Computers, Materials & Continua
DOI: 10.32604/cmc.2022.030698
Article

Tech Science Press

Cartesian Product Based Transfer Learning Implementation for Brain Tumor Classification

Irfan Ahmed Usmani[1,*], Muhammad Tahir Qadri[1], Razia Zia[1], Asif Aziz[2] and Farheen Saeed[3]

[1]Electronic Engineering Department, Sir Syed University of Engineering & Technology, Karachi, 75300, Pakistan
[2]Department of Computer Science, Bahria University Karachi Campus, Karachi, 75260, Pakistan
[3]Christus Trinity Clinic, Texas, USA
*Corresponding Author: Irfan Ahmed Usmani. Email: iausmani@ssuet.edu.pk
Received: 31 March 2022; Accepted: 12 May 2022

Abstract: Knowledge-based transfer learning techniques have shown good performance for brain tumor classification, especially with small datasets. However, to obtain an optimized model for targeted brain tumor classification, it is challenging to select a pre-trained deep learning (DL) model, optimal values of hyperparameters, and optimization algorithm (solver). This paper first presents a brief review of recent literature related to brain tumor classification. Secondly, a robust framework for implementing the transfer learning technique is proposed. In the proposed framework, a Cartesian product matrix is generated to determine the optimal values of the two important hyperparameters: batch size and learning rate. An extensive exercise consisting of 435 simulations for 11 state-of-the-art pre-trained DL models was performed using 16 paired hyperparameters from the Cartesian product matrix to input the model with the three most popular solvers (stochastic gradient descent with momentum (SGDM), adaptive moment estimation (ADAM), and root mean squared propagation (RMSProp)). The 16 pairs were formed using individual hyperparameter values taken from literature, which generally addressed only one hyperparameter for optimization, rather than making a grid for a particular range. The proposed framework was assessed using a multi-class publicly available dataset consisting of glioma, meningioma, and pituitary tumors. Performance assessment shows that ResNet18 outperforms all other models in terms of accuracy, precision, specificity, and recall (sensitivity). The results are also compared with existing state-of-the-art research work that used the same dataset. The comparison was mainly based on performance metric "accuracy" with support of three other parameters "precision," "recall," and "specificity." The comparison shows that the transfer learning technique, implemented through our proposed framework for brain tumor classification, outperformed all existing approaches. To the best of our knowledge, the proposed framework is an efficient framework that helped reduce the computational complexity and the time to attain optimal values of two important hyperparameters and consequently the optimized model with an accuracy of 99.56%.

Keywords: Deep transfer learning; Cartesian product; hyperparameter optimization; magnetic resonance imaging (MRI); brain tumor classification

1 Introduction

One of the most widely known imperative causes for the increase in deaths among adults and children is brain tumors, which emerge as a group of anomalous cells developing inside or around the brain [1]. A precise and early brain tumor diagnosis plays a significant role in successful therapy. Among imaging modalities, MRI is the most extensively utilized non-invasive approach that succor radiologists and physicians in the discernment, diagnosis, and classification of brain tumors [2–4]. The radiologist approaches brain tumor classification in two ways: (i) categorizing the normal and anomalous magnetic resonance (MR) images and (ii) Scrutinize types and stages of the anomalous MR images [2].

Since brain tumors show a high-level of dissimilarities related to size, shape, and intensity [5], and tumors from various neurotic types might show comparatively similar appearances [6], therefore classification into different types and stages has become quite a wide research topic [7,8]. Manual classification of comparatively similar appearing brain tumor MR images is quite a challenging task, which relies upon the accessibility and capability of the radiologists. Despite the radiologist's skills, the human visual system always bounds the analysis as the knowledge contained in an MR image surpasses the perception of the human visual system. Thus, the computer was used as the second eye to understand the MR images.

Computer vision-based image analysis methods usually encompass several sub-processes: pre-processing, segmentation, feature extraction, and classification. A low-level process, preprocessing consists of operations such as image sharpening, contrast enhancement, noise reduction, etc. [8]. Segmentation and classification are in the mid-level process domain. Image segmentation is the process used to extract the anomalous region from MR images, which helps to accurately locate the tumor and determine its size. The classification results depend upon suitable feature extraction from the delineated segmented region. Feature extraction is generally dependent on the knowledge of an expert in a particular domain, which makes it quite challenging for an unskilled person to use it in traditional image processing as well as in machine learning. One can eliminate the manual feature extraction problem by using DL approaches based on the self-learning hierarchical fashion principle.

Deep learning, especially convolutional neural networks (CNNs), outperforms many machine learning (ML) approaches in different areas, such as generating text [9], natural language processing [10], speech recognition [11], face verification [12], object detection [13], image description [14], machine translation [15], and the game of Go [16]. In particular, improved performance in computer vision boosted the utilization of DL methods for brain tumor MR image analysis [17,18]. For decades, CNN's have been utilized but, they gained popularity when Krizhevsky [19] participated and won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) by developing a DL model "AlexNet" trained on ImageNet dataset [20], in 2012. Another similar but deeper visual geometry group (VGG) network, known as VGGNet, was presented by Zisserman and Simonyan [21] for classification in the 2014 ILSVRC. In fact, DL progressed with the availability of big data [20,22,23], advanced learning algorithms [24–28], and powerful graphical processing units (GPUs).

Algorithms based on deep learning for a certain classification task are difficult to effectively reuse and generalize. Therefore, a new algorithm from scratch has to be rebuilt even for a similar task that requires considerable computational power and time. At the same time, if sufficient data

are not available for similar tasks, the developed algorithm may have difficulty in attaining the desired performance or might even fail to complete the tasks. In case of a shortage of data, for example, in brain tumor classification, the concept of knowledge-based transfer learning technique may be employed by using pre-trained DL models that are already trained for other classification problems. However, selection of a pre-trained DL model, hyperparameters' optimal values, and an optimization algorithm (solver) is challenging tasks to obtain an optimized model for targeted brain tumor classification. This research aims to provide a robust framework for implementing knowledge-based transfer learning techniques for brain tumor classification with a small-sized dataset. The research contributions of this study are as follows:

- A robust framework is proposed for brain tumor classification that describes complete approach to utilize the knowledge of a pre-trained DL model and re-train it for brain tumor classification with a small dataset.
- Following the framework, the knowledge transfer technique is deployed using 11 state-of-the-art pre-trained DL models to select an appropriate model for brain tumor classification.
- The concept of Cartesian Product Matrix is introduced to find the most suitable pair of two important hyperparameters, batch size and learning rate. The cartesian product matrix is formed from initialized set of hyperparameters vector. Each pre-trained DL model was evaluated for the three most popular solvers (SGDM, ADAM, RMSProp) to obtain the most appropriate set consisting of a solver, one batch size, and one learning rate.
- To investigate the model performance, a comprehensive comparative analysis of each model for brain tumor classification was conducted.

The rest of the paper is divided into five sections. Section 2 presents a comprehensive literature review related to brain tumor classification with the focus on their approaches. Section 3 describes the proposed framework to implement transfer learning technique with the used dataset, preprocessing, augmentation, pre-trained networks, fine-tuning and performance assessment. Section 4 discusses the results and analysis of the proposed frame work. At the end, conclusion and future work have been discussed in Section 5.

2 Literature Review

To classify brain tumors, many research efforts have contributed to different subfields of detection and classification processes. Different techniques have been presented for the segmentation of anomalous regions in MR images [29–32]. MRI(s) are classified into different types and grades after segmentation. In [33–35], binary classifiers were used for tumor classification into malignant and benign classes.

Abdolmaleki et al. [33] extracted 13 different features to differentiate malignant and benign tumors using a three-level neural network. The features were extracted with the help of the radiologists. The proposed methods were applied to a dataset of 165 patients' MRIs, which helped to achieve accuracies of 94% and 91% for benign and malignant tumors, respectively. In [34], the author categorized brain tumors as malignant or benign by using a hybrid scheme consisting of a genetic algorithm (GA) and a support vector machine (SVM). Furthermore, Papageorgiou et al. [35] introduced fuzzy cognitive maps (FCM) to classify tumors into low-grade and high-grade gliomas. Papageorgiou's FCM model used 100 cases and achieved 93.22% accuracy for high-grade gliomas and 90.26% accuracy for low-grade gliomas.

In addition to the binary classification of brain tumors, Zacharaki et al. [36] proposed a technique using SVM and K-nearest neighbor (KNN) for multi-classification of brain tumors into primary gliomas, meningiomas, metastases, and glioblastomas. This research encompasses sub-processes, such as segmentation and feature extraction, to perform multi-classification. Accuracies of 88% and 85% for binary classification and multi-classification, respectively, were achieved. Hsieh et al. [37] also proposed a technique based on extracted features from a dataset of 107 MR images, consisting of 73 and 34 low-grade and high-grade glioma MRIs respectively, to measure malignancy. The proposed technique produced accuracies of 83%, 76%, and 88% using local, global, and fused features, respectively. Sachdeva et al. [38] presented a technique that depends on optimal features, based on color and texture, extracted from segmented regions, and used GA in combination with SVM and artificial neural network (ANN). The technique achieved accuracies of 91.7% and 94.9% for GA-SVM and GA-ANN, respectively.

Cheng et al. [5] presented a framework based on the following approaches to extract features: intensity histogram, gray level co-occurrence matrix (GLCM), and bag-of-words (BoW). In the domain of classification, this research is considered as the first significant work towards brain tumor multi-classification using the challenging and largest publicly available dataset in figshare [39], consists of glioma, meningioma, and pituitary brain tumor types. The approach is dependent on a wide range of features extracted from a manually defined segmented region and applied as input to different classifiers. The author used sensitivity, specificity, and classification accuracy as measurement parameters and obtained the best results using SVM for a particular set of features. In [40], Ismael and Abdel Qadir worked on the same challenging publicly available dataset, as used in [5] and proposed an algorithm for brain tumor classification. The algorithm uses the Gabor filter and discrete wavelet transform (DWT) to extract the statistical features used to train the classifier. The authors randomly selected 70% and 30% of MR images for training and validation of the classifier, respectively. Feature extraction from a segmented region of interest (ROI) and its appropriate selection is significant in establishing the best learning of the classifier. These handcrafted features must be extracted by an expert with sound knowledge and skills to determine the most significant features. Furthermore, the feature extraction process requires a significant amount of time and is susceptible to errors when dealing with big data [41].

In contrast to ML, DL algorithms do not require handcrafted features. DL requires a preprocessed dataset and applies a self-learning approach to determine the significant features [42]. Eventually, many CNNs such as AlexNet [19], ResNet [43], and VGGNet [21] were deployed after being trained on a large ImageNet dataset in ILSVRC [20] for classification. These CNNs are recognized as state-of-the-art DL models [19,21,43]. Afshar et al. [44] proposed a method for the brain tumor classification based on the CapsNet model. The method relies on adopting CapsNets, capability of CapsNets, analysis of overfitting, and output visualization pattern setup. Furthermore, Zia et al. [45] presented a brain tumor classification technique that relied on rectangular window image cropping. The technique used DWT for feature extraction, principal component analysis for dimensionality reduction, and SVM for classification. Hossam et al. [46] presented a CNN-based DL model to classify different types of brain tumors using two datasets. They used one dataset for three-class brain tumor classification, while the other was used for glioma tumor classification into grades II, III, and IV. The proposed methodology proved to be useful for the multi-classification of brain tumors. Jia et al. [47] achieved an accuracy of 98.51% for normal and anomalous brain tissue classification while evaluating MR images. The author used a support vector machine to perform fully automatic heterogeneous segmentation of brain tumors based on the extreme learning machine (ELM) deep learning technique. They proposed

a fully automatic algorithm with the support of structural, relaxometry, and morphological details to obtain optimum features.

The increase in deep CNN performance, owing to the concepts of feature extraction in a hierarchical fashion, motivated researchers to transfer the pre-trained network knowledge acquired during training with millions of images into new classification tasks with a small amount of data, to take advantage of their learned parameters, specifically, weights. Tab. 1 summarizes a review of research performed in the area of brain tumor classification based on knowledge-based transfer learning for the period 2017–2022. In [48], the author fine-tuned ResNet [49] and VGG [21], the pre-trained classification models, to distinguish between high-grade and low-grade brain tumors. They implemented the concept of transfer learning and achieved an accuracy of 97.19%. Talo et al. [50] presented a binary classifier based on the transfer learning concept for brain MR image classification into normal and anomalous brain images. The authors fine-tuned a pre-trained DL model, ResNet34, and claimed that this is the first work on brain MRI classification using the deep transfer learning approach. They used a dataset containing 613 images [51] and obtained better performance in comparison with other DL-based approaches using the same dataset. Sajjad et al. [52] used a pre-trained DL model and customized it for grading of brain tumors. The authors evaluated the proposed system on both original and segmented data. The results show a convincing performance in comparison with benchmark systems. Swati et al. [53] used VGG-19 [21], a pre-trained DL model, to transfer knowledge. The author performed a manual block-wise fine-tuning approach for the classification of a more challenging task of multi-class brain tumors. Following his work in [53], Swati proposed a method to develop similar brain tumor images based on content retrieval [54]. For similarity measurements, the author used the VGG-19 [21] features. Deepak et al. [55] utilized the knowledge of GoogleNet (ImageNet) [56] as a pre-trained DL model for brain tumor classification into different classes.

Table 1: Literature summary related to MR image processing using transfer learning techniques

Reference	Year	Study Domain	Dataset	Used Model for Classification	Description
[57]	2017	Segmentation of Brain Tumor	Radboud University Nijmegen Diffusion Tensor and Magnetic Resonance Cohort (RUN DMC)	15 layers CNN without pooling layer	They trained a CNN model on brain MRI, followed by its assessment with different domains images.
[58]	2018	Brain Tumor Detection & Classification	The Cancer Genome Atlas Glioblastoma Multiforme (TCGA-GBM) & The Cancer Genome Atlas Low Grade Glioma (TCGA-LGG)	VGG 16 and ResNet 50	The authors implemented the concept of transfer learning by utilizing two pre-trained convolutional networks to distinguish high grade glioma (HGG) and low grade glioma (LGG).
[50]	2019	Brain abnormality classification	Harvard Medical School Data	ResNet34	In this work pre-trained ResNet model is used for classification of normal and anomalous brain MRI scans.
[53]	2019	Brain tumor classification	Brain tumor public dataset (Figshare)	VGG-19	Performed block-wise fine-tuning for a more challenging classification of multi-class brain tumors.

(Continued)

Table 1: Continued

Reference	Year	Study Domain	Dataset	Used Model for Classification	Description
[54]	2019	Content-based Image retrieval	Brain tumor public dataset (Figshare)	VGG-19	VGG-19 features were used to retrieve similar brain tumor images.
[55]	2019	Brain tumor classification	Brain tumor public dataset (Figshare)	GoogleNet	A pre-trained model GoogleNet was used for brain tumor classification into three different classes.
[59]	2020	Grading based Brain tumor classification	Multiple Brain tumor dataset	AlexNet	Brain tumor grading based on a pre-trained AlexNet model was proposed.
[60]	2020	Multi-grade Classification of Brain Tumor	Radiopedia dataset	VGGNet	The paper investigates the CNN models for the Classification of Brain Tumor.
[61]	2020	Brain MRI Reconstruction for classification	The Molecular Interactive Display and Simulation (MIDAS) Dataset	Custom CNN	The authors trained a customized CNN model using a public dataset. They fine-tuned the model for the reconstruction of brain MR Images.
[62]	2020	MRI based Brain tumor diagnosis	Brain tumor public dataset (Figshare) & Harvard Medical School Data	Inception V3, DensNet201	The author concatenated different features that are extracted from multiple layers of used pre-trained deep learning models and then utilized those features to classify the brain tumors using a softmax classifier.
[63]	2021	Brain Tumor Classification	Brain tumor public dataset (Figshare)	Inception V3, Xception, and multiple ML Algorithm	Author used Inception V3 & Xception models to extract the features and then different deep and ML classifiers for the classification. The main contribution of the author is to introduced ensemble model based on the extracted features from used deep learning models
[64]	2021	Brain Tumor Classification using GoogleNet features and ML algorithm	Brain tumor public dataset (Figshare) & Harvard Medical School Data	GoogleNet	In this paper, the authors used the Pre-trained GoogleNet to extract the features and input to the ML-based classifiers K-NN and SVM.
[65]	2021	Multimodal Brain Tumor Classification	Brain Tumor Segmentation (BraTS) 2018 Dataset	VGG-19	First, the authors extracted features from two different dense layers of VGG19 pre-trained deep learning model and fused them to get more informative features' knowledge. Second, they used IPSO algorithm to select the optimum features.
[66]	2021	Brain Tumor classification using Ensemble of Deep features and ML algorithm	02 different brain tumor dataset from Kaggle consists of MR Images with and without tumor & Brain tumor public dataset (Figshare).	DeneNet-160, InceptionV3, ResNet50 for feature extraction and Multiple ML algorithms for classification	Authors used ensemble features, extracted from DeneNet-160, InceptionV3, and ResNet50 and use multiple ML algorithms such as K-NN, SVM, AdaBoost, etc., to classify MR images into different classes,

(Continued)

Table 1: Continued

Reference	Year	Study Domain	Dataset	Used Model for Classification	Description
[67]	2022	Brain Tumor Classification	02 different brain tumor dataset from BraTS and Figshare	Features extracted from Xception model	Authors extracted features using Xception network and proposed multi-level attention network (MANet) by designing spatial attention & long short-term memory network (LSTM)-based cross-channel module.

3 Proposed Framework

This section explains all the details related to the proposed framework, presented in Fig. 1, for implementing the transfer learning technique utilizing pre-trained DL models for the classification task. Any pre-trained classification model with its learned parameters can be used after customization. The proposed framework is based on a novel idea to input hyperparameters in the form of ordered pairs (batch size and learning rate). The ordered pair can be defined as a 2-tuple element of a matrix constructed using the concept of the Cartesian product of two initialized sets of batch size and learning rate. The following subsections discuss the step-by-step implementation of the transfer learning technique using the proposed framework.

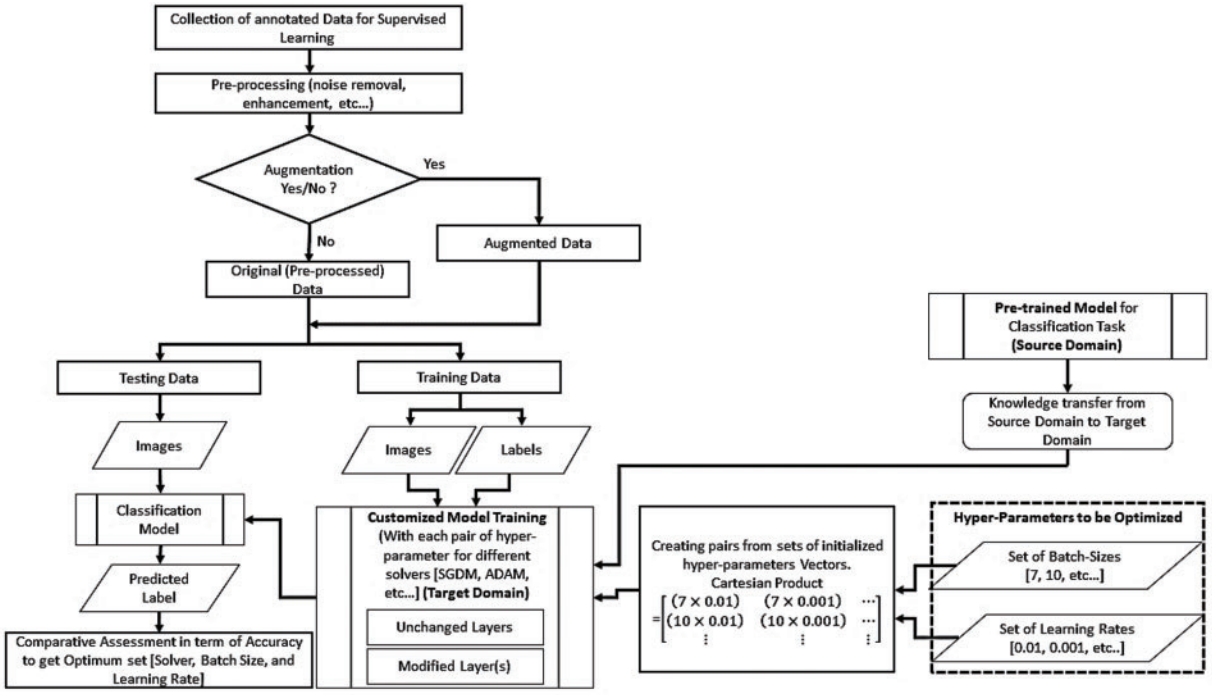


Figure 1: Proposed framework to implement transfer learning technique

3.1 Dataset

A publicly accessible brain tumor dataset of 233 patients consisting of 3064 MR images [39], was used. Three types of brain tumor MR images comprising 1426 slices of gliomas, 708 slices of meningioma, and 930 slices of pituitary tumors are available in this dataset. Fig. 2 illustrates each type of tumor percentage in the dataset. The data is available in .mat file format (Matlab data format), contains a label for the image, patient ID, image in the form of a 512 × 512 matrix, a tumor mask, and discrete points coordinate on tumor border.

Dataset statistics clearly depicts that the classes in the dataset are imbalance. Imbalance dataset is itself a challenging issue in which predicting correctly smaller size class is more critical than larger size class. In this research, model-based approach, such as transfer learning technique, is used to tackle this issue. Experiments in previous research [50,53–55,57–64] have supported the use of model-based approach provided that the tuning parameters are chosen carefully. In fact, transfer learning technique takes advantage of the features extracted from source domain at different levels and compensates for the overall lack of samples in the training data or targeted domain, ending up with a good trained model for imbalanced small dataset.

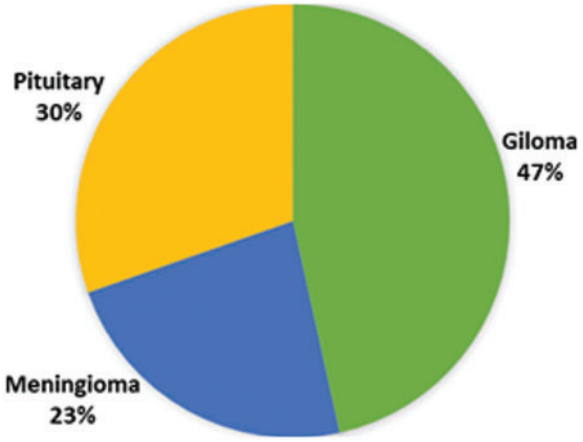


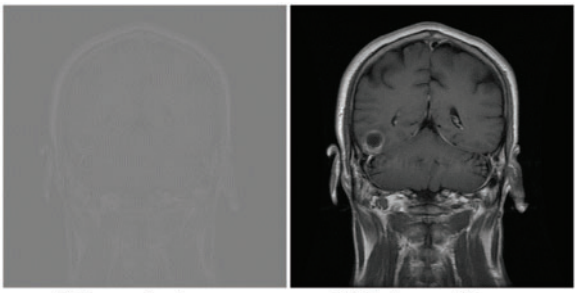
Figure 2: Percentage of different type of tumors in the dataset

3.2 Preprocessing

Medical image analysis requires data preprocessing, which includes contrast enhancement and standardization. First, the dataset was normalized to the intensity values and then mapped to grayscale's 256 levels using the relationship, as described in Eq. (1).

$$y_{(i,j)} = \frac{x_{(i,j)} - x_{min}}{x_{max} - x_{min}} \times 2^8 \tag{1}$$

where $y_{(i,j)}$ corresponds to any one of the 8-bit grayscale pixel values between 0 and 255 against x at position(i, j). x_{max} and x_{min} are the maximum pixel intensity and minimum pixel intensity in the original image, respectively. Fig. 3 shows the original and enhanced images.



(a) Image in dataset (b) Enhanced Image

Figure 3: Original and enhanced image

The enhanced resultant images are resized and concatenated three times, as per the standard input image size of the pre-trained DL models, to create channels. Tab. 2 lists the models used in this research and their standard input sizes.

Table 2: Used pre-trained models and input image size

Model Name	Standard Input Image Size
AlexNet, SqueezeNet	$227 \times 227 \times 3$
GoogleNet (ImageNet), GoogleNet (Places365), VGG 16, VGG 19, ResNet 18, ResNet 50, ResNet 101, MobileNet	$224 \times 224 \times 3$
Inception V3	$299 \times 299 \times 3$

3.3 Data Augmentation

High-quality big data plays a significant role in the effective training of any DL model. It is undoubtedly expensive to collect sufficient medical data available for training to reconstruct the classifier. In general, data augmentation techniques are used to enlarge the data size to provide a larger input sample space to achieve the desired accuracy and to reduce overfitting.

In this research, extensive data augmentation is performed not only to increase the data size but also to check the effectiveness of the knowledge-based transfer learning technique with and without data augmentation, particularly for our brain tumor classification task. A total of eight augmentation techniques with 32 different parameters were implemented to extend each data sample into 32 samples. Out of eight, four techniques–flipping, rotation, shears, and skewness–are for geometric transformation invariance, and the rest of the techniques, sharpening, Gaussian blur, emboss, and edge detection are used for noise invariance [52]. The details of the total dataset size and each class size before and after augmentation are listed in Tab. 3.

Table 3: Brain tumor dataset statistics with and without augmentation

Types of tumors	Dataset statistics	
	Original dataset	Augmented dataset
Glioma	1426	45632
Meningioma	708	22656
Pituitary	930	29760
Total	3064	98084

3.4 Pre-Trained Networks

For the classification task, there were many pre-trained CNN models. In this research, the idea behind the proposed framework is domain adaptation, in which transfer learning allows us to utilize the network and knowledge in terms of network weights of pre-trained DL models, from a source domain, to re-train it using new training data for another classification task in the target domain. The data size and similarity between the target and source domain tasks are important parameters for pre-trained model selection. Because almost all pre-trained existing DL models are trained on millions of natural images, choosing one pre-trained model directly to implement the transfer learning technique for the classification of brain tumors is quite difficult. For this reason, we used 11 contemporary pre-trained DL models, as shown in Tab. 2, and fine-tuned them to find the optimum model for our classification task. All 11 models were selected based on their learned rich feature representation as they were trained on the ImageNet database, consisting of 1000 object categories except one GoogleNet variant that was trained on images from the places 365 database, consisting of 365 image categories. Other than the "similarity" between source and target domain, the selected model's performance and efficiency differs because of many other characteristics such as nature of network architecture (sequential or Directed Acyclic Graph (DAG)), depth of the network (number of network layers), number of learned parameters (weights), etc. In this research, the selected models contain all types of networks, that is, sequential, DAG, and advanced compact CNN models.

3.5 Fine-Tuning

In the proposed framework, all 11 selected pre-existing deep learning models were fine-tuned for brain tumor multi-classification. In general, deep learning models consist of different layers, including convolutional, max pooling, FC, softmax layer, and a last cross-entropy-based classification layer. Using the concept of transfer learning, we retained all the network layers with their learned parameters, particularly weights, except that the last FC layer was replaced with a new FC layer with output size three, and the network's last cross-entropy-based classification layer is replaced with a new one. The retained layers help the network to use the low-level extracted features from the pre-trained model, while the replaced layers facilitate the network in high-level feature learning. The modified network is then fine-tuned to obtain an optimum model by training it with our brain tumor dataset. For training, the transfer learning technique uses two types of parameters: learned parameters (e.g., weights) from the original pre-existing deep learning model and hyperparameters, for example, batch size and learning rate, but the latter needs to be optimized. Because the choice of hyperparameters and their values depends on the targeted task, types in the dataset, and size of the dataset, it is quite difficult to select one specific hyper-parameter(s) optimal value that works for all pre-trained models.

In general, while training the CNN, a back-propagation algorithm is employed to minimize the cost function. Eq. (2) illustrates the cost function

$$C = -\frac{1}{m}\sum_{i}^{m} \ln\left(p\left(y^i|X^i\right)\right) \tag{2}$$

where m represents the total number of samples (images) for training, X^i represents i^{th} image sample with a y^i label, and $p\left(y^i|X^i\right)$ represents the probability of true classification. Optimization algorithms (solvers), such as stochastic gradient descent, are used to perform learning for mini-batches of size N using the gradient, computed using back-propagation, which results in minimizing the cost function C. Considering, for a convolutional layer l, the weights W_l^t at iteration t and mini-batch cost $\hat{C}$ to update the weights in the following iteration, using Eqs. (3), (4), and (5).

$$\gamma^t = \gamma^{\left\lfloor \frac{tN}{m} \right\rfloor} \tag{3}$$

$$V_l^{t+1} = \mu V_l^t - \gamma^t a_l \frac{\partial \hat{C}}{\partial W_l} \tag{4}$$

$$W_l^{t+1} = W_l^t + V_l^{t+1} \tag{5}$$

where a_l represents the layer l learning rate, γ represents the scheduling rate that affects the learning rate, and μ represents the momentum that affects the weights, updated previously, in the current iteration. It is clear from the above equations that among all hyper-parameters, batch size and learning rate are the two most important hyper-parameters and their optimal values, side by side, may help in solving different issues such as convergence problems, convergence time, and overfitting, and ultimately improve the accuracy of the pre-trained DL model while using them to implement the transfer learning technique. In this research, we used the following optimization strategy to obtain the optimal pair rather than the individual optimal values of the two most important hyper-parameters: learning rate and batch size.

We initialized two different sets X and Y for both hyperparameters, consisting of possible values based on their available values in various studies [46,50,53,62,68,69]. We define $X = [7, 10, 32, 128]$ and $Y = [0.01, 0.001, 0.0001, 0.00001]$ for the batch size and learning rate, respectively. A2 − dimensional matrix of size 4×4, containing 2−tuple elements, was generated by taking the Cartesian product of two initialized sets X and Y. The Cartesian product of two sets X and Y is the set of all ordered pairs (x, y), can be defined as

$$X \times Y = [(x, y)\,|x \in X \text{ and y} \in \text{Y}] \tag{6}$$

It can be generalized to n-ary Cartesian product over n sets $X_1, \cdots, X_n$ of differenthyperparameters

$$X_1 \times \cdots \times X_n = [(x_1, \cdots, x_n)\,|x_i \in X_i \text{ for every } i \in \{1, \cdots, \text{n}\}] \tag{7}$$

In our case, we transformed the Cartesian product vector into a matrix for a better understanding, as described in Eq. (8):

$$X \times Y = \begin{bmatrix} (7, 0.01) & (7, 0.001) & (7, 0.0001) & (7, 0.00001) \\ (10, 0.01) & (10, 0.001) & (10, 0.0001) & (10, 0.00001) \\ (32, 0.01) & (32, 0.001) & (32, 0.0001) & (32, 0.00001) \\ (128, 0.01) & (128, 0.001) & (128, 0.0001) & (128, 0.00001) \end{bmatrix} \tag{8}$$

Each element of the Cartesian product matrix is applied as a pair of inputs for two hyperparameters to retrain the modified network architecture against each pre-trained deep learning model with our dataset for the brain tumor classification task. Each modified network architecture was evaluated for the three most popular solvers: SGDM, ADAM, and RMSProp. An extensive comparative assessment was conducted in terms of accuracy to obtain the optimal values of batch size and learning rate, along with the most appropriate solver.

3.6 Performance Assessment

The performance assessment of the proposed framework was carried out using the same performance metrics used in related references [5,40,52,53,55,62,64,70]. A classifier can be tested using four parameters: true positive (TP): an outcome where model predicts the positive class correctly, true negative (TN): an outcome where model predicts the negative class correctly, false positive (FP): an outcome where model predicts the positive class incorrectly, and false negative (FN): an outcome where model predicts the negative class incorrectly. These parameters can be extracted from the confusion matrix to compute the performance metrics: accuracy, precision, specificity, and sensitivity (recall).

Precision:

Precision represents the ratio of correctly predicted positive data samples (instances) to the total predicted positive instances. Mathematically,

$$\text{Precision} = \frac{TP}{TP + FP} \tag{9}$$

Specificity:

Specificity measures the ratio of correctly predicted negative instances to the all instances in an actual class. Mathematically,

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{10}$$

Sensitivity:

Sensitivity measures the ratio of correctly predicted positive instances to the all instances in an actual class. Mathematically,

$$\text{Sensitivity (Recall)} = \frac{TP}{TP + FN} \tag{11}$$

Accuracy:

Accuracy measures the correctly classified instances with respect to the total number of instances.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

4 Results and Analysis of Framework

As discussed in Tab. 1, many researchers have explored transfer learning techniques for brain tumor classification by utilizing the knowledge of pre-trained DL models. The models were fine-tuned using different hyperparameters. However, the literature review reveals that pre-trained models have not been deeply investigated for brain tumor classification, particularly for multi-hyperparameters simultaneously. As discussed earlier, each hyperparameter has an impact on the model's performance, depending on the targeted domain and task. To analyze the proposed framework, we performed an

extensive comparative analysis to assess the optimal values of hyperparameters (Learning Rate and Batch Size) by applying their different values as 2-tuple input from a Cartesian product matrix, to 11 different deep learning models. The 11 models used for investigation were AlexNet, GoogleNet, GoogleNet (Places365) [71], ResNet18, ResNet50, ResNet101, VGG16, VGG19, SqueezeNet [72], MobileNet [73], and InceptionV3 [74]. The proposed framework was implemented and investigated for brain tumor classification using a system equipped with NVIDIA GEFORCE GTX 1080 – 8GB Graphics and MATLAB 2020. The dataset was divided into 70%, 15%, and 15% for training, validation, and testing of the model, respectively. After customizing the pre-trained deep learning model, a total of 435 experiments were performed with each pair of inputs from the Cartesian matrix of batch size and learning rate for the three most popular solvers. All the results in terms of accuracy are presented in Fig. 4.

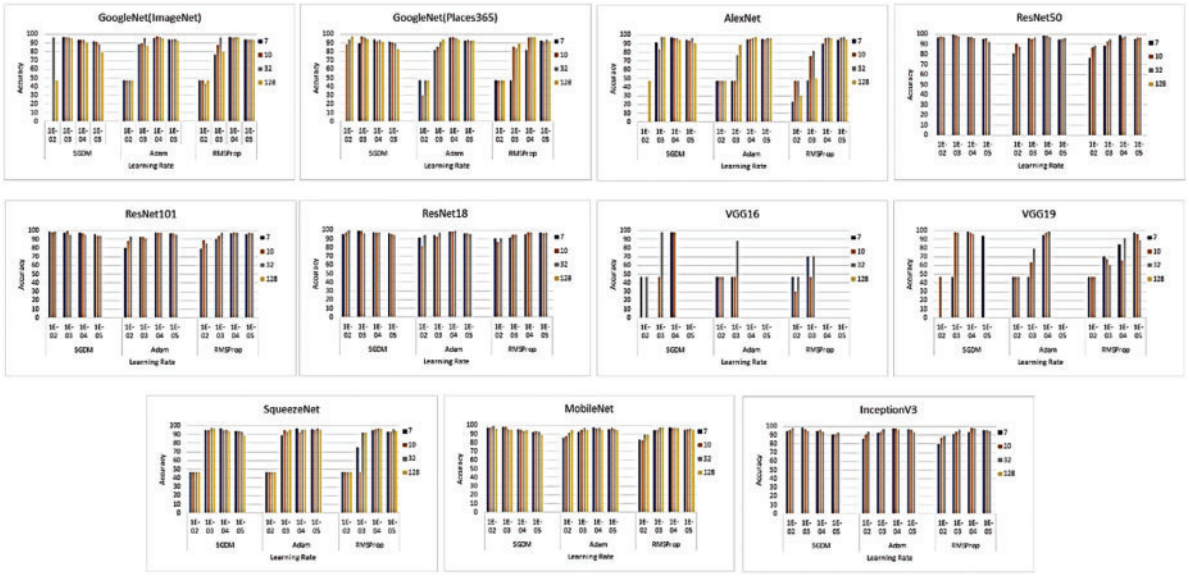


Figure 4: Performance evaluation in terms of accuracy for the pre-trained models using hyperparameters pair

To determine the optimal parameters, we first separated the maximum accuracy result for each pre-trained model and summarized it in Tab. 4 with other parameters observed during the training processes. The other parameters are the number of epochs utilized in convergence, number of iterations, validation accuracy, training time, and confusion matrix. To determine the optimized model for our brain tumor classification problem, all experimental results were carefully investigated. Starting with AlexNet, out of 48 combinations of {Solver type, Batch Size, Learning Rate}, our proposed framework performed well for the combination {SGDM, 32, 0.001} with an accuracy of 97.6%. AlexNet showed good performance for another combination {Adam, 128, 0.0001} with an accuracy of 97.82%, but its training time was much higher than the previous one, and could not converge until forced stop on completing 100 epochs. It is quite obvious from the experimental results that the training time for AlexNet is considerably less than that for other pre-trained networks because of its sequential nature. The results of GoogleNet (ImageNet), GoogleNet (Places365), and SqueezeNet are almost the same as AlexNet, even though they have a complex architecture based on a DAG network. The modified models VGG16, VGG19, MobileNet, and InceptionV3 when re-trained using our proposed framework performed better than the models discussed earlier. All three variants of ResNet, especially

ResNet18, outperformed all other networks with parameters {SGDM, 32, 0.01} by achieving 99.56% accuracy when using our proposed framework for brain tumor classification. This is due to the ResNet working principle of building a deeper network compared to other networks and its capability to solve the vanishing gradient problem simultaneously. Figs. 5a and 5b depict the training-validation accuracy and loss curve, and confusion matrix while training, validating, and testing ResNet18, the best-performing model. In addition to the ultimate accuracy measurement, utilizing three other measures: precision, recall, and specificity, the proposed framework was further evaluated. Tab. 5 summarizes the performance measures related to the above-mentioned measuring parameters for the average of all classes and each class separately as well for all deep learning networks presented in Tab. 4. The comparison shows that ResNet18 outperforms all the others in all the measuring fields. To the best of our knowledge, the strategy of our proposed framework has proven to be quite efficient, with an accuracy of 99.56%.

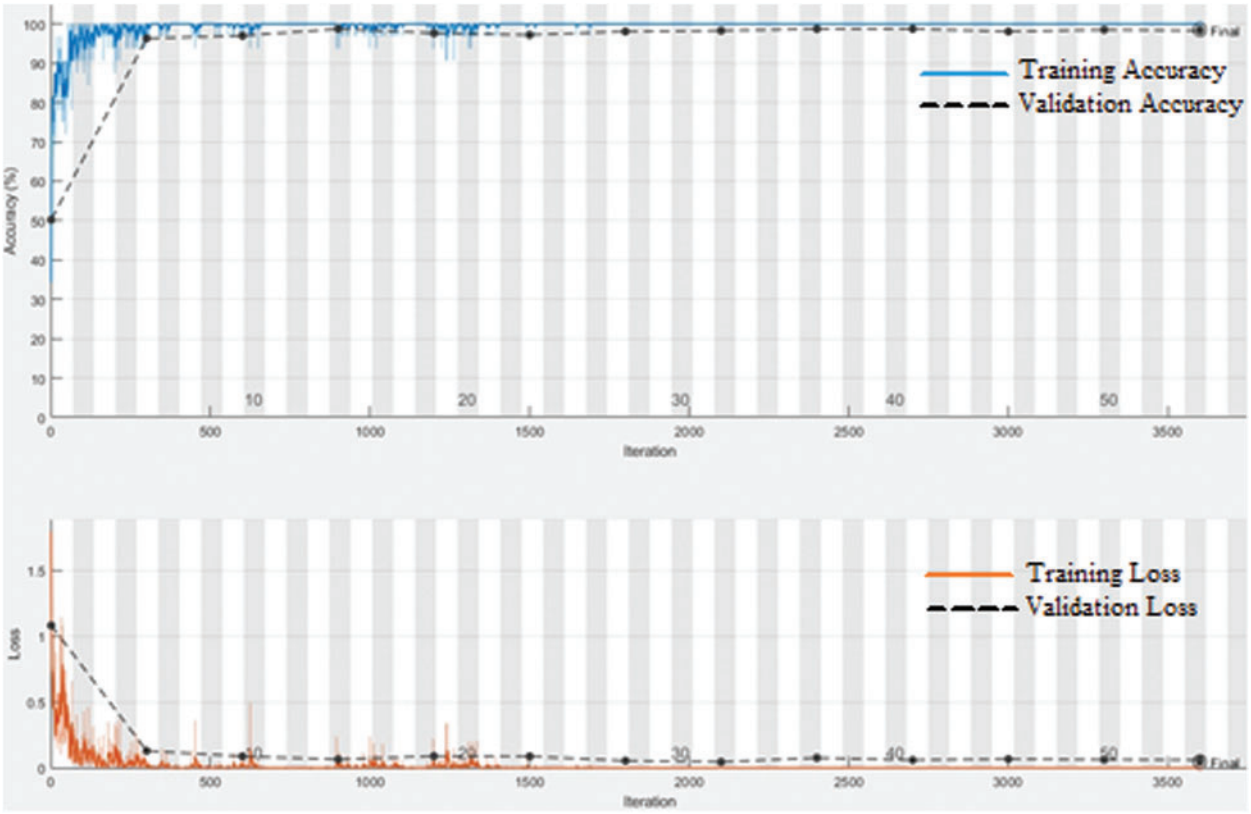
Table 4: Comparative study of models with their optimal parameters

Pre-Trained Model	Confusion Matrix	Predicted Class			Solver	Batch Size	Learning Rate	Epoch	Iterations	Validation Accuracy (%)	Testing Accuracy (%)	Training Time
		G	M	P								
AlexNet	True Class G	210	3	1	SGDM	32	0.001	54	3600	97.17	97.6	0:14:44
	M	2	101	3								
	P	0	2	137								
GoogleNet (ImageNet)	True Class G	210	4	0	Adam	10	0.0001	16	3300	98.4	97.39	00:16:03
	M	2	101	3								
	P	1	2	136								
GoogleNet (Places365)	True Class G	210	4	0	SGDM	10	0.001	20	4200	98.26	97.17	00:14:42
	M	6	99	1								
	P	1	1	137								
ResNet-50	True Class G	213	1	0	SGDM	7	0.001	17	5100	98.26	99.56	0:24:46
	M	1	105	0								
	P	0	0	139								
ResNet-101	True Class G	213	1	0	SGDM	10	0.001	23	4800	98.26	99.35	0:51:04
	M	1	105	0								
	P	1	0	138								
ResNet-18	True Class G	213	1	0	SGDM	32	0.01	54	3600	98.48	99.56	0:19:25
	M	0	105	1								
	P	0	0	139								
VGG16	True Class G	214	0	0	SGDM	7	0.0001	11	3300	96.74	98.26	0:14:41
	M	6	98	2								
	P	0	0	139								
VGG19	True Class G	211	3	0	SGDM	7	0.0001	15	4500	97.17	98.69	0:21:24
	M	1	105	0								
	P	0	2	137								

(Continued)

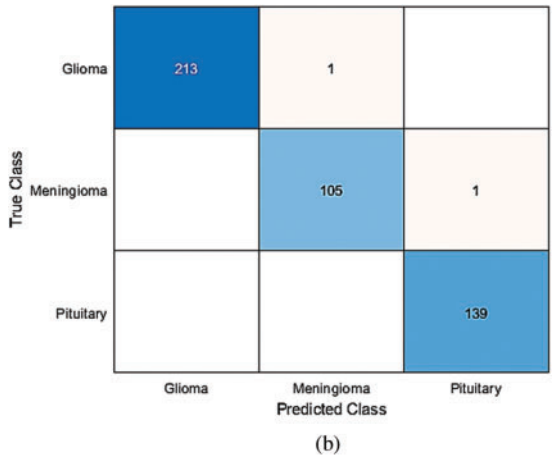
Table 4: Continued

Pre-Trained Model	Confusion Matrix		Predicted Class			Solver	Batch Size	Learning Rate	Epoch	Iterations	Validation Accuracy (%)	Testing Accuracy (%)	Training Time
		G	M	P									
SqueezeNet	True Class	G	208	5	1	SGDM	32	0.001	36	2400	97.39	97.39	0:11:18
		M	1	103	2								
		P	2	1	136								
MobileNet	True Class	G	213	1	0	SGDM	32	0.01	54	3600	97.61	98.91	0:43:46
		M	1	103	2								
		P	0	1	138								
Inception V3	True Class	G	211	3	0	RMS-Prop	10	0.0001	20	4200	98.04	98.26	0:58:39
		M	1	103	2								
		P	1	1	137								



(a)

Figure 5: (Continued)



(b)

Figure 5: (a) Training-Validation accuracy and loss for best performing model, (b) Confusion matrix for best performing model

Table 5: Comparative study of models in terms of performance metrics

Fine-Tune Models	Precision Per Class	Average Precision	Sensitivity Per Class	Average Sensitivity	Specificity Per Class	Average Specificity
AlexNet	99.06% 95.28% 97.16%	97.61%	98.13% 95.28% 98.56%	97.60%	99.18% 98.58% 98.75%	98.91%
GoogleNet (ImageNet)	98.11% 92.59% 98.56%	96.97%	97.20% 94.34% 98.56%	96.95%	98.37% 97.73% 99.38%	98.53%
GoogleNet (Places365)	96.77% 95.19% 99.28%	97.17%	98.13% 93.40% 98.56%	97.17%	97.14% 98.58% 99.69%	98.25%
ResNet50	99.53% 99.06% 100.00%	99.56%	99.53% 99.06% 100.00%	99.56%	99.59% 99.72% 100.00%	99.74%
ResNet101	99.07% 99.06% 100.00%	99.35%	99.53% 99.06% 99.28%	99.35%	99.18% 99.72% 100.00%	99.55%
ResNet18	100.00% 99.06% 99.29%	99.57%	99.53% 99.06% 100.00%	99.56%	100.00% 99.72% 99.69%	99.84%
VGG16	97.27% 100.00% 98.58%	98.30%	100.00% 92.45% 100.00%	98.26%	97.55% 100.00% 99.38%	98.67%

(Continued)

Table 5: Continued

Fine-Tune Models	Precision Per Class	Average Precision	Sensitivity Per Class	Average Sensitivity	Specificity Per Class	Average Specificity
VGG19	99.53% 95.45% 100.00%	98.73%	98.60% 99.06% 98.56%	98.69%	99.59% 98.58% 100.00%	99.48%
SqueezeNet	98.58% 94.50% 97.84%	97.41%	97.20% 97.17% 97.84%	97.39%	98.78% 98.30% 99.06%	98.75%
MobileNet	99.53% 98.10% 98.57%	98.91%	99.53% 97.17% 99.28%	98.91%	99.59% 99.43% 99.38%	99.49%
InceptionV3	99.06% 96.26% 98.56%	98.26%	98.60% 97.17% 98.56%	98.26%	99.18% 98.87% 99.38%	99.17%

Table 6: Comparative study of models with their optimal parameters using augmented dataset

Pre-Trained Model	Confusion Matrix		Predicted Class			Solver	Batch Size	Learning Rate	Epoch	Iterations	Validation Accuracy (%)	Testing Accuracy (%)	Training Time
			G	M	P								
AlexNet	True Class	G	6699	114	32	SGDM	32	0.001	4	8100	98.4	98.27	1:10:41
		M	45	3316	37								
		P	3	24	4435								
Google Net (ImageNet)	True Class	G	6664	142	39	Adam	10	0.0001	1	5100	97.6	97.97	1:04:27
		M	28	3320	50								
		P	10	30	4422								
GoogleNet (Places365)	True Class	G	6820	19	6	SGDM	10	0.001	2	6900	97.61	97.26	1:12:04
		M	238	3097	63								
		P	63	14	4385								
ResNet-50	True Class	G	6748	78	19	SGDM	7	0.001	01	7800	98.8	98.87	01:50:01
		M	19	3371	8								
		P	7	35	4420								
ResNet-101	True Class	G	6831	11	3	SGDM	10	0.001	2	9000	99.15	99.20	03:50:23
		M	58	3305	35								
		P	4	6	4452								

(Continued)

Table 6: Continued

Pre-Trained Model	Confusion Matrix		Predicted Class			Solver	Batch Size	Learning Rate	Epoch	Iterations	Validation Accuracy (%)	Testing Accuracy (%)	Training Time
			G	M	P								
ResNet-18	True Class	G	6806	35	4	SGDM	32	0.01	4	7200	99.23	99.23	1:29:40
		M	30	3363	5								
		P	8	31	4423								
SqueezeNet	True Class	G	6822	18	5	SGDM	32	0.001	4	7200	98.33	98.34	1:29:59
		M	153	3224	21								
		P	23	24	4415								
MobileNet	True Class	G	6825	18	2	SGDM	32	0.01	3	6000	99.21	99.31	1:56:17
		M	42	3341	15								
		P	9	16	4437								
Inception V3	True Class	G	6838	4	3	RMS-Prop	10	0.0001	2	7500	99.01	98.92	0:58:39
		M	73	3283	42								
		P	24	13	4425								

However, do we need an augmentation technique to increase the data size. In this research, as discussed above, to increase the data size, extensive data augmentation techniques are utilized, and the framework is evaluated only for the optimal hyperparameters that provide high accuracy for each pre-trained model without augmentation evaluation. Tab. 6 shows a minimum improvement of 0.09% using the GoogleNet (Places365) model and a maximum improvement of 0.95% using the SqueezeNet model, which shows quite similar results to the original data (without augmentation). This suggests that the transfer learning technique, if implemented with a proper framework for the selection of optimal hyperparameters, may not require data augmentation.

A performance comparison is presented in Tab. 7 between our work and other existing state-of-the-art research works that used the same brain tumor dataset for multi-type tumor classification. The comparison is mainly based on performance metric "accuracy" with support of three other parameters "precision," "recall," and "specificity." The comparison shows that the transfer learning technique, implemented through our proposed framework for brain tumor classification, outperformed all existing approaches based on traditional image processing [5,40], CNN [44,70], and transfer learning [52,53,55,60,62,64,69].

Table 7: Comparison of proposed framework with the related work based on same dataset

Related Work	Approach	Accuracy	Precision				Recall				Specificity			
			G	M	P	Average	G	M	P	Average	G	M	P	Average
[5]	BoW-SVM	91.28	–	–	–	–	96.4	86	87.3	–	96.3	95.5	95.3	–
[40]	DWT-Gabor-NN	91.90	–	–	–	–	95.1	86.9	91.2	–	96.3	96	95.7	–
[44]	CapsNet	90.89	–	–	–	–	–	–	–	–	–	–	–	–
[70]	CNN-ELM	93.68	91	94.5	98.3	–	97.5	76.8	100	–	–	–	–	–
[52]	fine-tuned VGG19	94.58	–	–	–	–	–	–	–	88.41	–	–	–	96.12
[53]	fine-tuned VGG19	94.82	93	87.97	87.34	89.52	95.97	89.98	96.81	94.25	93.79	96.42	93.93	94.69
[55]	GoogleNet-SVM	97.10	99	94.7	98	–	97.9	96	98.9	–	99.4	98.4	99.1	–
[69]	fine-tuned VGG16	98.69	–	–	–	–	–	–	–	–	–	–	–	–
[60]	VGGNet	94.00	–	–	–	–	–	–	–	–	–	–	–	–
[62]	DenseNet	99.51	99	99	100	–	100	99	99	–	–	–	–	–
[64]	GoogleNet-KNN	98.30	98	95.55	97.78	–	98.02	94.57	99.1	–	98.63	98.65	99.01	–
Proposed	ResNet18	99.56	100	99.06	99.29	99.45	99.53	99.06	100	99.53	100	99.72	99.69	99.8

5 Conclusion and Future Work

This research presents a comprehensive literature review, along with a robust framework for implementing the transfer learning technique. The comprehensive review reveals that there is a need for a solution to select an appropriate pre-trained deep learning model and optimal hyperparameter(s) values for such an implementation. Our proposed model not only solves the model selection issue, but also helps in determining the optimal hyperparameter values. To determine the appropriate pre-trained deep learning model, 11 state-of-the-art pre-trained models were used. A Cartesian product matrix is created to obtain all possible pairs from initialized sets of hyperparameters (batch size and learning rate). All pairs were applied as input one-by-one to each pre-trained deep learning model and re-trained with our brain tumor dataset for the three most popular solvers for the evaluation of the performance of the proposed framework. The simulation work for the framework's assessment reveals that the transfer learning technique is quite effective even with a small size imbalance dataset, and we may not need augmented data if it is implemented with a proper framework with an appropriate selection of hyperparameters and solvers. Further, the results reveal a tradeoff between batch size and learning rate, but it depends on the model architecture type and complexity. The assessment shows that the proposed framework is effective for radiologists and physicians in classifying diverse tumor types. The proposed framework can also be used for other classification issues.

The work can be broadened in the future to increase the dimensions of the Cartesian product matrix to obtain optimal values of other hyperparameters. Further, in-depth investigation is required for a few pre-trained DL models that failed to retrain on our dataset for selected pairs of the Cartesian product matrix.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] K. Selvanayaki and M. Karnan, "CAD system for automatic detection of brain tumor through magnetic resonance image—A review," *International Journal of Engineering Science and Technology*, vol. 2, no. 10, pp. 5890–5901, 2010.

[2] K. M. Brindle, J. L. Izquierdo-Garcia, D. Y. Lewis, R. J. Mair and A. J. Wright, "Brain tumor imaging," *Journal of Clinical Oncology*, vol. 35, no. 21, pp. 2432–2438, 2017.

[3] P. Y. Wen, D. R. Macdonald, D. A. Reardon, T. F. Cloughesy, A. G. Sorensen *et al.,* "Updated response assessment criteria for high-grade gliomas: Response assessment in neuro-oncology working group," *Journal of Clinical Oncology*, vol. 28, no. 11, pp. 1963–1972, 2010.

[4] A. Drevelegas and N. Papanikolaou, "Imaging modalities in brain tumors," in *Imaging of Brain Tumors with Histological Correlations*, 2nd ed., Berlin Heidelberg: Springer, pp. 13–33, 2011.

[5] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang *et al.,* "Enhanced performance of brain tumor classification via tumor region augmentation and partition," *PloS One*, vol. 10, no. 12, pp. e0144479, 2015.

[6] J. Cheng, W. Yang, M. Huang, W. Huang, J. Jiang *et al.,* "Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation," *PloS One*, vol. 11, no. 6, pp. e0157112, 2016.

[7] S. Kumar, C. Dabas and S. Godara, "Classification of brain MRI tumor images: A hybrid approach," *Procedia Computer Science*, vol. 122, pp. 510–517, 2017.

[8] G. Mohan and M. M. Subashini, "MRI based medical image analysis: Survey on brain tumor grade classification," *Biomedical Signal Processing and Control*, vol. 39, pp. 139–161, 2018.

[9] I. Sutskever, J. Martens and G. E. Hinton, "Generating text with recurrent neural networks," in *Proc. ICML*, Bellevue, WA, USA, 2011.
[10] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proc. 25th Int. Conf. on Machine learning*, Helsinki, Finland, pp. 160–167, 2008.
[11] N. Jaitly and G. E. Hinton, "Vocal tract length perturbation (VTLP) improves speech recognition," in *Proc. ICML Workshop on Deep Learning for Audio, Speech and Language*, vol. 117, Atlanta, Georgia, USA, pp. 21, 2013.
[12] Y. Taigman, M. Yang, M. A. Ranzato and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, pp. 1701–1708, 2014.
[13] C. Szegedy, A. Toshev and D. Erhan, "Deep neural networks for object detection," in *Advances in Neural Information Processing Systems*. Vol. 26, pp. 2553–2561, 2013.
[14] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proc. IEEE Computer Vision and Pattern Recognition*, Boston, MA, USA, pp. 3128–3137, 2015.
[15] J. Zhang and C. Zong, "Deep neural networks in machine translation: An overview," *IEEE Intelligent Systems*, vol. 30, no. 5, pp. 16–25, 2015.
[16] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre *et al.,* "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.
[17] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein *et al.,* "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping," *NeuroImage*, vol. 129, no. 869–877, pp. 460–469, 2016.
[18] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall *et al.,* "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
[19] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh *et al.,* "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Int. Conf. on Learning Representations (ICLR)*, 2015. https://arxiv.org/abs/1409.1556
[22] M. Everingham and J. Winn, "The pascal visual object classes challenge 2012 (voc2012) development kit," *Pattern Analysis, Statistical Modelling and Computational Learning*, vol. 8, pp. 5, 2011.
[23] L. Roux, "Mitosis Atypia 14 Grand Challenge," 2014 [Online]. *Available:* https://mitos-atypia-14.grandcha.
[24] G. E. Hinton, S. Osindero and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
[25] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[26] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. on Machine Learning*, Lille, France, 2015.
[27] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, Haifa, Israel, 2010.
[28] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
[29] T. Ateeq, M. N. Majeed, S. M. Anwar, M. Maqsood, Z. Rehman *et al.,* "Ensemble-classifiers-assisted detection of cerebral microbleeds in brain MRI," *Computers & Electrical Engineering*, vol. 69, pp. 768–781, 2018.
[30] M. Havaei, A. Davy, D. Farley, A. Biard, A. Courville *et al.,* "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, no. 4, pp. 18–31, 2017.

[31] B. H. Menze, A. Jakab, S. Bauer, J. Cramer, K. Farahni *et al.,* "The multimodal brain tumor image segmentation benchmark (BraTS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.
[32] M. Prastawa, E. Bullitt, N. Moon, K. V. Leemput and G. Gerig, "Automatic brain tumor segmentation by subject specific modification of atlas priors1," *Academic Radiology*, vol. 10, no. 12, pp. 1341–1348, 2003.
[33] P. Abdolmaleki, F. Mihara, K. Masuda and L. D. Buadu, "Neural networks analysis of astrocytic gliomas from MRI appearances," *Cancer Letters*, vol. 118, no. 1, pp. 69–78, 1997.
[34] A. Kharrat, K. Gasmi, M. B. Messaoud, N. Benamrane and M. Abid, "A hybrid approach for automatic classification of brain MRI using genetic algorithm and support vector machine," *Leonardo Journal of Sciences*, vol. 17, no. 1, pp. 71–82, 2010.
[35] E. I. Papageorgiou, P. P. Spyridonos, D. Glotsos, C. D. Stylios, P. Ravazoula *et al.,* "Brain tumor characterization using the soft computing technique of fuzzy cognitive maps," *Applied Soft Computing*, vol. 8, no. 1, pp. 820–828, 2008.
[36] E. I. Zacharaki, S. Wang, S. Chawla, D. S. Yoo, R. Wolf *et al.,* "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1609–1618, 2009.
[37] K. L. Hsieh, C. Lo and C. Hsiao, "Computer-aided grading of gliomas based on local and global MRI features," *Computer Methods and Programs in Biomedicine*, vol. 139, pp. 31–38, 2017.
[38] J. Sachdeva, V. Kumar, I. Gupta, N. Khandelwal and C. K. Ahuja, "A package-SFERCB-Segmentation, feature extraction, reduction and classification analysis by both SVM and ANN for brain tumors," *Applied Soft Computing*, vol. 47, no. 12B, pp. 151–167, 2016.
[39] J. Cheng, "Brain magnetic resonance imaging tumor dataset," in *Figshare MRI Dataset Version 5,* 2017. [Online]. Available: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427/5
[40] M. R. Ismael and I. Abdel-Qader, "Brain tumor classification via statistical features and back-propagation neural network," in *Proc. IEEE Int. Conf. on Electro/Information Technology (EIT)*, Rochester, Michigan, USA, pp. 252–257, 2018.
[41] S. Khalid, T. Khalil and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in *Proc. Science and Information Conf.*, London, UK, pp. 372–378, 2014.
[42] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[43] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 770–778, 2016.
[44] P. Afshar, A. Mohammadi and K. N. Plataniotis, "Brain tumor type classification via capsule networks," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, Athens, Greece, pp. 3129–3133, 2018.
[45] R. Zia, P. Akhtar and A. Aziz, "A new rectangular window based image cropping method for generalization of brain neoplasm classification systems," *International Journal of Imaging Systems and Technology*, vol. 28, no. 3, pp. 153–162, 2018.
[46] H. H. Sultan, N. M. Salem and W. Al-Atabany, "Multi-classification of brain tumor images using deep neural network," *IEEE Access*, vol. 7, pp. 69215–69225, 2019.
[47] Z. Jia and D. Chen, "Brain tumor identification and classification of MRI images using deep learning techniques," *IEEE Access*, pp. 1, 2020. https//org.10.1109/ACCESS.2020.3016319.
[48] S. Banerjee, S. Mitra, F. Masulli and S. Rovetta, "Brain tumor detection and classification from multi-sequence MRI: Study using convnets," in *Proc. Int. MICCAI Brainlesion Workshop*, Granada, Spain, pp. 170–179, 2018.
[49] C. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi, "Inception-V4, inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. on Artificial Intelligence*, San Francisco, California, USA, 2017.
[50] M. Talo, U. B. Baloglu, Ö. Yıldırım and U. R. Acharya, "Application of deep transfer learning for automated brain abnormality classification using MR images," *Cognitive Systems Research*, vol. 54, pp. 176–188, 2019.

[51] K. A. Johnson and J. A. Becker, "Braintumor datasets," in *Harvard Medical School Data* [Online]. Available: http://www.med.harvard.edu/AANLIB/
[52] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah *et al.,* "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *Journal of Computational Science*, vol. 30, pp. 174–182, 2019.
[53] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali *et al.,* "Brain tumor classification for MR images using transfer learning and fine-tuning," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 34–46, 2019.
[54] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali *et al.,* "Content-based brain tumor retrieval for MR images using transfer learning," *IEEE Access*, vol. 7, pp. 17809–17822, 2019.
[55] S. Deepak and P. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *Computers in Biology and Medicine*, vol. 111, no. 3, pp. 103345, 2019.
[56] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed *et al.,* "Going deeper with convolutions," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, MA, USA, pp. 1–9, 2015.
[57] M. Ghafoorian, A. Mehertash, T. Kapur, N. Karssemeijer, E. Marchiori *et al.,* "Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation," in *Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, Quebec, Canada, pp. 516–524, 2017.
[58] C. Haarburger, P. Langenberg, D. Truhn, H. Schneider, J. Thuring *et al.,* "Transfer learning for breast cancer malignancy classification based on dynamic contrast-enhanced MR images," in *Bildverarbeitung für die Medizin 2018*. Berlin, Heidelberg, 216–221, 2018.
[59] G. S. Tandel, A. Balestrieri, T. Jujaray, N. N. Khanna, L. Saba *et al.,* "Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm," *Computers in Biology and Medicine*, vol. 122, no. 1, pp. 103804, 2020.
[60] K. Muhammad, S. Khan, J. D. Ser and V. H. C. de Albuquerque, "Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 507–522, 2020.
[61] S. U. H. Dar, M. Özbey, A. B. Çatlı and T. Çukur, "A transfer-learning approach for accelerated MRI using deep neural networks," *Magnetic Resonance in Medicine*, vol. 84, no. 2, pp. 663–685, 2020.
[62] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, M. Imran *et al.,* "A deep learning model based on concatenation approach for the diagnosis of brain tumor," *IEEE Access*, vol. 8, pp. 55135–55144, 2020.
[63] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad and M. O. Alassafi, "Brain tumor classification based on fine-tuned models and the ensemble method," *Computers Materials & Continua*, vol. 67, no. 3, pp. 3967–3982, 2021.
[64] A. Sekhar, S. Biswas, R. Hazra, A. K. Sunaniya, A. Mukherjee *et al.,* "Brain tumor classification using fine-tuned GoogLeNet features and machine learning algorithms: IoMT enabled CAD system," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 3, pp. 983, 2022.
[65] A. B. T. Tahir, M. A. Khan, M. Alhaisoni, J. A. Khan, Y. Nam *et al.,* "Deep learning and improved particle swarm optimization based multimodal brain tumor classification," *Computers, Materials & Continua*, vol. 68, no. 1, pp. 1099–1116, 2021.
[66] J. Kang, Z. Ullah and J. Gwak, "MRI-based brain tumor classification using ensemble of deep features and machine learning classifiers," *Sensors*, vol. 21, no. 6, pp. 2222, 2021.
[67] N. S. Shaik and T. K. Cherukuri, "Multi-level attention network: Application to brain tumor classification," *Signal, Image and Video Processing*, vol. 16, no. 3, pp. 817–824, 2022.
[68] M. Yaqub, J. Feng, M. S. Zia, K. Arshid, K. Jia *et al.,* "State-of-the-art CNN optimizer for brain tumor segmentation in magnetic resonance images," *Brain Sciences*, vol. 10, no. 7, pp. 427, 2020.
[69] A. Rehman, S. Naz, M. I. Razzak, F. Akram and M. Imran, "A deep learning-based framework for automatic brain tumors classification using transfer learning," *Circuits, Systems, and Signal Processing*, vol. 39, no. 2, pp. 757–775, 2020.
[70] A. Pashaei, H. Sajedi and N. Jazayeri, "Brain tumor classification via convolutional neural network and extreme learning machines," in *Proc. Int. Conf. on Computer and Knowledge Engineering (ICCKE)*, Ferdowsi University of Mashhad, Iran, pp. 314–319, 2018.

[71] B. Zhou, A. Khosla, A. Lapedriza, A. Torralba and A. Oliva, "Places: An image database for deep scene understanding," *Journal of Vision*, vol. 17, no. 10, pp. 296, 2017.

[72] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally *et al.,* "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size," in *Proc. Int. Conf. on Learning Representations*, Toulon, France, 2017.

[73] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. C. Chen, "MobilenetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 4510–4520, 2018.

[74] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 2818–2826, 2016.