β = 0.001$
Epochs	20
Batch size	64
Optimizer	Adam
Learning rate	0.0001
learning rate decay factor	0.5
patience	2
Validation split	0.35
Loss function	Sparse categorical cross-entropy

Table 3 Model structure

Layer	Output Parameters
EfficientNetB3	10,783,535
Global_Average_Pooling	0
Flatten	0
Batch Normalization	6144
FC1(Dense layer (256) + ELU)	393472
FC2(Dense layer (128) + ELU)	32896
Dropout (0.7)	0
Classification layer (Dense layer (4) +softmax)	516
Total parameters	11,216,563

IV. PERFORMANCE METRICS

We evaluated and analyzed the performance of the DL models using metrics like accuracy, precision, recall, F1-score, and roc curve. The accuracy can be determined by dividing the number of correct predictions by the total number of instances. It can be calculated as:

$$Acc = \frac{t_n + t_p}{t_n + t_p + f_p + f_n} X 100\% \qquad (5)$$

where:
t_n= true negative
t_p= true positive
f_p= false positive
f_n= false negative

Another important metric for evaluating the models is recall. It can be calculated as:

$$Rec = \frac{t_p}{t_p + f_n} X 100\% \qquad (6)$$

In contrast, precision indicates how many of the predicted positive values are actually positive:

$$Pre = \frac{t_p}{t_p + f_p} X 100\% \qquad (7)$$

F1-score represents a harmonic average of the Precision and Recall outcomes. Using a harmonic mean instead of a simple mean prevents extreme instances from being ignored. It is calculated from the following equation:

$$F1 - score = \frac{2t_p}{2t_p + f_p + f_n} X 100\% \qquad (8)$$

The receiver operating characteristic (ROC) curve is another way to assess model performance. It graphically depicts the relation between true-positive rate (or recall) and false-positive rate (equivalent to 1 minus specificity) [30].

V. RESULTS AND DISCUSSION

This section explores the outcomes gained after training and fine tuning the suggested model and other model variants for multi-classification of brain tumor classes into normal, glioma, meningioma, and pituitary. Two distinct ablation scenarios were investigated: pretrained models, the proposed model with different optimizers. Table 4 displays results of the pretrained models which were applied on the used dataset. It is observed that Resnet101 scored high accuracy of 93.3% and loss of 0.19 while Inception V3 scored low accuracy of 81% and high loss of 1.04 on the used dataset. In addition, the EfficientNetB3 model attained an accuracy rate of 93.3% with a loss of 0.18 without adding layers. Table 5 demonstrates the results of the suggested fine-tuned EfficientNetB3 after training and validation with different optimizers. It shows that the proposed approach outperforms other variants with an accuracy of 97.7%, precision of 96.8%, specificity of 97.3%, and an F1-score of 96.6% with Adam optimizer.

Table 4 Results with pretrained models

Model	Acc	Pre	Rec	F1-score	loss	Execution time(s)
Vgg16	81.9%	82.8%	82.9%	82.86%	1.09	517
Resnet101	93.3%	94.7%	94.7%	95.2%	0.19	629
Inception V3	81%	82%	82.3%	82%	1.04	253
Xception	88%	87%	87%	87.4%	0.41	501
Efficient B3	93.3%	93.4%	93.4%	94%	0.18	418
DenseNet 121	86%	86.1%	86.3%	86.2%	0.4	358

These results demonstrated that the suggested approach for fine-tuning the pre-trained EfficientNetB3 performed significantly better across all metrics used for evaluation. Pituitary is the most effectively categorized class, with a test accuracy of 98%, while meningioma is the least correctly classified class, with an accuracy of 85.7% when compared to other classes. This is due to the fact that EfficientNetB3, the most basic variation of the EfficientNets family, has only 237 layers. Regardless of improved training and validation accuracy, the model was incapable of generalization well on previously hidden samples.

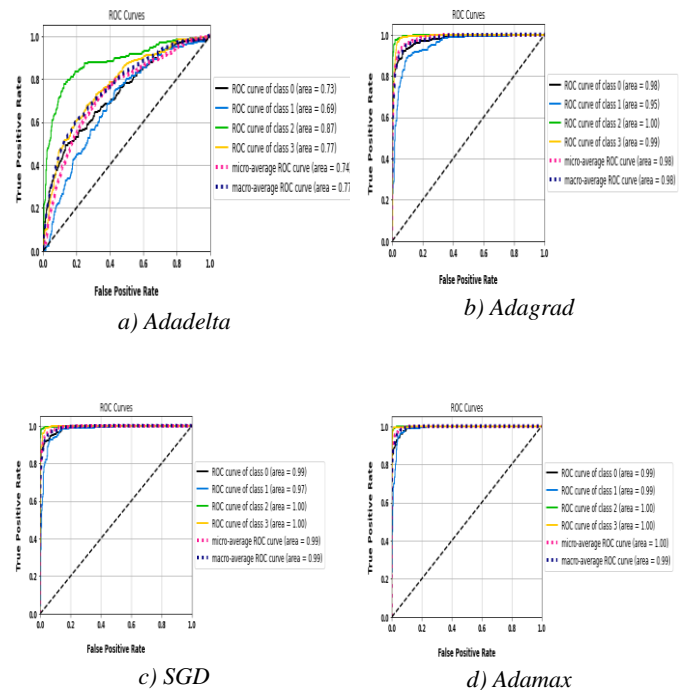Table 5 Model performance of the proposed model with different optimizers

Optimizer	Acc	Pre	Rec	F1-score	loss	Execution time(s)
Adam	97.7%	96.8%	95.9%	96.9%	0.17	390
RMSProp	94.5%	94%	94.2%	94.4%	0.2	392
SGD	92%	93.11%	93.11%	93%	0.22	387
Adagrad	88.5%	89.9%	89.9%	90%	0.3	397
Adadelta	53%	49.2%	49.2%	49.45%	1.16	401
Adamax	93%	93.2%	92.8%	93.4%	0.17	391

Fig. 6 indicates the ROC curves for the proposed model with different optimizers. Also, it depicts the total performance of the proposed modified EfficientNetB3 model with different optimizers when predicted and tested on the test data as shown in Fig. 6. Meningioma appears to be the only class with a high FNR, resulting in lower accuracy with the optimization algorithm Adam when compared to the other classes. In contrast, pituitary tumor has resulted in the highest levels of precision, recall/sensitivity, F1-score, and specificity, at 96.8%, 95.9%, 96.6%, and 97.3%, respectively, with minimum loss of 0.17. From Table 6, it observed that the enhancement of the model performance due to the added layers. As shown, the proposed model scored higher accuracy of 97.7% with the added layer than EfficientNetB3 pre-trained model without the added layers which achieved an accuracy of 93.3%. On the other hand, the second experiment was based on applying the pretrained models as it is on the used dataset. Fig. 7 indicates the ROC curves for different pretrained models. This research also examines the model's robustness with consideration of speed (training and inference), overall parameters, and model size. Table 6 compares the outcomes of the proposed modified EfficientNetB3 with state-of-the-art models in terms of accuracy. The ablated models' results show the performance of the proposed model without the GAP, BN layers, and ELU activation function respectively. From the table, it observed that the integration between the EfficientNetB3 pre trained model and the added layers with ELU activation function improves the performance of the proposed model. Furthermore, to illustrate the benefits of the proposed model, we compared the accuracy of the proposed system to its competitive current literature models that partially employed different public sources of the datasets including (Figshare, SARTAJ, and Br35H). In our study, we made a combination of the three sources of the datasets (Figshare, SARTAJ, and Br35H). The goal of combining the dataset from three sources is to get near balanced classes to prevent bias in model training and prediction. Further, some comparable models used one source from the mentioned three sources of the datasets only such as Figshare, SARTAJ, or Br35H only compared to our study. We additionally compared our method to similar approaches that used consolidation of diverse datasets to ensure that its performance and

generalizability were fairly evaluated. Some studies in the literature applied small sample of the dataset. This comparing technique enables a fair comparison while still providing useful insights into how our system works in the context of the datasets and methods utilized by others. The main criterion for comparing classification results is accuracy.

Table 6 Comparison of the classification accuracy with state-of-the-art models

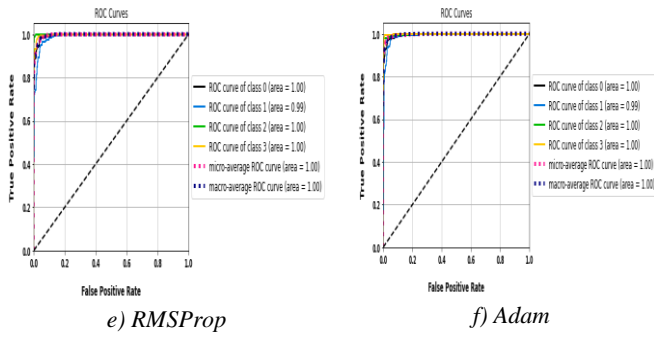| Study	No. of images	Method	Accuracy
Swati et al. [3]	3064	TL+ Vgg 19	94.82%
Ashfar et al. [17]	3064	BayesCap	78%
Sohaib and Zhao [18]	2045	TL	95.8%
Lu et al. [19]	291	Alexnet	89%
Sajjad et al. [20]	121	Vgg 19	87.4%
Chato and Latifi [21]	163	DL	90%
Alanazi et al. [22]	3000	CNN	96.89%
Irmak et al. [23]	3950	CNN	92.66%
Proposed without GAP, BN, and ELU.	5721	EfficientNet B3	93.3%
Proposed without (BN+ELU)	5721	EfficientNetB3 +GAP	94.1%
Proposed without ELU	5721	EfficientNet B3 +GAP + BN	95.3%
The proposed model on the first dataset	5721	TL based on modified Efficient B3 model	97.7%
The proposed model on the second dataset	1311	TL based on modified Efficient B3 model	97.2%

a) Adadelta

b) Adagrad

c) SGD

d) Adamax

e) RMSProp *f) Adam*

Fig. 6. ROC curves for the proposed model with different optimizers.

a) Inception V3 *b) Vgg 16*

c) DenseNet 121 *d) Xception*

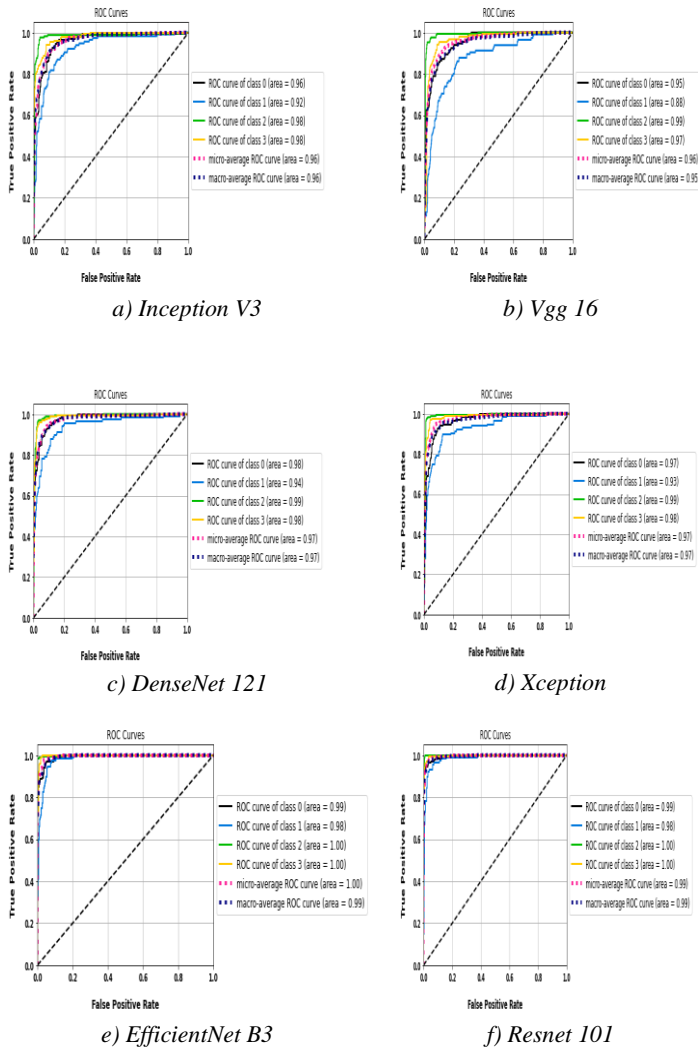*e) EfficientNet B3* *f) Resnet 101*

Fig. 7. ROC curves for the different pretrained models.

VI. CONCLUSION AND FUTURE WORK

This study uses TL with pre-trained model versions for performing multi-class identification of brain tumors through MRI images of three tumor types: normal, glioma,

meningioma, and pituitary tumor classes. EfficientNetB3 construction has been altered by initially importing the pre-trained ImageNet weights into the initial approach. The design's convolutional base is then supplemented with few top layers, including Global Average Pooling (GAP), batch normalization, Flatten, dropout, and a fully connected layer. The EfficientNetB3 pre-trained model convolutional base served as the feature extractor, along with the upper layers were incorporated into the classification algorithm for the brain tumor categories classification task. In order to develop the model on the brain's tumors dataset, various hyper parameter values for the improved EfficientNetB3 were precisely fine-tuned. A variety of experiments were performed to compare the reliability for the suggested fine-tuned EfficientNetB3 with other pre-trained models. The proposed model was tested on two datasets to ensure greater model generalization. The suggested technique of fine-tuning EfficientNeB3 pre-trained model as its core outperforms multiple cutting-edge methods for the same problems with classification, scoring 97.7% as a whole test accuracy with Adam optimizer. In addition, it achieved an accuracy of 97.2% when tested on the second dataset. In the future, a transformer-based design for brain tumor type classification could be proposed as an alternative for deep CNN-based techniques, allowing for the extraction of more information-rich feature maps while also reducing network complexity. Also, a feature fusion method can be applied to achieve higher accuracy.

REFERENCES

1. Mandloi, S., et al., *An explainable brain tumor detection and classification model using deep learning and layer-wise relevance propagation. Springer.* 2023: p. 1-31.
2. Waskita, A., et al. *EfficientNetV2 based for MRI brain tumor image classification.* in *2023 International Conference on Computer, Control, Informatics and its Applications (IC3INA).* 2023. IEEE. P. 171-176.
3. Swati, Z.N.K., et al., *Brain tumor classification for MR images using transfer learning and fine-tuning.* Elsevier. 2019. **75**: p. 34-46.
4. Anagun, Y.J.M.T. and Applications, *Smart brain tumor diagnosis system utilizing deep convolutional neural networks.* Springer. 2023. **82**(28): p. 44527-44533.
5. Sharif, M.I., et al., *A decision support system for multimodal brain tumor classification using deep learning.* Springer. 2021: p. 1-14.
6. Khan, M.A., et al., *Multimodal brain tumor classification using deep learning and robust feature selection: A machine learning application for radiologists.* MDPI. 2020. **10**(8): p. 565.
7. Celik, G., M.F.J.B.S.P. Talu, and Control, *A new 3D MRI segmentation method based on Generative Adversarial Network and Atrous Convolution.* 2022. **71**: p. 103155.
8. Goodenberger, M.L. and R.B.J.C.g. Jenkins, *Genetics of adult glioma.* 2012. **205**(12): p. 613-621.
9. Louis, D.N., et al., *The 2016 World Health Organization classification of tumors of the central nervous system: a summary.* Springer. 2016. **131**: p. 803-820.
10. Emam, M.M., et al., *Optimized deep learning architecture for brain tumor classification using*

improved Hunger Games Search Algorithm. Elsevier. 2023. **160**: p. 106966.

11. Rajesh, T. and R.S.M. Malar. *Rough set theory and feed forward neural network based brain tumor detection in magnetic resonance images*. in *International Conference on Advanced Nanomaterials & Emerging Engineering Technologies*. 2013. IEEE. P. 240-244.

12. Bakas, S., et al., *Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge.* 2018.

13. Milletari, F., N. Navab, and S.-A. Ahmadi. *V-net: Fully convolutional neural networks for volumetric medical image segmentation*. in *2016 fourth international conference on 3D vision (3DV)*. 2016. IEEE. p. 565-571.

14. Sultan, H.H., N.M. Salem, and W.J.I.a. Al-Atabany, *Multi-classification of brain tumor images using deep neural network.* 2019. **7**: p. 69215-69225.

15. Talo, M., et al., *Application of deep transfer learning for automated brain abnormality classification using MR images. Elsevier.* 2019. **54**: p. 176-188.

16. Khan, M.A., et al., *Gastrointestinal diseases segmentation and classification based on duo-deep architectures. Elsevier.* 2020. **131**: p. 193-204.

17. Afshar, P., A. Mohammadi, and K.N.J.I.S.P.L. Plataniotis, *Bayescap: A bayesian approach to brain tumor classification using capsule networks. IEEE.* 2020. **27**: p. 2024-2028.

18. Asif, S., et al., *An enhanced deep learning method for multi-class brain tumor classification using deep transfer learning.* Springer. 2023. **82**: p. 31709-31736.

19. Lu, S., Z. Lu, and Y.-D.J.J.o.c.s. Zhang, *Pathological brain detection based on AlexNet and transfer learning.* Elsevier. 2019. **30**: p. 41-47.

20. Sajjad, M., et al., *Multi-grade brain tumor classification using deep CNN with extensive data augmentation.* Elsevier. 2019. **30**: p. 174-182.

21. Chato, L. and S. Latifi. *Machine learning and deep learning techniques to predict overall survival of brain tumor patients using MRI images*. in *2017 IEEE 17th international conference on bioinformatics and bioengineering (BIBE)*. 2017. IEEE. p. 9-14.

22. Alanazi, M.F., et al., *Brain tumor/mass classification framework using magnetic-resonance-imaging-based isolated and developed transfer deep-learning model.* MDPI. 2022. **22**(1): p. 372.

23. Irmak, E.J.I.J.o.S. and T.o.E.E. Technology, *Multi-classification of brain tumor MRI images using deep convolutional neural network with fully optimized framework.* Springer. 2021. **45**(3): p. 1015-1036.

24. Deng, J., et al. *Imagenet: A large-scale hierarchical image database*. in *2009 IEEE conference on computer vision and pattern recognition*. 2009. IEEE.

25. Girshick, R., et al. *Rich feature hierarchies for accurate object detection and semantic segmentation*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014. p. 580-587.

26. Long, M., et al. *Learning transferable features with deep adaptation networks*. in *International conference on machine learning*. 2015. PMLR. P. 97-105.

27. Sharif Razavian, A., et al. *CNN features off-the-shelf: an astounding baseline for recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2014.

28. Alhichri, H., et al., *Classification of remote sensing images using EfficientNet-B3 CNN model with attention.* IEEE. 2021. **9**: p. 14078-14094.

29. Kingma, D.P. and J.J.a.p.a. Ba, *Adam: A method for stochastic optimization.* 2014.

30. Erickson, B.J. and F.J.R.A.I. Kitamura, *Magician's corner: 9. Performance metrics for machine learning models.* 2021, Radiological Society of North America. **3**(3) p. e200126.