ACTA INFOLOGICA
dergipark.org.tr/acin

İSTANBUL UNIVERSITY PRESS

DOI: 10.26650/acin.880918

RESEARCH ARTICLE

Performance Comparison of Different Pre-Trained Deep Learning Models in Classifying Brain MRI Images

Beyin MR Görüntülerini Sınıflandırmada Farklı Önceden Eğitilmiş Derin Öğrenme Modellerinin Performans Karşılaştırması

Onur Sevli[1]

ABSTRACT

A brain tumor is a collection of abnormal cells formed as a result of uncontrolled cell division. If tumors are not diagnosed in a timely and accurate manner, they can cause fatal consequences. One of the commonly used techniques to detect brain tumors is magnetic resonance imaging (MRI). MRI provides easy detection of abnormalities in the brain with its high resolution. MR images have traditionally been studied and interpreted by radiologists. However, with the development of technology, it becomes more difficult to interpret large amounts of data produced in reasonable periods. Therefore, the development of computerized semi-automatic or automatic methods has become an important research topic. Machine learning methods that can predict by learning from data are widely used in this field. However, the extraction of image features requires special engineering in the machine learning process. Deep learning, a sub-branch of machine learning, allows us to automatically discover the complex hierarchy in the data and eliminates the limitations of machine learning. Transfer learning is to transfer the knowledge of a pre-trained neural network to a similar model in case of limited training data or the goal of reducing the workload. In this study, the performance of the pre-trained Vgg-16, ResNet50, Inception v3 models in classifying 253 brain MR images were evaluated. The Vgg-16 model showed the highest success with 94.42% accuracy, 83.86% recall, 100% precision and 91.22% F1 score. This was followed by the ResNet50 model with an accuracy of 82.49%.The findings obtained in this study were compared with similar studies in the literature and it was found that it showed higher success than most studies.

Keywords: Brain MRI Classification, Transfer Learning, Convolutional Neural Networks

ÖZ

Beyin tümörleri, beyin hücrelerinin kontrolsüz bölünmeleri sonucu meydana gelen kitlelerdir. Tümörler zamanında ve doğru teşhis edilmezlerse ölümcül sonuçlara neden olabilir. Beyin tümörlerini tespit etmede yaygın olarak kullanılan tekniklerden biri olan MRI'dir. MRI, sağladığı yüksek çözünürlük ile beyindeki anormalliklerin kolay tespitine imkân verir. MR görüntüleri geleneksel olarak radyologlar tarafından incelenip yorumlanır. Ancak teknolojinin gelişmesi ile birlikte üretilen çok miktarda veriyi makul sürelerde yorumlamak daha zor hale gelmektedir. Bu nedenle bilgisayarlı yarı otomatik ya da otomatik yöntemlerin geliştirilmesi önemli bir araştırma konusu haline gelmiştir. Verilerden öğrenerek tahmin yapabilen makine öğrenmesi yöntemleri bu alanda yaygın olarak kullanılmaktadır. Ancak makine öğrenmesi için görüntü özelliklerinin çıkarımı özel bir mühendislik gerektirir. Makine öğrenmesinin bir alt dalı olan derin öğrenme, veri içerisindeki karmaşık hiyerarşiyi otomatik olarak keşfetmeye imkân sağlar ve makine öğrenmesinin sınırlılıklarını ortadan kaldırır. Transfer öğrenme ise eldeki eğitim verisinin az olması halinde ya da iş yükünü azaltmak için daha önceden eğitilmiş bir derin sinir ağının bilgilerinin benzer başka bir modele aktarılmasıdır. Bu çalışmada önceden eğitilmiş Vgg-16, ResNet50 ve Inception v3 modellerinin sınıflamadaki performansları değerlendirilmiştir. Vgg-16 modeli %94.42 doğruluk, %83.86 recall, %100 precision ve %91.22 F1 skoru ile en yüksek başarıyı göstermiştir. Bunu %82.49 doğrulukla ResNet50 modeli izlemektedir. Bu çalışmada elde edilen bulgular literatürdeki benzer çalışmalarla karşılaştırılmış ve çoğu çalışmadan daha yüksek başarı gösterdiği görülmüştür.

Anahtar kelimeler: Beyin MRI Sınıflama, Transfer Öğrenme, Evrişimsel Sinir Ağları

[1](Asst. Prof. Dr.), Mehmet Akif Ersoy University, Engineering and Architecture Faculty, Department of Computer Engineering, Burdur, Turkey

ORCID: O.S. 0000-0002-8933-8395

Corresponding author:
Onur SEVLİ
Mehmet Akif Ersoy University, Engineering and Architecture Faculty, Department of Computer Engineering, Burdur, Turkey
E-mail address: onursevli@mehmetakif.edu.tr

Submitted: 15.02.2021
Revision Requested: 02.04.2021
Last Revision Received: 24.04.2021
Accepted: 05.05.2021
Published Online: 29.06.2021

Citation: Sevli, O. (2021). Performance comparison of different pre-trained deep learning models in classifying brain MRI images. *Acta Infologica, 5*(1), 141-154. https://doi.org/10.26650/acin.880918

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International License

1. INTRODUCTION

The brain is one of the most complex organs with billions of cells. A brain tumor is an abnormal group of cells that results from uncontrolled cell division in the brain. If precautions are not taken in time, this abnormal group of cells leads to impaired brain activity and damage to healthy cells. Brain tumors, like all other types of cancer, are a serious problem that threaten human life and can be fatal if not diagnosed and treated in time. 24,000 new brain cancer cases were diagnosed in the United States in 2019 (Cancer.Net, 2020). Accurate and rapid diagnosis of brain tumors is important for the effective treatment of the disease. Computer Aided Diagnosis (CAD) assists neuro-oncologists in a variety of ways, and therefore CAD-based brain tumor detection is a popular research topic (Kumar, Dabas, and Godara 2017). Diagnostic systems using machine learning and deep learning are examples of CAD. CAD systems are based on magnetic resonance (MR) images.

Magnetic resonance imaging (MRI) is one of several techniques used in medical imaging and is widely used in imaging abnormal brain tissues. Among other techniques, MRI is the most popular and risk-free one. MRI provides higher contrast for soft tissues of the brain than CT images. MRI provides valuable information about the brain with high resolution. With the help of MRI, abnormalities in the brain can be detected easily.

Advances in the field of medical imaging make it possible to extensively investigate human organs using a large number of images produced. Traditionally, MR images are examined by radiologists, and abnormalities in the brain are detected. However, with advances in medical imaging techniques and hardware acceleration, manual interpretation of large amounts of data produced by MRI becomes a laborious task. For this reason, semi-automatic or automatic analysis of images with computer-aided systems has become an important research topic (Bauer et al. 2013). As the issue is directly related to human health, high accuracy and clarity are needed. At this point, computer-aided diagnostics help to obtain accurate results faster.

Brain tumor segmentation aims to detect the location and spread of tumors, active tumor tissue, and edema (Rajinikanth et al. 2017). Segmentation and quantitative lesion assessment are vital for treatment planning methods, disease monitoring, and control of disease progression. This is done by comparing abnormal tissues with normal ones. Brain tumor segmentation from MR images is of great importance for advanced diagnosis and treatment planning. Current automated and semi-automated brain tumor segmentation methods are generally categorized as either generative or discriminative model-based methods (Menze et al. 2014). Generative models require knowledge of probabilistic image atlases of both healthy and tumor tissues. Brain tumor segmentation based on probabilistic image atlases is an outlier detection problem (Prastawa et al. 2004). Discriminative models require less information about the anatomy of the brain and instead use the basic features of the image. These basic features are pixels, textures, local histograms (Kleesiek et al. 2014), region shape difference and symmetry analysis (Tustison et al. 2013). The discriminative model methods classify MR images as either tumors or normal tissue according to their characteristics, and the performance of the method depends on the image properties and the classification algorithms used. In the discriminative model, machine learning algorithms are used, which handle handcrafted features. Machine learning is a set of algorithms that allow computers to make predictions over big data. However, the extraction of features that enable classification in these techniques requires special expertise (Akkus et al. 2017).

Different MR image classification techniques have been used in the literature to address brain abnormalities. In these studies, the techniques which are typically used in the identification of normal and abnormal images are preprocessing, feature extraction, and classification steps. These studies include various methods of supervised machine learning such as wavelet transform (El-Dahshan, Hosny, and Salem 2010), Independent Component Analysis (ICA) (Moritz et al. 2000), Support Vector Machines (SVM) (Chaplot, Patnaik, and Jagannathan 2006), Random Forest Classifier (RFC) (Reza and Iftekharuddin 2014) and productive Gaussian Mixture Model (GMM) (Domingues et al. 2018). Recent studies use advanced machine learning techniques. Nayak et al. used Discrete Wavelet Transform (DWT) for feature extraction as well as the PCA for feature reduction and the AdaBoost algorithm with the RFC (Nayak, Dash, and Majhi 2016). In another study, the same researchers increased success by developing a model that uses the SVM with the AdaBoost (Ranjan Nayak, Dash, and Majhi 2017). Mohsen et al. used the properties obtained by the DWT and the PCA technique on segmented brain MR images to train the neural network used for the classification of brain tumors (Mohsen et al. 2018). Zhang et al. developed a brain MRI

classifier by combining the SVM and Particle Swarm Optimization (PSO) (Zhang et al. 2013). The same researchers used the wavelet entropy method instead of SVM in another study (Zhang et al. 2015). Wang et al. used PSO, artificial bee colony and feed-forward neural networks in their studies (Wang et al. 2015). Saritha et al. used probabilistic neural networks (PNN) with wavelet entropy method (Saritha, Joseph, and Mathew 2013).

In most automated brain tumor segmentation studies, the detection process is carried out with the help of machine learning classifiers using handcrafted features. However, traditional machine learning algorithms cannot make very good generalizations. A machine learning technique called deep learning which has attracted great attention in recent years eliminates the limitations of classical machine learning algorithms and enables automatic discovery of the features of MR images with its self-learning ability. In deep learning, features are automatically extracted by various feature processing layers, and feature engineering is less needed (LeCun, Bengio, and Hinton 2015). Deep learning techniques provide solutions to a wide range of problems.

Recent studies on computer-aided medical diagnosis show improved performances with the emergence of the concept of deep learning, and the success of deep learning in feature extraction is expanding the research field (Cao et al. 2018). Deep learning models were used in many biomedical applications such as diagnosis of lung cancer (Gu et al. 2018), breast cancer (Yousefi, Krzyżak, and Suen 2018), skin cancer (Zuo et al. 2017), diagnosis of pulmonary nodules (Cheng et al. 2016) and histopathological diagnosis (Litjens et al. 2016).

Recently, deep learning techniques have been used in the analysis of brain MRI images. Despite effortful studies on medical image analysis, automated detection of brain abnormalities remains an unsolved problem due to variations in brain morphology, imaging techniques and devices (Akkus et al. 2017). Studies mention that convolutional neural networks (CNN) perform better in deep learning-based brain tumor segmentation methods (Kamnitsas et al. 2017). The advantage of CNN-based systems is that they do not require manual segmentation and can perform fully automatic classification. In their study, Pashaei et al. automatically extracted the features from brain MR images using a CNN architecture and the model achieved 81% classification accuracy (Pashaei, Sajedi, and Jazayeri 2018). Charron et al. used a CNN-based model to track metastases (Charron et al. 2018). Havaei et al. proposed a model that simultaneously detects both local and global contextual features of brain tumors by trying a series of CNN architectures (Havaei et al. 2017). Agn et al. proposed a deep learning-based model that adjusts the parameters in brain tumor radiotherapy planning to minimize the risk for healthy tissues (Agn et al. 2019). Afshar et al. used a capsule neural network, a modified CNN architecture, to classify brain tumors, but the improvement in performance was not sufficient (Afshar, Mohammadi, and Plataniotis 2018).

Examining MR images requires large amounts of data processing. For this data processing task, using the graphics processing units (GPU) of computers provides higher performance than using the central processing unit (CPU). Another solution for increasing performance in deep learning is using transfer learning method. Transfer learning is simply a matter of transferring the knowledge of another pre-trained neural network to a new and similar model. Transfer learning has great potential for computer-aided detection of medical problems. Zhou et al. used a pre-trained Inception v3 model to detect kidney tumors (Zhou et al. 2019). Deniz et al. used VGG-16 and AlexNet model for breast cancer diagnosis and then they classified tumors using SVM (Deniz et al. 2018). Hussein et al. proposed a transfer learning-based model for lung and pancreatic tumor characterization (Hussein et al. 2019). Transfer learning also attracts attention in neurooncological studies. Kaur and Gandhi conducted a study that showed that transfer learning is more effective in brain tumor classification than traditional machine learning (Kaur and Gandhi 2020). Yang et al. used AlexNet and GoogLeNet in their study to determine the tumor grade with brain MR images (Yang et al. 2018). Jain et al. used a pre-trained VGG-16 network to diagnose Alzheimer's disease via MR images (Jain et al. 2019). Deepak and Ameer used pre-trained GoogLeNet to distinguish three different types of tumors via MR images (Deepak and Ameer 2019). Rehman et al. performed a study using GoogleNet, AlexNet and Vgg models for a similar task (Rehman et al. 2020). Chelghoum et al. conducted a study demonstrating the effectiveness of transfer learning with limited time and small number of epochs in brain tumor classification (Chelghoum et al. 2020).

In this study, three different pre-trained models were used to classify 253 brain MR images. The performances of the Vgg-16, ResNet50 and Inception V3 models were evaluated in terms of accuracy, recall, sensitivity and F1-score measurements. With a relatively limited dataset, the aim was to achieve high accuracy values in low epoch numbers, and a success of over 94%

was achieved as a result of 30 epochs. The results obtained in this study were compared with similar recent studies in the literature. It was revealed that the success achieved with the Vgg-16 model was higher than similar studies in the literature. In this study, an alternative and successful solution for tumor detection on brain MRI images was presented.

In the following sections, transfer learning models used in the study, dataset, data processing method, deep learning techniques used in the study, experiments and findings are presented .

2. MATERIAL AND METHOD

Most previous studies are based on manual extraction of tumor features before classification. This prevents them from being fully automated. Only a small number of studies show solutions being produced if the available data are relatively small.

In this study, a fully automated classification approach for brain MRI images was presented, using the transfer learning method in deep convolutional neural networks. A dataset consisting of 253 brain MRI images was used. Before the learning process in deep neural networks, the images were preprocessed. The basic steps of the process are as follows:

1. Preprocessing: Raw MRI images were cropped and the borders of the brain tissue were determined with the help of the OpenCV library.

2. Deep learning model implementations: Three different pre-trained models (Vgg-16, ResNet50 and Inception V3) were used.

3. Data augmentation: Data augmentation technique was carried out during the training to increase learning success and prevent overfitting.

4. Learning-rate optimization: The learning rates of the models were optimized using the adaptive momentum method.

5. The performances of the models were evaluated with accuracy, precision, recall, and F1 score metrics.

The workflow of the method used is given in Fig. 1.

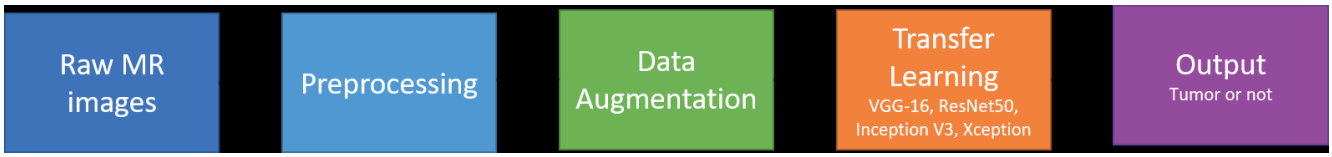


Figure 1. Flowchart of the study

2.1. Dataset

The dataset used in the study consists of 253 brain MRI images collected and shared publicly by Chakrabarty at the Jalpaiguri Government College of India (Chakrabarty, N. n.d.). 155 of these images were with tumor and 98 of them were without tumor. The graph showing the tumor and non-tumor image distributions within the dataset is given in Figure 2.

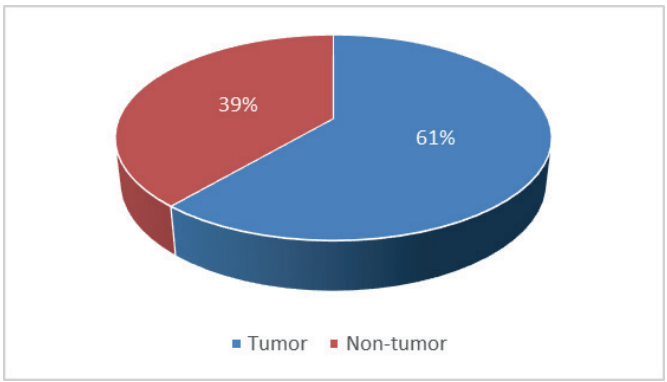


Figure 2. Image distributions within the dataset

The dataset is imbalanced in terms of the class distribution of the images. As a result of the dominance of a particular class in the dataset, the classifier becomes more prone to learning the dominant class. To eliminate this instability problem, the number of instances of the dominant class should be reduced or the instances of the less data class should be increased. Since the dataset used in this study was relatively limited, reducing the amount of data of the dominant class decreases the generalizability of the classification success. For this reason, data augmentation was applied to the class with a small number of samples.

Examples of MR images with tumors and without tumors in the dataset are given in Fig. 3. The dimensions of the images in the dataset and the black frames around them are different. For this reason, preprocesses were applied to standardize the images before the training.

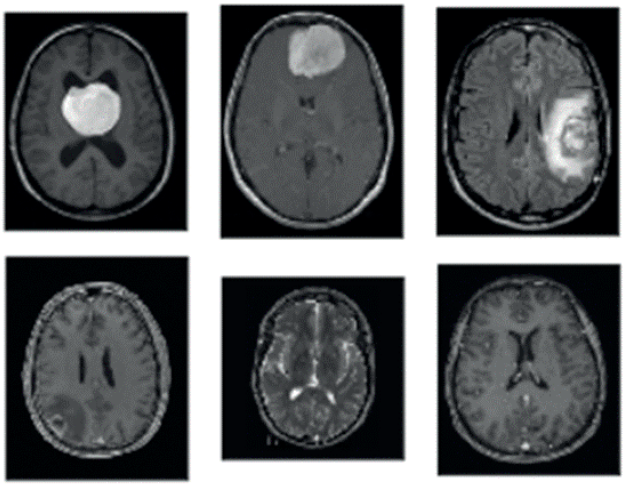


Figure 3. Sample images from the dataset

2.2. Image Preprocessing and Data Augmentation

During the preprocessing process, using the OpenCV library with Python, the boundaries of the brain tissue over the frame located around the image, namely, the endpoints in four separate poles, right, left, top, and bottom, were identified (Fig. 4).

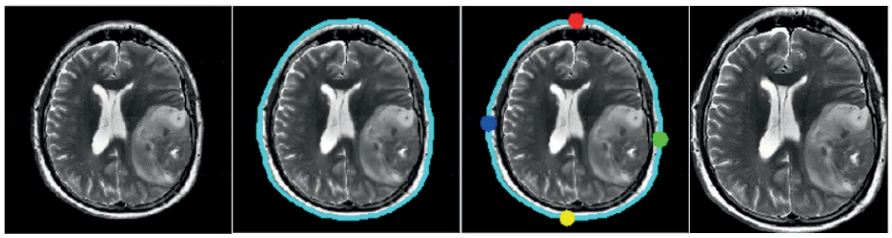


Figure 4. Preprocessing stages

The image was cropped according to the boundaries and polar points. In this way, the boundaries of the brain tissue were determined on the raw image and the part outside was discarded. This makes data processing easier. Images with different width and length values were resized to 224x224 pixel size after cropping.

Working with a relatively small amount of data and an imbalanced dataset leads to overfitting. The model tends to learn the dominant class and cannot generalize for other situations. Data augmentation technique is used to prevent this problem and to diversify the data. The data augmentation technique enables the generation of new data by applying processes such as zooming, reflecting, and rotating at varying rates on the existing dataset. The transformation methods and rates applied to the original data are obtained empirically, depending on the dataset and process. In this study, augmentation over the current dataset was done using 15% rotation, 5% shifting on the horizontal and vertical axis, 1/255 rescaling, and mirroring on the

horizontal and vertical axis. The rotation, shifting, and scaling ratios used for this study were also obtained experimentally in a way that maximizes the performance increase for the model.

Data augmentation was applied to reproduce tumor-free images with fewer samples to balance the dataset used. While the number of images with tumor remained the same, the number of images without tumor was increased by 50%. As a result of data augmentation, the number of images belonging to each class is given in Table 1.

Table 1

Number of images after data augmentation

Image class	Image count
Images with tumor	155
Images without tumor	147
Total	302

2.3. Deep Learning, CNN and Transfer Learning

Deep learning refers to the learning process performed with multiple layer neural networks to extract a feature hierarchy from raw input data. Contrary to manual feature extraction in classic machine learning, it is a popular technique that automatically extracts features from images. With the advances in hardware technology, large amounts of data can be processed quickly and high success and generalization can be achieved.

There are various deep learning techniques used for different purposes. CNN is a deep learning technique widely used in image classification and segmentation. CNN has also been widely used in medical image analysis to achieve better results (Belaid and Loudini 2020). CNN enables automatic extraction and definition of features through images. A typical CNN includes convolution, pooling, activation and classification layers. Images that are classified with specific tags are transferred to the CNN in pixels so that the trainable parameters in the network are optimized to increase the classification accuracy. The convolution layer applies a kernel onto the input pixels, thereby revealing the features. The pooling layer reduces the data by taking the largest or average of the data from previous layers. The fully connected layer enables the transition to the prediction phase. The general structure of a CNN architecture is shown in Fig. 5.

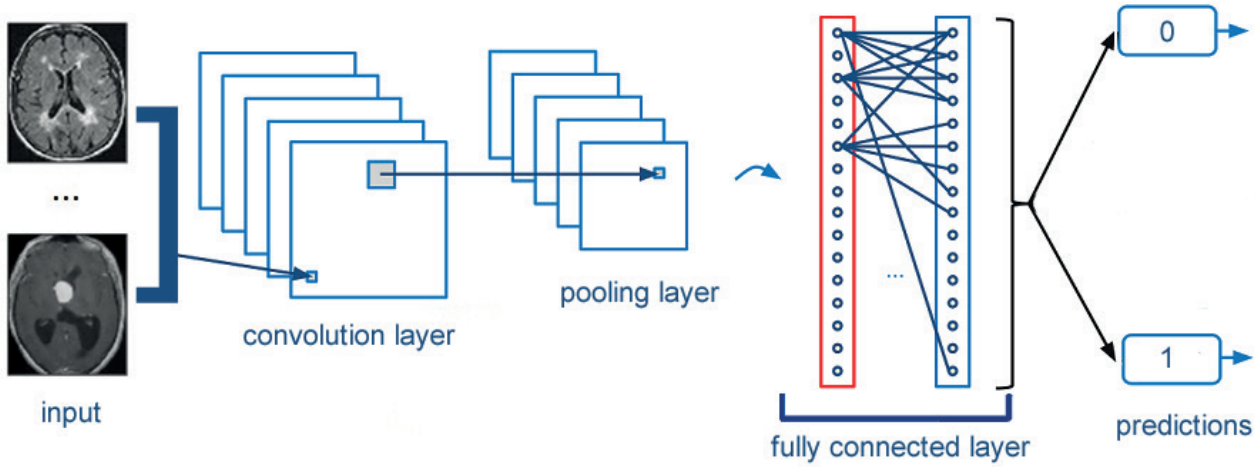


Figure 5. CNN architecture

Transfer learning is to transfer the weights of a network previously trained with large amounts of data to another model created to solve a similar problem. This method becomes important if there is not enough training data for the current problem. With a relatively small amount of data, an overfitting problem may occur in the network and successful generalizations cannot be made. If the data on which the pre-trained model was trained is comprehensive enough, the transferred network parameters ensure the correct classification of a small amount of data. The calculated weights of the pre-trained model are transferred to the new model and only the classifiers in the last part of the new model are trained.

Studies on medical image analysis also reveal that transfer learning is useful when a limited number of images are available (Deepak and Ameer 2020). In this study, a brain MRI classification task was performed using pre-trained VGG-16, ResNet50, Inception v3 transfer learning models, and then the success of the models was compared.

2.3.1. VGG-16 model

VGG network architecture is a CNN model introduced in 2014 by Simonyan and Zisserman (Simonyan and Zisserman 2020). VGG-16 is a special VGG type with 16 weighted layers. Its layers are convolution, maxpooling, activation, and fully connected layer. The general structure of the VGG-16 model is given in Fig. 6.



Figure 6. Structure of VGG-16 model

There are a total of 21 layers in the architecture: 13 convolution, 5 pooling, 3 dense layers. Only 16 of these are weighted layers. There are 64 filters in Convolution 1 layer, 128 filters in Convolution 2 layer, 256 filters in Convolution 3 layer, 512 filters in each of Convolution 4 and Convolution 5 layers.

The VGG-16 network was trained on the ImageNet dataset containing more than 14 million images and 1000 classes and achieved 92.7% accuracy.

2.3.2. ResNet-50 model

ResNet was introduced by He et al. in 2015 and the model became the winner of ImageNet competition with a 3.57% error rate (He et al. 2016). Unlike traditional sequential network architectures, ResNet has a structure based on microarchitecture modules. In theory, success is expected to increase as the number of layers in a model increases. However, increasing the number of parameters makes training and optimization difficult. During training, neurons that do not have high activations are ineffective in the deep neural networks, and in this case, residues appear in the network. ResNet is created by adding blocks that feed residual values to the next layers (Fig. 7).

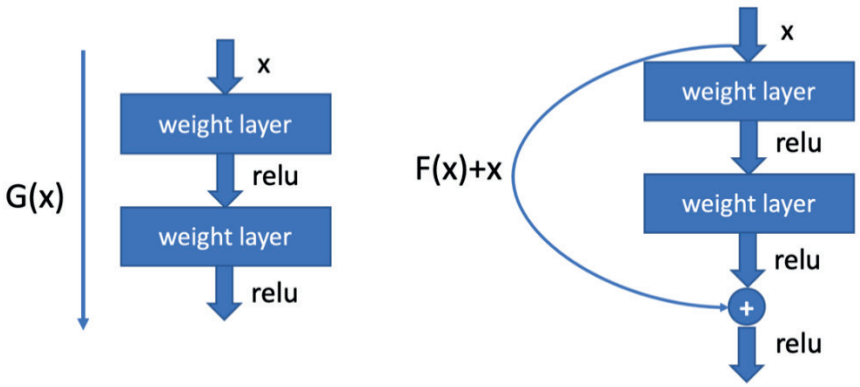


Figure 7. Typical CNN and ResNet

In a normal CNN, the model that runs sequentially from input to output is represented by a nonlinear G (x) function. In ResNet, by creating a shortcut from the input to the output, the input value (x) is added to the F (x) function arithmetically.

Then F (x) + x is passed through ReLU. With the aim of conveying the values in the past layers to the next layers more strongly, the input at the end of the 2nd layer in ResNet is added .

ResNet50 is a special version of the ResNet model with 50 weighted layers.

2.3.3. Inception v3 model

Inception networks constitute one of the important steps in the development of CNN architectures. There are 4 versions with differences in performance and accuracy. Inception v3 was designed by Szegedy et al. and the model was successful in ImageNet competition with a low error rate (Szegedy et al. 2016). The auxiliary classifiers used in the Inception architecture did not contribute until the end of the training process in the second version. Inception v3 was designed with minor changes in this version. Inception v3 consists of 42 layers. In addition to the previous versions, there are label smoothing, normalization for auxiliary classifiers, 7x7 convolution, and RMSProp optimizer. The structure of the Inception v3 model is shown in Fig. 8.

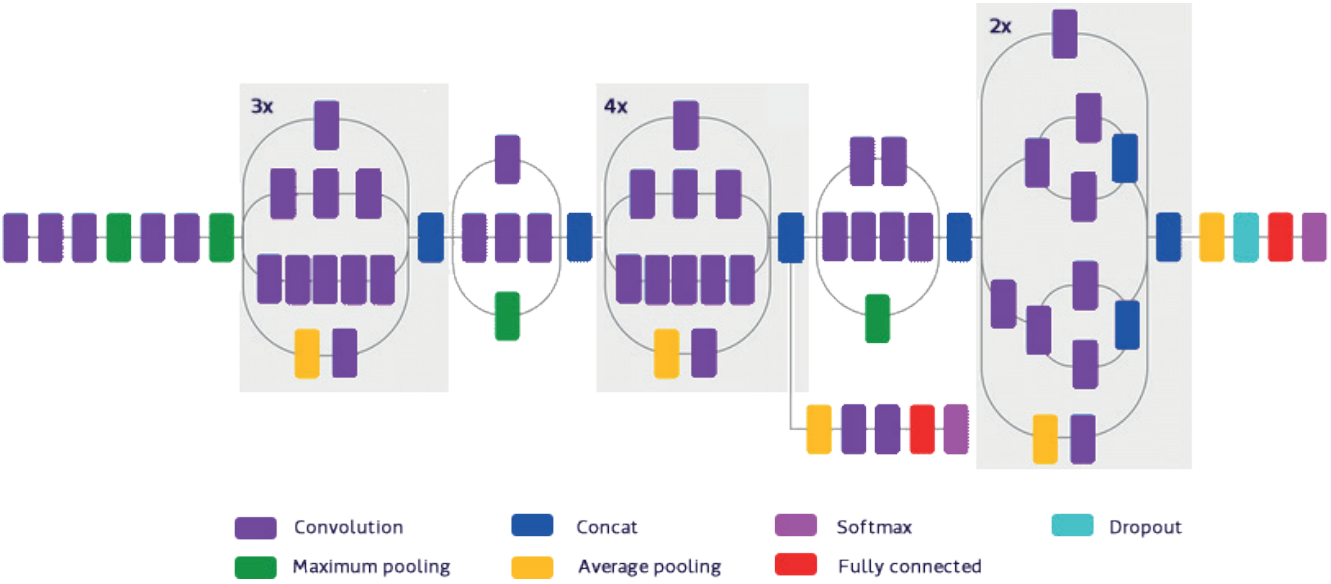


Figure 8. Structure of Inception v3 model (Agravat and Raval 2018)

3. EXPERIMENTAL STUDY

In this study, a fully automated classification approach for brain tumors was presented. First, preprocessing was performed on the raw MRI images, the boundaries of the brain tissue were determined automatically, and the images were cropped. Then, the dataset was expanded by applying the data augmentation method. The transfer learning method was used to reduce the processing load and achieve successful results with a small amount of data.

3.1. Performance Metrics

There are several metrics used to evaluate a classifier. Accuracy is the most widely used metric. Accuracy in classification is defined as the ratio of the number of correctly classified samples to the total number of data (Eq. 1).

$$accuracy = \frac{number\ of\ correctly\ classified\ samples}{total\ number\ of\ samples} \quad (1)$$

Classification accuracy is an effective parameter to characterize performance if the test dataset contains an equal number of samples from each class, i.e., it is balanced. However, when working with an unbalanced dataset, the system should be evaluated with more performance metrics. For this purpose, confusion matrices are used. A confusion matrix gives true and false classifications as a table. A confusion matrix contains values for four possible outcomes: When the original positive data is correctly classified as positive it is called true positive (TP), when the original positive data is classified incorrectly

as negative it is called false negative (FN), when the original negative data is classified correctly as negative it is called true negative (TN), when the original negative data is classified incorrectly as positive it is called false negative (FP). With these values obtained from the confusion matrix, different metrics are calculated to express the performance of the classifier.

Recall (sensitivity) which is calculated from these results is expressed in Eq. 2. Recall indicates the ability of the classification to detect true positives.

$$recall = \frac{TP}{TP + FN} \tag{2}$$

Precision is calculated as shown in Eq. 3. Precision indicates the ability of the classification to eliminate false positives.

$$precision = \frac{TP}{TP + FP} \tag{3}$$

The F1 score used to express the balance between recall and precision is the harmonic mean of these two metrics. The F1 score is calculated as expressed in Eq. 4.

$$F1\ score = 2 * \frac{precision * recall}{precision + recall} \tag{4}$$

For each model used in this study, recall, precision, and F1 score values were evaluated along with the accuracy parameter.

3.2. Findings and Discussion

In this study, the dataset consisting of 253 brain MR images was classified using three different pre-trained models. All other external parameters such as learning rate optimization, batch size, and number of epochs were the same for each model applied. The dataset was divided into 80% training and 20% test sets. 20% of the test set was used as a validation set.

The accuracy and the loss graphs on epoch basis were given for each model. Besides, the performance of each model was compared in terms of the specified metrics.

VGG-16 model: The accuracy and loss graphs obtained as a result of the training process are given in Fig. 9.

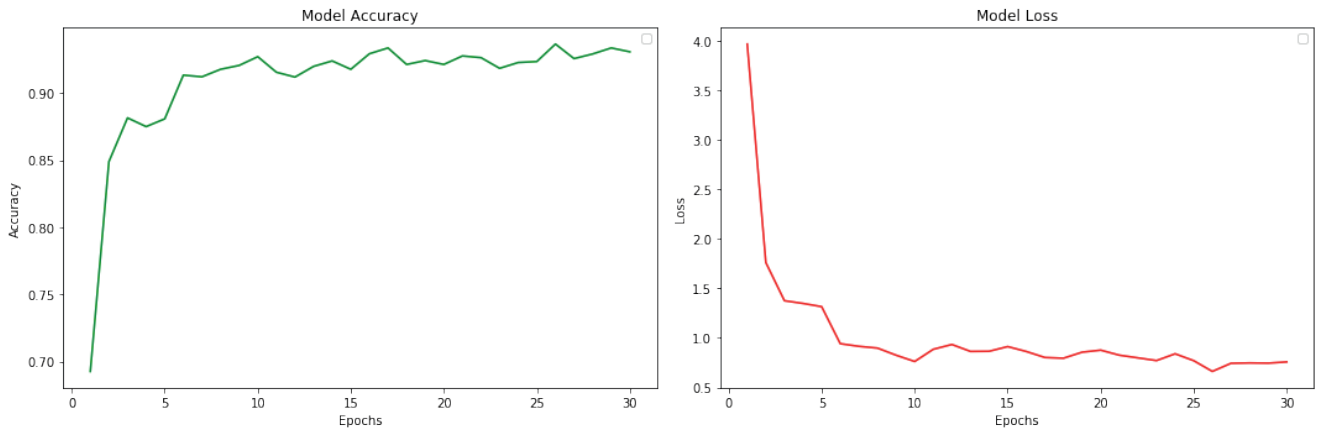


Figure 9. Accuracy and loss graphs of the VGG-16 model

The accuracy obtained at the end of the 30 epochs is 94.42%. Recall is 83.86%, precision is 100%, F1 score is 91.22%.

ResNet50 model: The accuracy and loss graphs obtained as a result of the training process are given in Fig. 10.

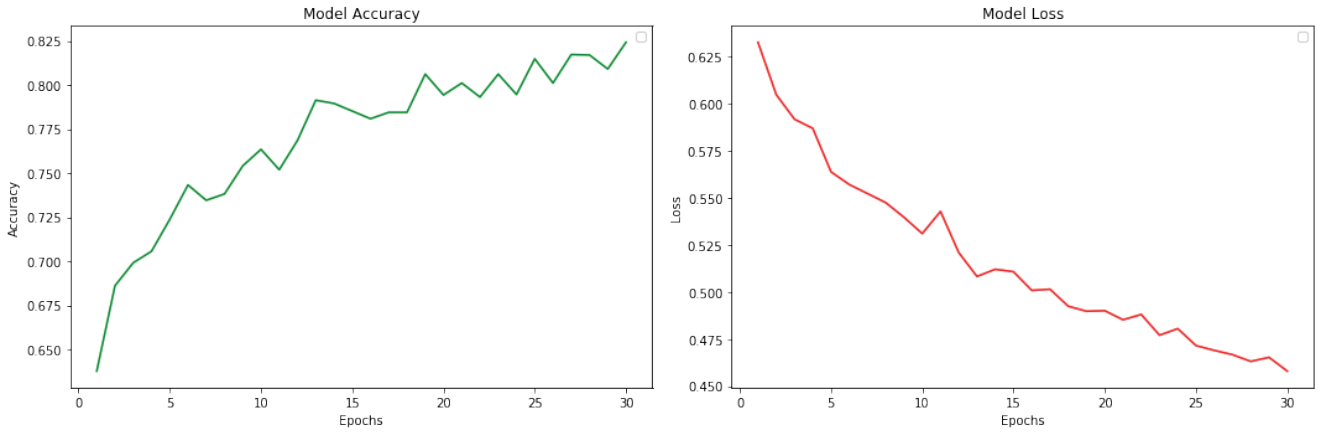


Figure 10. Accuracy and loss graphs of the ResNet50 model

The accuracy obtained at the end of the 30 epochs is 82.49%. Recall is 82.94%, precision is 100%, F1 score is 90.67%.

Inception v3 model: The accuracy and loss graphs obtained as a result of the training process are given in Fig. 11.

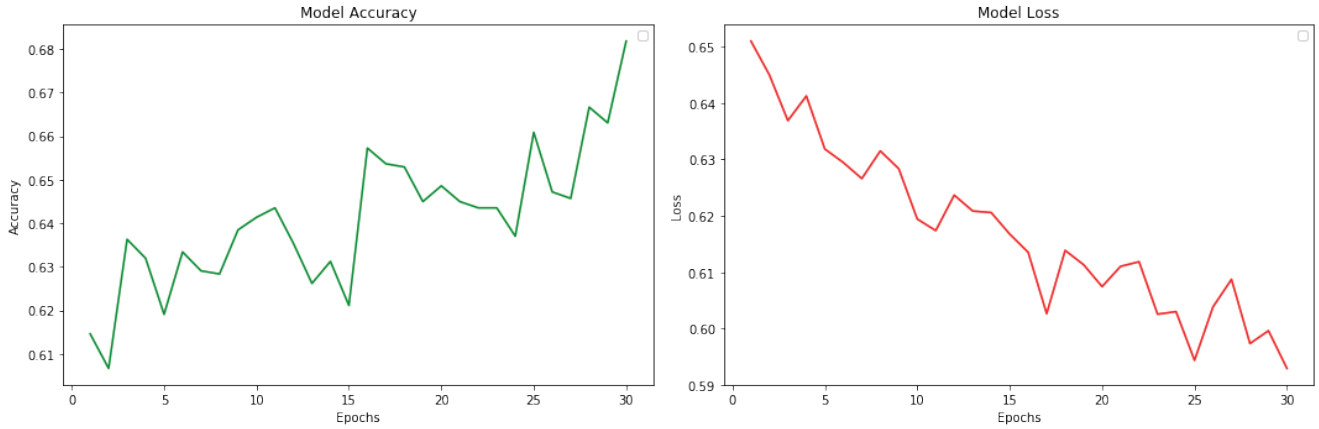


Figure 11. Accuracy and loss graphs of the Inception V3 model

The accuracy obtained at the end of the 30 epochs is 68.13%. Recall is 50.48%, precision is 40.12%, F1 score is 44.70%.

The performance results of the three pre-trained models in the same number of epochs are given in Table 2.

Table 2

Performance measurements

Model	Accuracy	Recall	Precision	F1 Score
Vgg-16	94.42%	83.86%	100%	91.22%
ResNet50	82.49%	82.94%	100%	90.67%
Inception V3	68.13%	50.48%	40.12%	44.70%

When the models are compared in terms of metric values, it is seen that the most successful model on the dataset used is Vgg-16 with an accuracy of 94.42%. The second successful model is ResNet50 with 82.49% accuracy. The Inception V3 model showed significantly lower success compared to these two models. The Vgg-16 and ResNet50 models are similar in terms of their ability to detect the right positives (recall) and to eliminate false positives (precision). However, Vgg-16 model was superior in terms of all metrics.

Since the dataset used in this study was imbalanced, we aimed to increase the success using the data augmentation technique. In order to reveal the effect of data increment on prediction success, each model was re-run without applying data augmentation,

keeping other parameters the same. The accuracy values obtained with and without data augmentation are comparatively given in Table 3.

Table 3

Model accuracies with and without data augmentation

Model	Without augmentation	With augmentation
Vgg-16	85.92%	94.42%
ResNet50	72.59%	82.49%
Inception V3	53.14%	68.13%

According to the measurements obtained, it was observed that the application of data augmentation brought an increase of up to 22% in the success of the model.

Most studies in the literature on the classification of brain MR images used machine learning techniques with handcrafted features. The manual extraction of features is a burden and increases the error rate. Deep learning enables automatic discovery of the features of MR images with its self-learning ability. A comparison of the findings obtained in this study with similar recent studies that classify brain MR images using deep learning methods in the literature is given in Table 4.

Table 4

Comparison of the results obtained in this study with the literature

Reference	Dataset	Method	The highest accuracy (%)
Shahamat and Abadeh 2020	140 MRI images	Custom CNN	70
Zhang et al. 2020	361 MRI images	R-CNN	79
Naser and Deen 2020	110 MRI images	CNN-UNet	89
Zhou et al. 2019	192 MRI images	InceptionV3	69
Saxena et al 2020	3064 MRI images	Vgg-16	90
Mlynarski et al. 2019	285 MRI images	CNN-UNet	80.92
· Afshar et al. 2019	3064 MRI images	CapsNet	90.89
· Abiwinanda et al. 2018	3064 MRI images	Custom CNN	84.19
Pashai et al. 2018	3064 MRI images	Custom CNN	81.09
Yang et al. 2018	113 MRI images	GoogLeNet	86.7
This study	253 MRI images	Vgg-16	94.42

Table 4 shows that the accuracy of the studies performed using various CNN models on different sizes of datasets in the literature varies between 70% and 90%. This study demonstrates that the best result of 94.42% obtained by using the Vgg-16 model was higher than similar studies in the literature and Saxena et al.'s study using the same model. Inception V3 was the least successful model of this study with 68.13% accuracy. In the literature, results obtained in other studies using the Inception V3 model in the classification of brain MR images were also between 55% and 69% (Saxena et al. 2020) (Zhou et al. 2019).

4. CONCLUSION

Due to the high image resolution it provides, MRI helps detect abnormalities in the brain with greater ease . Traditionally, MR images are interpreted by radiologists and tumors are diagnosed. However, with the advances in medical imaging techniques, it has become difficult to interpret manually large amounts of data produced in reasonable periods. As deep learning can effectively identify complex relationships, it has begun to attract great attention in the field of medical image analysis. Recently, deep learning has been used in the analysis of brain MRI images, but automatic detection of brain tumors is still a problem waiting to be solved due to variations in the morphological structure of the brain and differences in imaging technology. CNNs are commonly used in the deep learning-based brain tumor segmentation problems. One of the methods used to increase data processing performance is transfer learning. With the transfer of the learned parameters, high success is achieved while reducing the workload of the new model. Besides, transfer learning ensures successful results in case of low training data.

In this study, classifications were made using pre-trained Vgg-16, ResNet50, and Inception V3 models on a dataset consisting of 253 brain MRI images with and without tumors. Raw MR images were preprocessed to reduce the workload in the training

process. The boundaries of the brain tissue were determined over the raw image and the image was cropped and resized. Data augmentation was applied to balance the dataset class distributions. The classification process with three different models was evaluated with accuracy, recall, precision, and F1-score metrics. Vgg-16 model showed the highest success with 94.42% accuracy, 83.86% recall, 100% precision and 91.22% F1 score. This was followed by the ResNet50 model with an accuracy of 82.49%. Inception V3 model showed the lowest success in this study as in similar studies in the literature. Findings obtained in this study were compared with recent similar studies in the literature. The Vgg-16, which was the most successful model in this study, was shown to be more successful than the peer studies in the literature. In this study, it was shown that the transfer learning method provides successful results with a limited amount of data and fewer epochs. An alternative support system solution was presented for experts that will facilitate tumor detection from brain MR images.

Hakem Değerlendirmesi: Dış bağımsız.
Çıkar Çatışması: Yazar çıkar çatışması bildirmemiştir.
Finansal Destek: Yazar bu çalışma için finansal destek almadığını beyan etmiştir.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The author has no conflict of interest to declare.
Grant Support: The author declared that this study has received no financial support.

References/Kaynaklar

Abiwinanda, Nyoman, Muhammad Hanif, S. Tafwida Hesaputra, Astri Handayani, and Tati Rajab Mengko. 2019. "Brain Tumor Classification Using Convolutional Neural Network." Pp. 183–89 in World congress on medical physics and biomedical engineering 2018. Springer.

Afshar, Parnian, Arash Mohammadi, and Konstantinos N. Plataniotis. 2018. "Brain Tumor Type Classification via Capsule Networks." Pp. 3129–3133 in 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE.

Agn, Mikael, Per Munck af Rosenschöld, Oula Puonti, Michael J. Lundemann, Laura Mancini, Anastasia Papadaki, Steffi Thust, John Ashburner, Ian Law, and Koen Van Leemput. 2019. "A Modality-Adaptive Method for Segmenting Brain Tumors and Organs-at-Risk in Radiation Therapy Planning." Medical Image Analysis 54:220–237.

Agravat, Rupal R., and Mehul S. Raval. 2018. "Deep Learning for Automated Brain Tumor Segmentation in MRI Images." Pp. 183–201 in Soft Computing Based Medical Image Analysis, edited by N. Dey, A. S. Ashour, F. Shi, and V. E. Balas. Academic Press.

Akkus, Zeynettin, Alfiia Galimzianova, Assaf Hoogi, Daniel L. Rubin, and Bradley J. Erickson. 2017. "Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions." Journal of Digital Imaging 30(4):449–59.

Bauer, Stefan, Roland Wiest, Lutz-P. Nolte, and Mauricio Reyes. 2013. "A Survey of MRI-Based Medical Image Analysis for Brain Tumor Studies." Physics in Medicine & Biology 58(13):R97.

Belaid, Ouiza Nait, and Malik Loudini. 2020. "Classification of Brain Tumor by Combination of Pre-Trained VGG16 CNN." Journal of Information Technology Management 12(2):13–25.

Cancer.Net, "Brain Tumor - Statistics." Retrieved October 28, 2020 (https://www.cancer.net/cancer-types/brain-tumor/statistics).

Cao, Chensi, Feng Liu, Hai Tan, Deshou Song, Wenjie Shu, Weizhong Li, Yiming Zhou, Xiaochen Bo, and Zhi Xie. 2018. "Deep Learning and Its Applications in Biomedicine." Genomics, Proteomics & Bioinformatics 16(1):17–32.

Chakrabarty, N. n.d. "Brain MRI Images for Brain Tumor Detection." Kaggle. Retrieved May 28, 2020 (https://kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection).

Chaplot, Sandeep, Lalit M. Patnaik, and N. R. Jagannathan. 2006. "Classification of Magnetic Resonance Brain Images Using Wavelets as Input to Support Vector Machine and Neural Network." Biomedical Signal Processing and Control 1(1):86–92.

Charron, Odelin, Alex Lallement, Delphine Jarnet, Vincent Noblet, Jean-Baptiste Clavier, and Philippe Meyer. 2018. "Automatic Detection and Segmentation of Brain Metastases on Multimodal MR Images with a Deep Convolutional Neural Network." Computers in Biology and Medicine 95:43–54.

Chelghoum, Rayene, Ameur Ikhlef, Amina Hameurlaine, and Sabir Jacquir. 2020. "Transfer Learning Using Convolutional Neural Network Architectures for Brain Tumor Classification from MRI Images." Pp. 189–200 in Artificial Intelligence Applications and Innovations, edited by I. Maglogiannis, L. Iliadis, and E. Pimenidis. Cham: Springer International Publishing.

Cheng, Jie-Zhi, Dong Ni, Yi-Hong Chou, Jing Qin, Chui-Mei Tiu, Yeun-Chung Chang, Chiun-Sheng Huang, Dinggang Shen, and Chung-Ming Chen. 2016. "Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans." Scientific Reports 6(1):1–13.

Deepak, S., and P. M. Ameer. 2019. "Brain Tumor Classification Using Deep CNN Features via Transfer Learning." Computers in Biology and Medicine 111:103345. doi: 10.1016/j.compbiomed.2019.103345.

Deepak, S., and P. M. Ameer. 2020. "Automated Categorization of Brain Tumor from MRI Using CNN Features and SVM." Journal of Ambient Intelligence and Humanized Computing. doi: 10.1007/s12652-020-02568-w.

Deniz, Erkan, Abdulkadir Şengür, Zehra Kadiroğlu, Yanhui Guo, Varun Bajaj, and Ümit Budak. 2018. "Transfer Learning Based Histopathologic Image Classification for Breast Cancer Detection." Health Information Science and Systems 6(1):18.

Domingues, Rémi, Maurizio Filippone, Pietro Michiardi, and Jihane Zouaoui. 2018. “A Comparative Evaluation of Outlier Detection Algorithms: Experiments and Analyses.” Pattern Recognition 74:406–21.

El-Dahshan, El-Sayed Ahmed, Tamer Hosny, and Abdel-Badeeh M. Salem. 2010. “Hybrid Intelligent Techniques for MRI Brain Images Classification.” Digital Signal Processing 20(2):433–41.

Gu, Yu, Xiaoqi Lu, Lidong Yang, Baohua Zhang, Dahua Yu, Ying Zhao, Lixin Gao, Liang Wu, and Tao Zhou. 2018. “Automatic Lung Nodule Detection Using a 3D Deep Convolutional Neural Network Combined with a Multi-Scale Prediction Strategy in Chest CTs.” Computers in Biology and Medicine 103:220–231.

Havaei, Mohammad, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. 2017. “Brain Tumor Segmentation with Deep Neural Networks.” Medical Image Analysis 35:18–31.

He, K., X. Zhang, S. Ren, and J. Sun. 2016. “Deep Residual Learning for Image Recognition.” Pp. 770–78 in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Hussein, Sarfaraz, Pujan Kandel, Candice W. Bolan, Michael B. Wallace, and Ulas Bagci. 2019. “Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches.” IEEE Transactions on Medical Imaging 38(8):1777–1787.

Jain, Rachna, Nikita Jain, Akshay Aggarwal, and D. Jude Hemanth. 2019. “Convolutional Neural Network Based Alzheimer’s Disease Classification from Magnetic Resonance Brain Images.” Cognitive Systems Research 57:147–59. doi: 10.1016/j.cogsys.2018.12.015.

Kamnitsas, Konstantinos, Christian Ledig, Virginia FJ Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert, and Ben Glocker. 2017. “Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation.” Medical Image Analysis 36:61–78.

Kaur, Taranjit, and Tapan Kumar Gandhi. 2020. “Deep Convolutional Neural Networks with Transfer Learning for Automated Brain Image Classification.” Machine Vision and Applications 31:1–16.

Kleesiek, Jens, Armin Biller, Gregor Urban, U. Kothe, Martin Bendszus, and F. Hamprecht. 2014. “Ilastik for Multi-Modal Brain Tumor Segmentation.” Proceedings MICCAI BraTS (Brain Tumor Segmentation Challenge) 12–17.

Kumar, Sanjeev, Chetna Dabas, and Sunila Godara. 2017. “Classification of Brain MRI Tumor Images: A Hybrid Approach.” Procedia Computer Science 122:510–17.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. “Deep Learning.” Nature 521(7553):436–444.

Litjens, Geert, Clara I. Sánchez, Nadya Timofeeva, Meyke Hermsen, Iris Nagtegaal, Iringo Kovacs, Christina Hulsbergen-Van De Kaa, Peter Bult, Bram Van Ginneken, and Jeroen Van Der Laak. 2016. “Deep Learning as a Tool for Increased Accuracy and Efficiency of Histopathological Diagnosis.” Scientific Reports 6:26286.

Menze, Bjoern H., Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, and Roland Wiest. 2014. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS).” IEEE Transactions on Medical Imaging 34(10):1993–2024.

Mlynarski, Pawel, Hervé Delingette, Antonio Criminisi, and Nicholas Ayache. 2019. “Deep Learning with Mixed Supervision for Brain Tumor Segmentation.” Journal of Medical Imaging 6(3):034002.

Mohsen, Heba, El-Sayed A. El-Dahshan, El-Sayed M. El-Horbaty, and Abdel-Badeeh M. Salem. 2018. “Classification Using Deep Learning Neural Networks for Brain Tumors.” Future Computing and Informatics Journal 3(1):68–71.

Moritz, Chad H., Victor M. Haughton, Dietmar Cordes, Michelle Quigley, and M. Elizabeth Meyerand. 2000. “Whole-Brain Functional MR Imaging Activation from a Finger-Tapping Task Examined with Independent Component Analysis.” American Journal of Neuroradiology 21(9):1629–35.

Naser, Mohamed A., and M. Jamal Deen. 2020. “Brain Tumor Segmentation and Grading of Lower-Grade Glioma Using Deep Learning in MRI Images.” Computers in Biology and Medicine 121:103758.

Nayak, Deepak Ranjan, Ratnakar Dash, and Banshidhar Majhi. 2016. “Brain MR Image Classification Using Two-Dimensional Discrete Wavelet Transform and AdaBoost with Random Forests.” Neurocomputing 177:188–97.

Pashaei, Ali, Hedieh Sajedi, and Niloofar Jazayeri. 2018. “Brain Tumor Classification via Convolutional Neural Network and Extreme Learning Machines.” Pp. 314–319 in 2018 8th International conference on computer and knowledge engineering (ICCKE). IEEE.

Prastawa, Marcel, Elizabeth Bullitt, Sean Ho, and Guido Gerig. 2004. “A Brain Tumor Segmentation Framework Based on Outlier Detection.” Medical Image Analysis 8(3):275–83.

Rajinikanth, V., Suresh Chandra Satapathy, Steven Lawrence Fernandes, and S. Nachiappan. 2017. “Entropy Based Segmentation of Tumor from Brain MR Images–a Study with Teaching Learning Based Optimization.” Pattern Recognition Letters 94:87–95.

Ranjan Nayak, Deepak, Ratnakar Dash, and Banshidhar Majhi. 2017. “Stationary Wavelet Transform and Adaboost with SVM Based Pathological Brain Detection in MRI Scanning.” CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders) 16(2):137–49.

Rehman, Arshia, Saeeda Naz, Muhammad Imran Razzak, Faiza Akram, and Muhammad Imran. 2020. “A Deep Learning-Based Framework for Automatic Brain Tumors Classification Using Transfer Learning.” Circuits, Systems, and Signal Processing 39(2):757–75. doi: 10.1007/s00034-019-01246-3.

Reza, S., and K. M. Iftekharuddin. 2014. “Improved Brain Tumor Tissue Segmentation Using Texture Features.” Proceedings MICCAI BraTS (Brain Tumor Segmentation Challenge) 27–30.

Saritha, M., K. Paul Joseph, and Abraham T. Mathew. 2013. “Classification of MRI Brain Images Using Combined Wavelet Entropy Based Spider Web Plots and Probabilistic Neural Network.” Pattern Recognition Letters 34(16):2151–2156.

Saxena, Priyansh, Akshat Maheshwari, and Saumil Maheshwari. 2020. “Predictive Modeling of Brain Tumor: A Deep Learning Approach.” Pp. 275–85 in Innovations in Computational Intelligence and Computer Vision. Springer.

Shahamat, Hossein, and Mohammad Saniee Abadeh. 2020. “Brain MRI Analysis Using a Deep Learning Based Evolutionary Approach.” Neural Networks 126:218–34.

Simonyan, K., and A. Zisserman. 2020. “Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv 1409.1556 (09 2014).” URL Https://Arxiv. Org/Abs/1409.1556. Accessed: February.

Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. “Rethinking the Inception Architecture for Computer Vision.” Pp. 2818–2826 in Proceedings of the IEEE conference on computer vision and pattern recognition.

Tustison, Nick, Max Wintermark, Chris Durst, and Brian Avants. 2013. “Ants and Arboles.” Multimodal Brain Tumor Segmentation 47.

Wang, Shuihua, Yudong Zhang, Zhengchao Dong, Sidan Du, Genlin Ji, Jie Yan, Jiquan Yang, Qiong Wang, Chunmei Feng, and Preetha Phillips. 2015. “Feed-Forward Neural Network Optimized by Hybridization of PSO and ABC for Abnormal Brain Detection.” International Journal of Imaging Systems and Technology 25(2):153–164.

Yang, Yang, Lin-Feng Yan, Xin Zhang, Yu Han, Hai-Yan Nan, Yu-Chuan Hu, Bo Hu, Song-Lin Yan, Jin Zhang, Dong-Liang Cheng, Xiang-Wei Ge, Guang-Bin Cui, Di Zhao, and Wen Wang. 2018. “Glioma Grading on Conventional MR Images: A Deep Learning Study With Transfer Learning.” Frontiers in Neuroscience 12:804–804. doi: 10.3389/fnins.2018.00804.

Yousefi, Mina, Adam Krzyżak, and Ching Y. Suen. 2018. “Mass Detection in Digital Breast Tomosynthesis Data Using Convolutional Neural Networks and Multiple Instance Learning.” Computers in Biology and Medicine 96:283–293.

Zhang, Min, Geoffrey S. Young, Huai Chen, Jing Li, Lei Qin, J. Ricardo McFaline-Figueroa, David A. Reardon, Xinhua Cao, Xian Wu, and Xiaoyin Xu. 2020. “Deep-Learning Detection of Cancer Metastases to the Brain on MRI.” Journal of Magnetic Resonance Imaging 52(4):1227–36.

Zhang, Yu-Dong, Shuihua Wang, Zhengchao Dong, Preetha Phillip, Genlin Ji, and Jiquan Yang. 2015. “Pathological Brain Detection in Magnetic Resonance Imaging Scanning by Wavelet Entropy and Hybridization of Biogeography-Based Optimization and Particle Swarm Optimization.” Progress In Electromagnetics Research 152:41–58.

Zhang, Yudong, Shuihua Wang, Genlin Ji, and Zhengchao Dong. 2013. “An MR Brain Images Classifier System via Particle Swarm Optimization and Kernel Support Vector Machine.” The Scientific World Journal 2013.

Zhou, Leilei, Zuoheng Zhang, Yu-Chen Chen, Zhen-Yu Zhao, Xin-Dao Yin, and Hong-Bing Jiang. 2019. “A Deep Learning-Based Radiomics Model for Differentiating Benign and Malignant Renal Tumors.” Translational Oncology 12(2):292–300.

Zuo, Haiqiang, Heng Fan, Erik Blasch, and Haibin Ling. 2017. “Combining Convolutional and Recurrent Neural Networks for Human Skin Detection.” IEEE Signal Processing Letters 24(3):289–293.