Proceedings of CONF-SEML 2025 Symposium: Machine Learning Theory and Applications
DOI: 10.54254/2755-2721/154/2025.TJ23213

Brain Tumour Classification Algorithm Based on Multi-scale Convolution and Attention Mechanism

Yujie Xie

School of Sciences, Chang'an University, Xi'an, China
2023901763@chd.edu.cn

Abstract: This paper proposes a classification algorithm that integrates a multi-scale convolutional structure and adaptive attention mechanism (SEblock) for the automatic classification technology of MRI brain tumors, which is used to extract high-dimensional features and recognize images that cannot be observed by the human eye. The model mentioned in this paper is improved on the pre-trained resnet50 network, and two smaller and larger datasets are used to improve the generalization ability and anti-interference data ability of the model, and the convolutional kernel of 3*3, 5*5, and 7*7 is fused to obtain receptive fields of different scales and increase the feature extraction ability of different fine-grained features. The SEblock module is added to increase the perception of key feature channels and improve the accuracy of classification. On the two datasets published by Kaggle, the model achieves an accuracy of 79.93% and 85.35%, respectively, indicating that the model has a certain degree of accurate classification ability. The existing models still have many deficiencies in accuracy, and more improvements will be introduced in the attention mechanism in the future.

Keywords: Brain Tumor, Multi-scale convolutional structure, Attention mechanisms, Deep Learning, Classification

1. Introduction

A brain tumour is a collection or mass of abnormal cells in the human brain, which seriously threatens the lives and health of people around the world [1]. It is classified into benign brain tumors, including meningioma, pituitary adenoma, neurofibroma (Schwannoma), etc.) and malignant brain tumors (including glioma and primary brain cancer). There are significant differences in biological characteristics, aggressiveness, response to treatment, and prognosis among brain tumors, while benign and malignant tumors vary greatly in histological characteristics of growth rate and location [2]. Magnetic Induction Resonance (MRI) has achieved a diagnostic accuracy of 93% in diagnosing head injury, which is much higher than the 73% accuracy of CT [3]. MRI is an important means of modern medicine for the diagnosis of cranial diseases. If the type of disease of the patient can be quickly identified and treated in a timely and effective manner, it may provide better medical care to the patient. Therefore, the classification of different types of brain tumors based on MRI has become an important topic in the current medical field.

At present, the classification of brain tumors based on MRI images mainly uses CNN (Convolutional Neural Network) and transformer as the basic models. Ayesha Younis et al. used a model of CNN binding to VGG-16 to classify MRI images of brain tumors, which was superior to

DOI: 10.54254/2755-2721/154/2025.TJ23213

conventional detection methods at the time [4]. Amin ul Haq et al. used an improved multi-level CNN model and improved the classification results of the MCNN model using data augmentation and transformer methods, achieving an accuracy of 99.89% [5]. Yu et al. used the multi-scale feature extraction submodule to extract more fine-grained and local information and established the multi-scale feature relationship of the model through the attention mechanism, reaching an accuracy rate of 98.13% [6]. H Gholamalinejad et al. proposed a new pooling method based on Haar transform in real-time vehicle classification, which paid more attention to important details such as edges in the image and improved the recognition rate of the IRVD dataset from 97.12% to 99.06%. In addition, he used the SEBlock method to connect the traditional convolutional layer with the SE block, which helped to readjust the feature response of the channel and improve the accuracy [7]. Although the above methods can effectively improve the training accuracy, there are still some technical problems and challenges when the dataset becomes more mixed and the MRI scan angle changes. This paper aims to combine the multi-scale feature extraction method mentioned by Yu Zhiheng et al. and the method of connecting convolutional layers with SE blocks mentioned by H Gholamalinejad et al. to classify brain tumor MRI images in the dataset [6, 7].

In this paper, we use the model of mixed feature extraction and channel attention adaptation and use Resnet50 as a pre-trained network to classify glioma, meningioma, pituitary tumor, and no tumor on 2800 datasets and 24723 datasets published on Kaggle, respectively. The purpose of this paper is to effectively improve the work efficiency of doctors and reduce the impact of a lack of work experience on the MRI classification model of brain tumors based on deep learning. Assist remote areas and grassroots hospitals to improve diagnostic capabilities, achieve medical equity, reduce labor costs through automation, improve service efficiency, and save medical costs.

2. Method

2.1. Dataset

A total of 27,523 images from Kaggle's publicly available MRI brain tumor dataset were used in this experiment, which has extensive user reviews to ensure its quality, and the authors recently updated the dataset in September 2024. There are 6510 glioma and 6599 meninglioma, 7385 notumor, and 7029 pituitary, which are roughly divided into training, validation, and testing sets in a 7:2:1 ratio. I also selected a smaller dataset (about 2,800 photos) provided by the Eindhoven University of Technology to evaluate the model's classification ability with a small amount of data.

2.2. Materials and tools

In this paper, a convolutional neural network based on multi-scale mixed features and ResNet50 was used to classify brain tumors, and the framework of TensorFlow 2.9.1 was used. Resnet50 is a deep neural network that has been pre-trained on imagenet, which overcomes the gradient vanishing problem in the deep network through residual connection, which is convenient for us to connect more layers of neural networks in the future, and freeze its convolutional layer to focus on the training of subsequent layers on the basis of maintaining the pre-training weight. The convolution kernels of different scales in the multi-scale convolutional kernel can capture the image information of different sizes, extract richer features in the image, and finally stitch together the results of different scales, which enhances the perception ability of the model at different scales and is conducive to improving the expression of complex image features. SEBlock is an attention mechanism that captures information in the entire image dimension by pooling the average of each feature map and then assigns different weights to each feature map through a fully connected layer to improve the focus on key features. In this paper, different datasets are used for combination, and the ImageDataGenerator in Keras is used to normalize the images. Considering the problem that there are different views of

DOI: 10.54254/2755-2721/154/2025.TJ23213

MRI brain detection images in the dataset, this paper uses translation, rotation, and horizontal flipping data augmentation methods for the training set data. In this way, the translation, rotation, and left and right invariance of the model can be improved, the robustness and generalization ability of the model can be improved, and more attention can be paid to the essential characteristics of the image.

2.3. Method

This paper first uses the resnet50 network as the backbone structure and removes the classification layer, retaining only the convolution layer and the fully connected layer to pre-train the model. This pre-training saves about 20m parameters for the model, and it outputs a feature vector of shape (batch size, 7, 7, 2048) to the multi-scale convolution layer. In the multi-scale convolution layer, this paper uses three types of convolution kernels: 3*3, 5*5, and 7*7 for convolution. Each convolution kernel extracts features of different scales for output. Through these three different sizes of convolution kernels, the model can extract features from different scales of perspective, ensuring that the model has stronger robustness when processing images of different sizes and details. L2 regularization is added to each convolutional layer and connected to the Squeeze-and-Excitation module, which adaptively adjusts the weights of each channel so that different scales of data reflect different degrees of importance in the model. Compared with the Attention Mechanism, seblock is a more lightweight attention mechanism, which means that it only requires a smaller amount of computation to achieve the matching problem of each channel weight. After splicing the features of these outputs, GlobalAveragePooling2D is used to carry out global pooling for each channel, and a two-dimensional vector is connected to the pooling layer, flattened and connected to the fully connected layer, and output is carried out through a dropout regularization to prevent overfitting.

3. Result

In this paper, the convolutional neural network model CNN obtained by three convolutional layers is used, and an accuracy rate of 68.53% is obtained on the basis of 10 epochs. Later, the data augmentation method of mixup was added in this paper, but the accuracy rate decreased by about 14 percentage points (54.57%), After analysis, it is found that mixup data enhancement improves the generalization ability of the model, but it may cause the loss of key features such as the location, size, and morphology of the tumor, and the images generated by mixup are not necessarily pure category mixtures, which may cause the model to fail to learn consistent category information. In addition, in a smaller data set, the mixup method may make it difficult for the model to learn the details of each category. Intuitively speaking, it may cause different brain tumors to overlap or cause multiple brain tumors to appear in different locations in one image, thus affecting classification. Through the method proposed in the method, we finally get an accuracy rate of 79.93%, which is 10.18% higher than that of ordinary CNN networks. Table 1 shows the accuracy of different models under two datasets; as can be seen, the left column is the accuracy of 2800 datasets, and the right is the accuracy of 24723 datasets mentioned above. Compared with CNN and the common MultiScaleConv model, the model in this paper has a higher accuracy, and the obvious accuracy loss on a larger dataset is less (only 5.41% accuracy loss), indicating that the model can resist interference to a certain extent.

Table 1: The performance of the model under different datasets

Model	Accuracy	Model	Accuracy
CNN	69.80%	CNN	54.47%
Multi-scale convolution	68.53%	Multi-scale convolution	55.38%
Mixup Data augmentation	64.46%	Mixup Data augmentation	48.90%
this model	85.35%	this model	79.94%

4. Discussion

Compared with the standard Multi-Scale CNN, the accuracy rate is 96.62% [7]. Although the accuracy of this model is insufficient, the amount of computation required is relatively small. The existence of the SEBlock module provides about 2% accuracy for the model, and there may be some misrepresented distractors in a large number of datasets that affect the training of the model, in a smaller but more accurate training set with a total of 2800 images, the accuracy of the model reaches 85.35%, and we will introduce more Vision Transformer methods in the future, such as Convolutions to used by HaipingWu and others Vision Transformers to further improve accuracy on larger datasets [8]. A study by Chenjie Ge that can improve the semi-supervised learning method of the model has shown great advantages in the study, which can solve the problem of unlabeled MRI datasets in the current large number of MRI datasets. He proposed a graph-based semi-supervised learning method, which feeds both labeled and unlabeled training data into GAN, and the discriminator evaluates and provides feedback through the pseudo-label images generated by the generator. The generator is continuously optimized to produce more realistic images, and these pseudo-label images are eventually used for further training. It achieved an accuracy rate of 90.70% on the MICCAI dataset [9]. In addition, the current MIR dataset still has the problem of uneven number of images of different types of brain tumors, and Aninday Nag et al. proposed in their paper to use GAN to generate new images, so as to obtain pictures with similar data volume in different types of brain tumors, and GAN can simulate the real lesion structure, which can improve the model's ability to identify rare categories or complex features [10].

5. Conclusion

In this study, the MultiScaleConv model using resnet50 as pre-training and SEBlock for channel attention weight allocation finally achieved an accuracy of 79.94% in 24723 datasets and 85.35% in 2800 datasets. In this paper, it is experimentally tested that the mixup method may inhibit the training of the MRI brain tumor classification model. Although the model used in this paper does not have an accuracy advantage over common tumor classification models. However, the training of this model on a large dataset has certain universality, and the data contains multiple perspectives, which can classify MRI brain tumor pictures in a more comprehensive manner. In subsequent experiments, we will introduce more models that can be used for classification, adjust more reasonable parameters, and make some adjustments to the attention mechanism.

References

[1] Liu, M., & Zhang, G. (2021). Research progress on brain tumor imaging based on deep learning. Imaging Technology, 33(5), 50–55.

[2] Zhang, W. (2024). Research on brain tumour segmentation and classification method based on MRI images (Master's thesis, Southwest University). CNKI.

[3] Qian, D. (2025). The application value of cranial magnetic resonance imaging in the diagnosis of head injury after neonatal asphyxia. Journal of Practical Medical Imaging, 26(1), 57–59.

[4] Younis, A., Li, Q., Nyatega, C. O., & Kawuwa, H. B. (2022). Brain tumor analysis using deep learning and VGG-16 ensembling learning approaches. Applied Sciences, 12, 7282.

[5] Haq, A. U., Li, J. P., Kumar, R., Ali, Z., Khan, I., Uddin, M. I., & Agbley, B. L. Y. (2023). A multi-level CNN model for the classification of brain tumors in IoT healthcare system. Journal of Ambient Intelligence and Humanized Computing, 14(6), 4695–4706.

[6] Yu, Z., Li, X., Liu, Q., & Wei, B. (2025). Brain tumour classification algorithm based on mixed features and channel attention. Computer Engineering and Design, 46(2), 610–616.

[7] Gholamalinejad, H., & Khosravi, H. (2021). Vehicle classification using a real-time convolutional structure based on DWT pooling layer and SE blocks. Expert Systems with Applications, 183, 115420.

DOI: 10.54254/2755-2721/154/2025.TJ23213

[8] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 10012–10022).
[9] Wu, H., Xiao, B., Codella, N., Liu, M., Dai, X., Yuan, L., & Zhang, L. (2021). CvT: Introducing convolutions to vision transformers. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 22–31).
[10] Ge, C., Gu, I. Y. H., & Jakola, A. S. (2020). Deep semi-supervised learning for brain tumor classification. BMC Medical Imaging, 20, 87.
[11] Nag, A. (2024). TumorGANet: A transfer learning and generative adversarial network-based data augmentation model for brain tumor classification. IEEE Access, 12, 103060–103081.