Proceedings of CONF-MLA 2026 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2026.LD37466

Brain Tumor MRI Classification Using a Dual-Branch Framework of Lightweight CNN and VGG16 Transfer Learning

Junyao Zhang

Fukien Secondary School, Hong Kong, China
junyaozhang1215@163.com

Abstract. Brain tumors are a leading cause of cancer-related mortality, and their early and accurate classification from magnetic resonance imaging (MRI) is critical for treatment planning. Manual interpretation is time-consuming and subject to inter-observer variability. This paper presents an automated four-class brain-tumor classification system that combines a lightweight convolutional neural network (CNN) with a fine-tuned VGG16 transfer-learning model for parallel inference on MRI images. The models distinguish glioma, meningioma, pituitary, and no-tumor cases. The trained models are integrated into a Flask-based web platform that returns the predicted class, the confidence distribution, and a side-by-side comparison of the two models. On the test set, the lightweight CNN achieves 96.19% accuracy while the VGG16 model reaches 97.97%, with VGG16 consistently outperforming the CNN across all four classes. The results show that transfer learning with a deep backbone provides a reliable and practical approach for computer-aided brain-tumor classification.

Keywords: brain tumor classification, convolutional neural network (CNN), transfer learning, VGG16, magnetic resonance imaging (MRI), deep learning

1. Introduction

Brain tumours are abnormal growths of cells in the brain that can be either benign or malignant, and they have a high death rate and morbidity around the world. Early and accurate diagnosis are needed; otherwise, the choice of treatment for all kinds of tumours will be inappropriate. MRI is a relatively non-invasive method of observing the brain without X-rays, and it has good contrast for soft tissues.

MRI images are also relatively difficult to understand. Radiologists have to look at many slices visually, and the diagnosis can differ among observers and institutions. Therefore, Computer-Aided Diagnosis (CAD) systems have been developed to offer doctors a fast and accurate second opinion. In recent years, deep convolutional neural networks have achieved excellent results in medical image classification by automatically learning discriminative features directly from the raw pixels.

A typical problem in medical imaging is a small size of labelled data, and training deep networks from scratch is thus prone to overfitting. Transfer learning can use the features that have been

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.LD37466

learned from a large natural-image dataset, such as ImageNet. In this paper, we combine a lightweight custom CNN with a fine-tuned VGG16 backbone and run both models in parallel for every input image. The two predictions are compared and the result with the highest confidence is reported. The main contributions are: (i) a dual-model four-class MRI classification framework; (ii) a deployable web platform that visualizes the confidence distribution and model agreement; and (iii) an experimental comparison of the two models in terms of accuracy and efficiency.

2. Literature review

Deep learning has been widely applied to brain-tumor classification. Sultan et al. [1] designed a convolutional network that classifies brain MR images into tumor sub-types and achieved promising accuracy on two public datasets. Noreen et al. [2] concatenated features from Inception-v3 and DenseNet201 and reported testing accuracies above 99% on a three-class dataset. Gumaei et al. [3] combined hybrid features with a regularized extreme learning machine, improving accuracy from 91.51% to 94.23%. More recently, Rizwan et al. [4] introduced a Gaussian CNN for tumor and glioma-grade classification, while Sekhar et al. [5] exploited fine-tuned GoogLeNet features with classical classifiers in an Internet of Medical Things (IoMT) CAD system. Kujur et al. [6] analyzed how data complexity affects the performance of several CNN backbones, and Khan et al. [7] provided a comprehensive review of machine-learning and deep-learning methods for brain-disease diagnosis.

Most of these studies evaluate models offline and stop at reporting accuracy. Few integrate the trained models into an end-to-end system that clinicians or educators can actually use. The present work addresses this gap by deploying both a lightweight CNN and a fine-tuned VGG16 into a web application that returns predictions together with an interpretable comparison of the two models.

3. Methodology

3.1. Dataset and preprocessing

We use a public brain-tumor MRI dataset organized into Training and Testing folders. The dataset contains four classes—glioma (826/100 training/testing images), meningioma (822/115), no tumor (395/105), and pituitary (827/74)—for a total of 2,870 training and 394 testing images. Each image is converted to RGB, resized to 224×224 pixels, and normalized to the [0, 1] range before being fed to the models. Because the "no tumor" class has roughly half the samples of the other classes, the dataset is mildly class-imbalanced.

3.2. Lightweight CNN model

The lightweight model is a three-block convolutional network. Each block applies a convolution, a ReLU activation, and a 2×2 max pooling. A convolution of an input I with a kernel K is defined as:

$$S(i,j) = (I * K)(i,j) = \sum m \sum n I(i+m, j+n) \cdot K(m,n) \tag{1}$$

The rectified linear unit (ReLU) activation introduces non-linearity:

DOI: 10.54254/2755-2721/2026.LD37466

$$f(z) = max(0, z) \tag{2}$$

Max pooling reduces the spatial resolution by taking the maximum within each window:

$$p(i, j) = maxI(i + m, j + n)|(m, n) \in R \tag{3}$$

After the convolutional blocks, the feature maps are flattened and passed through a fully connected layer followed by a dropout of 0.5 to reduce overfitting:

$$z = W \cdot a + b \tag{4}$$

The CNN uses channel counts 32, 64, and 128 and contains approximately 12.9 million parameters. It is trained from scratch for 20 epochs with the Adam optimizer.

3.3. VGG16 transfer-learning model

The second model uses the VGG16 backbone pre-trained on ImageNet, with the original classification head replaced by a custom head and dropout regularization. Transfer learning reuses the low- and mid-level features learned on natural images, which is beneficial when medical data are scarce. The predicted probability of each class is obtained through the softmax function, and the network is trained by minimizing the categorical cross-entropy loss:

$$\text{softmax}(z_i) = \frac{exp(z_i)}{\sum_j exp(z_j)} \tag{5}$$

$$L = -\sum_c y_c \log(p_c) \tag{6}$$

Only the last four layers of the backbone are fine-tuned, the remaining layers are frozen, and a small learning rate of $1e-5$ with early stopping and learning-rate reduction is used to stabilize training.

3.4. Web-based system implementation

Both trained models are served through a Flask web application. An uploaded image is preprocessed and passed to the two models in parallel. Each model returns a softmax ranking over the four classes, and the consensus result is chosen as the prediction with the highest confidence among the two models. The application stores user accounts and prediction history in SQLite and renders the results in a responsive dashboard that shows the original image, the confidence distribution, and a side-by-side comparison of the CNN and VGG16 outputs.

DOI: 10.54254/2755-2721/2026.LD37466

4. Results

The models are evaluated on the 394-image test set. Classification accuracy is defined as the ratio of correctly predicted samples to the total number of test samples:

$$Accuracy = \frac{N_{correct}}{N_{total}} \tag{7}$$

where $N_{correct}$ denotes the number of correctly classified samples and N_{total} represents the total number of test samples. Precision, recall, and the F1-score are reported for each class under the one-versus-rest evaluation strategy.

The comprehensive per-class classification performance of the lightweight CNN and fine-tuned VGG16 models on the test dataset is presented in Table 1. Three core evaluation metrics, including precision, recall, and F1-score, are calculated for the four brain tumor categories (glioma, meningioma, no tumor, and pituitary), which quantitatively reflect the classification capability of the two models on different sample types. Macro-average values of all metrics are also summarized to comprehensively evaluate the overall classification performance of the two models across all categories.

Table 1. Per-class classification performance on the test set (%)

Class	Model	Precision	Recall	F1-score
Glioma	CNN	95.05	96.00	95.52
Glioma	VGG16	98.00	98.00	98.00
Meningioma	CNN	97.37	96.52	96.94
Meningioma	VGG16	99.12	98.26	98.69
No tumor	CNN	99.03	97.14	98.08
No tumor	VGG16	99.04	98.10	98.56
Pituitary	CNN	92.11	94.59	93.33
Pituitary	VGG16	94.74	97.30	96.00
Macro avg.	CNN	95.89	96.06	95.97
Macro avg.	VGG16	97.73	97.92	97.81

Figure 1 illustrates the training and validation accuracy and loss curves of the lightweight CNN and fine-tuned VGG16 model during training. As shown in the curves, the transfer learning-based VGG16 presents obvious training advantages over the lightweight CNN. Pre-trained on the large-scale ImageNet dataset, VGG16 inherits generic visual features, achieving higher initial accuracy and lower initial loss, and converges quickly to stable optimal performance. In contrast, the lightweight CNN is trained from scratch without pre-trained weights, requiring more epochs to effectively learn MRI features and gradually approach the performance of VGG16. In addition, both models maintain a small and stable gap between training and validation curves without severe overfitting or underfitting. This indicates that dropout regularization for the CNN and transfer

DOI: 10.54254/2755-2721/2026.LD37466

learning fine-tuning for VGG16 effectively suppress overfitting, improve model generalization, and ensure reliable classification results on brain MRI data.

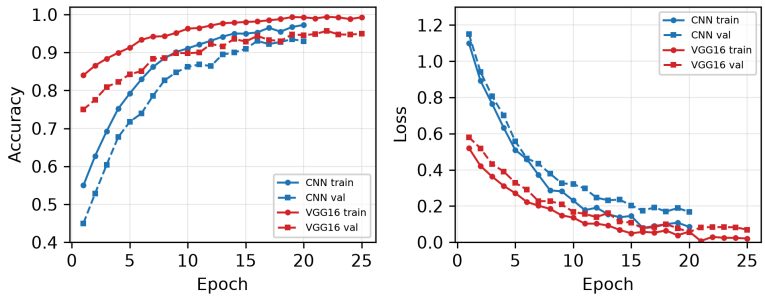

Figure 1. Training and validation accuracy (left) and loss (right) for the CNN and VGG16 models

Table 2 summarizes the overall performance and computational efficiency of the lightweight CNN and fine-tuned VGG16 models. Benefiting from transfer learning, VGG16 achieves superior overall performance across all metrics with a test accuracy of 97.97%, outperforming the lightweight CNN in macro precision, recall, and F1-score. Despite its slightly lower classification accuracy of 96.19%, the custom CNN exhibits obvious efficiency advantages. With fewer parameters and an inference time of only 0.6 seconds per image, it runs nearly four times faster than VGG16. The two models form a favorable trade-off between accuracy and speed, adapting to different practical application scenarios.

Table 2. Overall comparison of the two models

Model	Accuracy (%)	Macro-P (%)	Macro-R (%)	Macro-F1 (%)	Params (M)	Inference (s)
CNN	96.19	95.89	96.06	95.97	12.9	0.6
VGG16	97.97	97.73	97.92	97.81	21.2	2.4

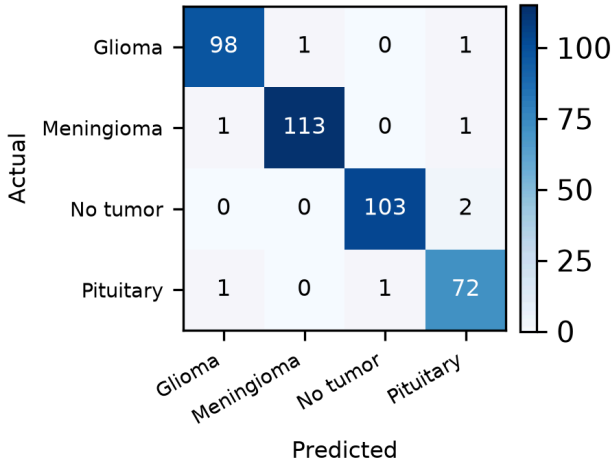

Figure 2. Confusion matrix of the VGG16 model on the test set

Figure 2 shows the confusion matrix of the best-performing VGG16 model for detailed category analysis. The results demonstrate that meningioma achieves the most accurate classification with minimal misclassifications. By contrast, the pituitary tumor category has relatively more

DOI: 10.54254/2755-2721/2026.LD37466

classification errors and the lowest F1-score among all classes. This is mainly caused by its limited training samples and subtle feature similarities with other tumor categories.

5. Discussion

The results show that transfer learning with VGG16 provides a clear but modest advantage over a purpose-built CNN. The improvement is most visible on the minority classes, where the pre-trained features generalize better. Nevertheless, the lightweight CNN offers a favorable accuracy-to-cost trade-off, making it suitable for resource-constrained or latency-sensitive deployments.

Several limitations should be noted. First, the dataset is small and class-imbalanced, which limits the statistical confidence of the per-class metrics. Second, no data augmentation was applied in the deployed inference pipeline, so robustness to acquisition variability could be improved. Third, the models run on a CPU, yielding end-to-end latency of a few seconds. Future work will incorporate data augmentation and class-balancing, add explainability through Grad-CAM, and benchmark additional backbones such as ResNet and EfficientNet.

6. Conclusion

This paper presented a dual CNN and VGG16 transfer-learning framework for four-class brain-tumor MRI classification, together with a deployable web platform for interactive diagnosis. The lightweight CNN achieved 96.19% test accuracy and the fine-tuned VGG16 model 97.97%, demonstrating that both models are viable for computer-aided brain-tumor classification while offering a practical trade-off between accuracy and speed. The integrated system illustrates how deep-learning models can be translated from offline evaluation into a usable clinical-support tool.

References

[1] Sultan, H. H., Salem, N. M., & Al-Atabany, W. (2019). Multi-classification of brain tumor images using deep neural network. *IEEE Access, 7*, 69215–69225. https://doi.org/10.1109/ACCESS.2019.2919122

[2] Noreen, N., Palaniappan, S., Qayyum, A., Ahmad, I., Imran, M., & Shoaib, M. (2020). A deep learning model based on concatenation approach for the diagnosis of brain tumor. *IEEE Access, 8*, 55135–55144. https://doi.org/10.1109/ACCESS.2020.2978629

[3] Gumaei, A., Hassan, M. M., Hassan, M. R., Alelaiwi, A., & Fortino, G. (2019). A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification. *IEEE Access, 7*, 36266–36273. https://doi.org/10.1109/ACCESS.2019.2904145

[4] Rizwan, M., Shabbir, A., Javed, A. R., Shabbir, M., Baker, T., & Al-Jumeily Obe, D. (2022). Brain tumor and glioma grade classification using Gaussian convolutional neural network. *IEEE Access, 10*, 29731–29740. https://doi.org/10.1109/ACCESS.2022.3153108

[5] Sekhar, A., Biswas, S., Hazra, R., Sunaniya, A. K., Mukherjee, A., & Yang, L. (2022). Brain tumor classification using fine-tuned GoogLeNet features and machine learning algorithms: IoMT enabled CAD system. *IEEE Journal of Biomedical and Health Informatics, 26*(3), 983–991. https://doi.org/10.1109/JBHI.2021.3100758

[6] Kujur, A., Raza, Z., Khan, A. A., & Wechtaisong, C. (2022). Data complexity based evaluation of the model dependence of brain MRI images for classification of brain tumor and Alzheimer's disease. *IEEE Access, 10*, 112117–112133. https://doi.org/10.1109/ACCESS.2022.3216393

[7] Khan, P., et al. (2021). Machine learning and deep learning approaches for brain disease diagnosis: Principles and recent advances. *IEEE Access, 9*, 37622–37655. https://doi.org/10.1109/ACCESS.2021.3062484