

RESEARCH ARTICLE

3D Multimodal Brain Tumor Segmentation and Grading Scheme based on Machine, Deep, and Transfer Learning Approaches

Erdal Tasci^{1,2*}, Aybars Ugur², Kevin Camphausen¹, Ying Zhuge¹, Rachel Zhao³, and Andra Krauze¹

¹Center for Cancer Research, National Cancer Institute, NIH, Building 10, Bethesda, MD, 20892, United States.

²Computer Engineering Department, Ege University, Bornova, Izmir, 35100, Turkey.

³University of British Columbia, Faculty of Medicine, 317 - 2194 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada.

Abstract

Glioma is one of the most common tumors of the brain. The detection and grading of glioma at an early stage are very critical for increasing the survival rate of the patients. Computer-aided detection (CADE) and computer-aided diagnosis (CADx) systems are essential and vital tools that allow for more accurate and systematic results to speed up the decision-making process of clinicians. By utilizing various deep learning models (e.g., CNN) and transfer learning strategy (e.g., fine-tuning), performance results for image classification have increased accuracy and improved effectiveness especially for novel large-scale data sets that share similarities. In this paper, we introduce a novel method consisting of combined variations of machine, deep, and transfer learning approaches for the effective brain tumor (i.e., glioma) segmentation and grading on the multimodal brain tumor segmentation (BRATS) 2020 dataset. We apply popular and efficient 3D U-Net architecture for the brain tumor segmentation phase. We also utilize 23 different combinations of deep feature sets and machine learning/fine-tuned deep learning CNN models based on Xception, IncResNetv2, and EfficientNet by using 4 different feature sets and 6 learning models for the tumor grading phase. The experimental results demonstrate that the proposed method achieves a 99.5% accuracy rate for slice-based tumor grading on BraTS 2020 dataset. Moreover, our approach results in competitive performance with similar recent works.

Key Words: Glioma tumor; Glioma grades; Segmentation; Feature extraction; Deep learning, Ensemble learning, MRI classification

*Corresponding Author: Erdal Tasci, Computer Engineering Department, Ege University, Bornova, Izmir, 35100, Turkey; E-mail: arif.erdal.tasci@ege.edu.tr

Received Date: May 25, 2022, Accepted Date: June 17, 2022, Published Date: July 15, 2022

Citation: Erdal Tasci, Aybars Ugur, Kevin Camphausen, Ying Zhuge, Rachel Zhao, and Andra Valentina Krauze. 3D Multimodal Brain Tumor Segmentation and Grading Scheme based on Machine, Deep, and Transfer Learning Approaches. *Int J Bioinform Intell Comput.* 2022;1(2):84-102.



This open-access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (CC BY-NC) (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits reuse, distribution and reproduction of the article, provided that the original work is properly cited, and the reuse is restricted to non-commercial purposes.

Abbreviations

ACC: Accuracy Rate; AFPNet: Atrous-convolution Feature Pyramid Network; AUC: Area Under the Receiver Operating Characteristics Curve; BraTS: Brain Tumor Segmentation Challenge; CAde: Computer-Aided Detection; CADx: Computer-Aided Diagnosis; CNN: Convolutional Neural Network CNS - Central Nervous System; CRF: Conditional Random Fields; DSC: Dice Coefficient; FC: Fully Connected; FCNN: Fully Convolutional Neural Network; FLAIR: Fluid-Attenuated Inversion Recovery; FN: False Negative; FP: False Positive; F1: F-(Measure) Score; GBM: Glioblastoma Multiforme; GPU: Graphics Processing Unit; HGG: High-Grade Glioma; ISLES: Ischemic Stroke Lesion Segmentation; KNN: K Nearest Neighbors; MLP: Multi-Layer Perceptron; MRI: Magnetic Resonance Imaging; NB: Naive Bayes; LGG: Low-Grade Glioma; LR: Logistic Regression; PRE: Precision; RAM: Random Access Memory; REC: Recall; RF: Random Forest; ROC: Receiver Operating Characteristics; SGDM: Stochastic Gradient Descent with Momentum; SPEC: Specificity; STD: Standard Deviation; SVM: Support Vector Machine; TN: True Negative; TP: True Positive; T1-CE: Postcontrast Enhanced T1 Weighted; 3D: Three Dimensional

1. Introduction

The term "brain tumor" often refers to a collection of intracranial neoplasms [1], representing an uncontrolled growth of abnormal cells in the brain [2]. According to Cancer.net, brain and other nervous system cancers are the 10th leading cause of death for men and women [3]. In 2021, an estimated 18,600 adults (10,500 men and 8,100 women) will die from a primary cancerous brain tumor (i.e., primary brain tumor arising in the brain) and central nervous system (CNS) tumors [3].

Glioma is a common brain tumor originating from glial cells [4]. Brain lesions are named based on the type of brain cell that transforms to generate the lesion. Examples include astrocytes/astrocytoma or glial cells/glioblastoma, oligodendrocytes/oligodendrogliomas, and ependymal/ependymomas [4,5]. Gliomas originating from glial cells comprise approximately 30% of all brain and CNS tumors and 80% of all malignant brain tumors [6]. According to the World Health Organization classification guidelines, gliomas can be grouped into four grades: I, II, III, and IV, based on histopathological analysis. Grades II and III represent low-grade gliomas (LGGs), whereas grade IV, also identified as glioblastoma multiforme (GBM), is a high-grade glioma. GBM is the most common glioma, responsible for approximately 15% of all primary brain and CNS tumors and 55% of all gliomas [6].

Glioma detection and grading are essential in cancer diagnosis and management, including radiation therapy planning and the assessment and prediction of outcomes [7,8]. Traditionally diagnosis is arrived at by utilizing invasive methods such as obtaining tissue diagnosis for pathological examination (e.g., resection), which carries the risk of pain, hemorrhage, and infection, or employing non-invasive methods automatically by using various medical imaging techniques.

The most important and commonly employed test/medical imaging technique involved in diagnosing and following glioma patients is Magnetic Resonance Imaging (MRI) since it accurately provides details on the size, type, and position of the investigated tumor region [1,9]. Furthermore, MRIs are capable of high soft-tissue resolution and are more sensitive in visualizing and detecting subtle changes in tissue density related to the tumor [9]. MRIs are

acquired without ionizing radiation; hence patients are not exposed to harmful effects of ionizing radiation [10]. MRI modalities consisting of T1-weighted (T1), post-contrast enhanced T1-weighted (T1-Gd, T1-CE), T2-weighted and T2-FLAIR (Fluid-attenuated inversion recovery) images are used for brain tumor diagnosis tasks (Figure 1).

With the advance and development of information technology and artificial intelligence (e.g., machine learning, deep learning, and pattern recognition) techniques, computer-aided detection (CADe) and computer-aided diagnosis (CADx) systems are growing in efficacy and significance to the medical field. With transfer learning, we focus on transferring the knowledge (e.g., pre-trained CNN models) obtained from one problem (e.g., source dataset) to the learning process of a related new task [11]. Transfer learning strategies yield generalized problem-solving ability for similar domain/problems. They are also crucial due to providing resource efficiency and more accurate results without needing large-scale data sets, to train a model from scratch.

These techniques can help theoretically improve outcomes for patients by decreasing the time and resources needed for the decision-making of clinicians/expert radiologists by automating tasks. Accordingly, brain tumor detection and grading are essential tasks to improve patient cases and allow clinicians to overcome existing drawbacks (e.g., the need for tissue diagnosis and waiting for the clinical decision-making process).

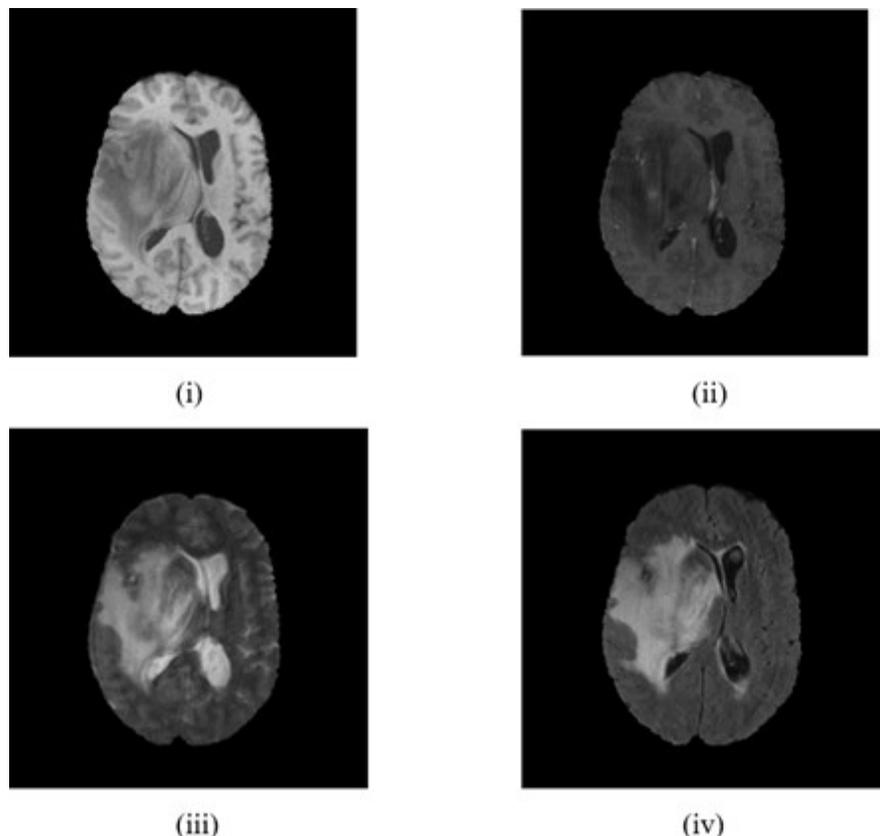


Figure 1: Sample MRI slice images with different modalities from BraTS2020 dataset [12]. (i) T1 (ii) T1-CE (iii) T2 (iv) T2-FLAIR.

In this study, we introduce a deep learning-based segmentation of glioma that combines machine and deep learning-based approaches to obtain effective tumor grading results on MRI images. Additionally, we compute the mean performance values for six evaluation metrics (accuracy rate (%) (ACC) and AUC (Area Under ROC Curve), precision (PRE), recall (REC), specificity (SPEC), and F-score (F1)) on different train-test sets. Finally, we examine all computational results to determine the most optimal methods for improved glioma detection and grading performance.

The significant contributions of this manuscript are summarized as follows:

- To our knowledge, our study consists of the first method that uses both variations of deep learning and machine learning models to evaluate deep feature sets for brain tumor grading.
- Our study provides a novel contribution by showcasing a method that allows for both glioma segmentation and tumor grading using deep learning and transfer learning-based methods and five commonly used machine learning models.
- We focus on combining the advantages of the machine learning, deep learning, transfer learning, and ensemble learning methods.
- In addition to 3D tumor segmentation, we also extracted deep features from three efficient fine-tuned CNN models, and the extensive effects of the features are analyzed on different learning models for the tumor grading phase.
- We also observe the effects of the ensemble of extracted deep features on the commonly used multimodal brain tumor image dataset (i.e., BraTS).
- We aim to figure out the best model with a related deep feature set to obtain more accurate and reliable results for tumor grading purposes.
- The performances of the learning models are measured with six performance metrics: accuracy, Area under the ROC curve (AUC), precision, recall, specificity, and F-score by using various seed values in detail.

The organization of this paper is structured as follows. In Section 1, we present general background information and statistics associated with brain tumors. We briefly discuss related studies and the methods employed to address this problem in Section 2. In Section 3, we describe our methodology and provide details on the proposed work, while Section 4 provides details on the experimental analysis and presented results. We also offer a comprehensive performance analysis of the proposed methodology. Finally, Section 5 concludes this paper with future directions.

2. Related Works

Image segmentation and classification are crucial for many medical applications such as CADe and CADx. Manual, semi-automatic, fully automatic, or hybrid approaches can be performed for brain tumor segmentation studies. In the manual method of segmentation, boundaries of regions can be extracted by medical experts (e.g., radiologists, radiation oncologists) by manually delineating tumors (i.e., time-consuming, tedious, and laborious process), whereas, in fully automatic segmentation, boundaries are identified automatically with the help of computational methods, such as traditional or deep learning-based approaches [13]. For brain tumor/medical image classification studies, the extraction and selection of handcrafted and/or deep features are carried out, followed by machine learning methods such as Naïve Bayes (NB) classifier, support vector machines (SVM), and K-nearest neighbors (K-NN), ensemble learning methods such as bagging (e.g., Random Forest),

boosting, and deep learning methods such as convolutional neural networks (CNNs) [11,14-17] are utilized.

In this section, we present the summary information about recent developments and studies in brain tumor segmentation and classification as follows (Table 1):

Table 1: Summary information about recent studies in brain tumor segmentation and classification.

Study	Year	Method	Dataset	Type of Abnormality
[18]	2021	Salient and Convolutional Features	BraTS (2015, 2018)	LGG, HGG
[19]	2021	Handcrafted + Deep Features	BraTS (2017)	LGG, HGG
[7]	2020	3DUNet+Mask R-CNN+ConvNet	BraTS (2018)	LGG, HGG
[20]	2018	CNN + CRF-RNN	BraTS (2013, 2015, 2016)	LGG, HGG
[21]	2020	AFPNET	BraTS (2013, 2015, 2018)	LGG, HGG
[22]	2018	Incremental CNN	BraTS (2017)	LGG, HGG
[23]	2020	Stacked Autoencoders	BraTS (2012-2015)	LGG, HGG
[24]	2019	Score Level Fusion with TL	BraTS (2013-2016)	LGG, HGG
[25]	2019	CNN with Texture Feature	BraTS (2015)	LGG, HGG
[26]	2018	Statistical and Wavelet Features	BraTS (2015)	LGG, HGG
[27]	2019	DRRNet	BraTS (2015)	LGG, HGG

Takacs *et al.* [18] proposed a fusion of the saliency-based model and CNN features for brain tumor segmentation on Multimodal Brain Tumor Image Segmentation Challenge (BraTS)2015 and BraTS2018 datasets consisting of MRI scans of patients with high-grade glioma (HGG), and low-grade glioma (LGG). Al-qazzaz *et al.* [19] presented a method that combines histogram-based (e.g., handcrafted) features with CNN-based features from the CIFAR network for image classification-based brain tumor segmentation. Their approach was input into a decision tree classifier, and it is evaluated on the BraTS 2017 dataset. Zhuge *et al.* [7] introduced two novel approaches, including the popular 3D U-Net model for brain tumor segmentation and 2D Mask R-CNN and 3D ConvNet for improving segmentation and tumor grading. They evaluated their fully automated proposed schemes on the BraTS 2018 dataset for survival predictions without requiring surgical biopsy.

Zhao *et al.* [20] integrated Conditional Random Fields (CRFs) and Fully Convolutional Neural Networks (FCNNs) in a combined framework to get the segmentation results accurately with spatial and appearance consistency on BraTS2013, BraTS2015, and BraTS2016 datasets. Their method segments brain images slice-by-slice faster than image patches. Zhou *et al.* [21] proposed a 3D fully connected CNN with atrous-convolution feature pyramid (AFPNet) to segment MRI images with brain tumors on BraTS2013, BraTS2015, and BraTS2018 datasets. Paoli *et al.* [22] introduced three end-to-end Incremental Deep Convolutional Neural Networks models (namely, 2CNet, 3CNet, and EnsembleNet) to obtain more accurate segmentation results. Their proposed models do not use any guided method for acquiring suitable hyper-parameters.

Amin *et al.* [23] constructed a stacked sparse autoencoders-based deep learning model to predict input slices as a tumor or non-tumor. They employed high-pass and median filters to preprocess slices. Then, seed growing algorithms are supplied to segment images. The model was then tested on the BraTS 2012-2015 datasets. Amin *et al.* [24] proposed score level fusion using transfer learning with AlexNet and GoogleNet CNNs for brain tumor detection and classification. The score vectors were given into multiple classifiers using softmax layer. They analyzed their method on BraTS 2013-2016 and ischemic stroke lesion segmentation (ISLES) 2018 datasets, respectively.

Deng *et al.* [25] presented an improved Fully Convolutional Neural Network (FCNN) and non-quantifiable texture feature to segment brain tumors with appearance and spatial consistency on the BraTS 2015 dataset. Latif *et al.* [26] extracted hybrid first-order, and second-order statistical features with discrete wavelet transform features (i.e., 152 features) for glioma MRI image classification. These features are provided to multi-layer perceptron (MLP) classifier for machine learning purposes. They also compared their results with other crucial predictors such as Random Forest, NB, and SVM on a partial BraTS 2015 dataset. Their proposed feature set produced relatively better results (i.e., 96.72% accuracy for high-grade glioma and 96.04% for low-grade glioma) than the existing studies. In another study, Sun *et al.* [27] designed an automatic 3D CNN architecture based on U-Net and replaced the simple skip connection with encoder adaptation blocks for brain tumor segmentation. They also used densely connected fusion block in the decoder part to improve the performance and reduce the computational time. Their method achieved results comparable to the state-of-the-art results on the BraTS 2015 dataset.

As stated in related works, there are many variations and applications of CNNs and their derivatives for brain tumor segmentation. However, to the best of our knowledge, limited published studies address both brain tumor segmentation and tumor grading/classification.

3. Methods

In this section, we describe the methods used in this study in detail. Firstly, a general overview of the proposed methods is given. Then, image preprocessing approaches, including image standardization and masking, brain tumor segmentation (namely, 3D U-Net), and tumor grading methods (namely, fine-tuned CNNs and deep feature extraction substages), are explained in the following subsections.

3.1. The overview of the proposed method

The flowchart of the proposed methods for glioma segmentation and grading is illustrated in Figure 2. As shown in Figure 2, the proposed methods consist of two stages: (i) brain tumor segmentation phase based on the 3D U-Net convolutional neural networks (CNN) model; (ii) brain tumor grading phase based on fine-tuning of three CNN models (namely, Xception, IncResNetv2, and EfficientNet) and/or deep feature extraction from these models with utilizing traditional machine learning classifiers (namely, Random Forest (RF), K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), Naïve Bayes Classifier (NB), and Logistic Regression (LR)).

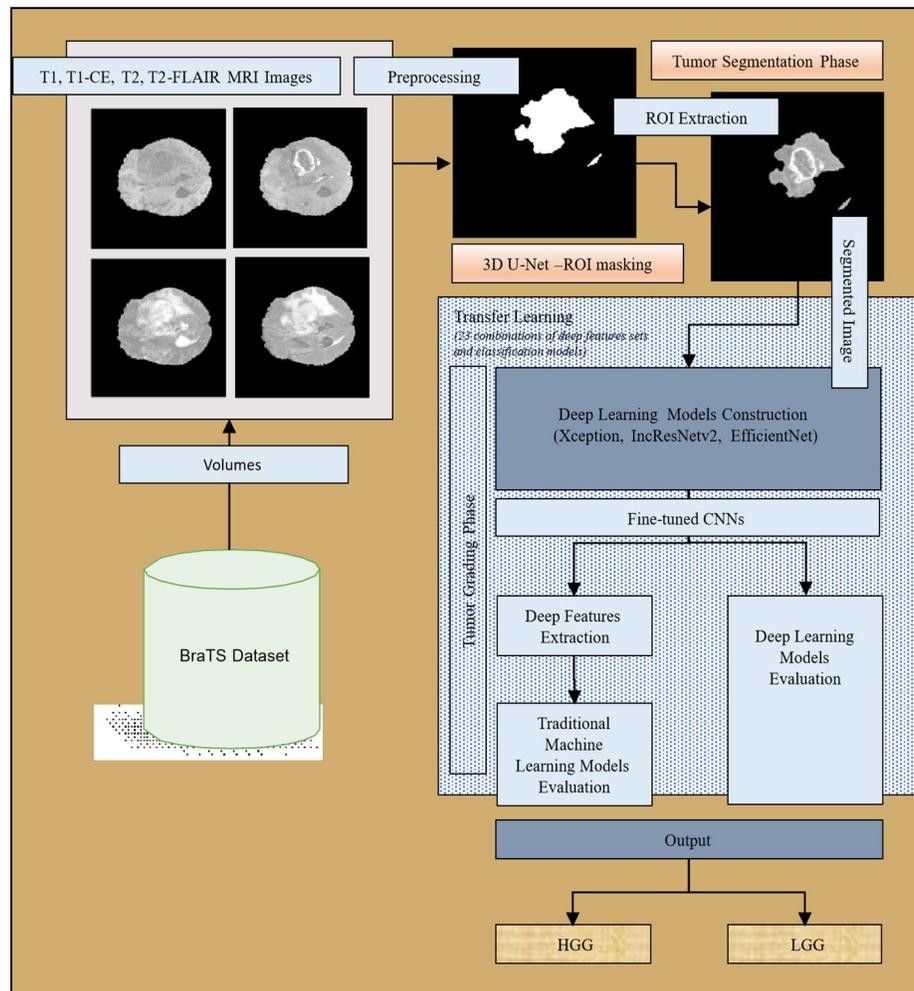


Figure 2: The flowchart of the proposed methods for glioma segmentation and grading.

Firstly, a multimodal brain tumor image segmentation dataset (BraTS) is obtained, and image preprocessing approaches (namely, image standardization and masking) are applied for the following stage (i.e., glioma segmentation). Then, multimodal image volumes are segmented with the popular 3D U-Net CNN model. Furthermore, the deep learning model construction stage is carried out using fine-tuning 3 popular and efficient CNN models (namely, Xception, IncResNetv2, and EfficientNet). Afterward, diverse deep features are extracted from deep learning models constructed. Finally, they are classified into high-grade glioma or low-grade glioma based on their characteristics and the learning models employed. Each substage is explained in the following subsections in detail.

3.2. Image processing

Image preprocessing is the step that aims to improve the shape and quality of an image, reduce noises, and make the following stages, such as segmentation and classification easier. In this study, the brain region is determined, masked and the intensity of each image is standardized by subtracting the mean and dividing by the standard deviation of the cropped

brain region to efficiently preprocess the MRI brain tumor data with the segmentation model applied (i.e., 3D U-Net) [28] for glioma detection and grading.

3.3. Brain tumor segmentation

Brain tumor segmentation/detection is related to the set of operations that partition corresponding images as tumor and non-tumor regions. Our study employs a popular and efficient, patch-based, and pre-trained 3D U-Net segmentation model for glioma segmentation that allows for adaptation to GPU memory and RAM limitations. The overview of the 3D U-Net model is illustrated in Figure 3. The input and output patch sizes of the network model are 132-by-132-by-132, and 44-by-44-by-44 voxels, respectively.

3D U-Net architecture includes an encoder and decoder subnetworks connected by a bridge section. These subnetworks in the 3D U-Net model consist of multiple stages. Each encoder stage comprises two sets of batch normalization, convolutional, and ReLU layers [29,30]. The name U-Net comes from the letter U since the network can be modeled with this symmetric shape.

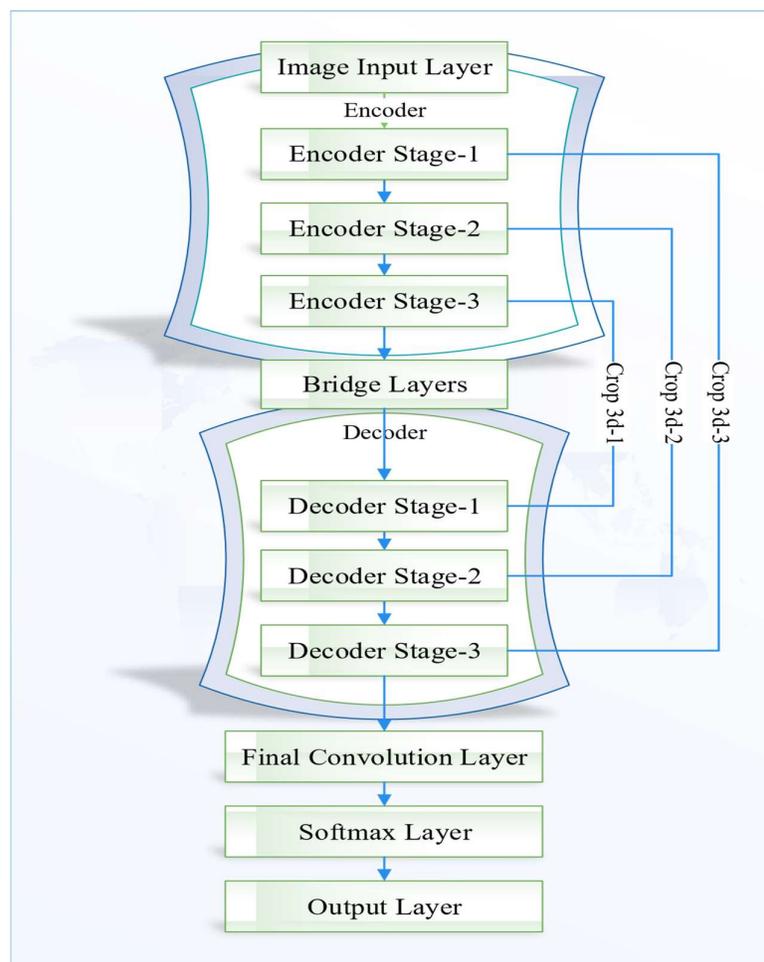


Figure 3: The overview of the 3D U-Net model for glioma segmentation [28].

3.4. Brain tumor grading

Brain tumor grading is a classification of the degree of malignancy of tumor image in the brain into low-grade glioma or high-grade glioma. In this study, binary classification models consisting of traditional machine learning models and three fine-tuned CNN models (namely, Xception, IncResNetv2, and EfficientNet) constructed from pre-trained CNNs are employed. Deep features are extracted from related CNN models separately for machine learning models, and they are given into classifiers for grading. These models are explained in detail in the following subsections as follows:

3.4.1. Pre-trained CNN models

A pre-trained CNN model is a model that is trained on other problems and can be used for similar problems without training from scratch. This approach represents a type of transfer learning technique. This model is particularly advantageous when data samples are limited or lacking as it assists with network generalization and can accelerate the convergence of the system [31].

To construct fine-tuned CNN models, we used three popular and commonly employed pre-trained CNN models, including Xception, IncResNetv2, and EfficientNet. These pre-trained CNN models have been trained on the large-scale ImageNet datasets (i.e., more than a million images) for image recognition purposes. Xception and IncResNetv2 networks have an image input size of 299- by-299, whereas the EfficientNet network has an image input size of 224-by-224. Xception and IncResNetv2 networks are 71 and 164 layers deep, respectively [32].

3.4.2. Fine-tuning

Fine-tuning is one of the most common approaches to transfer learning and improves the ability to generalize the selected model by applying backpropagation operation to the weights of the related pre-trained CNN models [16,33]. Fine-tuning a pre-trained CNN model is often much easier and faster than training a CNN model from scratch with randomly initialized weights [34].

For the fine-tuning process, we remove the last fully connected (FC) layer of corresponding pre-trained CNN models and change them with our new FC layer as a primary approach (i.e., the number of classes is equal to the size for classification in our new dataset) [16]. In this paper, we set the number of classes as 2 since we have two output categories, HGG and LGG.

3.4.3 Feature Extraction

The feature extraction stage is responsible for obtaining informative and discriminative data from images. This study extracts deep (learned) features from the last pooling layers of fine-tuned CNNs. The total number of features extracted from Xception, IncResNetv2, and EfficientNet are 2048, 1536, and 1280, respectively. These features are provided to five machine learning models used for training and testing tumor grading performance.

3.4.4 Classification

The classification phase assigns the related categories to the patterns regarding labeled data by employing supervised learning. In this study, five different supervised models are utilized

such as Random Forest, K-Nearest Neighbors, Support Vector Machine, Naïve Bayes Classifier, and Logistic Regression.

Random Forest: The Random Forest proposed by Leo Breiman is a bagging (i.e., bootstrap aggregation) ensemble learning method that consists of three combinations growing in randomly selected data subspaces [35–37]. It is considered one of the most popular, powerful, and accurate general-purpose learning techniques for dealing with a vast number of input variables without overfitting [37].

K-Nearest Neighbors: K-Nearest Neighbors is one of the most popular non-parametric classification and lazy learning algorithms that groups test instances into the category of the closest instances for distance measure and the number of neighbors. In other words, the instance is assigned to the class most common among its k nearest neighbors.

Support Vector Machine: Support Vector Machine, based on statistical learning theory, is a supervised learning algorithm proposed by Vapnik et al. [38,39] in 1992. SVM model explores and finds a hyperplane such that the margin (i.e., the width of the gap between the two groups) is maximized. SVM is thus beneficial and effective for two-class (i.e., binary) and non-linear classification problems on medical datasets.

Naïve Bayes Classifier: Naïve Bayes classifier, based on Bayes' theorem, is a probabilistic, simple, highly scalable, and extremely fast classifier that makes strong assumption conditional independency between all features given the scope of the class. This model assigns the most likely class to a given instance described by its pattern/feature vector [40].

Logistic Regression: Logistic Regression is one of the simplest and most used statistical models for binary classification, wherein probability is positive or negative depending on a linear measure of the instances [41].

4. Experimental Work

This section explains the experimental processes, brain tumor image dataset, and evaluation measures. Afterward, we give comprehensive computational results regarding the performance metrics of segmentation and classification/grading in the following subsections in detail.

4.1. Experimental process

All the experiments were conducted on a system running Windows 10 based on Matlab R2021a software and a desktop computer with the hardware configuration of Intel Core i7 8700K processor operating at 3.70 GHz with 64 GB RAM and 8 GB NVIDIA GeForce GTX 1080 GPU Memory. The train-test ratio is set to 80-20 for all the experiments (namely, fine-tuning, and deep feature extraction methods). We adjust the CPU random number generator seed to 1 for fine-tuning all CNN models to get the same computational results with the same seed value on the dataset used. We also utilized 5 different seeds (i.e., from 1 to 5) to evaluate the mean predictive performance for train-test splits.

In this study, the implementation of 3D brain tumor segmentation in MATLAB has been utilized for glioma detection [28]. We set the number of trees to 50 for the Random Forest

classifier and adjusted default parameters for the five machine learning models used. We analyzed every fine-tuned CNN individually to assess deep learning model performance. We also evaluated the performance of each machine learning classifier for each deep feature sets extracted from CNN models.

We used T1-CE, T2, and T2-FLAIR modalities of images for tumor grading because pre-trained CNN models can handle 3 channels. We also considered only slices that include the tumor region after the segmentation phase by counting white pixel intensities on each image separately. We employed stochastic gradient descent with a momentum (SGDM) optimizer for fine-tuning the networks. To accommodate GPU memory limitations, we identified the minibatch size as 16 for Xception, IncResnetv2, and EfficientNet CNNs. The maximum number of epochs is adjusted to 30 for all selected CNN models. The default values are assigned to the corresponding parameter values for other training options.

4.2. Image dataset

To analyze the effectiveness of the proposed methods, we employed the most used and recognized MRI dataset for this task named MICCAI brain tumor image segmentation (BraTS) 2020[12,42,43]. BraTS2020 dataset consists of 293 multimodal MRI scans of glioblastoma (GBM/HGG) and 76 lower-grade gliomas (LGG) image volumes in training test with labels and segmented annotations. Each image volume has 152 slices and an image size of 240*240 pixels approximately. This brain tumor dataset has been segmented manually by one to four raters, and their annotations were approved by experienced neuro-radiologists [12,42,43].

After brain tumor segmentation, we convert multimodal 3D brain tumor images into 2D tumor images for tumor grading purposes. To this end, we obtained total 22287 images consisting of 17806 HGG and 4481 LGG images for learning model construction and testing for each modality.

4.3. Evaluation metrics

To evaluate the tumor grading performance of the proposed method, we have utilized 6 different evaluation metrics consisting of classification accuracy (ACC), Area Under the ROC Curve (AUC), precision, sensitivity, specificity, and F-Measure [44]. We have also computed the Dice coefficient of images for the brain tumor segmentation task.

Dice coefficient (DSC), known as Sørensen–Dice index, is used to measure the similarity coefficient of two image samples. It is described as twice the number of elements (cardinalities) common to both image sets divided by the sum of the number of elements in each set. The equation of DSC is defined in Equation 1.

$$\text{Dice Coefficient} = \frac{2 |X \cap Y|}{|X| + |Y|} \quad [1]$$

Classification accuracy is calculated by dividing the total number of true positives and true negatives by the total number of instances (i.e., the total number of true positives, false positives, false negatives, and true negatives). The equation is defined in Equation 2.

$$ACC = \frac{TN + TP}{TN + TP + FN + FP} \quad [2]$$

where TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively.

AUC, the Area under the Receiver Operating Characteristic (ROC) curve, is constructed by plotting the true positive rate against the false-positive rate for the performance of the binary classifier model. An Area of 1 (i.e., maximum AUC value) indicates a perfect test, whereas an Area of 0 (i.e., minimum AUC value) indicates that the predictive model miscategorizes all instances.

Precision means the positive predictive value. It is computed by dividing the number of true positives by the total number of false positives and true positives. The equation is shown in Equation 3.

$$Precision = \frac{TP}{TP + FP} \quad [3]$$

Sensitivity is the true positive rate/hit rate or recall. It is computed by dividing the number of true positives by the total number of false negatives and true positives. The equation is defined in Equation 4.

$$Sensitivity = \frac{TP}{FN + TP} \quad [4]$$

Specificity is expressed as true negative rate. It is calculated by dividing the number of true negatives by the total number of false positives and true negatives. The equation of specificity is defined in Equation 5.

$$Specificity = \frac{TN}{FP + TN} \quad [5]$$

F-Measure is the harmonic mean of precision and recall. It is represented in Equation 6.

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad [6]$$

4.4. Computational results

This subsection presents the comprehensive results for brain tumor segmentation and grading purposes. Testing a typical image volume takes only around 1 minute for the brain tumor segmentation. We obtained 0.99784, and 0.74518 mean Dice coefficient values using 3D patch-based U-Net for semantic segmentation of brain tumor images on non-tumor and tumor region of training tests, respectively. Fine-tuning time of CNN models is shown in Table 2.

Table 2: Fine-tuning time of CNN models.

CNN Model	Fine-Tuning Time
Xception	6 hours 33 mins
IncResNetv2	16 hours 26 mins
EfficientNet	6 hours 33 mins

As preliminary testing, we aimed to elicit the most efficacious combination of methods by selecting the three best fine-tuned CNN models with the five common most popular supervised learning models. To accomplish this, we employed AlexNet, VGGNet, GoogleNet, Xception, IncResNetv2 and EfficientNet CNN models in diverse combinations on the same dataset to assess accuracy rate. Xception, IncResNetv2, and EfficientNet were chosen given that they represent the best three fine-tuned CNN models. The selected CNN models contain more deep layers, provide more speed, and better accuracy secondary to their architectural designs. RF, KNN, SVM, NB, and LR (given their popularity as classification/learning algorithms tools in the pertinent literature) were then employed for the problem at hand and are supported based their performance based on our results (i.e., accuracy rate) [16]. To further improve the results, we carried out additional analysis constructing different combination schemes to determine the most appropriate combination including analysis of the performance of various combinations on the tumor grading phase. Firstly, we analyzed the effects of five supervised learning models by utilizing three different deep features from fine-tuned CNN models for slice-based tumor grading performance. The results are reported in Table 3. Bold values indicate the highest-valued results. As shown in Table 3, mean tumor grading performance results of models with standard deviations according to BraTS 2020 datasets with respect to five seed values show that the best accuracy value of 99.2% is obtained from deep features from fine-tuned EfficientNet by using the Random Forest model. The best AUC value of 0.998 is obtained from deep features from fine-tuned EfficientNet by using the Logistic Regression model. The highest precision value of 0.995 is obtained from deep features from fine-tuned EfficientNet by using the K-Nearest neighbors classifier. The highest recall value of 0.998 is obtained from deep features from fine-tuned EfficientNet by using the Random Forest model. The highest specificity value of 0.979 is acquired from deep features from fine-tuned EfficientNet by using the K-Nearest neighbors classifier. The highest F-score value of 0.995 is obtained from fine-tuned EfficientNet by using the Random Forest model.

After analyzing deep features individually on the learning models, we observed the effects of the ensemble of deep features on these five classifiers. It is presented in Table 4. According to Table 4, the best accuracy and AUC values are obtained from the Logistic Regression model as 99.5% and 0.999, respectively. The highest precision value of 0.996, recall value of 0.999, specificity value of 0.983, and F-score value of 0.997 are obtained from K- Nearest neighbors, Naive Bayes, K-Nearest Neighbors, Logistic Regression, respectively.

Tables 3 and 4 show that the best ACC and AUC values result from the ensemble of deep features from three fine-tuned CNNs using the Logistic Regression classifier. The fusion of these features contributes to improving the tumor grading performance. The overview of the ensemble of deep features on the tumor grading scheme is illustrated in Figure 4.

In addition to slice-based image results, we also performed the patient-based results. To this end, we randomly selected (i.e., the seed value as 1) 20 GBM patients and 10 LGG patients for

evaluation using the patient- based image results. These slice images of patients were not used in fine-tuned CNN models and supervised machine learning model construction. We then obtained the experimental results in detail. In the training set, 20540 slices were obtained for each modality, with 1747 in the test set, and a total number of slices of 22287 for each modality. The computational results are illustrated in Tables 5 and 6. As can be seen from the results in Tables 5 and 6, the best accuracy value is obtained from the IncResNetv2 CNN model with a 74.9% accuracy value. The difference between patient-based and section-based performance results is due to learning from similar cross-sectional images, the small number of LGG samples, and the image differences between patients. Our method resulted in excellent slice-based results with some significant distinguishing features as compared to similar work in this space e.g. Amin *et al.* [24], including notably: 1) the goal of advancing the tumor grading task as compared to classifying benign vs. malignant; 2) multistep classification via 3D U-net (segmentation) + fine-tuned CNN as compared to the use of Alex and Google networks for classification and 3) the use of a guided method for feature extraction in transfer learning vs generated score vectors [24].

Table 3: Mean tumor grading performance results of models with standard deviations according to BraTS 2020 datasets with respect to five seed values.

Fine-tuned Classifier		ACC	std	AUC	std	PRE	std	REC	std	SPEC	std	F1	std
CNN	Model												
Xception	RF	0.989	0.009	0.995	0.008	0.991	0.007	0.995	0.003	0.965	0.029	0.993	0.005
	KNN	0.988	0.008	0.981	0.013	0.993	0.006	0.992	0.004	0.971	0.022	0.992	0.005
	SVM	0.990	0.009	0.998	0.003	0.992	0.006	0.995	0.005	0.970	0.025	0.994	0.005
	NB	0.988	0.009	0.975	0.020	0.988	0.010	0.996	0.001	0.954	0.040	0.992	0.005
	LR	0.987	0.008	0.998	0.003	0.991	0.007	0.993	0.003	0.964	0.028	0.992	0.005
	CNN	0.987	0.008	0.997	0.004	0.991	0.006	0.993	0.003	0.963	0.025	0.992	0.005
IncResNet v2	RF	0.988	0.010	0.994	0.008	0.991	0.006	0.994	0.006	0.966	0.025	0.993	0.006
	KNN	0.987	0.009	0.980	0.012	0.992	0.004	0.992	0.007	0.967	0.017	0.992	0.006
	SVM	0.990	0.009	0.998	0.002	0.993	0.006	0.994	0.005	0.973	0.024	0.994	0.006
	NB	0.982	0.009	0.978	0.017	0.992	0.008	0.985	0.004	0.967	0.030	0.988	0.006
	LR	0.987	0.010	0.998	0.003	0.990	0.007	0.994	0.005	0.962	0.029	0.992	0.006
	CNN	0.986	0.009	0.997	0.003	0.986	0.008	0.997	0.004	0.945	0.031	0.991	0.006
EfficientNet	RF	0.992	0.008	0.997	0.004	0.992	0.007	0.998	0.003	0.968	0.029	0.995	0.005
	KNN	0.991	0.006	0.987	0.010	0.995	0.005	0.994	0.003	0.979	0.018	0.994	0.004
	SVM	0.991	0.007	0.998	0.002	0.994	0.006	0.995	0.003	0.975	0.024	0.994	0.004
	NB	0.991	0.007	0.986	0.015	0.993	0.007	0.996	0.002	0.972	0.029	0.994	0.005
	LR	0.991	0.008	0.998	0.002	0.994	0.006	0.995	0.004	0.976	0.023	0.995	0.005
	CNN	0.991	0.008	0.998	0.002	0.994	0.006	0.994	0.005	0.977	0.022	0.994	0.005

Table 4: Mean tumor grading performance results of models with standard deviations according to BraTS 2020 datasets for five seed values using ensemble of deep features from three fine-tuned CNN models.

Classifier Model	ACC	std	AUC	std	PRE	std	REC	std	SPEC	std	F1	std
RF	0.993	0.007	0.996	0.005	0.995	0.005	0.997	0.003	0.978	0.022	0.996	0.004
KNN	0.994	0.006	0.990	0.010	0.996	0.004	0.997	0.003	0.983	0.017	0.996	0.004
SVM	0.994	0.005	0.999	0.002	0.995	0.005	0.997	0.002	0.980	0.020	0.996	0.003
NB	0.991	0.008	0.979	0.020	0.990	0.010	0.999	0.000	0.959	0.040	0.994	0.005
LR	0.995	0.005	0.999	0.001	0.995	0.005	0.998	0.002	0.981	0.018	0.997	0.003

Table 5: Patient-based tumor grading performance results of models according to BraTS 2020 datasets.

Fine-tuned CNN	Classifier Model	ACC	AUC	PRE	REC	SPEC	F1
Xception	RF	0.700	0.699	0.696	0.955	0.227	0.805
	KNN	0.710	0.607	0.705	0.952	0.263	0.810
	SVM	0.718	0.793	0.711	0.956	0.278	0.815
	NB	0.669	0.540	0.669	0.971	0.108	0.792
	LR	0.726	0.780	0.716	0.959	0.294	0.820
	CNN	0.709	0.768	0.705	0.950	0.261	0.809
IncResNetv2	RF	0.709	0.729	0.708	0.939	0.283	0.808
	KNN	0.718	0.625	0.717	0.935	0.315	0.811
	SVM	0.715	0.822	0.710	0.948	0.283	0.812
	NB	0.697	0.595	0.699	0.938	0.250	0.801
	LR	0.733	0.836	0.728	0.941	0.346	0.821
	CNN	0.749	0.823	0.745	0.933	0.407	0.828
EfficientNet	RF	0.716	0.758	0.708	0.959	0.265	0.815
	KNN	0.737	0.649	0.731	0.943	0.356	0.823
	SVM	0.725	0.781	0.715	0.960	0.289	0.820
	NB	0.709	0.623	0.718	0.911	0.335	0.803
	LR	0.736	0.797	0.729	0.944	0.348	0.823
	CNN	0.744	0.783	0.757	0.893	0.467	0.819

Table 6: Patient-based tumor grading performance results of models according to BraTS 2020 datasets using an ensemble of deep features from three fine-tuned CNN models.

Classifier Model	ACC	AUC	PRE	REC	SPEC	F1
RF	0.699	0.719	0.694	0.962	0.212	0.806
KNN	0.748	0.661	0.737	0.952	0.371	0.831
SVM	0.738	0.829	0.729	0.950	0.346	0.825
NB	0.663	0.529	0.664	0.977	0.082	0.790
LR	0.744	0.842	0.734	0.949	0.363	0.828

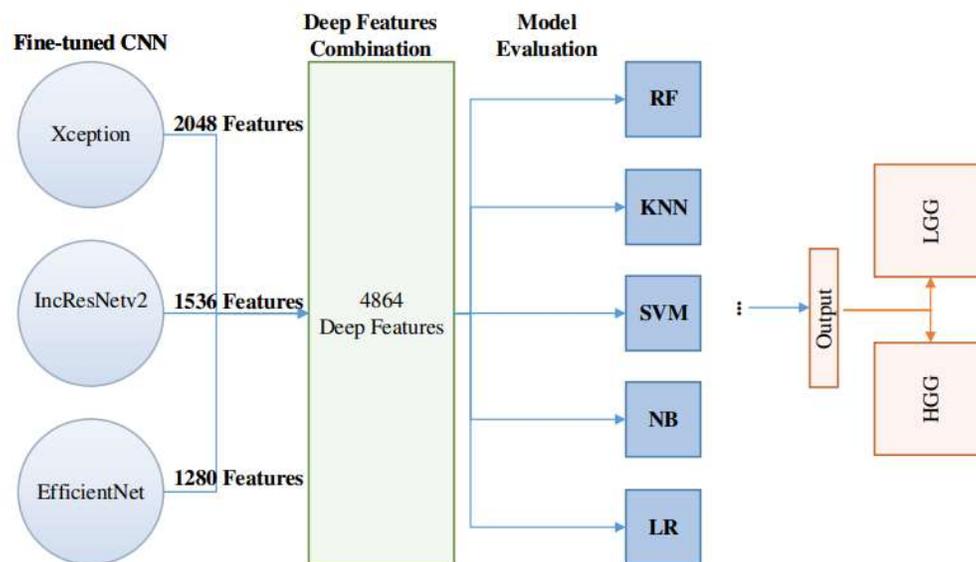


Figure 4: The overview of the ensemble of deep features on machine learning models for tumor grading.

5. Conclusion

This study introduces the variations of the machine, deep, and transfer learning approaches for the effective brain tumor (i.e., glioma) segmentation and grading. Proposed methods are based on 23 different combinations of deep feature sets and classification models on brain tumor grading. We also observed the extensive effects of these combination sets in terms of six different performance metrics.

From a clinical and biological standpoint, glioma grading is a crucial aspect of clinical decision making and treatment selection. From a technical standpoint grading is difficult due to significant data volumes and tumor and data heterogeneity, compounded by GPU and memory limitations. Developing effective CADe and CADx applications will play an essential role in selecting optimal treatment options, reducing biopsy, and providing suitable guidance to expert radiologists and clinicians. Future directions of this research include examining additional MRI sequence types with other data types and constructing user-defined datasets

based on large-scale data sets allowing for the proposed methods in this study to be efficiently tested and validated on additional datasets.

Funding

Funding was provided in part by NCI NIH intramural program (ZID BC 010990).

References

1. DeAngelis LM. Brain tumors. *N Engl J Med*. 2001;344:114-23.
2. Tazin T, Sarker S, Gupta P, et al. A robust and novel approach for brain tumor classification using convolutional neural network. *N Engl J Med*. 2021;2021:1-11.
3. <https://www.cancer.net/cancer-types/brain-tumor/statistics>
4. <https://about-cancer.cancerresearchuk.org/about-cancer/brain-tumours/types/glioma-adults>
5. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-Oncol*. 2021;23:1231-51.
6. Goodenberger ML, Jenkins RB. Genetics of adult glioma. *Cancer Genet*. 2012;205:613-21.
7. Zhuge Y, Ning H, Mathen P, et al. Automated glioma grading on conventional MRI images using deep convolutional neural networks. *Med Phys*. 2020;47:3044-53.
8. Zhao R, Krauze A. Survival prediction in gliomas: current state and novel approaches. In: Debinski W (eds) *Gliomas*. Exon Publications, Brisbane, Australia. 2021;pp. 151-69.
9. Jalan HA, Hasan AM. Magnetic resonance imaging segmentation techniques of brain tumors: a review. *Arch Neurosci*. 2019;6.
10. <https://www.fda.gov/radiation-emitting-products/mri-magnetic-resonance-imaging/benefits-and-risks>
11. Tasci E. Voting combinations-based ensemble of fine-tuned convolutional neural networks for food image recognition. *Multimed Tools Appl*. 2020;79: 30397-418.
12. Menze BH, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans Med Imaging*. 2014;34:1993-2024.
13. Rao CS, Karunakara K. A comprehensive review on brain tumor segmentation and classification of MRI images. *Multimed Tools Appl*. 2021;80: 17611-43.
14. Wijethilake N, Meedeniya D, Chittaranjan C. et al. Glioma survival analysis empowered with data engineering – a survey. *IEEE Access*. 2021;9:43168-91.

15. Tasci E, Ugur A. Image classification using ensemble algorithms with deep learning and handcrafted features. 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey. 2018.
16. Tasci E, Uluturk C, Ugur A. A voting-based ensemble deep learning method focusing on image augmentation and preprocessing variations for tuberculosis detection. *Neural Comput Appl.* 2021;33:15541-55.
17. Tasci E, Ugur A. Shape and texture-based novel features for automated juxtapleural nodule detection in lung CTs. *J Med Syst.* 2015;39:1-13.
18. Takacs P, Kovacs L, Manno-Kovacs A. A fusion of salient and convolutional features applying healthy templates for MRI brain tumor segmentation. *Multimed Tools Appl.* 2021;80:22533-50.
19. Al-qazzaz S, Sun X, Yang H, et al. Image classification-based brain tumour tissue segmentation. *Multimed Tools Appl.* 2021;80:993-1008.
20. Zhao X, Wu Y, Song G, et al. A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Med Image Anal.* 2018;43:98-111.
21. Zhou Z, He Z, Jia Y. Afpnet: A 3d fully convolutional neural network with atrous-convolution feature pyramid for brain tumor segmentation via MRI images. *Neurocomputing.* 2020;402:235-44.
22. Saouli R, Akil M, Kachori R, et al. Fully automatic brain tumor segmentation using end-to-end incremental deep neural networks in MRI images. *Comput Methods Programs Biomed.* 2018;166:39-49.
23. Amin J, Sharif M, Gul N, et al. Brain tumor detection by using stacked autoencoders in deep learning. *J Med Syst.* 2020;44:1-12.
24. Amin J, Sharif M, Yasmin M, et al. A new approach for brain tumor segmentation and classification based on score level fusion using transfer learning. *J Med Syst.* 2019;43:1-16.
25. Deng W, Shi Q, Luo K, et al. Brain tumor segmentation based on improved convolutional neural network in combination with non-quantifiable local texture feature. *J Med Syst.* 2019;43:1-9.
26. Latif G, Iskandar DA, Alghazo JM, et al. Enhanced MR image classification using hybrid statistical and wavelets features. *IEEE Access.* 2018;7:9634-44.
27. Sun J, Chen W, Peng S, et al. Dense residual refine networks for automatic brain tumor segmentation. *J Med Syst.* 2019;43:1-9.
28. <https://www.mathworks.com/help/images/segment-3d-brain-tumor-using-deep-learning.html>

29. <https://www.mathworks.com/help/vision/ref/unet3dlayers.html>
30. Cicek O, Abdulkadir A, Lienkamp SS, et al. 3D U-net: learning dense volumetric segmentation from sparse annotation. International Conference on Medical Image Computing and Computer- assisted Intervention, Springer, France. 2016.
31. Alzubaidi L, Zhang J, Humaidi AJ, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021. 8:1-74.
32. <https://www.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html>
33. Cayamcela MEM, Lim W. Fine-tuning a pre-trained convolutional neural network model to translate American sign language in real-time. 2019 International Conference on Computing, Networking and Communications (ICNC), Hawaii, USA. 2019.
34. Deniz E, Sengur A, Kadiroglu Z, et al. Transfer learning-based histopathologic image classification for breast cancer detection. Health Inf Sci Syst. 2018;6:1-7.
35. Breiman L. Random forests. Mach Learn. 2001;45:5-32.
36. Tasci E. A meta-ensemble classifier approach: Random rotation forest. Balkan Journal of Electrical and Computer Engineering. 2019;7:182-7.
37. Biau G. Analysis of a random forests model. J Mach Learn Res. 2012;13:1063-95.
38. Cristianini N, Ricci E. Support Vector Machines. In: Kao MY (eds) Encyclopedia of Algorithms. Springer, Boston. 2008;pp. 928 – 32.
39. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, Pennsylvania, USA. 1992.
40. Rish I. An empirical study of the naive Bayes classifier. IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence. 2001;3:41-6.
41. Feng J, Xu H, Mannor S, et al. Robust logistic regression and classification. Adv Neural Inf Process Syst. 2014;27:253-61.
42. Bakas S, Akbari H, Sotiras A, et al. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci Data. 2017;4:1-13.
43. Bakas S, Reyes M, Jakab A, Bauer S. et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint arXiv:1811.02629. 2018.
44. Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett. 2006;27:861-74.