Infinite Science
Publishing

On the search of CNN parameters

Max Bindemann[1*] and Thomas Schanze[1]

[1] *IBMT, FB Life Science Engineering, Technische Hochschule Mittelhessen, Gießen, Germany*
[*] *Corresponding author, email: max.bindemann@lse.thm.de*

Abstract: The design of neural network architectures often relies on heuristic decisions. In this study, we systematically varied the number of convolutional filters, convolutional blocks, and dense-layer widths in convolutional neural networks (CNNs) and evaluated their performance on a brain tumor classification task. Using 682 model configurations and repeated training runs, we observed that classification accuracy depends strongly on the chosen architectural design. Multiple network configurations achieved competitive performance, indicating that efficient models can be identified without relying on large architectures. These findings highlight the importance of structured architecture exploration for constructing effective and resource-efficient neural networks in biomedical imaging.

I. Introduction

Deep learning has become a central tool in biomedical image analysis, supporting tasks such as tumor classification, lesion detection, and organ segmentation. Recent developments have brought powerful architectures such as Vision Transformer or State Space Models [1]. However, these models often come with high resource requirements and long training times. In practical biomedical settings, where datasets are limited in size and availability, images often contain more noise and artifacts, and hardware resources are constrained, classical convolutional neural networks (CNNs) remain highly relevant and frequently achieve competitive performance compared to more complex architectures [2]. Despite the continued success of CNNs, the process of designing an efficient architecture remains largely intuition-driven. Against this background, this study analyzes how variations in convolutional depth, filter width, and hidden dense-layer size affect classification performance in MRI-based brain tumor classification.

II. Material and methods

Experiments were conducted using the BRISC2025 [3] brain tumor classification dataset, which contains images of T1-weighted MRI scans across different anatomical planes and four diagnostic classes. The dataset provides a predefined, stratified, patient-level split (6,000 training, 1,000 test images). A stratified 85/15 split of the training data was used to create the validation set. All images were resized to 256 x 256 pixels and normalized to [0,1]. Data augmentation consisted of random rotations up to ±15°. Multiple CNN models were implemented following the base architecture illustrated in Fig. 1. The number of convolutional blocks, the number of filters per layer, the number of dense blocks, and the width of the dense layers were systematically varied to assess the influence of architectural depth and width. Within each model configuration, the number of convolutional filters increased with depth, while the number of neurons in the dense layers decreased towards the output.
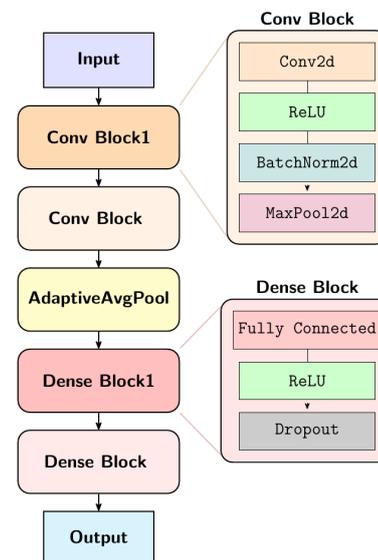

Figure 1: CNN structure with variable layer configuration and detailed structure of building blocks.

The design space included convolutional filter sizes {16, 32, 64, 256} and dense-layer widths {64, 128, 256, 512} as the available parameter choices for generating the model configurations. Dropout was set to either 0.0 or 0.4 across all dense blocks. Aside from these controlled variations, all remaining architectural components, including activation functions, pooling strategy, batch normalization, and the classifier head, were kept identical across all models to ensure fair comparability. All networks were trained with a batch size of 32 for up to 200 epochs. Early stopping with a patience of 25 was used as well as a cosine-annealing learning rate schedule (0.001 to 0.0001) with warm-up. All training runs were repeated three times with different fixed seeds, and model performance was assessed using accuracy, precision, recall, F1-score, cross-entropy loss, and total trainable parameters, reported as mean values across the three runs. The experiments were performed using Python 3.12.12 and PyTorch 2.9.1 with CUDA 12.8 on NVIDIA L40S GPUs.
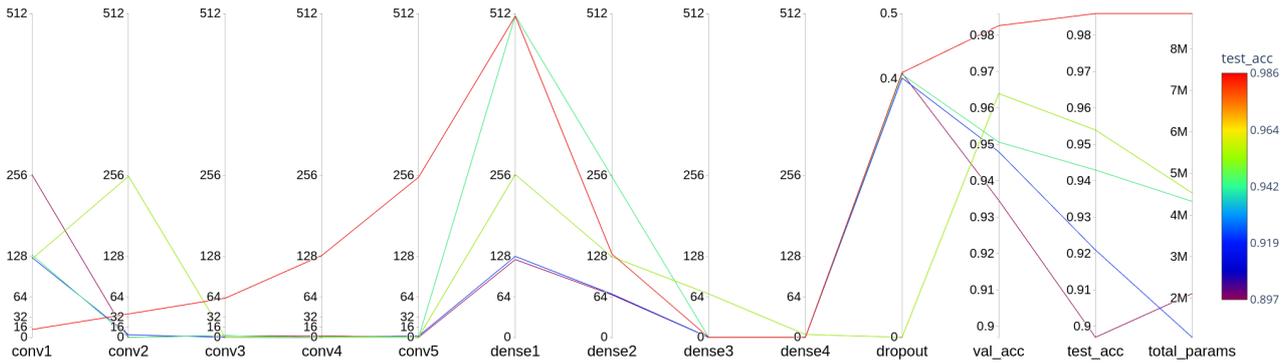
Figure 2: The five configurations represent linearly spaced accuracy levels between the highest-performing model (100%) and the lowest-performing one (0%), with intermediate points at 75%, 50%, and 25%. The term 'conv' refers to the number of convolutional filters in each layer, while 'dense' indicates the number of neurons in the fully connected layers; a value of 0 denotes the absence of a block. Line color represents the respective mean test accuracy. Red: 0.986; Purple: 0.897.

III. Results and discussion

In total, we trained and evaluated 682 different network configurations. Fig. 2 presents the models achieving the highest and lowest mean test accuracy, along with three intermediate configurations corresponding to 25%, 50%, and 75% of the accuracy range. The results show that final classification performance depends strongly on the chosen architectural design. The highest mean test accuracy of 0.9853 ± 0.0031 (precision 0.9846 ± 0.0037, recall 0.9873 ± 0.0023, F1-score 0.9858 ± 0.0031) was obtained by a model comprising five convolutional blocks {16, 32, 64, 128, 256}, two dense blocks {512, 128}, and a dropout rate of 0.4, resulting in 8,848,612 trainable parameters. The third-best model, using four convolutional blocks {16, 64, 128, 256}, two dense blocks {256, 128}, and the same dropout rate, achieved a mean test accuracy of 0.9813 ± 0.0068 (precision 0.9785 ± 0.0091, recall 0.9837 ± 0.0059, F1-score 0.9809 ± 0.0076) while requiring almost 50% fewer parameters (4,607,364). For comparison, previously reported results on the same dataset [3] show that EfficientNetB0 achieved a mean test accuracy of 0.9920 (F1-score: 0.9920), whereas ResNet50 reached 0.9820 mean accuracy with an F1-score of 0.9820.

Fig. 3 illustrates the distribution of mean test accuracies across all configurations. Notably, high-performing models emerged across a range of model sizes, indicating that no single depth-width relationship universally dominated. This suggests that multiple architectural configurations can achieve competitive performance within the tested design space.
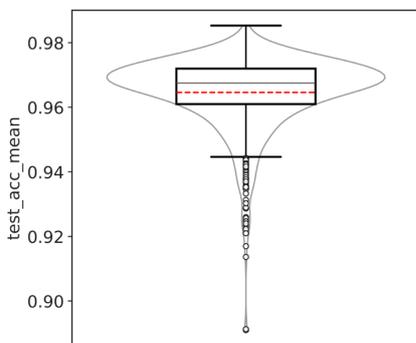

Figure 3: Distribution of mean test accuracy across all model configurations. The boxplot indicates the mean (0.9646) by a red line and median (0.9675) by a gray line. Outliers are marked with circles. The violin shape illustrates the estimated density distribution.

Although the best-performing configurations achieved solid results, performance could likely be improved further through additional fine-tuning, hyperparameter optimization, or expanded data augmentation strategies [4]. Moreover, because this study was conducted on a single dataset, future work should evaluate the identified architectural trends on other biomedical imaging domains to assess their generalizability.

IV. Conclusion

This study explored the influence of convolutional depth and filter width, as well as dense-layer complexity on CNN performance for brain tumor classification. The results show that multiple network configurations can achieve high and competitive accuracy, and that deeper models with progressively increasing filters tend to perform more consistently. At the same time, efficient architectures with fewer parameters can match the performance of larger models when appropriately configured. These findings underline the value of structured design space exploration for developing reliable and computationally efficient neural networks. In medical engineering applications, such efficiency is particularly relevant, as diagnostic models must operate under hardware-specific constraints. While a direct quantitative comparison with other researchers' results on the same dataset or with other datasets would further contextualize these outcomes, the presented systematic architecture optimization provides valuable insights that can facilitate future benchmarking, ensuring both predictive performance and practical implementation in clinical settings.

AUTHOR'S STATEMENT

The authors state no funding involved and no conflicts of interest.

REFERENCES

[1] Q. Xia *et al.*, "A comprehensive review of deep learning for medical image segmentation," *Neurocomputing*, vol. 613, p. 128740, Jan. 2025, doi: 10.1016/j.neucom.2024.128740.

[2] F. Isensee *et al.*, "nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation," July 25, 2024, *arXiv*: arXiv:2404.09556. doi: 10.48550/arXiv.2404.09556.

[3] A. Fateh *et al.*, "BRISC: Annotated Dataset for Brain Tumor Segmentation and Classification," Nov. 05, 2025, *arXiv*: arXiv:2506.14318. doi: 10.48550/arXiv.2506.14318.

[4] F. Garcea, A. Serra, F. Lamberti, and L. Morra, "Data augmentation for medical imaging: A systematic literature review," *Computers in Biology and Medicine*, vol. 152, p. 106391, Jan. 2023, doi: 10.1016/j.compbiomed.2022.106391.