Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/18/20230986

Diagnosing brain tumors based on Edge-Impulse platform

Tianyou Xie

Shinyway Education, Zhejiang, 310063, China

yxie@arcvideo.com

Abstract. Brain tumors pose a serious threat to human health, and timely diagnosis is crucial for effective treatment. Conventional methods for diagnosing brain tumors, such as manual diagnosis, are often inaccurate and comparatively expensive. In this paper, a new method of diagnosis that can replace classic methods in the future is proposed. Ideas of machine learning is used to train a model so that it can classify images of brains into 2 categories: with tumors and with no tumors. Process of training is based on Edge Impulse, an AI platform on which model can be created conveniently. Technology used in the training is Convolutional Neural Network, which uses convolution as a mathematical operation instead of traditional matrix multiplication. It comprises convolutional layers, pooling layers and fully connected layers. Convolutional layers and pooling layers are used to operate feature maps, transforming them into a 1-dimensional matrix, followed by being multiplied by matrices of weights in fully connected layers. Cross entropy loss function is utilised to assess accuracy and modify weights. Transfer learning is also used to improve effectiveness and efficiency of training. Additionally, experiments are taken to determine the best image size suiting the model. This is accomplished by changing sizes of input image while keeping other parameters constant. Results of the experiments indicate that the model has a fantastic performance on brain tumor diagnosis.

Keywords: machine learning, deep learning, CNN.

1. Introduction

A brain tumour denotes an abnormal growth of cells in the brain or its vicinity. Such tumors can arise within the brain tissue or in the tissues surrounding the brain, including nerves, the pituitary gland, the pineal gland, and the meninges. Brain tumors are typically associated with high mortality rates. The 5-year relative survival rates for individuals across different age groups were investigated in this study. Specifically, the rates for those under 15 years of age, those between 15 to 39 years of age, and those aged 40 years or older were examined. The results show that the 5-year relative survival rate for the first group was approximately 75%, while the rate for the second group was around 72%. In contrast, the survival rate for individuals aged 40 and above was notably lower, at 21% [1]. Brain tumors account for close to 20,000 deaths annually. Nowadays, the method of diagnosing brain tumors mainly used is manual diagnosis by doctors. However, doctors cannot diagnose brain tumors very efficiently due to factors like technology limitations, and they are likely to misdiagnose. In addition, labour cost is usually too high, so that some poor families cannot afford diagnosis. Therefore, there is a need to identify alternative approaches to aid in manual diagnosis, and Artificial Intelligence (AI) is a promising technology for creating machines and instruments that can perform specific tasks, including aiding in the diagnosis of brain tumors.

AI has a rich history that can be traced back to the efforts of classical philosophers in explaining the mechanics of human thought processes through mechanical symbol processing. In the 1940s, the emergence of programmable digital computers based on abstract mathematical reasoning provided an impetus to seriously explore the possibility of building an electronic brain. Machine learning, as a subset of AI, has been developed as a technique to achieve artificial intelligence by employing it to address problems related to the field.

In the realm of machine learning, a neural network is a computational or mathematical model that mimics the structure and function of the biological neural network found in animals' central nervous systems, especially the brain. Neural networks are composed of a multitude of artificial neuron connections and are utilized to estimate or approximate functions. The adaptive nature of the artificial neural network enables it to alter its internal structure based on external information, making it an effective learning system. Neural networks are statistical data modeling tools that are nonlinear, and their optimization typically involves a type of learning method based on mathematical statistics. By utilizing standard mathematical statistical techniques, numerous local structure spaces that are expressed as functions can be obtained. In addition, within the field of machine perception, neural networks can exhibit basic decision-making and judgment abilities through statistical methods, which offer more advantages than formal logical reasoning and calculus.

Deep learning, a subfield of machine learning, focuses on learning and creating models for data representation using neural networks that contain multiple layers. The primary benefit of deep learning is its ability to utilize unsupervised or semi-supervised feature learning, as well as hierarchical feature extraction algorithms that enhance the effectiveness of the manual feature acquisition process. In various fields, such as computer vision, speech recognition, natural language processing, audio recognition, and bioinformatics, several deep learning frameworks have emerged, such as deep neural network, convolutional neural network, deep belief network, and recurrent neural network.

Automated machine learning (AutoML) refers to the process of automating the application of machine learning to practical problems. In this research, automated machine learning is used to diagnose brain tumors. In order to solve the problems mentioned above, this study employs the data set uploaded from Kaggle platform, to train a model by means of Edge Impulse, an AI platform. Results of the experiments imply effectiveness of this method.

2. Method

2.1. Data set preparation

This data set used in this study is sourced from Kaggle [2], a platform providing various machine learning data sets. The data set consists of 2 categories, which are images with brain tumors and images without brain tumors respectively. Each of those images has distinct size, but their ratios of sizes are all around 4:5. All images are based on the gray format instead of RGB color space since employed CT and MRI technologies cannot provide color information. The aim of the data set is to classify images into 2 categories, tumorous and non-tumorous. Tumorous images are labeled ‘yes’, while non-tumorous images are labeled ‘no’. Figure 1 provides sample images of the collected dataset.

Figure 1. The sample images on the collected dataset.

2.2. *Edge Impulse-based brain tumor recognition model*

The Edge Impulse platform is a tool for enterprise teams involved in the development of novel products leveraging edge artificial intelligence [3]. Within this platform, the creation of machine learning models aimed at achieving specific tasks, such as brain tumor recognition, is readily achievable. A common technique employed in training within this framework is the use of convolutional neural networks (CNNs) [4-6]. These networks utilize convolution as a mathematical operation in place of traditional matrix multiplication, typically within at least one of their layers. CNNs are constructed from several essential layers, including convolutional layers, pooling layers, and fully connected layers.

In the field of deep learning, CNNs utilize convolutional layers to generate a set of feature maps in parallel [7-9]. These feature maps are generated by sliding different convolution kernels over the input image and applying a product-sum operation at each position. The stride, a parameter that controls the size of the output feature map, determines the amount of sliding between the kernels and the input image. The convolution kernel, which is generally much smaller than the input image, overlaps or operates on the input image in parallel. A feature map is associated with a set of shared weight and bias terms to capture specific features of the input data.

Pooling layers are form of downsampling, so that size of data is decreased, and calculation can proceed more quickly [10]. There are 2 ways to operate this process: One is to select the maximum value from a patch of feature map, putting it into the downsampled feature map, and the other is to calculate the average of values in that patch. Max pooling is used more widely nowadays as it performs better in practice. Through the pooling layer, the sensitivity of the convolutional layer can be reduced to the edge.

After convolutional and pooling layers, advanced inference in neural networks is done through fully connected layers. The feature map is expanded into a 1-dimentional matrix and then is multiplied by a weight matrix and activated, forming a new matrix. Finally, a vector representing final scores is obtained. Cross entropy loss function is utilised to assess accuracy and modify weights.

Construction of the model based on the Edge Impulse platform is simple to operate. The data set is uploaded firstly, and then impulse is created by adding processing blocks and learning blocks. After that, parameters and training settings are determined, followed by transfer training. Figure 2 provides the interfaces of each operation in this platform.

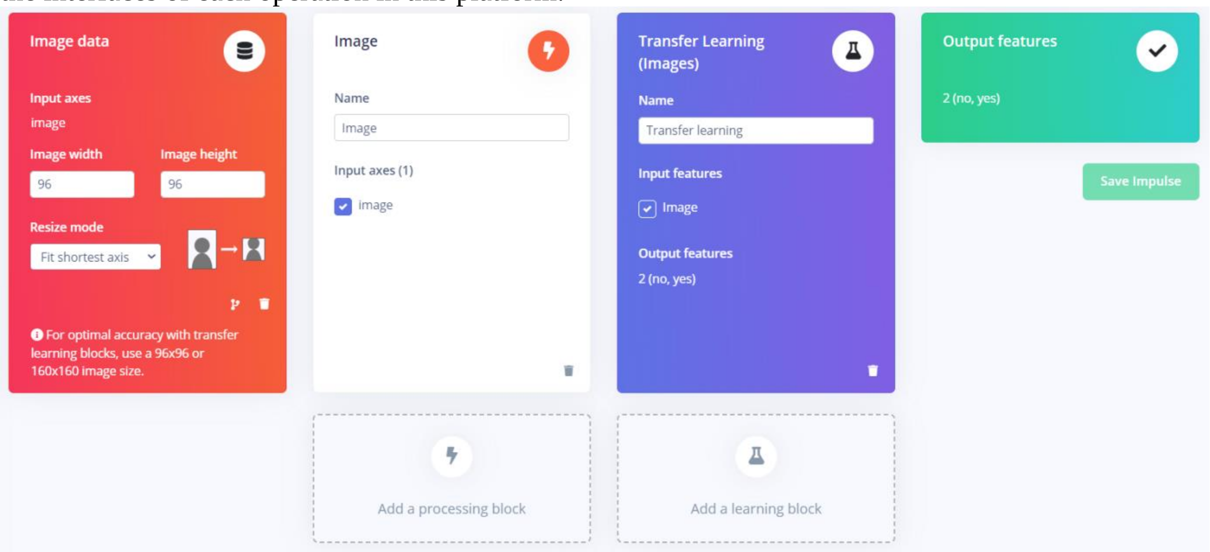

Figure 2. The interfaces of each operation in Edge-Impulse platform.

2.3. *Implementation details*

Majority of setting of training is default values and not modified because their performance is already really nice. Exceptionally, data argumentation is turned on to improve the accuracy. Figure 3 show default values of some of the most significant parameters.

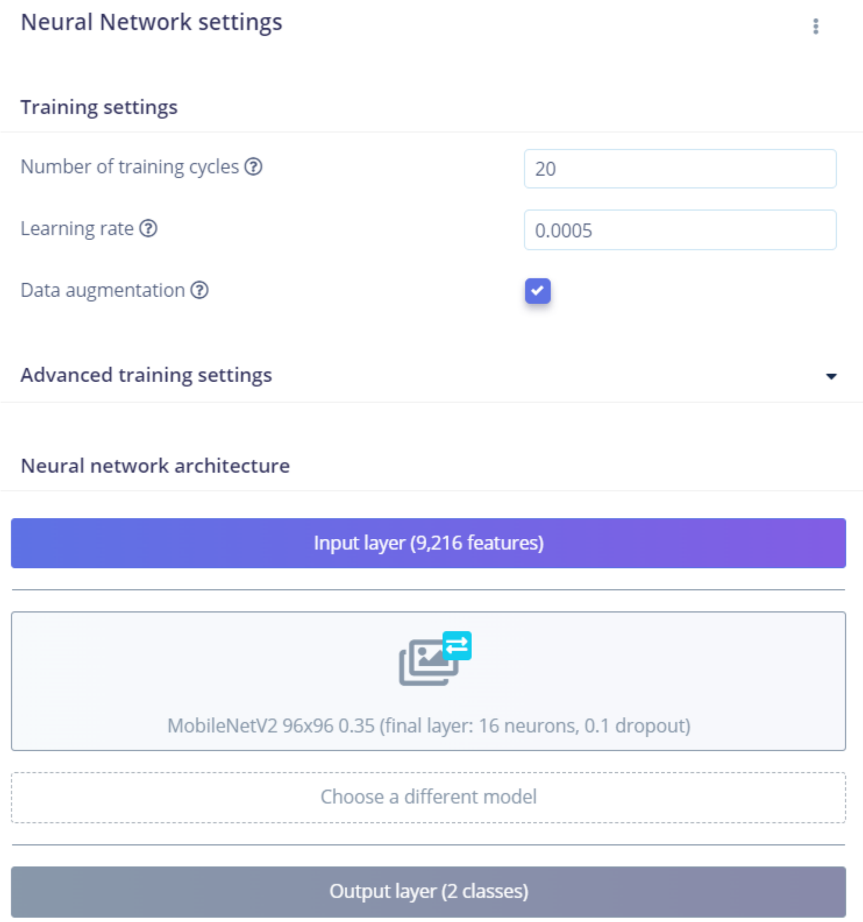

Figure 3. Default values of parameters.

3. Result and discussion

By applying aforementioned methods and parameters, the performance of the model can be obtained. Additionally, it is essential to determine the best image size suiting the model. So, image sizes are adjusted while other parameters are not changed, and different results are obtained for each size. The corresponding results are provided in Table 1.

Table 1. The performance of the model based on various image sizes.

	Accuracy	*Loss*	*F1 Score*	*Inferencing Time*	*Peak RAM Usage*	*Flash Usage*
96×96	96.70%	0.24	0.97	102ms	333.7K	579.5K
128×128	97.30%	0.16	0.97	171ms	503.2K	579.8K
160×160	97.70%	0.15	0.98	271ms	720.7K	580.1K

Table 1 demonstrates that performances for all of image sizes are notably outstanding, showing the advantage of detection of brain tumors by means of machine learning algorithms. As image sizes getting greater, the accuracy increase, and the loss decrease a bit. Meanwhile, time, RAM and flash

DOI: 10.54254/2755-2721/18/20230986

usage consumed also increase a lot, requiring higher-quality instruments. In summary, image size of 96×96 pixels are the best choice for most cases due to its low space and time consumption compared with other two.

According to research, the degree of agreement among experts in manually diagnosing brain tumors ranges from 90% to 95%. However, for mixed categories of tumor, such as mixed glioma and medulloblastoma, the level of disagreement among experts decreases further to 77% and 58%, respectively [11]. Therefore, the implementation of machine learning for brain tumor diagnosis is deemed to be an efficient and reliable approach. It is believed that this technology will be utilised more widely in fields of clinical medicine, assisting doctors to select the best and most suitable way to treat patients.

4. Conclusion

In this study, an automated brain tumor diagnosis system is developed using a machine learning approach. Specifically, transfer learning is employed to enhance the performance of the model. Experiments are conducted to assess the feasibility and benefits of this method. The findings demonstrate that the diagnostic accuracy achieved by the model is superior to that of manual diagnosis. Future research can focus on optimizing the model by reducing its space and time complexity without compromising accuracy. This holds great potential for clinical applications, as the proposed model can potentially replace traditional diagnostic methods and facilitate more efficient and accurate brain tumor diagnosis.

References

[1] Cancer Net 2023 Brain tumor: Statistics https://www.cancer.net/cancer-types/brain-tumor/statistics#:~:text=The%205%2Dyear%20relative%20survival%20rate%20for%20people%20younger%20than,brain%20tumor%20every%205%20years.
[2] Kaggle 2022 Br35H:: Brain tumor detection 2020 https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection?select=Br35H-Mask-RCNN
[3] Mihigo I N Zennaro M Uwitonze A et al 2022 On-Device IoT-Based Predictive Maintenance Analytics Model: Comparing TinyLSTM and TinyModel from Edge Impulse Sensors 22(14): 5174.
[4] Qiu Y Yang Y Lin Z et al 2020 Improved denoising autoencoder for maritime image denoising and semantic segmentation of USV China Communications 17(3): 46-57
[5] Arena P Basile a Bucolo M et al 2003 Image processing for medical diagnosis using CNN Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 497(1): 174-178
[6] Kayalibay B Jensen G van der Smagt P 2017 CNN-based segmentation of medical imaging data arXiv preprint arXiv:1701.03056
[7] O'Shea K Nash R 2015 An introduction to convolutional neural networks arXiv preprint arXiv:1511.08458
[8] Gu J Wang Z Kuen J et al 2018 Recent advances in convolutional neural networks Pattern recognition 77: 354-377
[9] Albawi S Mohammed T An Al-Zawi S 2017 Understanding of a convolutional neural network 2017 international conference on engineering and technology (ICET) Ieee 1-6
[10] Sun M Song Z Jiang X et al 2017 Learning pooling for convolutional neural network Neurocomputing 224: 96-104
[11] Naseer A Yasir T Azhar A et al 2021 Computer-aided brain tumor diagnosis: performance evaluation of deep learner CNN using augmented brain MRI International Journal of Biomedical Imaging