Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU33590

Research on Lightweight CNN Model for MRI Image Analysis Based on Edge Impulse Platform

Changhao Chen

Department of Electronic and Information Engineering, Wuyi University, Jiangmen, China
a81428293@gmail.com

Abstract. Brain tumors are an extremely dangerous disease that is lethal to mental health. MRI is one of the most significant methods for diagnosing this disease. To diagnose brain tumors accurately, doctors need to maintain a high level of concentration for a long time to check MRI images, which may cause visual fatigue and lead to misdiagnosis. Computer vision, machine learning, and deep learning are being introduced to medicine in a gradual way due to their development, which alleviates visual fatigue for doctors and decreases the chances of misdiagnosis. However, there are still deficiencies in lightweight model deployment and resolution adaptability. In this paper, the research is based on the Edge Impulse platform, and MobileNetv2 lightweight convolutional neural networks are used to construct an MRI images classification models. In the process of dataset preparation, the images were adjusted, and intensity normalization was performed to eliminate the impact of different MRI image parameters. To discover the impact of model capacity and input resolution, diverse input size and width multiplier (96×96 0.05, 96×96 0.1, 96×96 0.35, 160×160 0.35 and 160×160 0.5)were set to conduct a comparative test. The training accuracy, training loss, peak RAM, and test accuracy were regarded as the key evaluation indicators for the models. Training results show that the model with input size of 160×160 and width multiplier of 0.5 yield the best classification performance, which achieved an ideal balance between lightweight and accuracy.

Keywords: MRI, Machine Learning, Deep Learning, Edge Impulse

1. Introduction

The brain is the organ with the most importance and complexity in the human body, which composed of billions of cells, controlling humans' awareness and neural activation [1]. The growth of uncontrolled and irregular brain tissue contributes to a brain tumor, which is a serious disease. The mortality of brain tumors is very high. There are approximately 7 to 11 cases of brain tumors per 100000 people of all ages each year [2]. Only in the US, a brain tumor is diagnosed annually by more than 88000 adults and 5500 children, and only 35.6% of adults survive five years after being diagnosed with a malignant brain or other CNS tumor [3]. The brain tumors can be classified into benign tumors with low aggressiveness and malignant tumors with high aggressiveness [4]. Primary brain tumors originate in the brain tissues, whereas secondary tumors are metastases from other organ cancers through the blood flow. In primary brain tumors, meningioma, glioma, and pituitary

DOI: 10.54254/2755-2721/2026.GU33590

are three typical detrimental categories of brain tumors, which are intractable in early diagnosis and effective treatment; untimely treatment can lead to serious consequences [5].

The diagnosis of a brain tumor is challenging due to the complexity of the brain. Diagnosing the brain tumor precisely is the crucial part of the early treatment. At the present stage, various applications can be achieved by MRI, including imaging the musculoskeletal system, cardiovascular system, and specifically the central nervous system and its neural subsystems [6]. MRI is the preferred technique for confirming brain tumors because it has excellent contrast resolution and can obtain anatomical details without exposure to ionizing radiation [7].

Mostly, magnetic resonance imaging (MRI) visual judgment by doctors determines the identification of a brain tumor; the doctor's prolonged diagnosis is likely to result in an erroneous judgment because of visual fatigue. In order to reduce the occurrence of such situations, computer-aided algorithms are gradually employed in diagnosis. The development of deep learning (DL) and artificial intelligence (AI), particularly convolutional neural networks (CNNs), has given medical imaging analysis a distinct advantage [8].

This study proposes a deep learning method for classifying brain tumors using MRI scans. The model is trained on the edge impulse platform, which offer tools to preprocess data, train models, and evaluate performance. The dataset includes four classifications of brain MRI images: glioma, meningioma, pituitary tumor, and normal ones. CNNs is employed to learn and identify different features of four classes. The trained model is evaluated through accuracy and a confusion matrix to verify its classification capability.

2. Methods

2.1. Dataset preparation

The MRI images shown in Figure 1 used in this research are provided by the dataset on Kaggle. Glioma, meningioma, pituitary, and no-tumor are among the four categories in the dataset. Each category contains 400 images, for a total of 1600 images in the dataset. Four-fifths of the images in each category are used for the training set, and the rest of the images serve as the test set. As MRI images, the original size of each image is 512×512 and the color depth of the dataset is RGB and gray.

In the preprocessing section, the direction of the images in the dataset were adjusted. In the original dataset, some of the images were reversed, which would make the model associated the direction with the feature of the tumor. To avoid the probability of erroneous recognition, the direction of images in accordance were adjusted with the standard orientation in medical regulation. In addition, this paper performed intensity normalization on the dataset. To eliminate the impact of diverse MRI collection parameters, Z-score normalization was made on each image: the mean pixel intensity was subtracted, then divided by the standard deviation. Then the normalized values were mapped linearly to the range of [0,255] and saved as an 8-bit JPEG file. The images after normalization have a coincident gray-scale distribution, which reduces the impact of intensity difference caused by the scanner.

DOI: 10.54254/2755-2721/2026.GU33590

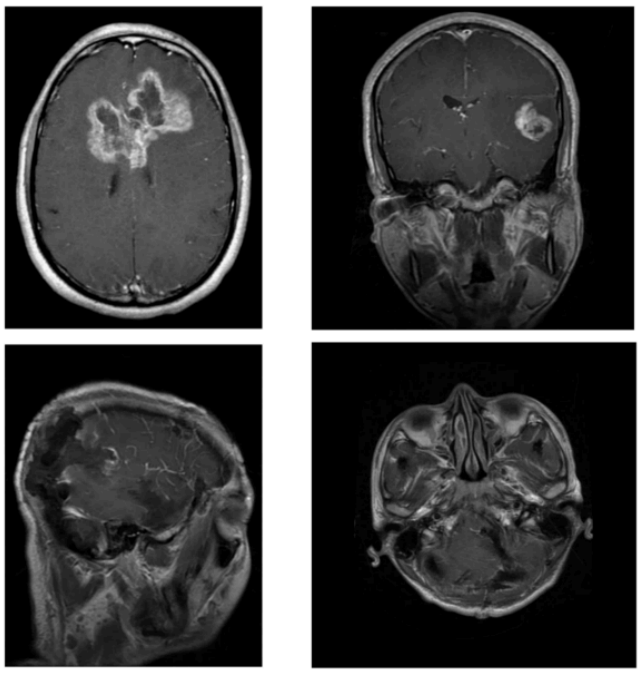

Figure 1. Four MRI images from different shooting directions (picture credit: original)

2.2. Introduction of CNN-based framework on Edge Impulse platform

In deep learning, CNN plays the key role in computer vision. CNN has been employed as a powerful tool in the domain of image processing [9,10]; it demonstrates an outstanding performance in several parts of computer vision [6]. In the process, CNN is able to identify visual patterns in images, extract features from them, and use those characteristics to categorize the images.

Convolutionary, activation function, pooling, and fully connected layers make up a conventional CNN. Among them, the convolutional layer is the most crucial part. Convolutional layer works as a feature extractor to capture the presence of specific patterns, such as edges, textures, and other complicate structures. Following convolution, to fix more complex problems, non-linearity is introduced by applying an activation function, typically ReLU. Next, the pooling layer is employed to reduce and compress the feature images extracted by the convolutional layer, which reduces computational load. After several convolutional layers and pooling layers, the refined features extracted are classified by fully connected layers and eventually map to the final output.

Edge Impulse is a machine learning operations(MLOps) platform in the cloud designed for creating embedded and edge TinyML(ML) systems that can be deployed on multiple hardware devices. To address the challenge existed in the TinyML system, the edge impulse platform simplified the design process of micro-machine learning. An extensible and portable software stack for numerous embedded systems is made possible by its support for multiple software and hardware optimizations. Edge Impulse has 118,185 projects hosted by 50,953 developers as of October 2022 [7].

2.3. Implementation details

The model was implemented and trained on the Edge Impulse platform. MobileNetV2 was used as the backbone model to efficiently infer on edge devices through a lightweight convolutional neural network. RGB is used to input the model's color depth, and the result is the probability distribution

DOI: 10.54254/2755-2721/2026.GU33590

of four MRI categories. In this research, this paper investigated 2 input revolutions: 96×96 pixels and 160×160 pixels, and the input layer was configured.

To discover the trade-off between model capacity and computational cost, various width multipliers of MobileNetV2 were adopted. For the 96×96 revolution, this paper evaluated the width multiplier of 0.05, 0.1, and 0.35. For the 160×160 revolution, this paper evaluated the width multiplier of 0.35 and 0.5. These configuration resulted in a total of 5 experimental setups, which allows to evaluate the impact of input size and model width on classification performance.

3. Results and discussion

3.1. Results

As the experimental setup referred to in the method before, classification performance of the MobilenetV2 was evaluated under different input revolutions and width multipliers shown in Table 1.

For the input of 96×96 resolution, with a width multiplier of 0.05, the accuracy achieve 69.3% with the loss of 0.8 and the peak RAM usage is 159.3K, its test accuracy was 58.02%. When the width multiplier increase to 0.1, the accuracy improved to 75.7% with a loss of 0.66, the peak RAM usage reached 169.1K, and the test accuracy is 56.41%. With the width multiplier of 0.35, the 96×96 model achieved the highest accuracy at 78.4% with a loss of 0.65 and peak RAM usage of 214.6K.

For the input of 160×160 resolution, with a loss of 1.69, the model with a width multiplier of 0.35 attained an accuracy of 83.9%, and its peak RAM usage is 416.1 K. The test accuracy was 80.77%. The accuracy rose to 88.2% with the loss of 1.37 when the width multiplier increased to 0.5, the peak RAM usage was 484.5K, and the test accuracy was 84.29%.

Table 1. MobileNetV2 performance under distinct input size and width multiplier

Model	Training Accuracy	Training Loss	Peak RAM Usage	Test Accuracy
MobileNetV2 96×96 0.05	69.40%	0.8	159.3K	58.02%
MobileNetV2 96×96 0.1	75.70%	0.66	169.1K	56.41%
MobileNetV2 96×96 0.35	78.40%	0.65	214.6K	66.99%
MobileNetV2 160×160 0.35	83.90%	1.69	416.1K	80.77%
MobileNetV2 160×160 0.5	88.20%	1.37	484.5K	84.29%

3.2. Discussion

For the 96×96 resolution, while the width multiplier increased from 0.05 to 0.35, the training accuracy improved to 78.4% from 69.4%, with the peak RAM usage rose to 214.6K from 159.3K. This performance demonstrates that with a wider network, the model can extract more feature details, though the input size was limited. The test accuracy shows a different trend with model accuracy, while the width multiplier of 0.05 reached 58.02%, the width 0.1 model dropped to 56.41%, though the low input model had the highest test accuracy, improved to 66.99% with a width multiplier of 0.35. This phenomenon implies that the overfitting has probably occurred in 96×96 model. This item may be caused by the insufficient details of features, which low input resolution brought through the wilder multiplier increase consistently.

For the 160×160 resolution, the model performs better than the model with 96×96 resolution. Under the width multiplier of 0.35, the model showed an accuracy of 83.9% and the test accuracy of

DOI: 10.54254/2755-2721/2026.GU33590

80.77%. Compared to 96×96 0.35 model(training accuracy 78.4%, test accuracy 66.99%), 160×160 width model represent a preferable performance. When the width multiplier increased to 0.5, the model accuracy and test accuracy had a relatively significant improvement. This demonstrates that with higher input resolution, better anatomical feature details are extracted in MRI images, which is the fundamental part of the categorization of brain tumors.

From the perspective of deployment, the restriction of specific resources determines the choice of models. To the resource-constrained device, 96×96 model is still viable to accomplish the classification with appropriate width multiplier, though it would still lead to low accuracy of models. 160×160 models outperform in accuracy, in this research, 160×160 model with width multiplier shows the best performance(accuracy 88.2%, test accuracy 84.29%), while the peak RAM usage is still reasonable, which makes it suitable for most edge deployment scenarios.

4. Conclusion

This research proposed to discover recognition accuracy of four different categories of brain tumor on the Edge Impulse platform, the objective is to enhance and improve identification accuracy while reducing the burden on doctors. The research employed the CNNs model and the lightweight architecture MobileNetV2 is adopted as the backbone network. During the training process, image sizes of 96×96 and 160×160 were used to train the model. For 96×96 resolution, the width multiplier of 0.05, 0.1 and 0.35 were evaluated and for 160×160, with width multiplier of 0.35 and 0.5. As the result, 160×160 resolution with 0.5 width multiplier achieve the best performance. So far, the training results were only achieved through 30 epochs because of the limited equipment conditions. In the future, the training process will employ larger width multipliers and higher resolution to enhance the training accuracy and recognition capacity.

References

[1] Amin, J., et al. (2019). Brain tumor detection using statistical and machine learning method. Computer Methods and Programs in Biomedicine, 177, 69–79.
[2] Anantharajan, S., Gunasekaran, S., & Subramanian, T. (2024). MRI brain tumor detection using deep learning and machine learning approaches. Measurement: Sensors, 31, 101026.
[3] Bootorabi, F., Haapasalo, J., Smith, E., Haapasalo, H., & Parkkila, S. (2011). Carbonic anhydrase VII—A potential prognostic marker in gliomas. Health, 3, 6–12.
[4] Dorfner, F. J., Patel, J. B., Kalpathy-Cramer, J., et al. (2025). A review of deep learning for brain tumor analysis in MRI. npj Precision Oncology, 9(2).
[5] El Sakka, M., Ivanovici, M., Chaari, L., & Mothe, J. (2025). A review of CNN applications in smart agriculture using multimodal data. Sensors, 25, 472.
[6] Hymel, S., et al. (2022). Edge impulse: An MLOps platform for tiny machine learning. arXiv preprint arXiv: 2212.03332.
[7] Missaoui, R., Hechkel, W., Saadaoui, W., Helali, A., & Leo, M. (2025). Advanced deep learning and machine learning techniques for MRI brain tumor analysis: A review. Sensors, 25, 2746.
[8] Rastogi, D., et al. (2025). Deep learning-integrated MRI brain tumor analysis: Feature extraction, segmentation, and survival prediction using replicator and volumetric networks. Scientific Reports, 15(1), 1437.
[9] Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems, 33(12), 6999-7019.
[10] Wu, J. (2017). Introduction to convolutional neural networks. National Key Lab for Novel Software Technology. Nanjing University. China, 5(23), 495.