

Brain tumor diagnoses based on VGG-16 and MobileNet

Yuheng Liu

School of Ocean and Marine Science, Tongji University, Shanghai, 200092, China

2050337@tongji.edu.cn

Abstract. Brain tumor, recognized as one of the most formidable and aggressive diseases globally, continues to pose significant challenges for medical practitioners in clinical diagnosis and treatment. Addressing the burden on doctors and addressing resource limitations in certain hospitals necessitates the development of efficient and dependable alternative approaches. Convolutional Neural Networks (CNNs), renowned for their prowess in image recognition, hold immense potential in addressing this pressing issue. Leveraging transfer learning, the capabilities of established models such as VGG-16 and MobileNet can be harnessed to construct superior models within a comparatively abbreviated timeframe. This research paper aims to construct and evaluate VGG-16 and MobileNet-based models, employing transfer learning, to explore the applicability of these two classical models in the context of brain tumor diagnosis. The ultimate goal is to assist doctors and hospitals in alleviating the challenges associated with brain tumor diagnoses. The results demonstrated the effectiveness of brain tumor recognition based on CNNs.

Keywords: brain tumor, CNN, transfer learning, VGG-16, MobileNet.

1. Introduction

Brain tumors, renowned for their dismal prognosis and high fatality rates, manifest as abnormal cell accumulations within the brain, encompassing various classifications such as benign, malignant, and pituitary tumors. Notably, this disease exerts its aggressive nature across both pediatric and adult populations, as even benign tumors can elicit elevated intracranial pressure, culminating in irreparable life-threatening impairments. Given the gravity of the condition, early detection and classification of brain tumors have emerged as imperative research areas within the realm of medical imaging. These endeavors hold promise in facilitating the identification of optimal treatment modalities, thus salvaging patients' lives. Consequently, the meticulous planning and accurate diagnosis enabled by such advancements can significantly enhance patients' life expectancies.

The World Health Organization (WHO) emphasizes the significance of precise brain tumor diagnosis, which encompasses the detection of tumors, comprehensive localization within the brain, and accurate classification based on malignancy, grade, and type. To meet these essential diagnostic criteria, contemporary medical practices rely on Magnetic Resonance Imaging (MRI) techniques. MRI enables the detection and classification of brain tumors by capturing detailed images that facilitate the assessment of tumor grade, type, and precise localization within the brain. Through the utilization of MRI scans, a substantial volume of image data is generated, requiring careful examination by skilled radiologists to extract pertinent information from the acquired images. This multimodal imaging

approach serves as a crucial method for acquiring comprehensive patient data, thus contributing to effective brain tumor diagnosis and subsequent treatment planning.

However, the diagnosis of brain tumors presents a formidable challenge to medical practitioners due to the complex and diverse characteristics inherent in these neoplasms. The irregularities in tumor sizes and locations further complicate the comprehensive understanding of their nature. Moreover, the analysis of MRI scans necessitates the expertise of a trained neurosurgeon, a resource that may be scarce, particularly in developing countries where limited knowledge about tumors exists. Consequently, the generation of MRI reports becomes a laborious and time-consuming process. To address these issues stemming from human examination limitations and resource constraints, there is a pressing need to develop an automated system that exhibits high accuracy and consistency while operating with minimal resource requirements.

The advent of deep neural networks (DNN), particularly convolutional neural networks (CNNs), has attracted considerable attention due to their exceptional capacity for automated feature extraction and analysis. These networks have witnessed widespread adoption in image classification tasks, achieving notable performance improvements since 2012. Notably, CNNs applied to medical image classification have demonstrated remarkable performance, often on par with or surpassing human experts. For instance, the CNN-based CheXNet, comprising 121 layers and trained on a dataset of over 100,000 frontal-view chest X-rays (ChestX-ray 14), outperformed the average performance of four radiologists. Similarly, Kermany et al. proposed a transfer learning system utilizing DNNs to classify 108,309 Optical Coherence Tomography (OCT) images [1], where the weighted average error equaled the average performance of six human experts. In the realm of DNN models, transfer learning has emerged as a prevalent and efficient technique. Transfer learning entails leveraging the knowledge acquired from a prior task to enhance generalization on a new problem. By capitalizing on the capabilities of pre-existing powerful models and augmenting them with specialized task-oriented abilities, model designers can save time and effort that would otherwise be required to construct and train models from scratch. Moreover, this approach holds significant practical applicability, particularly when confronted with limited labeled data, as real-world problems often lack millions of labeled data points necessary for training complex models.

Based on the DNN models and transfer learning methods, in this paper, MRI dataset with around 3,000 pictures for training and 400 pictures for testing will be employed for machine learning to train powerful models, a combination of different pre-trained models including VGG-16, Mobilenet will be used as the foundation for the solution. The result demonstrated great accuracy achieved by both models with considerably low computational cost and training time, and that the potential of transfer learning on small medical datasets to yield desired results.

2. Method

2.1. Dataset description and preprocessing

The dataset is collected from Kaggle website which is composed of 3,264 images that can be further divided into 2,870 training images and 394 testing images in gray scale form with the size of 150×150 from four categories [2]: Glioma Tumor, Meningioma Tumor, Pituitary Tumor and No tumor. The sample image of each category is displayed in Figure 1.

In terms of data preprocessing, considering that the dataset is rather small, especially great imbalance of the number of images exists in different categories, data augmentation is carried out through the class of imagedatagenerator which conducts certain type of operations during each time of training as specified in advance, therefore enlarging the number of images for training based on original dataset without actually adding more images which helps to yield more accurate result. To be specific, methods including horizontal and vertical shift, zooming in or out and flipping vertically or horizontally are applied to different extent, such operations are chosen to a certain extent with deliberation that the overall feature and pivotal essence wouldn't be harmed and conspicuous variations are able to be made. And before feeding the data into training process, it is normalized by dividing each pixel by 255.

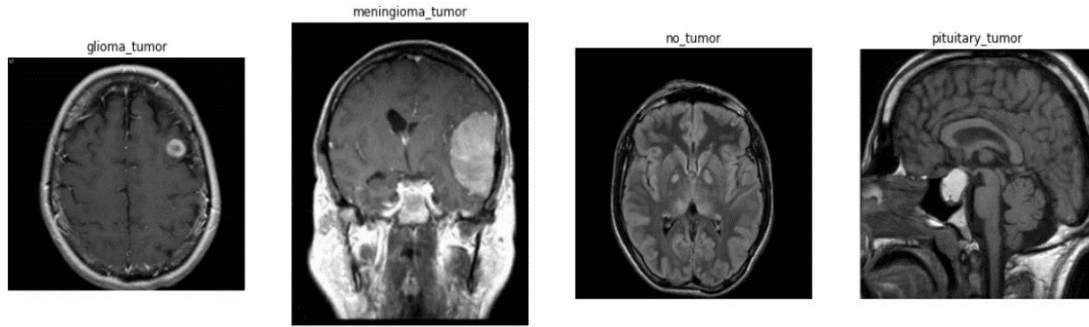


Figure 1. Sample images from each category [2].

2.2. CNN-based brain tumor recognition

CNN, famous for its ability to visual imagery analytical ability [3-6], is a kind of artificial neural network consisting of an input layer, an output layer and various hidden layers that features convolution layers in which shared-weight convolution filters slide along input features, summing up the results of pixel values multiplied by corresponding weight in each step during sliding and eventually forming a feature map representative of the original image or the feature map of previous convolution layer. As layer increases, the following layer gradually learns more profound features compared with the previous ones and is eventually able to grasp information to the depth that is adequate for accurate recognition. Between convolution layers, pooling layers are applied as well with similar kernel-based architecture aiming to reduce the dimension of feature maps for the sake computational cost. At the end of a convolution neural network, full-connected layers in which every neuron is fully connected to every neuron in the proceeding layer are applied to transform knowledge based on feature map into numeric matrixes and further output the result of classification based on that.

VGG16, a convolutional neural network (CNN), has earned recognition as one of the most accomplished computer vision models to date. The architects of this model conducted a meticulous evaluation of various network configurations, ultimately augmenting the network depth by employing an architecture featuring small (3×3) convolution filters. This architectural modification yielded a substantial improvement over previous iterations. Notably, the depth of the network was elevated to encompass 16–19 weight layers, as depicted in Figure 2 [7]. MobileNet, on the other hand, represents a category of CNNs designed with a mobile-first approach and was made available as an open-source solution by Google. It serves as a valuable foundation for training classifiers through its lightweight model [8]. MobileNet implements depthwise separable convolutions, a technique that significantly reduces the number of parameters in comparison to conventional convolutional networks while maintaining a similar depth. This results in the creation of lightweight deep neural networks that offer exceptional training speed without compromising the efficacy of the models.

In the convolution neural network of this paper, the methodology of transfer learning is applied to inherit the image recognition ability from models trained on imagenet, pre-trained model including MobileNet and VGG16 are separately applied as base model for convolution layers trained upon ImageNet dataset, based on the base model, fully-connected layers are customized along with batch normalization layers and dropout layers to avoid overfitting, the architecture of the model is shown in Figure 3.

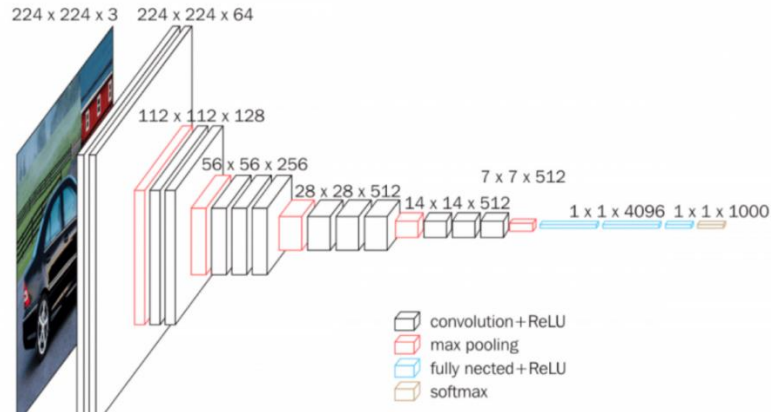


Figure 2. The architecture of VGG16 [7].

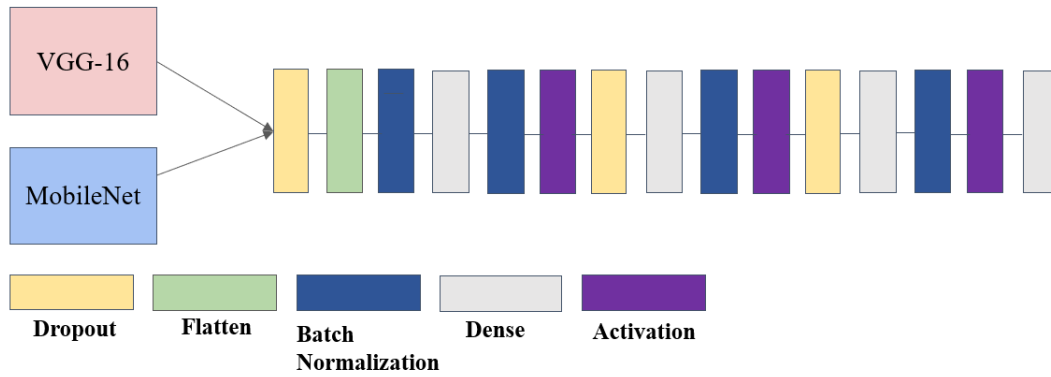


Figure 3. Architecture of the modified model.
Photo/Picture credit: Original

2.3. Implementation details

The model is primarily based on Tensorflow [9, 10], and the hyper-parameters can be found as following Table 1.

Table 1. The hyperparameter configuration of the model.

Configurations	Hyper-parameters
Output Activation	Softmax
Loss Function	Categorical Cross Entropy
Optimizer	Adam
Metrics	Accuracy
Batch size	64
Epochs	100
Callbacks	Early Stopping(patience=10)

3. Results and discussion

3.1. The performance of the model

Regarding the training process, both models were subjected to identical settings involving a maximum of 100 epochs, however, in practice, both models stopped training much earlier due to early stopping

with patience of 10. The early stopping criteria dictated that training be halted when no discernible improvement in test accuracy had been observed for a duration spanning 10 epochs.

In terms of the Mobilenet-based model, both training accuracy and training loss versus epochs demonstrated a rapid convergence according to Figure 4. and Figure 5., as for the test accuracy and test loss versus epochs curve, despite moderate fluctuation, both curves present upward trend. Eventually, the best training and test accuracy reach 95.7% and 91.8% respectively, indicating the robustness and efficiency of the Mobilenet-based model. Further analysis of the model's performance across different categories indicates near-flawless performance on the meningioma_tumor and no_tumor categories, while exhibiting comparatively poorer performance on the glioma_tumor and pituitary_tumor categories, particularly in relation to the former. Notably, there is a substantial likelihood of misclassifying glioma_tumor as meningioma_tumor, suggesting room for improvement in the model's classification accuracy for these specific tumor types.

Regarding the vgg-16 based model, it is evident that the overall performance is notably inferior shown in Figure 6 and Figure 7. Not only does the model has much worse ability in the task and performs almost equally bad in each category, but also exists serious overfitting problem. Though from the training accuracy versus epochs curve, there still exists potential for vgg16-based model, the training was stopped forcefully due to tremendous fluctuation in test accuracy.

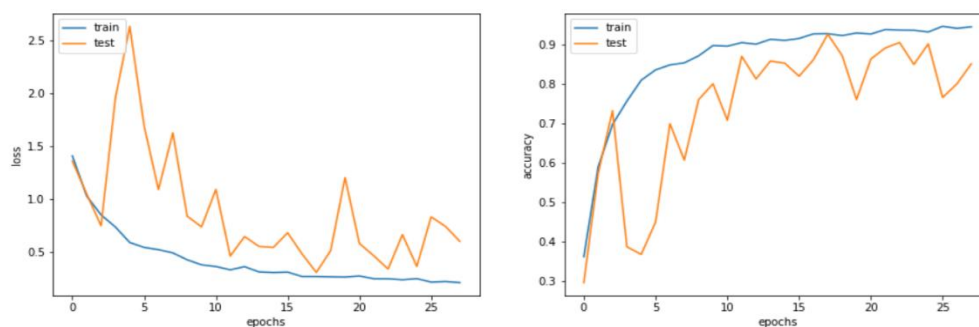


Figure 4. Training and test accuracy as well as loss versus epochs curve of MobileNet-based model.
Photo/Picture credit: Original

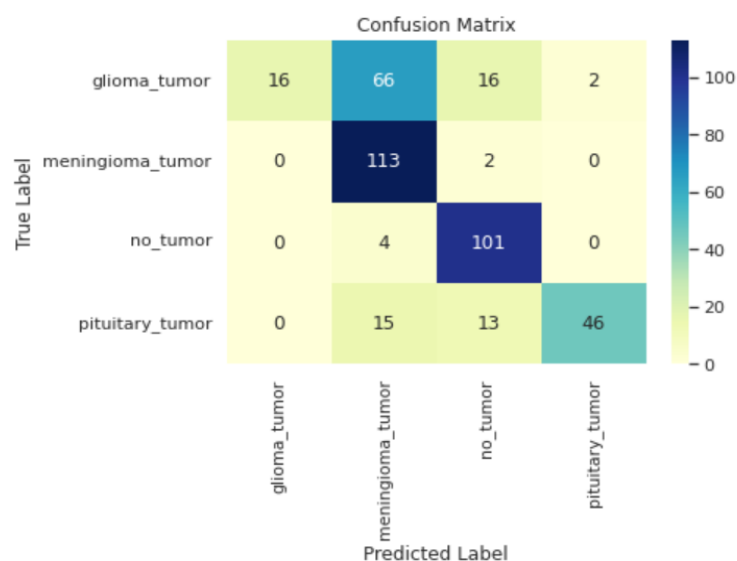


Figure 5. Confusion matrix of MobileNet-based model.
Photo/Picture credit: Original

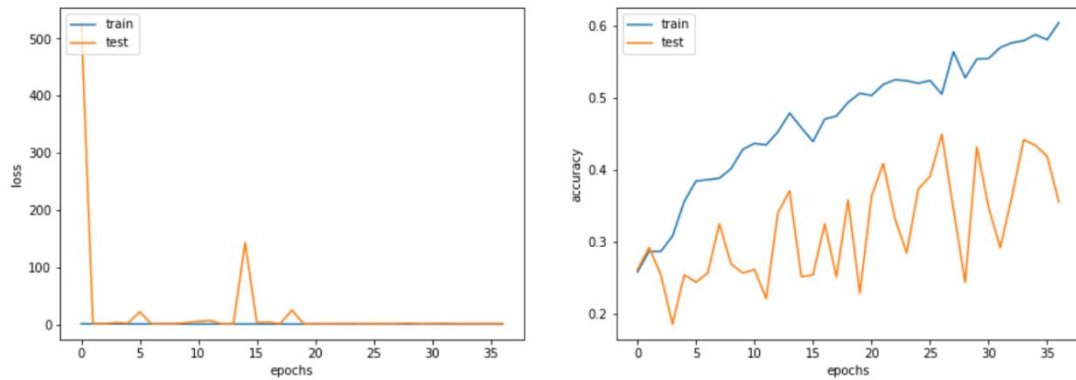


Figure 6. The training and test accuracy as well as loss versus epochs curve of VGG-16 based model.
Photo/Picture credit: Original

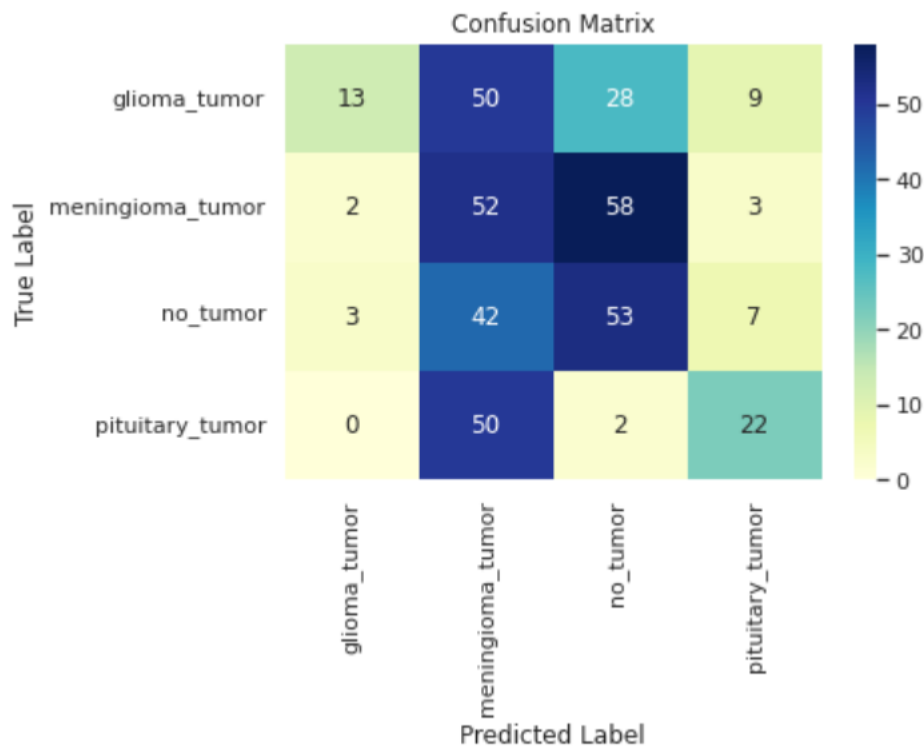


Figure 7. Confusion matrix of vgg-16 based model.
Photo/Picture credit: Original

3.2. Discussion

To account for the enormous difference between MobileNet-based model and vgg16-based model, the number of parameters must be taken into consideration. In this paper, the MobileNet-based model has 3.2 million parameters while the vgg16-based model has over 14 million. The small number of parameters and the depth wise design of MobileNet allow the MobileNet-based model to be trained at a higher speed, therefore optimizing faster to achieve better accuracy. Furthermore, the limited parameter count inherently mitigates the risk of overfitting by discouraging excessive model specialization that could potentially impede performance.

As is indicated in this case that huge model tends not to perform well in this tumor classification task with small amount of data, small yet robust model such as MobileNet should be recommended, to further improve the performance of the MobileNet-based model, two methods can be applied. First, it can be drawn from the confusion matrix, the model tends to mistake glioma_tumor for meningioma_tumor, therefore, more training needs to be done to further distinguish between these two categories, this can be achieved by unleashing the model from earlystopping as it can be inferred that overfitting problem is not obvious. Moreover, the number of data can be increased within this dataset to provide more meaningful data for the model to train, though in this model data augmentation is already carried out, other ways including oversampling can be taken into consideration.

4. Conclusion

In this paper, in response to the urgent need for assisting doctors to diagnose brain tumor, the MRI dataset is utilized to discover the power of CNN in this task. Two typical models including VGG-16 and MobileNet were used respectively by means of transfer learning to construct two models, the results demonstrated the robustness and efficiency of the MobileNet-based model compared to the huge models like VGG-16 in transfer learning for this brain tumor classification task. To further improve the performance of the MobileNet-based model to better distinguish between two certain categories, simply adding more epochs to train or using other data augmentation ways including oversampling are suggested. This paper hopes to contribute to solving the practical issue and other researches in related field for easing the great clinical pressure in brain tumor classification and scarcity of resources in the hospital.

References

- [1] Kermany DS et al 2018 Identifying medical diagnoses and treatable diseases by image-based deep learning *Cell* 172(5) pp 1122–1131.
- [2] Kaggle 2020 Brain Tumor Classification (MRI) <https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri>.
- [3] Simonyan K Zisserman A 2014 Very deep convolutional networks for large-scale image recognition *arXiv preprint arXiv:1409.1556*.
- [4] Howard A G Zhu M Chen B et al 2017 Mobilenets: Efficient convolutional neural networks for mobile vision applications *arXiv preprint arXiv:1704.04861*.
- [5] Zhou X Li Y Liang W 2020 CNN-RNN based intelligent recommendation for online medical pre-diagnosis support *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 18(3) pp 912-921.
- [6] Yu Q et al 2019 Semantic segmentation of intracranial hemorrhages in head CT scans 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS).
- [7] Jiang X Yang S Wang F et al 2021 OrbitNet: A new CNN model for automatic fault diagnostics of turbomachines *Applied Soft Computing* 110: 107702.
- [8] Sun Y Xue B Zhang M et al 2019 Completely automated CNN architecture design based on blocks *IEEE transactions on neural networks and learning systems* 31(4) pp 1242-1254.
- [9] Abadi M 2016 TensorFlow: learning functions at scale *Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming* 2016 pp 1-1.
- [10] Pang B Nijkamp E Wu Y N 2020 Deep learning with tensorflow: A review *Journal of Educational and Behavioral Statistics* 45(2) pp 227-248.