

Brain tumor classification using CNN: Difference between optimizers

Yizhou Li

Shanghai Starriver Bilingual School, Shanghai, 201108, China

15010340135@xs.hnit.edu.cn

Abstract. The classification of brain tumor using artificial intelligence has long been a heated topic and never failed to arouse public attention, being both efficient and useful. Each year, around nearly 12000 people are diagnosed with brain tumor, and the AI approach allows a number of specific cases to be identified quickly with few errors, contributing in saving millions of lives. This paper uses CNN (ResNet50 architecture) for classification and evaluates the performances of three kinds of optimizers – adaptive moment estimation (Adam), stochastic gradient descent (SGD), and genetic algorithm (GA) – when being applied to the model. The resulting accuracy scores are, respectively, 93%, 90%, and 95%, which demonstrates that genetic algorithm performs the best, suggesting a fine choice of utilization in practical scenarios. The results as well show the most precise diagnosis on pituitary tumor and the least on meningioma tumor, providing a direction for future improvement on dataset and training parameters.

Keywords: convolutional neural network, brain tumor classification, optimizers, comparison.

1. Introduction

As artificial intelligence (AI) is being applied to medical imaging for diagnosis, prognosis, and clinical workflow [1], artificial intelligence (AI) is serving a progressively more important role in clinical area.

AI-based computer-aid diagnosis (CAD), which is to use AI technology on diagnostic imaging, is already widely applied. Eye diseases as well as diabetic retinopathy and hypertensive retinopathy could be diagnosed from fundus photographs; images of skin lesions have been used by a research group from Stanford University in 2017 to diagnose skin cancer; CAD for breast imaging and chest imaging have as well reached impressive accuracy equivalent to or even better than an expert [2].

A widely used method in CAD is machine learning, a significant road to AI, predicting results through learning from input examples (features) that are labelled with expected category names [1]. Supervised machine learning is one commonly used subcategory, in which designed algorithms are used to learn from databases and generate correct results with new inputs in the same category with the learned ones [1].

Deep learning, a branch of machine learning, is a specific method that performs well when dealing with CAD. It involves using neural networks to explore sophisticated structures and relationships between layers of features and labels in large datasets [3]. Through multiple processing layers, deep learning brings advances to tasks from object recognition to genomic works [3].

Numerical cases of CAD have reached outcomes of high quality due to the use of deep learning. Ting et al. used this method to recognize diabetic macular edema (DME), choroidal neovascularization

(CNV), and drusen from optical coherence tomography (OCT) images [4]; and pediatric pneumonia from children chest X-rays (CXR) images [4]. Jin et al. as well used deep learning to screen COVID-19 from chest images obtained by computed tomography (CT) [5].

This study's goal is to compare the effects three different optimization methods on using deep learning, specifically, CNN, to diagnose and classify no-tumor brain and the three types of brain tumor: meningioma tumor, glioma tumor, and pituitary tumor. Results will be evaluated according to accuracy, confusion matrix and the AUC score for the ROC curve.

In the following section, the report will introduce the dataset used, how data are preprocessed, specific algorithms, optimizations and equations used for classification, and definitions for evaluation metrics. In the third section, training results will be explained, including a detailed comparison of the three optimizers. The fourth section will discuss possible reasons for the results, the study's contribution, and improvements that could be made in the future. The last section will demonstrate the overall conclusion of this study.

2. Method

The whole process of this study is illustrated in Figure 1. The input dataset is preprocessed, and a CNN (specifically ResNet50 architecture) model is trained using various optimization methods. After that, the trained model will predict the labels of the test dataset, and results will be evaluated to compare the effects of optimization methods.

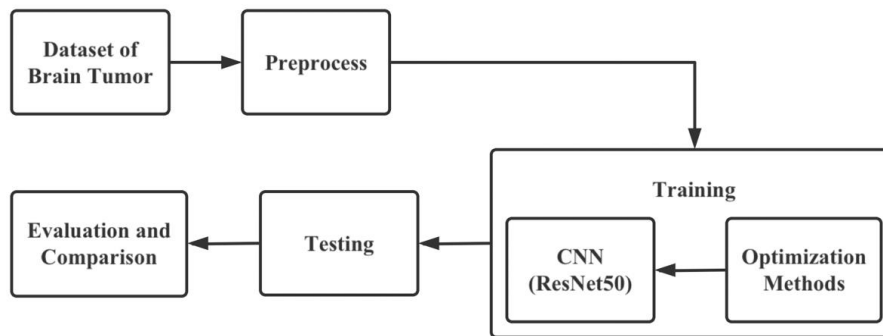


Figure 1. Project pipeline.

2.1. Dataset

This research uses the train dataset from Kaggle (accessed [here](#)), which contains 2870 images. They include four types of brain tumors detected by Magnetic Resonance Imaging (MRI): meningioma tumor, glioma tumor, pituitary tumor, and no tumor. The following four subsections will describe each category of brain tumor images in detail.

2.1.1. Meningioma Tumor. A meningioma tumor is one that exists in the membranes that surround the brain and spinal cord [6]. Being the most common type of brain tumor, it can cause severe effects on nearby tissues, nerves, or vessels [6]. An example is shown in Figure 2.

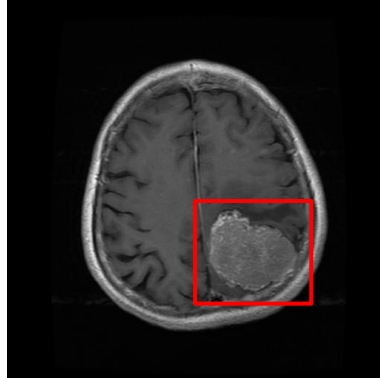


Figure 2. This is an MRI image of meningioma tumor, in which the place of the tumor is framed by a red box.

2.1.2. Glioma Tumor. The brain and spinal cord are where a glioma tumor might exist. The glioma tumor begins in the glial cells that surround nerve cells [7]. Being the one of the most common types of brain tumor, it can influence brain functions and even be life-threatening depending on its location and growth rate [7]. Examples of it are shown in Figure 3.

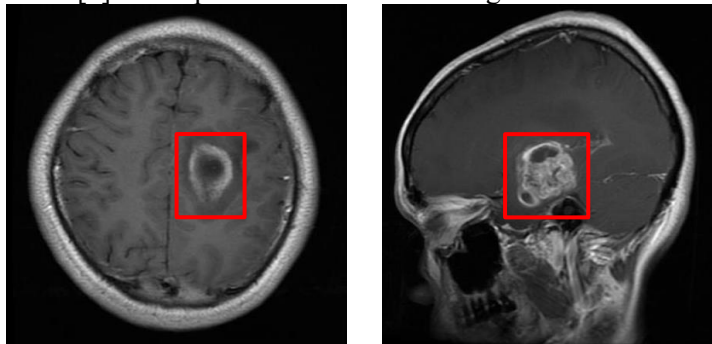


Figure 3. These are two kinds of MRI images of glioma tumor contained in the dataset, in which the places of the tumors are framed by red boxes.

2.1.3. Pituitary Tumor. A pituitary tumor is one that occurs in the pituitary gland [8]. Sometimes it could cause abnormal increases or decreases the amount of hormone that the pituitary gland produces, affecting body functions [8]. On the other hand, most pituitary tumors are benign – they do not spread to other body parts [8]. An example of it is shown in Figure 4.

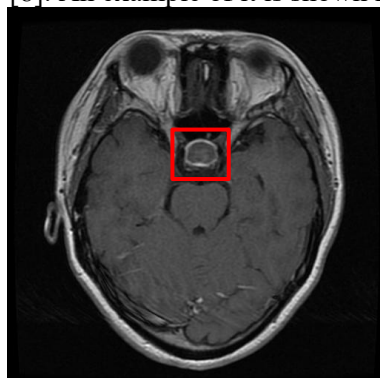


Figure 4. This is an MRI image of pituitary tumor, in which the place of the tumor is framed by a red box.

2.1.4. *No Tumor*. An example of a brain with no tumor is shown in Figure 5.

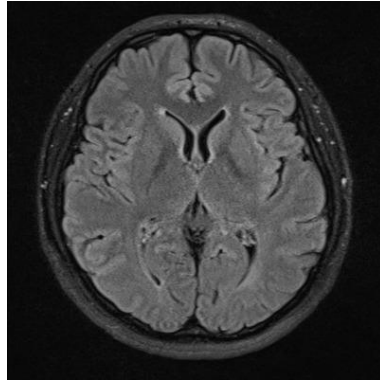


Figure 5. This is an MRI image of a normal brain with no tumor.

2.2. Preprocessing

Images are read, resized into 100*100 pixels, and converted into numpy array type. Below are two examples of the resizing process.

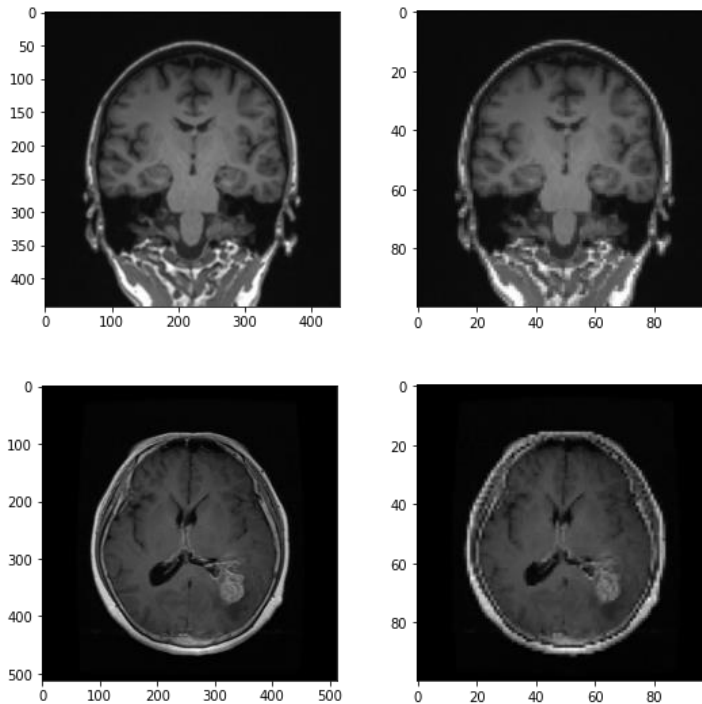


Figure 6. An example of processing an image that contains a brain with no tumor. The raw image is shown left, with size of (442, 442), the processed image shown right, with a size of (100, 100).

Figure 7. An example of processing an image that contains a brain with glioma tumor. Raw image is shown left, with a size of (512, 512), the processed image shown right, with a size of (100, 100).

Other than this, the train dataset is split into train data and test data with a ratio of 3:1.

2.3. Convolutional Neural Network (CNN)

Convolutional neural network (CNN) is a widely used deep learning algorithm that could automatically and gradually learn characteristics of data according to the input. In the input, an image will appear as a two-dimensional grid of pixel values [9]. Through minimizing the loss value, which demonstrates the extent of the model agreeing with trained data, after each training round, the training effect is improved.

A CNN architecture typically consists of convolution layers, pooling layers, and dense layers (fully connected layers), one layer's output serving as the next layer's input. The first two types extract significant features from the input, and the third produces output according to these features [9]. An activation function is applied at the end of each convolution layer and fully connected layer.

In a convolution layer, a kernel, a small matrix of values, is applied to the input grid, and they perform matrix multiplications to form feature maps (outputs), representing the input's various characteristics [9].

A pooling layer reduces the size of inputs through a filter, which, resembling a kernel, is a small grid of values. As it moves in a stride assigned by the model builder, each area covered by it returns one sole value to represent its most typical characteristic. Max pooling is the most popular type of pooling layer, and it returns the maximum value in each area.

Finally, a one-dimensional array will be transformed from the outputs, then being connected to dense layers. This process involves the association of inputs with outputs through learnable weights [9]. In classification tasks, the size of final outputs is equal to the number of categories; the outputs represent the possibilities of each class being identified as the one that the input belongs to.

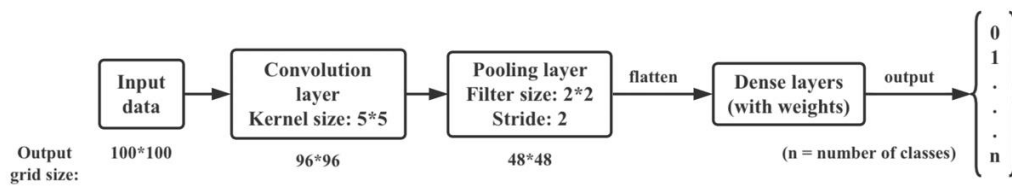


Figure 6. An example of CNN model.

2.3.1. ResNet50. This study uses ResNet50 (residual network) architecture, one of the most used, for CNN model training. This architecture aims to prevent the problems caused by vanishing gradient (which updates weights too little and makes the model cease to improve) or exploding gradient (which updates weights too much and makes the model unstable). Whenever the layers in the middle make the model performance worse, the input x will pass through a shortcut path, skipping the layers and directed added to the output. This ensures that the effect does not lessen as the layer becomes higher.

2.4. Optimization Algorithms

Optimizers are known as the “engine” of neural networks. They help adjust the model according to the output of the loss function each iteration in order to improve accuracy. In this study, three kinds of optimizers are utilized and compared.

2.4.1. Stochastic Gradient Descent (SGD) Optimizer. The SGD optimizer randomly selects one sample in every iteration and move parameters in the direction of the loss function's gradient so that the minimum of the loss function could be found. The small size of the samples the optimizer uses effectively reduces the algorithm's computational cost.

2.4.2. Adaptive Moment Estimation (Adam). Adam accelerates SGD optimizer, adapting to different learning rates. To reach its goal, it stores both exponentially decreasing weighted average of past gradients and exponentially decreasing weighted average of past squared gradients.

2.4.3. Genetic Algorithm (GA). Genetic algorithm is inspired by the evolution process in nature, motivated by Darwin's theory of natural selection [10], which involves mutations and crossing over the fittest chromosomes. The initial population is a random group of individuals (“chromosomes”); then, each is applied with a fitness function that trains the model, using the accuracy score for optimization. It then applies selection, crossover, mutation (all mimicking the biological process in reproduction) to this population in order to produce the next generation [10], and this whole process is repeated until the best accuracy value is achieved.

2.5. Evaluation Metrics

The evaluation metrics introduced here are accuracy, values in the confusion matrix, and receiver operating characteristic (ROC) curve. The meaning will be explained, and related formulas will be listed.

2.5.1. Accuracy. The accuracy score is the percentage of correct predictions among all predictions.

2.5.2. Confusion Matrix: Precision, Recall, and F1 Score. The precision score is the percentage of correct (TP) predictions among the sum of correct and false positive (FP) ones, expressed as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

The recall score (true positive rate, TPR) is the percentage of correct predictions among the sum of correct and false negative (FN) ones, expressed as

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

The equation of F1 score is

$$\text{F1} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

2.5.3. Receiver Operating Characteristic (ROC) Curve and AUC Score. The ROC curve helps visualize the relationship between the rate of correct (TP) results (plotted on y-axis) and the rate of FP results (plotted on x-axis) under different decision thresholds, which determine if a predicted result should be “true” or “false.” Model performance could be rated by the area under the curve (AUC): a higher AUC corresponds to a better model.

3. Results

In general, genetic algorithm (GA) functions the best, and stochastic gradient descent optimizer (SGD) functions better than adaptive moment estimation (Adam). Details of comparison are illustrated below.

3.1. Accuracy

Table 1. A chart for the resulting accuracy of the three optimizers.

	SGD	Adam	GA
Accuracy	0.93	0.90	0.95

As shown in Table 1, the optimizer that uses genetic algorithm yields the highest accuracy, which implies that it most effectively contributes to the model’s performance.

3.2. Confusion Matrix

Confusion matrixes resulting from the three optimizers are shown in the tables below.

Table 2. Confusion matrix after the use of SGD optimizer.

	Precision	Recall	F1 score
Glioma tumor	92%	97%	94%
Meningioma tumor	92%	84%	88%
No tumor	98%	85%	91%
Pituitary tumor	92%	100%	96%

Table 3. Confusion matrix after the use of adaptive moment estimation.

	Precision	Recall	F1 score
Glioma tumor	96%	87%	91%
Meningioma tumor	77%	95%	85%
No tumor	100%	68%	81%
Pituitary tumor	96%	99%	97%

Table 4. Confusion matrix after the use of genetic algorithm.

	precision	Recall	F1 score
Glioma tumor	94%	95%	95%
Meningioma tumor	93%	90%	91%
No tumor	93%	92%	92%
Pituitary tumor	98%	100%	99%

Among the confusion matrixes, the one after the utilization genetic algorithm contains the best result, in which all data are above 90%. Moreover, in the one with SGD and GA, the recall scores of the pituitary tumor class are both 100%, showing that in both cases, all pituitary tumors are correctly identified, which illustrates a quality performance of these two optimizers.

However, the precision score of meningioma tumor class in the one with Adam is much lower than that of other classes, the same case as the recall score of no tumor class in the same chart, which indicates that many other types of brain tumor are misidentified as meningioma type and that many images that actually show no brain tumor are wrongly diagnosed as existence of brain tumor, implying for poor performance of Adam. What is more, Adam optimizer appears to be relatively unstable as the F1 scores of the categories in its result vary from 81% to 97%.

3.3. AUC Score

Table 5. A chart for the resulting AUC score of the three optimizers.

	SGD	Adam	Genetic Algorithm
AUC score	94.57%	91.90%	96.22%

From the AUC scores shown in Table 5, it can be inferred that genetic algorithm brings about the best performance of the model since it has the highest AUC score.

4. Discussion and Future Work

Some of the low scores in the results might have been caused by features of the input data, such as ambiguity of characteristics. However, an overall trend of performance could still be observed. Through evaluation of three optimizers' performance on brain tumor MRI image classification, this study points out a possible direction of optimizer selection in practical cases.

In the future, more effort could be put into preprocessing data, for instance using the MRI image transformation method devised by Nyúl et al[11], so that potential errors that exist in the raw data could be eliminated or reduced. Furthermore, parameters involved in the training process could be adjusted to improve the accuracy.

5. Conclusion

The study classifies brain tumor types using MRI images of meningioma tumor, glioma tumor, pituitary tumor, and normal brain without tumor. The model training process involves using convolutional neural network (ResNet architecture specifically), three optimization algorithms (stochastic gradient descent, adaptive moment estimation, and genetic algorithm), which are evaluated and compared in the end. The

accuracy results are 93%, 90%, and 95%, respectively. After a comparison of accuracy, confusion matrixes, and AUC scores, it is decided that genetic algorithm best helps the model to function. This implies a viable optimizer choice when using artificial intelligence to diagnose brain tumors in clinical scenarios, especially if the dataset is well-preprocessed and training parameters ideally set.

References

- [1] Rajkomar A, Dean J and Kohane I 2019 Machine learning in medicine *N Engl J MedNew* **380** 1347-58.
- [2] Fujita, H 2020 AI-based computer-aided diagnosis (AI-CAD): The Latest Review to read first Radiological Physics and Technology 13 6-19.
- [3] LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* 521 pp 436-44.
- [4] Ting D S, Liu Y, Burlina P, Xu X, Bressler N M and Wong T Y 2018 AI for Medical Imaging Goes Deep *Nature Medicine* 24 pp 539-40.
- [5] Wang B, Jin S, Yan Q, Xu H, Luo C, Wei L, Zhao W, et.al. 2021 AI-assisted CT Imaging Analysis for COVID-19 screening: Building and deploying a medical AI system *Applied Soft Computing* 98 106897.
- [6] Mayo Foundation for Medical Education and Research 2022 Meningioma Mayo Clinic.
- [7] Mayo Foundation for Medical Education and Research 2020 Glioma Mayo Clinic.
- [8] Mayo Foundation for Medical Education and Research 2021 Pituitary Tumors Mayo Clinic.
- [9] Yamashita R, Nishio M, Do R K and Togashi K 2018 Convolutional Neural Networks: An overview and application in Radiology *Insights into Imaging* 9 pp 611-29.
- [10] Weile D S and Michielssen E 1997 Genetic algorithm optimization applied to electromagnetics: a review *IEEE Transactions on Antennas and Propagation* 45 pp 343–53.
- [11] Nyúl L G, Udupa J K and Zhang X 2000 New variants of a method of MRI scale standardization *IEEE Transactions on Medical Imaging* 19 pp 143–50.