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 Karşilaştırması

Onur Sevli

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

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

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.

![Flowchart of the study](image)

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.

![Image distributions within the dataset](image)
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.

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).

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

<table>
<thead>
<tr>
<th>Image class</th>
<th>Image count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Images with tumor</td>
<td>155</td>
</tr>
<tr>
<td>Images without tumor</td>
<td>147</td>
</tr>
<tr>
<td>Total</td>
<td>302</td>
</tr>
</tbody>
</table>

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.

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](image)

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](image)

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.

![Inception v3 model](image)

*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).

\[
\text{accuracy} = \frac{\text{number of correctly classified samples}}{\text{total number of samples}}
\]

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.

\[
recall = \frac{TP}{TP + FN} \quad (2)
\]

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

\[
precision = \frac{TP}{TP + FP} \quad (3)
\]

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.

\[
F1 \text{ score} = 2 \times \frac{precision \times recall}{precision + recall} \quad (4)
\]

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](image)

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](image)
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.

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>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Recall</th>
<th>Precision</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vgg-16</td>
<td>94.42%</td>
<td>83.86%</td>
<td>100%</td>
<td>91.22%</td>
</tr>
<tr>
<td>ResNet50</td>
<td>82.49%</td>
<td>82.94%</td>
<td>100%</td>
<td>90.67%</td>
</tr>
<tr>
<td>Inception V3</td>
<td>68.13%</td>
<td>50.48%</td>
<td>40.12%</td>
<td>44.70%</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Model</th>
<th>Without augmentation</th>
<th>With augmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vgg-16</td>
<td>85.92%</td>
<td>94.42%</td>
</tr>
<tr>
<td>ResNet50</td>
<td>72.59%</td>
<td>82.49%</td>
</tr>
<tr>
<td>Inception V3</td>
<td>53.14%</td>
<td>68.13%</td>
</tr>
</tbody>
</table>

Table 3
Model accuracies with and without data augmentation

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

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

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

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.

References/Kaynaklar


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


