(049)

High Accuracy Brain Tumor Classification with EfficientNet and Magnetic Resonance Images

J. M. Medina and J. Sánchez
University of Las Palmas de Gran Canaria, Computer Science Department,
Centro de Tecnologías de la Imagen (CTIM), 35017 Las Palmas de Gran Canaria, Spain
E-mail: jsanchez@ulpgc.es

Summary: The automatic detection of brain tumors is important for efficiently processing large amounts of data. This is a complex task due to the great variety that exists, and the inherent challenges associated with processing medical images. The aim of this work is to assess the performance of recent neural networks for the classification of brain tumors. We study the EfficientNet model, which has provided good results in many classification problems. We use two standard datasets with more than 3000 magnetic resonance images each. The classification includes four different classes with three tumor types (glioma, meningioma, pituitary tumors), and an additional class for brains without tumors. The experiments analyze three models of the EfficientNet architecture, using several techniques, such as transfer learning, early stopping and fine-tuning. The results show that the models attain an accuracy of 98.4 % and 97.5 % with the two datasets, which is on par with state-of-the-art methods.

Keywords: Brain tumor classification, Magnetic resonance imaging (MRI), Deep learning, Convolutional neural network (CNN).

1. Introduction

During the last years, many works have explored the application of deep neural networks for the classification of brain tumors. This is an intricate task due to the difficulty in the diagnosis of several types of tumors and the implications for the health of patients. Consequently, the quest for a neural network capable of delivering exceptional accuracy in tumor detection holds paramount importance for effective patient treatment.

Many works have obtained promising results with standard architectures, such as ResNet [13], Xception [11] or Vision Transformers [24], and a few works have compared the performance of multiple neural networks, like in [17-19].

The study presented in [18] showed that EfficientNet [23] provided good performance among many different models, with one of the best trade-offs between accuracy and model complexity. However, the authors did not search for the best hyper-parameters, using a general setting for all the models.

In this work, we are interested in deepening in the capabilities of this neural network. The aim is to analyze several techniques, such as transfer learning, fine tuning, or early stopping, and find the best configuration. We use two datasets of more than 3000 magnetic resonance images each. The first one, from Figshare [7], contains 3064 images with three types of tumors, whereas the second one, from Kaggle [4], has 3264 images with an additional label for images without tumors.

The experiments analyze the behavior of different models of the EfficientNet family, obtaining an accuracy of 98.4 % with the Figshare dataset and 97.5 % with the Kaggle dataset. In particular, we study three models with a high disparity in the number of parameters, so we may understand the behavior with respect to the size of the network. The use of transfer learning is important for attaining high levels of accuracy and fine-tuning is key for further improving the results. Comparing with state-of-the-art methods, our results rank in the top of the classification for both datasets.

Section 2 discusses the main works in the literature based on neural networks. Section 3 details the datasets that we use in our study, then it summarizes the main characteristics of the EfficientNet model and details the experimental setup. Section 3 analyzes the experimental results and compares with state-of-the-art methods. Some concluding remarks are given in Section 4.

2. Related Work

Over the past few years, many methods have addressed the problem of brain tumor classification with neural networks and a few studies have compared the performance of several convolutional neural networks (CNNs) for this task. For example, in [19], the authors trained five models (Xception, ResNet50, InceptionV3, VGG16, and MobileNet), achieving accuracies of 98.75 %, 98.50 %, 98.00 %, 97.50 %, and 97.25 %, respectively. However, this preliminary work lacks sufficient details to reproduce the results and it is not clear which dataset was used in the experiments.

The authors of the work presented in [17] investigated the performance of the VGG16, VGG19, ResNet50, and DenseNet121 models using the Figshare dataset, relying on transfer learning and adding three fully connected layers on top of the neural networks. They compared results using several optimization methods (Adam, Adadelta, RMSprop, and SGD), and obtained the best results with ResNet50 and DenseNet121, with an accuracy of 99.0 % and 98.9 %, respectively.

Nevertheless, these works are not directly comparable due to the use of different datasets. The work presented in [18] analyzes most of these CNNs and includes EfficientNet and the recent ConvNext, using the two datasets. They explored different versions of each family and provided sufficient details for reproducibility. They also obtained good performance for these models and draw some interesting conclusions. The EfficientNet model presented the best trade-off between accuracy and complexity, being one of the best candidates for this problem.

Many other articles have studied the behavior of other models, like in [25], where the authors showed the superiority of GoogLeNet [21] over AlexNet [12] using transfer learning.

The EfficientNet architecture has also been previously used in [10] and [14]. The first work was based on the EfficientNetB0 model and obtained an accuracy of 96.9 % using the Kaggle dataset. In this study, we obtain better results with both EfficientNetB0 and EfficientNetB3. For the work presented in [14], it is not clear which dataset was used.

A study on ResNet50 [13] provided results with an accuracy of 97.5 % using the Figshare dataset. They obtained better results without data augmentation, which is in line with the results obtained in [18].

More recently, in [11], the authors analyze several networks and propose an ensemble of three models based on VGG16, InceptionV3, and Xception, obtaining an accuracy of 96.9 % with the Kaggle dataset. The authors also studied three different Vision Transformers, but they did not obtain good results. In our work, we obtain higher accuracy with a single model. Although the accuracy of these methods is typically above 95 %, it is not easy to compare the results because they use different datasets and configurations.

3. Material and Methods

3.1. Datasets

In this work, we use two MRI datasets for brain tumor classification. The first one is available on Figshare [7] and was initially proposed in [6]. It is one of the first datasets for this purpose and many works have used it for evaluating their methods. It contains 3064 images with the three tumor types: 1426 images of gliomas, 708 images of meningiomas, and 930 images of pituitary tumors. In this case, there are no images of healthy brains, thus there is not any label for the no tumor class. The information is stored in MATLAB format in a data structure that contains the image, the label, a unique identifier, and the segmentation of the tumor region, including a mask and a polygon of the tumor contour. We converted the images into PNG format and split the images into the training and testing directories, with three subdirectories corresponding to the three tumor types.

The Kaggle dataset [4] contains images with the same types of tumors and an additional classification for images without tumors. It contains 3264 images, with 926 gliomas, 937 meningiomas, 901 pituitary tumors, and 500 without tumors. In this case, the dataset is organized in two directories (training and testing) and four subdirectories corresponding to each label. In a preprocessing step, we unified the size of the images to 512×512 by scaling and centering the original images. This dataset contains more variety than the previous one, because it was created by collecting images from other sources. For instance, it includes more than 2000 images from the Figshare dataset (as reported in [18]), as well as images from other datasets. Fig. 1 depicts some slices of this dataset from different views.

3.2. The EfficientNet Neural Network

EfficientNet [22] is a recent CNN that maintains a trade-off between accuracy and FLOPS by performing a compound scaling of the depth and width of the network. It also adapts the resolution of the input images for this purpose. This family of networks defines eight different models, from B0 to B7, with increasing sizes in width, depth, and image resolution. This network relies on the Mnasnet architecture, which uses inverted bottlenecks, and squeeze-and-excitation optimization.

In our study, we analyze the second version of this family [23], which introduces several improvements over the previous version. The main differences are the introduction of progressive learning, so that the size of the images and strength of regularization is gradually increased during training; the replacement of depthwise convolutions in the first layers to increase the speed; and the scaling of the network starting from later stages.

We report the results for the B0, the B3, and the Small configurations. The number of parameters is significantly smaller than in other neural networks with similar performance. The accuracy that we obtain with the models of this version is consistently higher than with the models of the first one.

For transfer learning, we remove the top layer and include a dense layer with the number of classes (three for the Figshare dataset and four for the Kaggle dataset) and SoftMax activation. Before the dense layer, we include dropout. In our experiments, we obtained slightly better results with a dropout rate of 70 % in comparison with 50 % and 30 %, therefore, we selected 70 % by default.

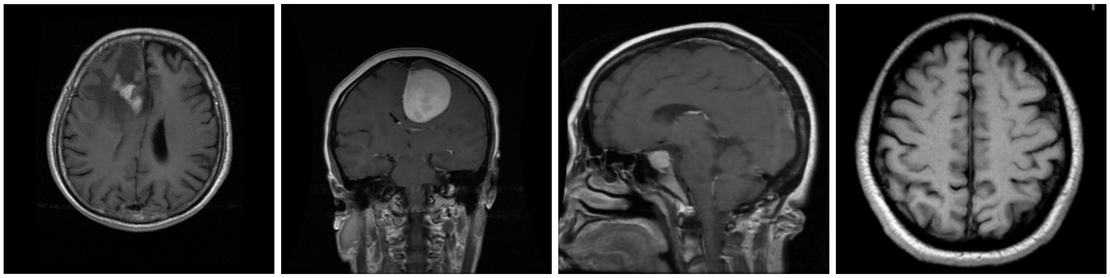

Fig. 1. Images of brain tumors from the Kaggle dataset: the first image on the left is an axial view of the brain with a *glioma*; the second image on the left is a coronal view with a *meningioma*; the third one is a sagittal view with a *pituitary* tumor; and the image on the right is a brain without tumor.

3.3. Experiments Setting

We use the `TensorFlow` and `Keras` libraries for the training of the neural networks, and the `scikit-learn` library to evaluate the models. The images are scaled down to 256×256 pixels with three channels. The range of the original pixel values is between 0 and 255, and we normalized the images between -1 and 1, dividing by 127.5 and subtracting 1. The experiments were conducted on an Intel Core i9-10940X CPU @3.30 GHz processor with 32 GB RAM, an NVIDIA GeForce RTX 2060 GPU with 8 GB RAM and an NVIDIA Geforce RTX 3060 GPU with 12 GB RAM, under Windows 10.

The models were trained using transfer learning with the weights corresponding to the ImageNet [8] dataset. We used the Adam optimizer with default parameters and early stopping with a patience parameter of 10. The maximum number of epochs was stablished to 50, although most executions finished before 25 epochs. For fine-tuning, we used the Adam optimizer with a learning rate of 10^{-4}, a decay rate of 10^{-5}, and 20 epochs. In this case, we unfroze the last 15 layers, comprising five convolutional layers.

During the experiments, we used several batch sizes, such as 32, 64, and 128, although we did not appreciate significant differences in accuracy but more stability during the optimization process.

We use one-hot encoding for coding the labels and categorical cross-entropy for the loss function. In the numerical results, we use the *accuracy*, *precision*, and *recall* metrics for comparing the methods, which are given by:

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN},$$

$$precision = \frac{TP}{TP+FP},$$

$$recall = \frac{TP}{TP+FN},$$

with *TP* the true positive values, *TN* the true negatives, *FP* the false positives, and *FN* the false negatives.

4. Results

The results for the Figshare dataset are given in Table 1. We observe that the best model is the EfficientNetB0, with an accuracy of 98.4 %, and a high precision and recall. The second-best method is EfficientNetB3, with a slightly inferior performance, and the third one is EfficientNetSmall.

Table 1. Results for the Figshare dataset.

Method	Accuracy	Precision	Recall
EfficientNetB0	98.4 %	98.1 %	98.0 %
EfficientNetB3	98.0 %	97.9 %	97.6 %
EfficientNetSmall	97.7 %	97.8 %	97.0 %

We observe a similar behavior when we use the Kaggle dataset; see Table 2. The accuracy is lower in this case for the three models. The best metrics are again for the EfficientNetB0 model. The training with this dataset is somehow more difficult as it has one more classification and image variety.

Table 2. Results for the Kaggle dataset.

Method	Accuracy	Precision	Recall
EfficientNetB0	97.5 %	97.3 %	97.6 %
EfficientNetB3	96.9 %	97.0 %	96.7 %
EfficientNetSmall	96.0 %	96.1 %	95.9 %

The number of parameters of the models is 6,2M for EfficientNetB0, 13,2M for EfficientNetB3, and 20,6M for EfficientNetSmall. The accuracy, precision, and recall are in general high for the three models. We note that the model with the lowest number of parameters provided the best accuracy in general. Fine-tuning, on the other hand, usually allowed to improve the accuracy by more than 2 %.

4.1. Precision per Tumor Type

In the following, we analyze the performance of the models with respect to each tumor type. In Figs. 2 and 3, we show the confusion matrices for the best model, *i.e.*, EfficientNetB0, with the Figshare and Kaggle datasets, respectively.

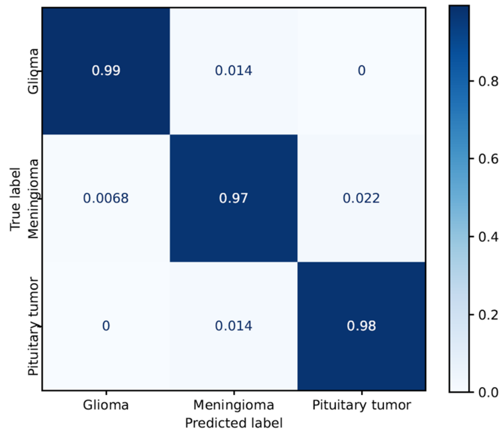

Fig. 2. Confusion matrix of the EfficientNetB0 model using the Figshare dataset.

In the case of the Figshare dataset, *gliomas* and *pituitary* tumors are correctly classified with high precision, with only one misclassification out of 146 for *gliomas* and two out of 87 for *pituitary* tumors. *Meningiomas* obtain the worse results with two errors out of 68 samples.

Looking at the confusion matrix corresponding to the Kaggle dataset, we observe that EfficientNetB0 classifies *pituitary* tumors with an accuracy of 100 %, whereas the *no tumor* class presents the worst results with an accuracy of 95 %. The performance for *gliomas* and *meningiomas* is similar, although we also observe that, for other experiments, *meningiomas* are slightly more difficult to classify than *gliomas*.

Tables 3 and 4 show the precision of the three methods for the three tumor types, the first one corresponding to the Figshare dataset and the second one to the Kaggle dataset. We observe that the precision for the *pituitary* tumor is the highest in general, followed by the *glioma* class. The *no-tumor* class also has a high classification rate for the Kaggle dataset. *Meningiomas* obtained the worst results in general.

Table 3. Precision of the methods with respect to each tumor type (Figshare dataset).

Method	Glioma	Meningioma	Pituitar
EfficientNetB0	99.3 %	97.1 %	97.8 %
EfficientNetB3	97.7 %	95.9 %	100 %
EfficientNetSm	97.3 %	97.1 %	98.9 %

Table 4. Precision of the methods with respect to each tumor type (Kaggle dataset).

Method	Glioma	Mening.	NoTum.	Pituit.
EfficientNetB0	97.0 %	97.1 %	95.1 %	100 %
EfficientNetB3	97.9 %	94.9 %	98.1 %	96.9 %
EfficientNetSm	98.9 %	88.5 %	98.1 %	98.9 %

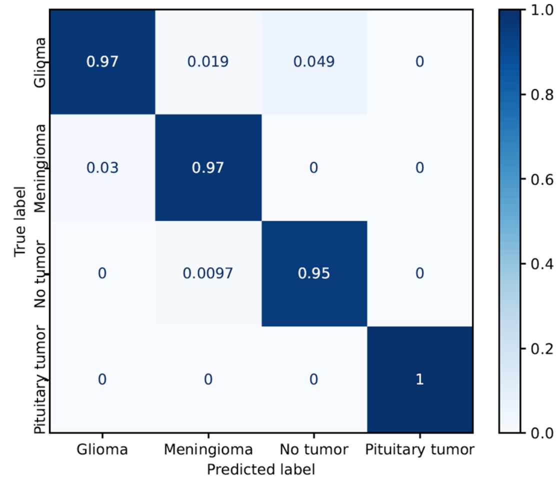

Fig. 3. Confusion matrix of the EfficientNetB0 model using the Kaggle dataset.

4.3. Comparison with State-of-the-art Methods

If we compare the results of the EfficientNet family with state-of-the-art methods, the models rank in the top of the classification.

Table 5 shows the best methods in the literature that have used the Figshare dataset. EfficientNetB0 ranks in the second position, whereas EfficientNetB3 ranks in the fourth position. The first work in the ranking [19] obtains a high precision with the ResNet model, only using transfer learning. We note that ResNet50 has 23,2M parameters, thus it is much larger than the EfficientNetB0 model. They also obtained good accuracy with other models, such as DenseNet and VGG, which are also much larger than EfficientNetB0. They add three fully connected layers on top of the network and analyze the performance using several optimizers. These models have also been analyzed in [18], also with transfer learning, although the results are not so accurate.

The third work [24] utilizes Vision Transformers for the classification. They obtain a slightly better result with an ensemble of Transformers. Nevertheless, they used an image resolution of 384×384 and the size of these networks is much bigger than EfficientNet.

In the case of the Kaggle dataset in Table 6, we observe that the EfficientNetB0 model provides the second-best result and EfficientNetB3 ranks in the third position. In this case, there are just a few works because this dataset is more recent.

The results presented in the first method [19], with an accuracy of 98.8 %, are, however, contradictory because it seems that they are using this dataset, but the number of images does not coincide, with more than 4000 images. On the other hand, it is not clear the configuration of the top layers they use after the CNN backbone and there are not enough details for reproducibility.

Table 5. Comparison with state-of-the-art methods using the Figshare dataset.

Method	Accuracy
Pashaei et al. [15]	93.7 %
Ayadi et al. [2]	94.7 %
Phaye et al. [16]	95.0 %
Ghassemi et al. [9]	95.6 %
Shaik et al. [20]	96.5 %
Badža et al. [3]	96.6 %
Kumar et al. [13]	97.1 %
EfficientNetSmall	**97.7 %**
Amin et al. [1]	98.0 %
Bodapati et al. [5]	98.0 %
EfficientNetB3	**98.0 %**
Tummala et al. [24]	98.2 %
EfficientNetB0	**98.4 %**
Polat et al. [17]	99.0 %

Table 6. Comparison with state-of-the-art methods using the Kaggle dataset.

Method	Accuracy
EfficientNetSmall	**96.0 %**
Hossain et al. [11]	96.5 %
Goutham et al. [10]	96.9 %
EfficientNetB3	**96.9 %**
EfficientNetB0	**97.5 %**
Saleh et al. [19]	98.8 %

5. Conclusion

This work has shown that EfficientNet can provide high accuracy for the classification of brain tumors. We selected two of the most important datasets and tried to search the best hyperparameters.

The best results for the two datasets were obtained with EfficientNetB0, which has the lowest number of parameters (6,2M). The second-best model was EfficientNetB3 with 13,2M parameters. This means that it is not necessary a neural network with a large capacity to deal with these datasets.

Transfer learning and fine-tuning are important for increasing the accuracy of the models. It is interesting to note that transfer learning was based on the ImageNet dataset, which is composed of natural images, very different from MR images. This denotes the generalization capacity of CNNs. On the other hand, fine-tuning allowed to further increase the accuracy by more than 2 %. The results that we have obtained in the experiments are on par with state-of-the-art methods, with the benefit that the EfficientNet models have less parameters.

Acknowledgements

This work was partly supported by the Servicio Canario de Salud, Gobierno de Canarias, under Contract F2022/03.

References

[1]. J. Amin, M. Sharif, N. Gul, M. Raza, M. A. Anjum, M. W. Nisar, S. A. C. Bukhari, Brain tumor detection by using stacked autoencoders in deep learning, *Journal of Medical Systems*, Vol. 44, Issue 2, 2020, pp. 1-12.

[2]. W. Ayadi, W. Elhamzi, I. Charfi, M. Atri, Deep CNN for brain tumor classification, *Neural Processing Letters*, Vol. 53, Issue 1, 2021, pp. 671-700.

[3]. M. Badža, M. Č. Barjaktarović, Classification of brain tumors from MRI images using a convolutional neural network, *Applied Sciences*, Vol. 10, Issue 6, 2020, 1999.

[4]. S. Bhuvaji, A. Kadam, P. Bhumkar, S. Dedge, S. Kanchan, Brain Tumor Classification (MRI), *Kaggle*, 2020.

[5]. J. D. Bodapati, N. S. Shaik, V. Naralasetti, N. B. Mundukur, Joint training of two-channel deep neural network for brain tumor classification, *Signal, Image and Video Processing*, Vol. 15, Issue 4, pp. 753-760, 2021.

[6]. J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang, Q. Feng, Enhanced performance of brain tumor classification via tumor region augmentation and partition, *PLOS ONE*, Vol. 10, Issue 12, 2015, e0144479.

[7]. J. Cheng, Brain Tumor Dataset, *Figshare*, 2017.

[8]. J. Deng, W. Dong, R. Socher, L. Li, K. Li, L. Fei-Fei, ImageNet: A large-scale hierarchical image database, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09)*, 2009, pp. 248-255.

[9]. N. Ghassemi, A. Shoeibi, M. Rouhani, Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images, *Biomedical Signal Processing and Control*, Vol. 57, 2020, 101678.

[10]. V. Goutham, A. Sameerunnisa, S. Babu, T. B. Prakash, Brain tumor classification using EfficientNet-B0 model, in *Proceedings of the 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE'22)*, 2022, pp. 2503-2509.

[11]. S. Hossain, A. Chakrabarty, T. R. Gadekallu, M. Alazab, M. J. Piran, Vision transformers, ensemble model, and transfer learning leveraging explainable AI for brain tumor detection and classification, *IEEE Journal of Biomedical and Health Informatics*, 2023, pp. 1-14.

[12]. A. Krizhevsky, I. Sutskever, G. E Hinton, ImageNet classification with deep convolutional neural networks, in Advances in Neural Information Processing Systems (F. Pereira, C. J. Burges, L. Bottou, K. Q. Weinberger, Eds.), Vol. 25, *Curran Associates, Inc.*, 2012, pp. 1-9.

[13]. R. L. Kumar, J. Kakarla, B. V. Isunuri, M. Singh, Multi-class brain tumor classification using residual network and global average pooling, *Multimedia Tools and Applications*, Vol. 80, Issue 9, 2021, pp. 13429-13438.

[14]. S. D. Kumar, G. Kumar, S. Kumar, Multi-classification of brain MRI using EfficientNet, in *Proceedings of the International Conference for Advancement in Technology (ICONAT'22)*, 2022, pp. 1-6.

[15]. A. Pashaei, H. Sajedi, N. Jazayeri, Brain tumor classification via convolutional neural network and extreme learning machines, in *Proceedings of the 8th*

International Conference on Computer and Knowledge Engineering (ICCKE'18), 2018, pp. 314-319.

[16]. S. R. Phaye, A. Sikka, A. Dhall, D. R. Bathula, Dense and diverse capsule networks: Making the capsules learn better, *CoRR*, 2018, abs/1805.04001.

[17]. Ö. Polat, C. Güngen, Classification of brain tumors from MR images using deep transfer learning, *The Journal of Supercomputing*, Vol. 77, Issue 7, 2021, pp. 7236-7252.

[18]. D. Reyes, J. Sánchez, Performance of convolutional neural networks for the classification of brain tumors using magnetic resonance images, *Preprint*, 2023.

[19]. A. Saleh, R. Sukaik, S. S. Abu-Naser, Brain tumor classification using deep learning, in *Proceedings of the International Conference on Assistive and Rehabilitation Technologies (iCareTech'20)*, 2020, pp. 131-136.

[20]. N. S. Shaik, T. K. Cherukuri, Multi-level attention network: application to brain tumor classification, *Signal, Image and Video Processing*, Vol. 16, Issue 3, 2022, pp. 817-824.

[21]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15)*, 2015, pp. 1-9.

[22]. M. Tan, Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in *Proceedings of the* 36^{th} *International Conference on Machine Learning*, 2019, pp. 6105-6114.

[23]. M. Tan, Q. Le, Efficientnetv2: Smaller models and faster training, in *Proceedings of the* 38^{th} *International Conference on Machine Learning*, 2021, pp. 10096-10106.

[24] S. Tummala, S. Kadry, S. A. C. Bukhari, H. T. Rauf, Classification of brain tumor from magnetic resonance imaging using vision transformers ensembling, *Current Oncology*, Vol. 29, Issue 10, 2022, pp. 7498-7511.

[25]. Y. Yang, L. Yan, X. Zhang, Y. Han, H. Nan, Y. Hu, B. Hu, S. Yan, J. Zhang, D. Cheng, et al., Glioma grading on conventional MR images: A deep learning study with transfer learning. *Frontiers in Neuroscience*, Vol. 12, 2018, 804

.