Classification of Brain Tumors with CNN by using an Augmented Dataset

Elvan Karasu
Department of Computer Engineering
Bogazici University
Istanbul, Turkey, 34342
elvan.karasu@boun.edu.tr

Can Koban
Department of Computer Engineering
Bogazici University
Istanbul, Turkey, 34342
burakcan.koban@boun.edu.tr

Abstract

The aim of this study is to classify MRI images of brain tumors with convolutional neural networks (CNN) by using different data augmentation techniques. The effect of applying these techniques is investigated by increasing the size of the dataset gradually. Firstly, 6 basic methods focusing on small modifications like shifting, zooming, flipping, rotation, shearing, and changing the brightness are applied and the amount of data is increased by 24%. Then, random displacement is applied and the amount of data is increased by 30%. Lastly, generative adverserial networks (GANs) are trained for generating MRI images but the generated images are not combined with training dataset because of quality concerns. In total, the size of the training data set is increased from 2946 number of images to 4526. Accuracy of the model trained with the original data set is found to be 91%. After applying basic methods, it changed to 87%, and after adding a deformable method it remained unchanged as 91% compared to the original dataset. Since the expected increase in performance is not observed, the reasons behind this are analyzed.

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

1 Introduction

Medical imaging techniques are indispensable part of prevention, diagnosis and treatment in medicine. The use of deep learning in medical imaging especially in MRI and CT scans has risen significantly in recent years [1]. Convolutional Neural Networks (CNN) are highly researched subtopic of deep learning in medical imaging. The reasons behind this are the excellent proven success of CNNs in image classification and image segmentation which can be quite useful for diagnosis of many diseases. As an example, "Automated Detection of Vertebral Fractures [2]" can be given. In this article, spinal MRI data set comes from 1019 patients. They use YOLOv3 for detection of vertebra and an ensemble model of ResNet34, DenseNet121 and DenseNet201 for classification of fractures. The reported accuracy is higher than 90%.

Data access is very restricted for medical data. Although there are promising results which shows medical imaging analysis can be improved by using CNNs, limited amount of medical images is the biggest obstacle to train powerful models. Especially, privacy and data protection issues restricts sharing medical data. As a result of this, most of the researchers use either anonymous open data sets or local data sets that come from researcher's institution for deep learning applications in medical imaging [3]. Another reason of limited data is acquiring labeled and organized medical data can be expensive and time consuming since doctors and radiologists are busy and hourly wages of them are quite high [4].

Limited data contradicts with deep learning's biggest source of strength which is large amount of data. Therefore, increasing amount and quality of data is a big problem for use of deep learning in medical imaging. Since collecting more data is still a problem, improving and making the best use of the existing data is necessary and obligatory. For this purpose different techniques are applied such as transfer learning and data augmentation. In transfer learning, the model is pre-trained on big data set bu updating model weights. Then, trained model can be used for similar task on different datasets. Sometimes, it is required to fine tune the model on new dataset to get more accurate predictions. Another well suited method for this problem is data augmentation where the goal is creating new samples from the currently existing data by using various techniques.

Data augmentation is a technique where the data is enriched by adding modified versions of the original data. It is a commonly used regularization technique to overcome overfitting [5]. Mostly used techniques are summarized in Table 1 in three categories: basic methods, deformable augmentation methods and methods with deep learning. In basic methods, the main goal is mapping the points to different location in the image and playing with the intensity. In deformable augmentation techniques, the aim is adding some deformation to images without exceeding medically acceptable limits. One example is randomised displacement field where each pixel is relocated by a random value in the vertical and horizontal lines. Another example is spline interpolation where a polynomial function is used for transforming actual values to new values in the pixels of image. The most complex augmentation techniques are grouped under the name of methods with deep learning. Generative Adversarial Networks (GAN) can be given as an example for this category where a generator and discriminator compete against each other. The generator network learns to generate new images similar to the original images in the data set and discriminator tries to separate synthetic images from the real images. This competition increases performance of both discriminator and generator [4].

Considering all the motivations above, more specific research is conducted to touch one particular problem in medical imaging domain and to use the various data augmentation methods on a dataset of brain tumor MR images to show possible contributions of augmentation techniques in this domain.

2 Related Work

In this section, the state of the art is analyzed. Especially, brain tumor detection and classification by using CNN models are the main focuses. Mostly used data augmentation techniques in the medical imaging data sets are also given place to make the survey complete.

The study of Sajjad et al. has approached to multi-grade brain tumor classification by using CNN in a similar way aimed in this study. They firstly segmented and separated tumors from brain MR images to images of segmented tumors saving labels of tumor images. For segmentation, they have used another CNN that is pre-trained for this job called InputCascadeCNN [6]. After segmented tumors are attained, they have used 8 basic augmentation techniques, which are rotation, flip, Gaussian blur, sharpen, edges detection, emboss, skew, and shear. This approach drastically increased amount of data. Lastly, they gave the augmented data to a fine-tuned deep neural network called VGG-19 to classify the tumor [7]. They have reported an overall accuracy increase from 87.38% to 90.67% after data augmentation [8]. The study of Sajjad et al. can be discussed with another study in this domain which is published by Swati et al. They again have used VGG-19 based network on the very popular data set called CE-MRI. They have applied transfer learning and fine-tuning. They have reported a precision of 96.13% [9].

Another successful study was published by Deepak and Ameer in 2019. They have proposed to use transfer learning on GoogLeNet [10].They have modified the last three layers of GoogLeNet for adapting the deep network to target classification. The data set consists of 3064 brain MRI from 233 patients, and three labels for tumor types: glioma tumor, meningioma tumor, and pitutiary tumor. They used transfer learning and haven't used any data augmentation technique. The last layer of the modified GoogLeNet is a softmax layer for different categories. Classification accuracy is 92.3 % for this study [11]. The success of transfer learning can be observed from this study but there is still room for improvements.

In another work, Ayadi et. al. have created deep convolutional neural network with 10 convolutional layers, 5 pooling layers, 1 dense layer and a softmax layer for classification of brain tumor MRIs. With convolution layers, ReLU activation function is used. Their model classifies the tumors into three groups, which are meningioma, pituitary and glioma. Their data set consists of 2132 MRIs which are increased to 10660 by using rotation, flip, Gaussian blur, and sharpening which are basic data augmentation techniques. They have used 5-fold cross validation during training. Their accuracy for grade I, II, III and IV are increased from 95.61 to 96.32 %, 92.98 to 95.31 %, 93.85 to 96.18 %, and 98.24 to 99.61 %, respectively. Overall accuracy increased from 90.35 to 93.71 % which shows the success of data augmentation in this domain [12]. This paper motivated this work a lot.

Allah et. al used the same data set with Deepak and Ameer with a VGG-19 feature extactor and a progressive growing generative adversarial network (PGGAN). They reached an accuracy of 98.54%. [13] Pesteie et. al used independent conditional variational autoencoder (ICVAE) in order to generate synthetic data for two different data sets, one is for classifying spinal transverse ultrasound (US) images and the other one is for segmantation of brain tumors in MRI images. In classification, they used a residual model and obtained an accuracy of 83%. When they also used ICVAE, accuracy increased to 92%. In brain segmentation, they used U-net and using ICVAE increased their accuracy from 83% to 88%. [14]

The work of Bayoumi et al. in 2021 is another successful study [15]. As a summary, they have trained 4 different CNN models in two datasets. First one includes 349 and second one 120 brain MRI images. They have trained 5 different deep networks, namely AlexNet, VGG16, GooLeNet, Resnet50 and Inceptionv3. Higher than 95 % accuracy is attained. Another work by Trong et al. in 2021 again achieves higher than 90 % by using 4 different CNN architectures namely DenseNet201, ResNet152V2, MobileNetV3 and VGG19 [16]. These successful resulted studies and promising results of data augmentation lead us to conduct a study that is focused on basic to advanced data augmentation techniques for brain tumor MR images to contribute classification of brain tumor problem.

3 Method

The proposed method is applying different data augmentation techniques on brain MRI images. The motivation behind this is that medical images are not easy to reach. Therefore, reproducing and increasing the size of the data set is an important subject for medical imaging. There are not any papers using this data set for research. However, there are many notebooks on Kaggle to train deep neural networks and achieve good results for this data set [17]. The previous contributions on this data set do not take data augmentation as the main focus. Therefore, applying extensive data augmentation techniques on this dataset helps to improve the state-of-the-art in this domain.

Brain Tumor Classification (MRI) dataset on Kaggle is used. [17] The dataset contains 4 classes, which are meningioma tumor, pituitary tumor, glioma tumor, and no tumor. The tumors are classified according to their places on the brain. Meningioma tumor is the tumor in meninges, which are layers surrounding the brain and keeping it safe. [18] Pituitary tumor occurs in pituitary gland (hypophysis) which secretes many hormones important for human life. [19] [20] Glioma tumor occurs on glial cells which surround the neurons. [21] No tumor class is self explanatory, these are the MRI images of brains without any tumor. To classify tumors, CNN is trained and the details of the training process is given in the next section.

Table 1: Data augmentation methods to be applied

Category	Methods
Basic methods	Rotation, Horizontal and Vertical Shift, Horizontal Flipping, Zoom, Brightness, Shear
Deformable augmentation method	Random displacement
Method with deep learning	GAN

The steps of this study are summarized in Table 1. The methods shown here are applied by starting from the most basic one and continuing with more complex methods. Basic augmentation principles depend on mapping pixels of image to a different position or playing with intensity values. Basic augmentation techniques are summarized in Figure 1. In the figure, they are applied on images one by one and exaggerated in order to clarify. However, during experiments, the techniques are applied to images randomly and sometimes together. As it can be seen from Figure 1, shifting, zooming, rotating, shearing and brightness require some hyperparameters to decide on limits of augmentation. To decide on those hyperparameters, different hyperparameters are tried for tuning. In the figure, basic augmentations are a little exaggerated to show the effects better. In real experiments, it is mindfully monitored that created images need to be medically meaningful. After commpleting image generation, the images are saved to a separate file directory in order to combine with the original dataset later.

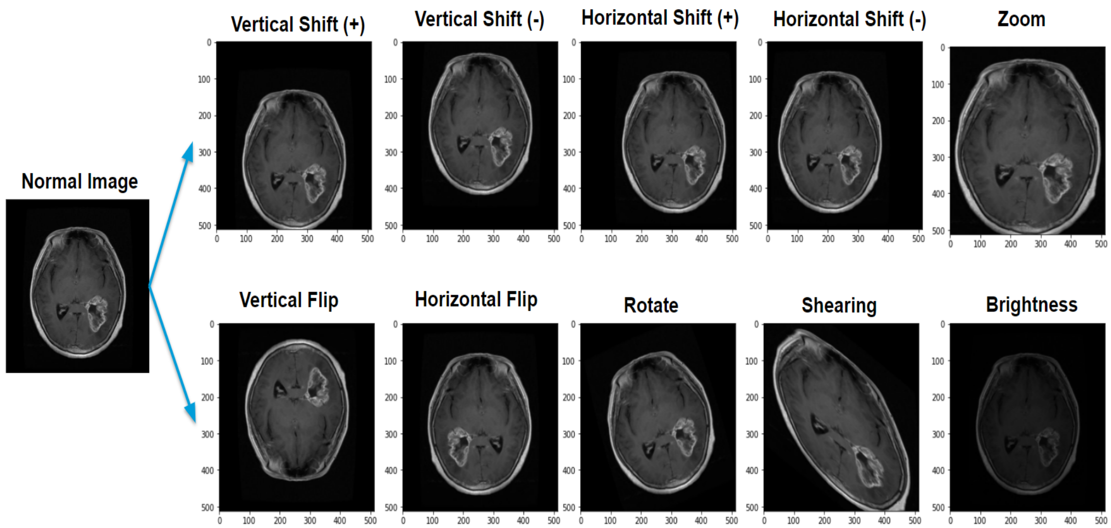
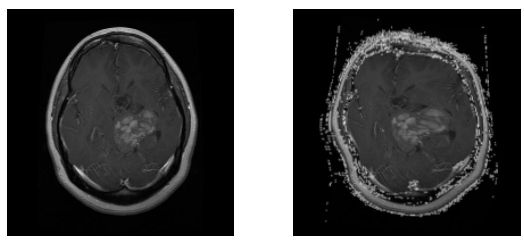


Figure 1: How basic augmentation techniques are affecting images



(a) Original Image

(b) Random displacement applied

Figure 2: How deformable augmentation techniques are affecting images

In Figure 2, on the other hand, random displacement technique is applied on MR images using elasticdeform library. Here, there are two hyperparameters that control the amount of displacement. They are named as sigma and points by its developers. [22] Points controls deformation intensity and sigma controls displacement field smoothing during augmenting image. They are tuned as same as in the basic methods. In Figure 2.b, it can be observed effect of sigma 6 and point 5. There are two reasons to create deformed brain tumor MR images. First, this level of distortion can occur in real life. Therefore these images are imitation of real life failures. Second, it is aimed to train a more robust model for various situations.

The last step of the augmentation is generating new images using GAN. The designed architecture and details of GAN are given in the next section

The last step of the proposed method is constructing a CNN to classify 4 different brain tumor images. Original dataset is concatenated with augmented dataset except the generated images by GAN. The results are discussed in depth through the following section.

4 Experiments

The dataset consists 3274 image files of four classes which are no tumor (500 images), glioma tumor (926 images), meningioma tumor (937 images), and pituary tumor (901 images). After shuffling the data, 10% (328 images) of the data is splitted for test and 10% is used for validation. The CNN architecture used is given in Table 2. As it can be seen from the table , softmax is applied as the last activation function to efficiently categorize 4 different labels. Categorical cross entropy is chosen as loss function which is defined in equation 1.

Table 2: CNN Architecture

Layer	**Parameters**	**Activation**
Convolution	Kernel Size = 5x5, Channels = 32, Padding = Same	ReLU
Pooling	Filter Size = 2x2	
Convolution	Kernel Size = 3x3, Channels = 32, Padding = Same	ReLU
Pooling	Filter Size = 2x2	
Batch Normalization		
Convolution	Kernel Size = 3x3, Channels = 32, Padding = Same	ReLU
Pooling	Filter Size = 2x2	
Convolution	Kernel Size = 3x3, Channels = 64, Padding = Same	ReLU
Pooling	Filter Size = 2x2	
Dense	Output Neurons = 128, Dropout rate = 0.5	ReLU
Dense	Output Neurons = 4	Softmax

$$Loss = \sum_{i=1}^{N} y_i \log \hat{y_i} \tag{1}$$

The experiments are done using Keras and TensorFlow.[1] The CNN described in Table 2 is trained with batches of size 32 and using Adam optimizer. After 8 epochs, accuracies of validation and test sets are 87% and 88%, respectively. The classification report for the test set is obtained using sklearn.metrics and it is shown in Table 3. [23]

Table 3: Classification performance on original and augmented datasets

Average metric	Original dataset	After basic augmentations	After deformable augmentations
Accuracy	0.91	0.87	0.91
Precision	0.91	0.86	0.91
Recall	0.91	0.88	0.92
F1 Score	0.91	0.86	0.91

Basic augmentation techniques are applied to full training data set randomly. To be precise, rotation range is 20^o, shift range for vertical and horizontal is 10%, zoom range is 15%, shear range is 6%,

[1]The implementation is available here. In order to run the code in Colab, you should add the shortcut of this folder to your Drive.

and brightness range is [0.1, 1.5], and horizontal flipping is applied. If there are empty remaining points in the image because of rotations or other transformations, these remaining parts are filled with constant colors so that they are similar to normal MR images where the color surrounding the brain is constant and dark. Training data is increased in different amounts for different classes. Considering that no tumor class has the least number of (500 images) training data, the number of no tumor images are increased by 50% in the training set. The number of images for other classes are increased by 20%. As a result, the number of training data points is increased from 2946 to 3643.

Random displacement is applied as a deformable method. The size of the original dataset is increased by 30% with these generated images. Firstly, the images generated using basic methods are added to the dataset. After 9 epochs, validation accuracy is 73% and test accuracy is 87%. Then, the images generated using random displacement are also added to the dataset. After 11 epochs, validation accuracy is 86% and test accuracy is 91%. The other metrics for these three datasets are given in Table 3.

Table 4: GAN Architecture

Generator		
Layer	**Parameters**	**Activation**
Dense	Output Neurons = 125000	Leaky ReLU with coefficient 0.2
Transpose Convolution	Kernel Size = 4x4, Channels = 100, Stride = 2x2, Padding = Same	Leaky ReLU with coefficient 0.2
Transpose Convolution	Kernel Size = 4x4, Channels = 50, Stride = 2x2, Padding = Same	Leaky ReLU with coefficient 0.2
Transpose Convolution	Kernel Size = 4x4, Channels = 1, Stride = 2x2, Padding = Same	Leaky ReLU with coefficient 0.2
Discriminator		
Layer	**Parameters**	**Activation**
Convolution	Kernel Size = 3x3, Channels = 64, Stride = 2x2, Padding = Same, Dropout rate = 0.4	Leaky ReLU with coefficient 0.2
Convolution	Kernel Size = 3x3, Channels = 32, Stride = 2x2, Padding = Same, Dropout rate = 0.4	Leaky ReLU with coefficient 0.2
Convolution	Kernel Size = 3x3, Channels = 16, Stride = 2x2, Padding = Same, Dropout rate = 0.4	Leaky ReLU with coefficient 0.2
Convolution	Kernel Size = 3x3, Channels = 8, Stride = 2x2, Padding = Same, Dropout rate = 0.4	Leaky ReLU with coefficient 0.2
Dense	Output Neurons = 1	Sigmoid

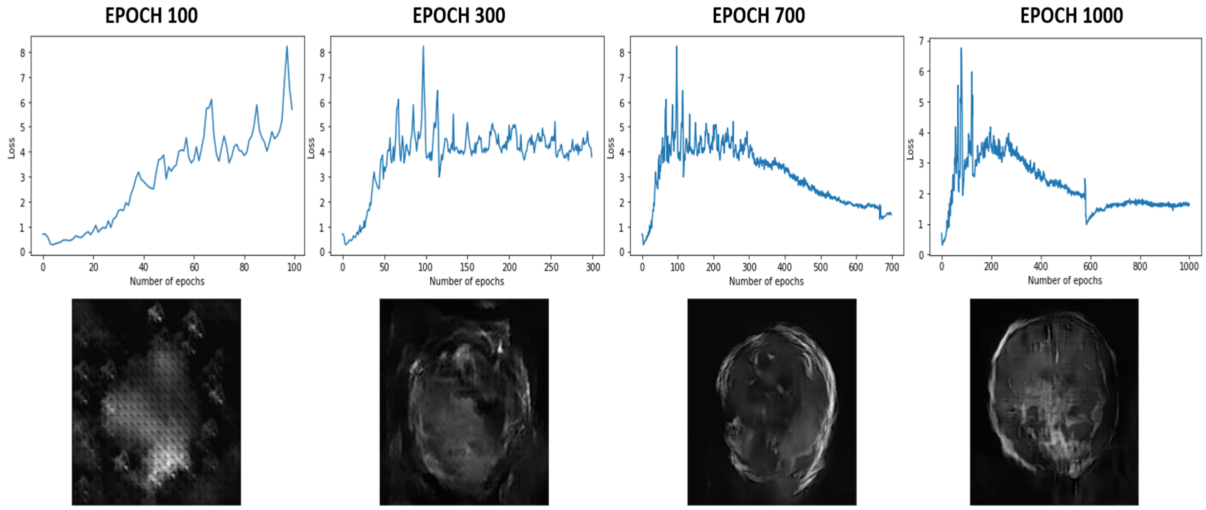


Figure 3: Loss functions and generated images using GANs

The architecture used for GAN is given in Table 4. Loss is defined to be cross entropy loss and Adam optimizer is used for optimization. In Figure 3, examples of the generated images after 100, 300, 700, and 1000 epochs are shown together with the plots of loss function. Unlike the images generated using basic and deformable methods, the images generated by GANs do not preserve their meaning and properties of their label. Even though GANs are trained for 1000 epochs, the images hardly look like a brain and they are not realistic. Therefore, these are not added to the dataset.

The aim of applying data augmentation was to help the model generalize and increase the classification performance by increasing the size of the dataset. After applying various augmentation methods, it is seen that basic augmentation methods do not provide an improvement on the performance of the model. Actually, they cause a decrease in performance metrics as seen in Table 3. After applying deformable methods and adding those images to the dataset, the performance increases. However, compared to performance metrics of the original dataset, there is no significant improvement. As a more advanced solution, training GANs is also tried. However, the generated images do not contain valuable information. It is not possible by looking at an image (such as the ones in Figure 3) and classifying the tumor. This may be due to the small size of the dataset, it is not enough for training the GANs of desired performance. The model is not able to generalize and capture the characteristics of this data. The MRI images are from very different angles and the model cannot handle this variety with the limited amount of data. It also cannot learn that there are 4 different classes, the classes of the generated images are indistinguishable.

A better approach would be first obtaining the tumors using segmentation, then applying data augmentation to these tumors as in the study of Havaei et. al, which is mentioned in the related work section. By applying data augmentation only to the tumors, it is possible to obtain more diverse patients while preserving the clinical plausibility. Another approach would be using transfer learning together with data augmentation as mentioned in the related work section. In transfer learning, there are no images generated but the model is able to learn from the pretrained models, which is also a useful tool for handling with limited amount of data. [24] [25] [26]

5 Conclusion

In this study, data augmentation methods are applied to a brain tumor classification problem. A dataset consisting of 3274 brain MRI images is used and new images are generated. The data augmentation methods are considered under three categories and the contribution of these different categories to the dataset and performance of the model are investigated. The first group is basic augmentation methods and images are generated by modifying the original images in various ways: rotation, height and width shifts, zoom, shear, brightness and flipping. With these methods, the size of the original dataset is increased by 24%. The second group is deformable augmentation methods, random displacement is applied and the size of the dataset is increased by 30%. The third group is advanced methods containing deep learning solutions and GANs are trained for generating new MRI images. After applying basic methods, a decrease in model's performance is observed, classification accuracy dropped from 91% to 87%. After adding images generated with deformed methods, the accuracy changed back to 91%. The images generated using GANs are not added to the dataset since these images have lower quality than expected. One reason behind this is thought to be the small size of the dataset; there are not enough images for GAN to capture the characteristics of brain MRIs. Overall, these groups did not have a significant contribution to the model's performance.

References

[1] L. Cai, J. Gao, and D. Zhao, "A review of the application of deep learning in medical image classification and segmentation," *Annals of translational medicine*, vol. 8, pp. 713–713, June 2020.

[2] Y.-C. Li, H.-H. Chen, H. Lu, H.-T. Wu, M. C. Chang, and P.-H. Chou, "Can a deep-learning model for the automated detection of vertebral fractures approach the performance level of human subspecialists?," *Clinical orthopaedics and related research*, vol. Publish Ahead of Print, 02 2021.

[3] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on mri," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019. Special Issue: Deep Learning in Medical Physics.

[4] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications," *Journal of Medical Imaging and Radiation Oncology*, vol. 65, no. 5, pp. 545–563, 2021.

[5] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, p. 60, Jul 2019.

[6] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, 2017.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.

[8] M. Sajjad, S. Khan, S. Khan, S. Khan, S. Khan, S. Khan, K. Muhammad, W. Wu, W. Wu, A. Ullah, and S. W. Baik, "Multi-grade brain tumor classification using deep cnn with extensive data augmentation," *Journal of Computational Science*, 2019.

[9] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. Lu, "Content-based brain tumor retrieval for mr images using transfer learning," *IEEE Access*, vol. 7, pp. 17809–17822, 2019.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014.

[11] S. Deepak and P. Ameer, "Brain tumor classification using deep cnn features via transfer learning," *Computers in Biology and Medicine*, vol. 111, p. 103345, 2019.

[12] W. Ayadi, W. Elhamzi, I. Charfi, I. Charfi, and M. Atri, "Deep cnn for brain tumor classification," *Neural Processing Letters*, 2021.

[13] A. M. Gab Allah, A. M. Sarhan, and N. M. Elshennawy, "Classification of brain mri tumor images based on deep learning pggan augmentation," *Diagnostics*, vol. 11, no. 12, 2021.

[14] M. Pesteie, P. Abolmaesumi, and R. N. Rohling, "Adaptive augmentation of medical data using independently conditional variational auto-encoders," *IEEE Transactions on Medical Imaging*, vol. 38, no. 12, pp. 2807–2820, 2019.

[15] E. Bayoumi, M. K. Abd-Ellah, A. A. M. Khalaf, and R. R. Gharieb, “Brain tumor automatic detection from mri images using transfer learning model with deep convolutional neural network,” *Journal of Advanced Engineering Trends*, vol. 41, no. 2, pp. 19–30, 2021.

[16] T. Han-Trong, H. Nguyen Van, H. Nguyen Thi Thanh, V. Tran Anh, D. Nguyen Tuan, and L. Vu Dang, “An efficient method for diagnosing brain tumors based on mri images using deep convolutional neural networks,” *Applied Computational Intelligence and Soft Computing*, vol. 2022, p. 2092985, Mar 2022.

[17] S. Bhuvaji, A. Kadam, P. Bhumkar, S. Dedge, and S. Kanchan, “Brain tumor classification (mri),” 2020.

[18] O. C. Mimenza, “Meninges: Anatomía, partes y funciones en el cerebro.” `https://psicologiaymente.com/neurociencias/meninges`, Mar 2022.

[19] “Pituitary gland: What it is, function amp; anatomy.” `https://my.clevelandclinic.org/health/body/21459-pituitary-gland`.

[20] “Brain tumor types.” `https://www.hopkinsmedicine.org/health/conditions-and-diseases/brain-tumor/brain-tumor-types#variable`.

[21] “Glioma.” `https://www.mayoclinic.org/diseases-conditions/glioma/symptoms-causes/syc-20350251`, Apr 2020.

[22] G. van Tulder, “elasticdeform: Elastic deformations for n-dimensional images,” Mar 2021.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[24] J. Brownlee, “A gentle introduction to transfer learning for deep learning,” Sep 2019.

[25] L. Torrey and J. Shavlik, “Transfer learning,” *Handbook of Research on Machine Learning Applications*, 01 2009.

[26] “Cs231n convolutional neural networks for visual recognition.” `https://cs231n.github.io/transfer-learning/`.