



OPEN Dilated SE-DenseNet for brain tumor MRI classification

Yuannong Mao^{1✉}, Jiwook Kim¹, Lena Podina² & Mohammad Kohandel^{1✉}

In the field of medical imaging, particularly MRI-based brain tumor classification, we propose an advanced convolutional neural network (CNN) leveraging the DenseNet-121 architecture, enhanced with dilated convolutional layers and Squeeze-and-Excitation (SE) networks' attention mechanisms. This novel approach aims to improve upon state-of-the-art methods of tumor identification. Our model, trained and evaluated on a comprehensive Kaggle brain tumor dataset, demonstrated superior performance over established convolution-based and transformer-based models: ResNet-101, VGG-19, original DenseNet-121, MobileNet-V2, ViT-L/16, and Swin-B across key metrics: F1-score, accuracy, precision, and recall. The results underscore the effectiveness of our architectural enhancements in medical image analysis. Future research directions include optimizing dilation layers and exploring various architectural configurations. The study highlights the significant role of machine learning in improving diagnostic accuracy in medical imaging, with potential applications extending beyond brain tumor detection to other medical imaging tasks.

Despite extensive research efforts over many years, brain tumors remain among the deadliest forms of cancer¹. Malignant primary brain tumors, in particular, present formidable challenges for treatment, evidenced by a five-year overall survival rate not exceeding 35%². Modern neuroimaging, especially Magnetic Resonance Imaging (MRI), plays a vital role in the clinical management of brain tumors due to its high-resolution structural imaging capabilities³. It is integral to various stages of treatment, from preoperative diagnosis and grading to intraoperative planning and postoperative monitoring.

The advent of deep learning in medical imaging has led to significant advancements in the classification of diagnostic categories. However, the acquisition of medical images is fraught with challenges. Strict legal and ethical requirements to protect patient privacy mean that there is often a scarcity of available data⁴. This limitation necessitates modern techniques that address both data utilization and protection.

In the state-of-the-art models, transformer-based models like the Vision Transformer (ViT) have been employed in medical image classification tasks due to their superior performance. For instance, the ViT architecture has been used in identifying pulmonary diseases from medical images⁵. However, these models often require large datasets to achieve their full potential. In scenarios with smaller datasets, such as in medical fields where data is limited, Convolutional Neural Networks (CNNs) are more efficient. Their local inductive bias proves to be crucial in analyzing smaller images⁶.

We aim to classify brain tumors into four categories: glioma, meningioma, pituitary tumors, and no tumor (healthy), using T1-weighted MRI images, each captured from various angles. Taking advantage of both CNNs and transformer-based models, our research introduces an innovative approach to MRI brain tumor classification. We advance upon the traditional DenseNet-121 architecture, integrating dilated convolution⁷ and an attention mechanism through Squeeze-and-Excitation (SE) networks⁸. This enhancement not only improves the model's representation learning capabilities but also addresses the challenge of limited data availability in medical imaging.

Our findings reveal that it is not necessarily larger models that lead to better performance in the specific field of medicine. Instead, proper optimization of dilation rate and kernel size in traditional convolution networks can achieve excellent performance, outperforming many existing models in brain tumor MRI classification. In addition to being beneficial in a clinical setting, this model can provide valuable diagnostic assistance to radiologists.

Related works

Deep learning has significantly advanced radiological medical imaging, particularly in cancer detection and classification tasks. For instance, in breast cancer detection using mammography, an end-to-end training approach has been emphasized, which ensures the entire network is jointly trained and avoids sub-optimal

¹Department of Applied Mathematics, University of Waterloo, Waterloo, ON N2L 3G1, Canada. ²David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada. ✉email: y64mao@uwaterloo.ca; kohandel@uwaterloo.ca

solutions from individual steps⁹. This method has been effectively compared with two CNN structures: the VGG network¹⁰ and the ResNet¹¹. Similarly, in the domain of skin cancer classification, CNNs have been employed using a dataset of 129,450 clinical images, comprising 2032 different diseases¹². The CNN demonstrated a capability to classify skin cancer at a level comparable to dermatologists, using only pixels and disease labels as inputs.

Moreover, DenseNet-based architectures have been utilized in diverse medical classification tasks beyond cancer detection. A notable example is the application of DenseNet for liver lesion prediction and classification using CT images¹³. In this study, the team conducted data pre-processing with region-growing segmentation. They also leveraged the DenseNet architecture for efficient feature extraction and advanced data augmentation, addressing challenges such as overfitting and dataset imbalance. Additionally, CNNs have been deployed for pneumonia detection in chest X-ray images¹⁴, where the 121-layer CNN model, CheXNet, was trained on over 100,000 frontal-view X-ray images encompassing 14 different thoracic diseases. Notably, CheXNet surpassed the average F1 score of four radiologists, highlighting the potential of deep learning in medical imaging.

In the broader context of image classification tasks, dilational layers have been applied to convolutional networks, effectively aggregating multi-scale contextual information without losing resolution, particularly in semantic segmentation tasks⁷. Furthermore, the integration of Squeeze-and-Excitation (SE) modules enables dynamic channel-wise feature recalibration, significantly enhancing the network's representational power⁸. The incorporation of SE modules contributed to the winning of the ILSVRC 2017 classification competition, demonstrating a remarkable 2.251% top-5 error on the ImageNet test set¹⁵. Specifically, as a precedent of using SE modules on DenseNet, coniferous wood's compressive strength has been predicted by the SE-DenseNet model combined with near-infrared spectroscopy¹⁶. The team has achieved an accuracy of 88.89% and an F1 score of 0.8831, with correlation coefficients (R) of 0.9144 and root mean square errors (RMSE) of 1.2389 MPa on the larch test set.

Building on these advancements, our work combines dilated layers and SE modules within the DenseNet framework. The DenseNet framework is known for its robust approach to medical imaging classification¹⁷, while SE modules enhance feature recalibration capabilities⁸, and dilated convolutions provide a broader contextual view⁷. This combination provides improved discriminative power for classifying brain tumors, pushing the boundaries of what is achievable in medical image analysis.

Methods

In this study, we present an adaptation of the DenseNet-121 architecture¹⁸, specifically tailored for enhanced brain tumor classification. To enhance the model's focus on the tumor site, we incorporated the Squeeze-and-Excitation (SE) module⁸ after each Dense block layer. The SE module is designed to adaptively recalibrate channel-wise feature responses by explicitly modeling interdependencies between channels, and suppressing less useful ones, akin to an attention mechanism. However, owing to the inclusion of the SE module, which adds complexity, we have replaced the traditional 3×3 convolutional layer with a 3×3 depth-wise separable layer¹⁹. This layer divides the convolution operation into two parts: a depthwise convolution and a point-wise convolution (shown in Fig. 1). Such an approach reduces both the number of parameters and the computational load by separating the spatial and channel-wise computations.

Dilated convolutional

We replaced the standard 3×3 convolutional kernel in the model with a 3×3 dilated convolutional kernel (shown as in Fig. 2), a modification that strategically introduces a $d-1$ space between the elements of the convolutional kernel, where d represents the dilation rate. This adjustment enables the model to encompass a larger region of the input image while using the same number of parameters as a traditional kernel. The dilated convolution operation is formulated as follows:

$$(F *_{\text{d}} k)(p) = \sum_{s+t=p} F(s)k(t) \quad (1)$$

where $F: \mathbb{Z}^2 \rightarrow \mathbb{R}$ denotes the input feature map function, and k represents the discrete kernel function. Here, p symbolizes a point in the discrete space \mathbb{Z}^2 . The operator $*_{\text{d}}$ signifies the dilated convolution with a dilation factor d , and k_{d} is the dilated kernel that introduces $d-1$ zeros between each element of the original kernel k . We have assigned the different dilation rate $d = d_i$ ($i \in \mathbb{Z}^+$) to each Enhanced DenseNet block as Fig. 2.

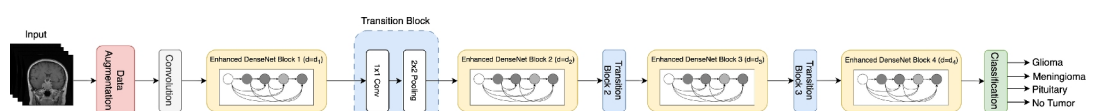


Fig. 1. Dilated convolutions and Squeeze-and-Excitation (SE) modules are incorporated into the original DenseNet-121 to form the Enhanced DenseNet Block, indicated in orange. The dilation convolutions are denoted by $d = d_i$ ($i \in \mathbb{Z}^+$), indicating the dilation rate at different stages of the model. Each transition block is shown in blue, and the final classification layer is shown in green.

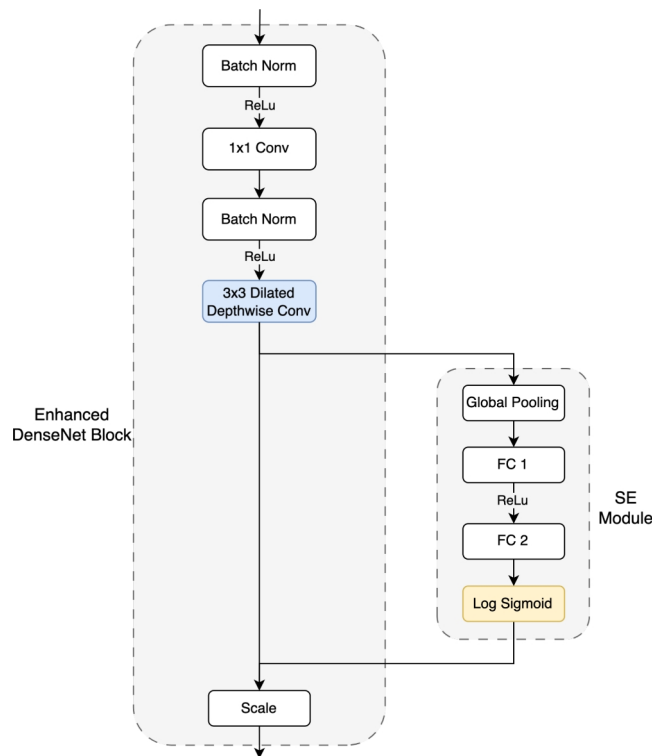


Fig. 2. This illustration depicts an enhanced DenseNet block, featuring 3×3 dilated depthwise convolutions—represented in blue—for efficient parameter use and expanded receptive fields. Highlighted in yellow are the LogSigmoid activation functions, which introduce non-linearity and facilitate the learning of complex patterns in the data.

The gridding effect can be a common problem when multilayer dilation convolution is applied. This happens because dilated convolutions introduce gaps between the active sampling pixels in the kernel. As a result, this can lead to significant information loss from the input data, particularly when the dilation rate is high. To avoid this theoretical problem we follow the design rules proposed by hybrid dilated convolution²⁰. Suppose there are N convolutional layers with kernel size $K \times K$ that have dilation rates of $[r_1, \dots, r_i, \dots, r_n]$:

1. The maximum distance between two non-zero values of the second layer should be less than or equal to the size of the convolution kernel of that layer. This maximum distance can be defined as:

$$M_i = \max[M_{i+1} - 2r_i, M_{i+1} - 2(M_{i+1} - r_i), r_i]$$

with $M_n = r_n$, and the goal is to let $M_2 \leq K$, where K is the kernel size of that layer.

2. Avoiding a common factor relationship between dilations (e.g. 2, 4, 8).
3. Incorporating a sawtooth wave-like pattern for dilation rates, where the dilation rate for groups of layers increases between groups. Based on the above rules, we tested the combination of different expansion rates. The numbers represent the dilation rates in each enhanced DenseNet block. The results are shown in Table 1.

The superior performance across all metrics with the “1-2-7-9” dilation rate sequence confirms its efficacy when incrementally increased. Nonetheless, opting for a high dilation rate in the initial blocks does not yield favorable outcomes.

Incorporating dilated convolution effectively broadens the receptive field, which is instrumental in gathering extensive contextual information. Such an enhancement is especially beneficial for detecting tumors that vary in size and location.

Squeeze-and-excitation module

The Squeeze-and-Excitation (SE) Module⁸ is a component that can be added after a convolutional layer in a neural network which performs similarly to a self-attention function on the channels. It acts on the C output channels $U_1, \dots, U_C, U_i \in \mathbb{R}^{H \times W}$. The module allows one to model the interdependencies between channels

Dilations	Test error (%)	Accuracy	F1 score
1-1-1-1	5.80	0.942	0.942
1-1-2-3	6.41	0.936	0.935
1-1-5-9	8.85	0.912	0.911
1-2-7-9	4.42	0.956	0.956
2-5-7-11	5.34	0.947	0.946
5-9-10-13	9.08	0.909	0.908

Table 1. Our model's performance on different combinations of dilation rates with the Sigmoid activation function in SE module.

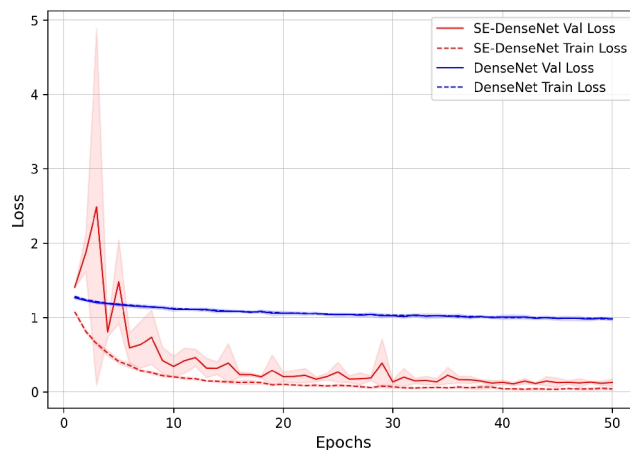


Fig. 3. Comparative loss trends over 50 epochs for SE-DenseNet (red) and DenseNet (blue), both using a Sigmoid activation function, with solid lines representing validation losses and dotted lines representing training losses. The shaded regions indicate the confidence intervals derived from 10-fold cross-validation.

globally, incorporating information from outside the receptive field of the convolutional filters. First, the *squeeze* operation performs average global pooling on each of the output channels as follows:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (2)$$

where $u_c(i, j)$ is the element at position (i, j) of the c th channel. This yields a vector $z = [z_1, z_2, \dots, z_C]$. The *excitation* operation yields a weight between 0 and 1 for each channel as follows:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (3)$$

where σ and δ refers to the Sigmoid and ReLU function respectively, and the weight matrices $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ are learned parameters.

To limit model complexity, the channel information is projected into dimension C/r by the first weight matrix and brought back to dimension C by the second weight matrix. As a final step in this component, the original channel outputs U_c are weighted using the learned vector s . This nonlinear gating mechanism allows multiple channels to be emphasized at once, capturing inter-channel dependencies.

Integrating the SE module into DenseNet architecture has markedly improved our model's ability to classify brain tumor labels. This improvement is visually evident in the provided Fig. 3, where the performance of our SE-enhanced model shows a decreasing trend in both training and validation loss, unlike the original model. The SE module's dynamic channel-wise recalibration of features addresses a critical aspect of deep learning—focusing on salient features for the task at hand. It adapts the network's activations in a data-driven manner, enabling the model to converge effectively, whereas the original model, without this recalibration, fails to refine its focus. This contrast suggests that the SE module provides a significant boost in the representational capacity of the network, allowing for more nuanced feature extraction, crucial for the subtle variances inherent in medical imaging diagnostics.

Class	Training set	Testing set
Glioma	1321	300
Meningioma	1339	306
Pituitary	1457	300
No-tumor	1595	405
Total	5712	1311

Table 2. T1-weighted brain MRI images organization.

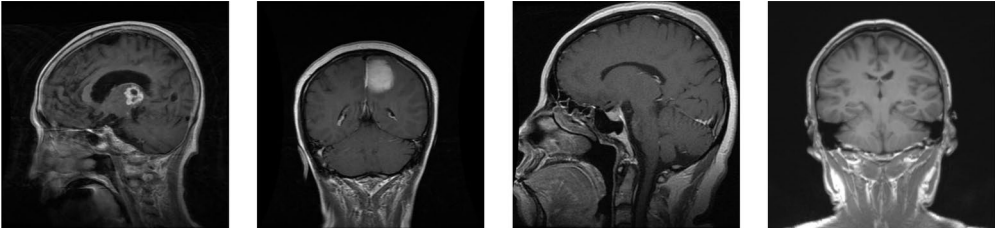


Fig. 4. Sample images from brain tumor MRI dataset. From left to right: Glioma, Meningioma, Pituitary, No-tumor.

Methodology
Data

We have used a publicly available image dataset from Kaggle²¹, which contains T1-weighted brain MRI images classified into four categories: glioma, meningioma, pituitary, and no-tumor. The dataset was carefully balanced for both the training and testing sets for a total of 7023 images in Table 2, and contains a variety of images taken from the front, side, and top of the head, providing a comprehensive picture of the anatomical variability and unique morphological features of the lesions as some samples in Fig. 4.

Data augmentation

To account for variability in image sizes and improve generalization, all images were preprocessed by cropping to remove black edges, resizing to 224 × 224 pixels, and normalizing based on the dataset’s overall mean and standard deviation. We applied Gaussian blur (kernel size 5, $\sigma = 2$) to reduce noise, as this technique efficiently attenuates high-frequency noise while preserving important low-frequency image signals. Affine transformations, including shear mapping, were employed to enhance perspective invariance. Additionally, common augmentation techniques such as random horizontal flipping, rotation, and zoom were applied to the training set to improve model robustness.

Activation function

As mentioned above, although the SE Module originally employed a sigmoid activation function¹⁸, our findings indicate an enhancement in performance upon substituting the traditional Sigmoid activation with the LogSigmoid function

$$\text{LogSigmoid}(x) = \log \left(\frac{1}{1 + e^{-x}} \right) \tag{4}$$

As depicted in Fig. 5, it is evident that activation functions involving logarithmic operations contribute to a more efficient gradient convergence rate since the LogSigmoid model attains the lowest validation loss overall. This underscores the activation function’s effectiveness in our model’s context.

It is not difficult to show that LogSigmoid satisfies the Lipschitz continuity condition along with continuity and smoothness. Simply find the first-order derivative of LogSigmoid, and as x goes to negative infinity, the derivative approaches 1. Hence, the derivative of LogSigmoid is bounded by 1 for all x , and thus with a Lipschitz constant $L = 1$. Here the Lipschitz constant is a metric for evaluating the robustness of a neural network against perturbations, which sets an upper limit for the distance between input perturbations and output changes²². We have therefore used Log Sigmoid in the SE block, as shown in Fig. 2.

Training

We employed five models as benchmarks: ResNet¹¹, VGG¹⁰, DenseNet¹⁸, MobileNet²³, ViT²⁴ and Swin-B²⁵. All models were initially pre-trained on the ImageNet1K dataset²⁶, except for DenseNet-121, which was not pre-trained and was used for direct comparison with our model. These models were then fine-tuned on our dataset following the described data preprocessing methods, and for our model, we selected a dilation rate of “1-2-7-9”.

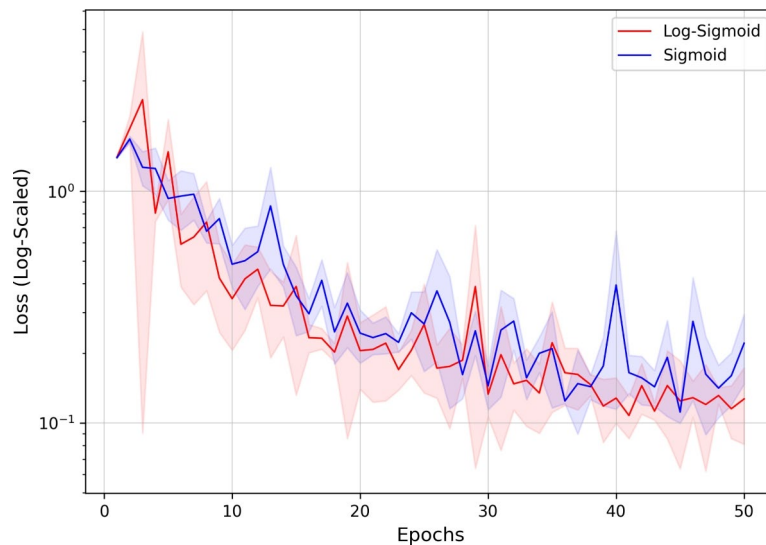


Fig. 5. Comparison of activation functions using a dilation rates combination of “1-2-7-9”: Sigmoid vs Log-Sigmoid based on validation loss. The Log-Sigmoid achieves the lowest overall validation loss, indicating enhanced performance with the specified dilation rates. The shaded regions indicate the confidence intervals derived from 10-fold cross-validation.

All the network was trained using the AdamW optimizer²⁷ with standard parameters ($\beta_1 = 0.9$ and $\beta_2 = 0.999$) and a weight decay of 0.001. We initially set the learning rate to 0.001 and applied the Cosine Annealing learning rate scheduler²⁸, as follows:

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min}) \left(1 + \cos \left(\frac{T_{\text{cur}}}{T_{\max}} \pi \right) \right) \quad (5)$$

where η_{\max} and η_{\min} are the initial and minimum learning rates, respectively, T_{\max} is the maximum number of iterations after which the learning rate is reset, and T_{cur} is the current iteration number. We set T_{\max} to 5, and η_{\min} to 0. Furthermore, we employed the general cross-entropy loss function during the training

$$L(y, \hat{y}) = - \sum_{i=1}^4 y_i \cdot \log(\hat{y}_i) \quad (6)$$

here $L(y, \hat{y})$ represents the loss function. The true label in one-hot encoded form is y , where y_i is 1 for the correct class and 0 otherwise. The predicted probability of the class is \hat{y}_i , which is the output of the Softmax function for class i . The summation runs over all four classes. This loss function is summed over all the training examples and typically averaged.

The training process consisted of 50 epochs with a batch size of 256. and we employed 10-fold cross-validation to ensure the robustness and generalizability of our model. The dataset was split into ten subsets, where in each iteration, one subset was used for validation and the remaining nine were used for training. The final model performance was averaged over all ten folds.

Testing

Our evaluation used a 10-crop testing method, which involves resizing each image to 256×256 pixels and producing ten distinct crops. These crops include the four corners and the center of the image, derived from both the original image and its horizontally flipped counterpart, resulting in a total of ten variations per image. The final test report is obtained by averaging the results over the ten crops.

Results

In our work, we present a novel brain tumor classification method utilizing a customized DenseNet-121 architecture, now augmented with dilation layers and Squeeze-and-Excitation (SE) modules^{7,18}. Initially, our model's performance fluctuated due to the absence of pre-trained weights, unlike its counterparts pre-trained on the ImageNet1K dataset²⁶. However, it quickly surpassed the performance of other models by effectively learning from raw images, attributed to the advanced feature extraction from SE modules⁸ and the expanded receptive fields due to dilated convolutional layers⁷.

Table 3 demonstrates that our model has achieved remarkable results despite its relatively lightweight design with only 8.08M parameters and 1.86B Floating-Point Operations Per Second (FLOPs). The FLOPs metric is

Model	Params	FLOPs	Test error (%)	Accuracy	F1 score	Precision	Recall
ResNet-101	42.51 M	7.83B	15.179	0.813	0.848	0.851	0.848
VGG-19	139.59 M	19.63B	17.96	0.808	0.817	0.826	0.820
MobileNet-V2	2.2 3M	0.31B	14.46	0.833	0.855	0.859	0.855
ViT-L/16	303.31 M	39.86B	8.93	0.909	0.911	0.913	0.911
Swin-B	86.75 M	15.47B	10.94	0.870	0.891	0.893	0.891
DenseNet-121	6.96 M	2.86B	45.92	0.499	0.536	0.589	0.541
NeuroNet19 ²⁹	139.59 M	19.63B	–	0.993	0.993	0.993	0.993
Darknet19-SVM ³⁰	–	–	–	0.95	0.96	0.96	0.96
Our model	8.08 M	1.86 B	3.85	0.962	0.965	0.965	0.965

Table 3. Models’ performance metrics.

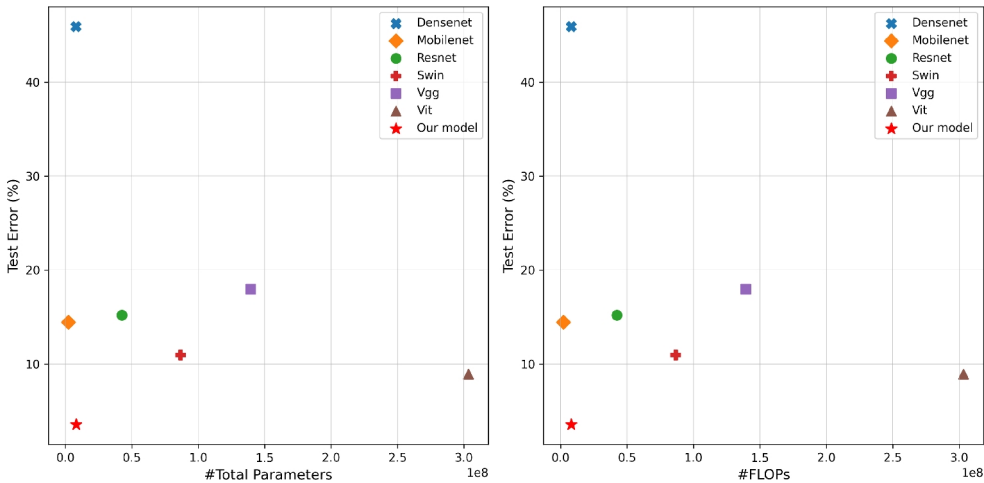


Fig. 6. Comparative analysis of brain tumor classification models: The top graph illustrates the trade-off between the number of parameters and test error rate, while the bottom graph displays the relationship between floating point operations (FLOPs) and test error rate.

used to measure the computational complexity of deep learning models, with a lower value being preferable. In comparison, VGG-19, with a hefty 139.59M parameters and 19.63B FLOPs, only managed an accuracy of 80.82% and an F1 score of 0.817. Our model not only proved to be more efficient but also significantly more effective, achieving a high accuracy of 96.2% and an F1 score of 0.965. This is a remarkable achievement, especially considering heavier models like ViT-L/16, which, with a massive 303.31M parameters and 39.86B FLOPs, reached an accuracy of 90.91% and an F1 score of 0.911. We also included two proposed models, NeuroNet19²⁹ and DarkNet19-SVM³⁰, for comparison. These models allow us to contextualize our model’s performance against established architectures. While both NeuroNet19 and DarkNet19-SVM achieved high accuracy and F1 scores, our model demonstrated superior performance with fewer parameters and lower computational cost (FLOPs), highlighting its efficiency and effectiveness. Further insights into this relationship between model complexity and performance can be seen in Fig. 6, which graphically depicts the trade-off between the number of parameters (and FLOPs) and the resultant accuracy and F1 scores across different models.

The distinction of our model’s test classification performance is clearly illustrated in the provided confusion matrix (Fig. 7), which reveals the accuracy of our model across various classes, showcasing the practical efficacy of our architectural choices in a clinical setting. These results bolster the effectiveness of our model, underlining that the synergy of Squeeze-and-Excitation (SE) modules, dilated layers, and perfectly sized kernels proved to be a powerful combination for accurate medical image classification. Our approach opens the door for wider applications in medical image analysis, demonstrating that a well-designed, lighter architecture can eclipse the capabilities of more cumbersome, pre-trained models, providing a resource-efficient, yet highly accurate, alternative.

Future work

To address challenges such as data scarcity in medical imaging, we plan to implement targeted image augmentation techniques that have demonstrated efficacy in enhancing dataset diversity. One promising method is MedAugment, which applies pixel and spatial transformations to simulate a broader range of anatomical variations³¹. Although the dataset used in this study is balanced, class imbalance is a common challenge in real-world medical datasets, as the distribution of images encountered by radiologists is typically non-uniform.

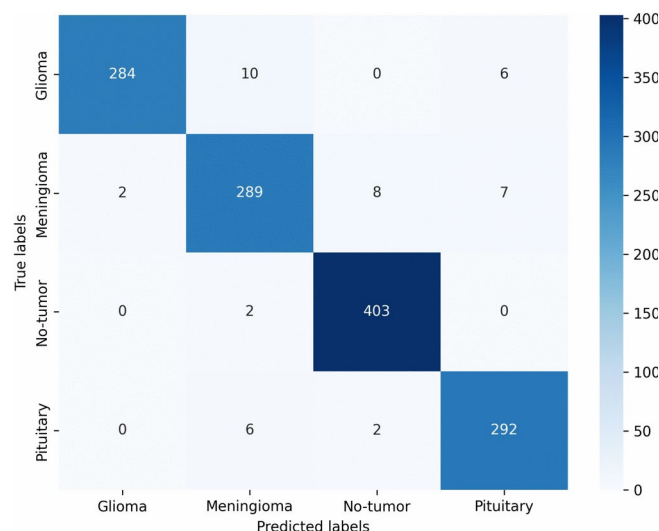


Fig. 7. Confusion matrix of the brain tumor classification performance by our model. The matrix shows the number of true positive predictions for each class: Glioma, Meningioma, No-tumor, and Pituitary tumors. The diagonal cells represent accurate predictions, with darker shades indicating higher counts. Off-diagonal cells indicate misclassifications, with our model exhibiting high precision for No-tumor predictions and good overall accuracy across all classes.

To mitigate this potential challenge, we aim to explore approaches such as Generative Adversarial Networks (GANs), which can generate synthetic images reflecting real data distributions and are particularly useful for addressing label imbalances in imbalanced datasets³². GANs can also help alleviate the issue of limited dataset availability, often restricted due to privacy concerns⁴. Additionally, diffusion models, which generate high-quality, diverse synthetic data, present a promising avenue for further exploration³³.

Expanding the application scope, we plan to integrate our model with various data types, such as patient medical records, alongside medical images. A survey of radiologists revealed that 87% believe clinical information significantly impacts interpretation³⁴. By incorporating comprehensive patient data, our model could offer more personalized predictions. We also envision an ensemble approach that simultaneously processes multiple data types. Beyond MRI images, we intend to extend our model's applicability to other medical imaging modalities, including CT and X-ray images.

While our current work focuses on 2D MRI images, extending the model to handle 3D MRI data is an important area for future exploration, and networks are trained using 3D whole-brain structures, their performance may be improved over 2D images due to the deeper architecture³⁵. The model could access additional spatial context by capturing volumetric information from 3D MRI, so convolutional layers and pooling layers in the network architecture and more image preprocessing are the areas we need to consider improving to adapt to 3D image features. Furthermore, transfer learning is also a good choice since traditional convolutional neural networks are based on 2D architectures, and using 3D inputs requires fewer pre-processing steps such as slice correction, slice selection, and slice extraction. Therefore, the processing steps can be reduced, making the system more robust³⁶.

While our model demonstrates promising performance on the dataset used in this study, its generalizability remains a theoretical prospect at this stage. We acknowledge that further validation of additional datasets is necessary to substantiate its adaptability and clinical applicability. Future work will focus on applying the model to a wider range of datasets, encompassing diverse medical conditions and imaging techniques. Such efforts aim to enhance the robustness and generalizability of the model, ultimately contributing to the development of more reliable diagnostic tools for clinical practice.

Another limitation of this study is the absence of clinical data, such as patient demographics, medical histories, or additional biomarkers, in the dataset used. The dataset available for this study is restricted to imaging data, limiting the ability to explore the integration of multimodal information. This constraint impacts the diagnostic relevance and applicability of the model in real-world clinical workflows. Future research will aim to address this limitation by utilizing datasets that include both imaging and clinical data. Such integration could bridge the gap between imaging analysis and clinical decision-making, providing a more comprehensive diagnostic framework.

Additionally, while the dataset used in this study relies on radiologist-provided annotations for training and testing, future evaluations aim to test the model in prospective clinical settings. In such settings, radiologists would interact with the model's outputs in real time on unseen cases, providing a practical assessment of its utility and reliability. Evidence suggests that machine learning-assisted tools can improve diagnostic accuracy beyond the capabilities of either models or radiologists alone³⁷. For example, models can identify radiographs often overlooked by radiologists, while radiologists can detect findings missed by the model, highlighting the complementary strengths of human and machine analysis. Future work will focus on leveraging this synergy in real-world applications to maximize the clinical impact of the model.

Conclusion

In conclusion, our study has effectively integrated dilated convolution layers into the DenseNet-121 architecture¹⁸, marking a significant advancement in brain tumor classification. This approach has not only improved upon the DenseNet-121 model but also outperformed other benchmark models: ResNet-101¹¹, VGG-19¹⁰, original DenseNet-121¹⁸, MobileNet-V2²³, ViT-L/16²⁴, and Swin-B²⁵ in all these metrics: F1-score, accuracy, precision, and recall. The enhanced feature representation, attributable to the innovative use of dilation layers⁷ and SE modules⁸, has been pivotal in achieving superior performance in complex medical diagnostics.

Due to its lighter architecture, our model, with an F1-score of 0.967, shows significant advantages compared to the model NeuroNet19²⁹. NeuroNet19, which obtained an F1-score of 0.973, employs VGG19 as its backbone network and utilizes over 143.7 million parameters³⁸. In contrast, our model leverages DenseNet-121 as its backbone, incorporating just over 8.083 million parameters. This means our model has only 5.62% of the parameters compared to NeuroNet19 but has performed with an F1-score difference of 0.006, rendering it particularly well-suited for smaller datasets. Furthermore, we do not leverage any pre-training for our model and yet manage to reach comparable performance. Our study not only paves the way for future research in optimizing dilated convolution layers and exploring hyperparameter spaces but also shows promise for extension to other medical and radiographic imaging tasks.

Data availability

The source of the data is stated in this article, and the code required to reproduce the model and analysis is available at <https://github.com/YuannongMao01/Improved-DenseNet-for-Brain-tumor-MRI>

Received: 24 April 2024; Accepted: 13 January 2025

Published online: 28 January 2025

References

1. Aldape, K. et al. Challenges to curing primary brain tumours. *Nat. Rev. Clin. Oncol.* **16**, 509–520 (2019).
2. Lapointe, S., Perry, A. & Butowski, N. A. Primary brain tumours in adults. *Lancet* **392**, 432–446 (2018).
3. Mabray, M. C., Barajas, R. F. & Cha, S. Modern brain tumor imaging. *Brain Tumor Res. Treat.* **3**, 8–23 (2015).
4. Kaissis, G. A., Makowski, M. R., Rückert, D. & Braren, R. F. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat. Mach. Intell.* **2**, 305–311 (2020).
5. Zhou, H.-Y. et al. A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. *Nat. Biomed. Eng.* 1–13 (2023).
6. Hassani, A. et al. Escaping the big data paradigm with compact transformers. arXiv preprint [arXiv:2104.05704](https://arxiv.org/abs/2104.05704) (2021).
7. Yu, F. & Koltun, V. Multi-scale context aggregation by dilated convolutions. arXiv preprint [arXiv:1511.07122](https://arxiv.org/abs/1511.07122) (2015).
8. Hu, J., Shen, L., Albanie, S., Sun, G. & Wu, E. Squeeze-and-excitation networks (2019). [arXiv: 1709.01507](https://arxiv.org/abs/1709.01507).
9. Shen, L. et al. Deep learning to improve breast cancer detection on screening mammography. *Sci. Rep.* **9**, 12495 (2019).
10. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014).
11. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (2016).
12. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
13. Prakash, N. N., Rajesh, V., Namakhwa, D. L., Pande, S. D. & Ahammad, S. H. A DenseNet CNN-based liver lesion prediction and classification for future medical diagnosis. *Sci. Afr.* **20**, e01629 (2023).
14. Rajpurkar, P. et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint [arXiv:1711.05225](https://arxiv.org/abs/1711.05225) (2017).
15. Russakovsky, O. et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).
16. Li, C., Chen, X., Zhang, L. & Wang, S. Determination of coniferous wood's compressive strength by SE-DenseNet model combined with near-infrared spectroscopy. *Appl. Sci.* **13**, 152 (2022).
17. Zhou, T. et al. Dense convolutional network and its application in medical image analysis. *BioMed Res. Int.* **2022** (2022).
18. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4700–4708 (2017).
19. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1251–1258 (2017).
20. Wang, P. et al. Understanding convolution for semantic segmentation. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* 1451–1460 (IEEE, 2018).
21. Nickparvar, M. Brain tumor MRI dataset. <https://doi.org/10.34740/KAGGLE/DSV/2645886> (2021).
22. Virmaux, A. & Scaman, K. Lipschitz regularity of deep neural networks: Analysis and efficient estimation. *Adv. Neural Inf. Process. Syst.* **31** (2018).
23. Howard, A. G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861) (2017).
24. Dosovitskiy, A. et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. arXiv preprint [arXiv:2010.11929](https://arxiv.org/abs/2010.11929) (2020).
25. Liu, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* 10012–10022 (2021).
26. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25** (2012).
27. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. arXiv preprint [arXiv:1711.05101](https://arxiv.org/abs/1711.05101) (2017).
28. Loshchilov, I. & Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint [arXiv:1608.03983](https://arxiv.org/abs/1608.03983) (2016).
29. Haque, R., Hassan, M. M., Bairagi, A. K. & Shariful Islam, S. M. Neuronet19: An explainable deep neural network model for the classification of brain tumors using magnetic resonance imaging data. *Sci. Rep.* **14**, 1524 (2024).
30. Celik, M. & Inik, O. Development of hybrid models based on deep learning and optimized machine learning algorithms for brain tumor multi-classification. *Expert Syst. Appl.* **238**, 122159 (2024).
31. Liu, Z., Lv, Q., Li, Y., Yang, Z. & Shen, L. Medaugument: Universal automatic data augmentation plug-in for medical image analysis. [arXiv: 2306.17466](https://arxiv.org/abs/2306.17466) (2023).
32. Kebaili, A., Lapuyade-Lahorgue, J. & Ruan, S. Deep learning approaches for data augmentation in medical imaging: A review. *J. Imaging* **9**, 81 (2023).

33. Trabucco, B., Doherty, K., Gurinas, M. & Salakhutdinov, R. Effective data augmentation with diffusion models. arXiv preprint [arXiv:2302.07944](https://arxiv.org/abs/2302.07944) (2023).
34. Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I. & Lungren, M. P. Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines. *NPJ Digit. Med.* **3**, 136 (2020).
35. Liu, M., Cheng, D., Yan, W. & Initiative, A. D. N. Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. *Front. Neuroinform.* **12**, 35 (2018).
36. Khagi, B. & Kwon, G.-R. 3D CNN design for the classification of Alzheimer's disease using brain MRI and pet. *IEEE Access* **8**, 217830–217847 (2020).
37. Majkowska, A. et al. Chest radiograph interpretation with deep learning models: Assessment with radiologist-adjudicated reference standards and population-adjusted evaluation. *Radiology* **294**, 421–431 (2020).
38. Himel, G. M. S. & Islam, M. M. Benchmark analysis of various pre-trained deep learning models on ASSIRA cats and dogs dataset. arXiv preprint [arXiv:2401.04666](https://arxiv.org/abs/2401.04666) (2024).

Author contributions

Y.M. and J.K. designed the project and performed the simulations; all authors helped in analyzing the results and contributed to writing the manuscript; all authors reviewed the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.M. or M.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com