International Journal of Electrical Engineering, Mathematics and Computer Science
E-ISSN: 3048-1945
P-ISSN: 3048-1910

Research Article Article

A Statistically Validated Comparison of CNN Architectures and Optimizers for Brain Tumor Classification on MRI Images

I Made Dharma Yoga Pratama [1], Syadia Nabilah Binti Mohd Safuan [2*]

[1] School of Graduate Studies, Management and Science University, Shah Alam, Selangor, Malaysia; e-mail: yogapratama0201@gmail.com
[2] Department of Information Technology, Faculty of Information Sciences and Engineering, Management and Science University, Shah Alam, Selangor, Malaysia; e-mail: syadia_nabilah@msu.edu.my
* Corresponding Author: I Made Dharma Yoga Pratama; e-mail: yogapratama0201@gmail.com

Abstract: Brain tumors are a major global health burden, and Magnetic Resonance Imaging (MRI) is the primary modality for their detection. Manual interpretation of MRI is time-consuming and subject to inter-observer variability, motivating automated classification using Convolutional Neural Networks (CNNs). A recurring weakness in existing studies is the comparison of CNN architectures and optimizers using single-run accuracy without formal statistical testing, leaving it unclear whether reported differences are genuine or due to random training variation. This study presents a statistically validated comparison of four CNN architectures (AlexNet, VGG-16, InceptionV3, and MobileNetV2) for four-class brain tumor classification (glioma, meningioma, pituitary, and no tumor) on the Kaggle Brain Tumor MRI dataset (7,023 images). AlexNet was trained from scratch, whereas the other three used feature-extraction transfer learning with frozen ImageNet-pretrained backbones. All models were evaluated using 5-fold stratified cross-validation, and architecture differences were assessed using one-way ANOVA with Tukey HSD post-hoc testing and Cohen's d effect sizes. The best architecture was then optimized with Adam and Stochastic Gradient Descent with Momentum (SGDM), compared using a paired t-test at the 0.05 level. AlexNet achieved the highest mean test accuracy of 96.63%, significantly outperforming VGG-16 (92.60%), InceptionV3 (88.89%), and MobileNetV2 (88.25%), with large effect sizes. Adam and SGDM produced statistically equivalent accuracy (97.38% versus 96.52%). The results indicate that learning domain-specific features from scratch can surpass non-fine-tuned transfer learning, and that the choice between Adam and SGDM does not significantly affect accuracy.

Keywords: Brain Tumor; CNN Architecture; Paired t-Test; Statistical Validation; Transfer Learning.

Received : April 30, 2025
Revised : May 27, 2025
Accepted : June 22, 2025
Published : June 30, 2025
Current Ver.: June 30, 2025

1. Introduction

The brain is the control center for cognitive, motor, and sensory functions, and any pathology affecting it can have severe consequences. Brain tumors, defined by the abnormal and neoplastic growth of cells within the brain, are categorized into primary tumors originating from brain tissue and metastatic tumors that spread from other organs, with gliomas, meningiomas, and pituitary tumors among the most common primary types (Louis et al., 2021; Ostrom et al., 2020).Epidemiological data indicate that brain and central nervous system tumors constitute a substantial disease burden worldwide, with Asia accounting for the largest share of new cases, underscoring the need for early and accurate diagnosis (Global Cancer Observatory, 2022).

Magnetic Resonance Imaging (MRI) is the modality of choice for brain tumor detection because it produces high-contrast soft-tissue images across multiple planes without ionizing radiation (Booth et al., 2020). Nevertheless, the interpretation of MRI scans is performed

manually by radiologists, a process that is time-consuming and prone to inter-observer variability, which can delay diagnosis and reduce consistency (Bouget et al., 2022). These limitations have driven the adoption of automated image analysis methods, particularly deep learning. Convolutional Neural Networks (CNNs) are especially well suited to this task because they learn hierarchical features directly from image data, eliminating the dependence on handcrafted features that characterizes conventional methods (Saeedi et al., 2023; Sarvamangala & Kulkarni, 2022). CNN-based systems have accordingly been reported to classify brain MRI with high accuracy across a range of architectures (Abdusalomov et al., 2023; Gómez-Guzmán et al., 2023).

Despite these advances, a persistent methodological gap remains in how CNN architectures and optimizers are compared for this task, as detailed in the following section. The objective of this study is therefore to compare four CNN architectures and two optimizers for multi-class brain tumor classification on MRI under identical experimental conditions with formal statistical validation. The main contributions of this work are: (1) a controlled comparison of four CNN architectures under an identical hyperparameter configuration; (2) formal statistical validation of architecture differences using one-way ANOVA, Tukey HSD post-hoc testing, and Cohen's d effect sizes; and (3) a paired statistical comparison of the Adam and SGDM optimizers on the best architecture. The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 describes the proposed method, Section 4 presents the results and discussion, Section 5 provides a comparison with the state of the art, and Section 6 concludes the study.

2. Related Work

CNN-based methods have become the dominant approach for automated brain tumor classification on MRI. Many studies adopt transfer learning from ImageNet-pretrained backbones to reduce training cost and improve accuracy on limited medical datasets (Deepak & Ameer, 2023; Srinivas et al., 2022). On the same publicly available Kaggle Brain Tumor MRI dataset of 7,023 images used in this study, (Disci et al., 2025) fine-tuned six pretrained architectures and reported that Xception achieved the highest weighted accuracy, while (Islam et al., 2023) similarly found that fine-tuned transfer learning yielded strong performance. Comparable results have been reported by (Mathivanan et al., 2024; Rajput et al., 2023), who applied deep and transfer learning to brain MRI classification. Reported results generally favor deeper modern architectures; however, the way these comparisons are conducted has an important weakness.

A recurring limitation is that many studies compare architectures using accuracy from a single experimental run and omit statistical hypothesis testing (Anwar et al., 2025; Iliyasu et al., 2025; Prabhas et al., 2025). Because neural network training involves stochastic processes such as weight initialization, mini-batch sampling, and data shuffling, two architectures may differ by one to three percentage points purely from training variability. Without formal statistical validation, it is impossible to determine whether an observed difference reflects a genuine advantage or random fluctuation. The same limitation applies to optimizer comparisons, which are frequently reported without a paired statistical test under identical conditions (Amin, Anjum, et al., 2022; Chandni et al., 2023).

In addition, the transfer learning strategy itself is often applied without examining whether frozen feature extraction is appropriate for the medical domain. Prior work has reported that fine-tuning, rather than fixed feature extraction, is often required for strong transfer to medical imaging because of the domain gap between natural images and MRI (Anwar et al., 2025; Islam et al., 2023; Swati et al., 2019). These observations motivate a controlled, statistically validated comparison that isolates the architecture and optimizer as the only variables, which is the focus of this study.

3. Proposed Method

This section presents the proposed methodology for comparing the four CNN architectures and two optimizers under identical conditions. The workflow comprises dataset preparation, preprocessing and augmentation, architecture design and transfer learning, model training with 5-fold stratified cross-validation, and formal statistical analysis. Each component is described in the following subsections and is structured so that the architecture and optimizer remain the only variables under study.

Dataset

The experiments used the publicly available Kaggle Brain Tumor MRI dataset (Masoud Nickparvar, 2021), which contains 7,023 T1-weighted contrast-enhanced MRI images across four classes: glioma, meningioma, pituitary, and no tumor. The dataset was partitioned into a training and cross-validation set of 5,712 images and an independent test set of 1,311 images reserved exclusively for final evaluation. The class distribution across both partitions is relatively balanced, as shown in Figure 1.

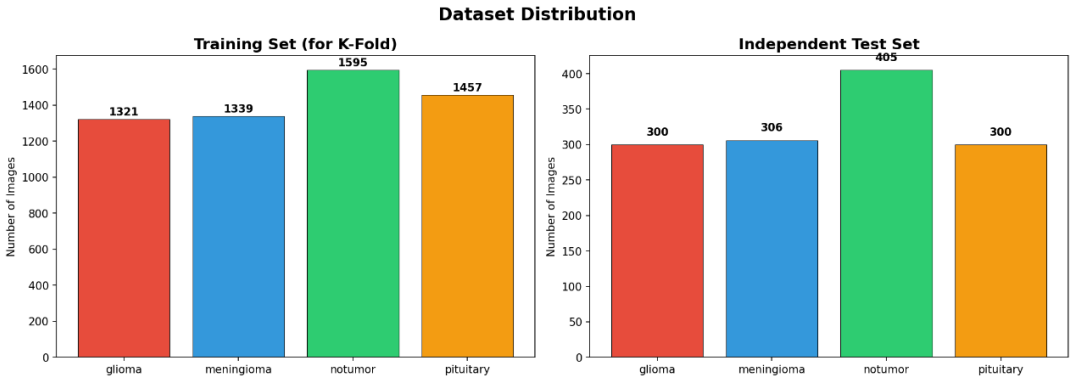

Figure 1. Class distribution of the training and independent test partitions.

Preprocessing and Data Augmentation

All images were resized to 224×224 pixels for a uniform input dimension across the four architectures. Pixel intensities were normalized using architecture-specific functions: AlexNet used a rescaling layer dividing pixel value by 255, while VGG-16, InceptionV3, and MobileNetV2 used their respective preprocess_input functions corresponding to ImageNet training. Class folder names were encoded to integer labels (0–3) and converted to one-hot vectors for the Softmax output and categorical cross-entropy loss. Five augmentation operations were applied to the training subset only: random horizontal flip, rotation up to ±10.8°, translation up to ±5%, zoom up to ±10%, and contrast adjustment up to ±15%. Validation and test data were not augmented. Representative original and augmented samples are shown in Figure 2.

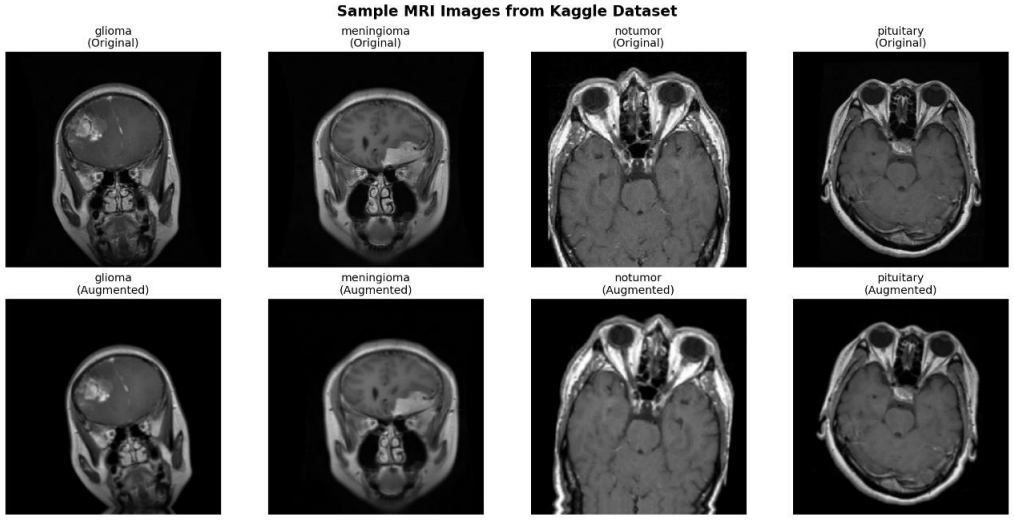

Figure 2. Original and augmented MRI samples for each tumor class.

CNN Architecture and Transfer Learning Strategy

Four architectures were evaluated to span a range of design philosophies. AlexNet was implemented as a custom CNN and trained from scratch as a classical baseline. VGG-16, InceptionV3, and MobileNetV2 were deployed using feature-extraction transfer learning: their convolutional backbones were initialized with ImageNet-pretrained weights and frozen, so only the appended classifier head was trained (Deepak & Ameer, 2023). The classifier head, identical across the three pretrained models, consisted of a global average pooling layer, a

dense layer of 512 units with ReLU activation, a dropout layer (rate 0.5), a dense layer of 256 units with ReLU activation, a dropout layer (rate 0.3), and a final dense layer of four units with Softmax activation.

Experimental Setup

All experiments were conducted using TensorFlow 2.16 on Google Colab (Google LLC, Mountain View, CA, USA) with an NVIDIA T4 GPU (NVIDIA Corporation, Santa Clara, CA, USA). A fixed random seed (seed = 42) was applied across NumPy, TensorFlow, and the Python random module for reproducibility. All four architectures were trained under the identical hyperparameter configuration summarized in Table 1, isolating architecture as the only variable. A learning-rate warmup over the first five epochs ramped the rate from 1×10^{-4} to 1×10^{-3} to stabilize early training.

Table 1. Experimental hyperparameters applied uniformly to all four architectures.

Parameter	Value
Input image size	224 × 224 × 3
Batch size	32
Maximum epochs	50
Learning rate	1×10^{-3} (warmup from 1×10^{-4} over 5 epochs)
Optimizer (Objective I)	Adam (β_1 = 0.9, β_2 = 0.999)
Cross-validation	5-fold stratified (seed = 42)
Early stopping	Patience = 15, after epoch 5
Loss function	Categorical cross-entropy

Cross-Validation Protocol

The 5,712-image training set was divided using 5-fold stratified cross-validation (scikit-learn StratifiedKFold, seed = 42). In each fold, approximately 4,570 images were used for training and 1,142 for validation, with stratification preserving class proportions. The best-weight model from each fold, saved on the highest validation accuracy, was evaluated on the independent 1,311-image test set, producing five test accuracy values per configuration that served as input for all statistical analyses.

Statistical Analysis

For the architecture comparison, a one-way ANOVA was applied to the five test accuracy scores of each architecture at the 0.05 level. When significant, Tukey HSD post-hoc testing identified which pairs differed, and Cohen's d quantified practical significance. For the optimizer comparison, normality of the paired fold-wise differences between Adam and SGDM was assessed with the Shapiro-Wilk test; because normality held, a parametric paired t-test was applied, with the Wilcoxon signed-rank test reserved as the non-parametric alternative. The significance threshold was set at 0.05 for all tests. All tests were performed using SciPy in Python.

4. Results and Discussion

Architecture Comparison

Table 2 reports the mean test accuracy, macro-averaged F1-score, and cumulative training time across the five folds. AlexNet achieved the highest mean test accuracy of 96.63% ± 0.96%, followed by VGG-16 (92.60% ± 1.07%). InceptionV3 and MobileNetV2 performed comparably at 88.89% ± 0.48% and 88.25% ± 0.43%. MobileNetV2 was the most efficient (39.2 minutes across five folds), while VGG-16 required the longest training time (88.1 minutes).

Table 2. Five-fold cross-validation results for the four CNN architectures.

Architecture	Test accuracy (Mean ± Std)	Macro F1 (Mean ± Std)	Training time
AlexNet	96.63% ± 0.96%	0.9645 ± 0.0104	53.8 min
VGG-16	92.60% ± 1.07%	0.9200 ± 0.0116	88.1 min
InceptionV3	88.89% ± 0.48%	0.8800 ± 0.0049	70.2 min
MobileNetV2	88.25% ± 0.43%	0.8729 ± 0.0049	39.2 min

AlexNet, the oldest and architecturally simplest model, outperformed the three modern pretrained architectures by a substantial margin despite being trained entirely from scratch,

whereas the others relied on frozen, non-fine-tuned ImageNet backbones. MobileNetV2 exhibited the lowest inter-fold variance. Figure 3 compares the four architectures across accuracy, per-class F1-score, and training time.

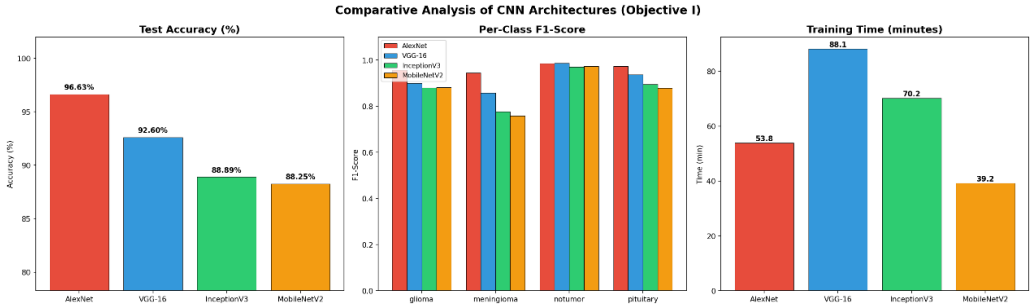

Figure 3. Comparative analysis of the four CNN architectures.

Per-Fold Accuracy Analysis

Table 3 presents the per-fold test accuracy. AlexNet remained consistently high across folds (95.27%–97.71%). VGG-16 showed a declining trend, highest in Fold 2 (93.90%) and lowest in Fold 5 (90.85%). InceptionV3 and MobileNetV2 were flat but consistently lower, as illustrated in Figure 4.

Table 3. Per-fold test accuracy across all four architectures (%).

Fold	AlexNet	VGG-16	InceptionV3	MobileNetV2
1	95.27	93.52	89.24	88.10
2	97.71	93.90	89.40	89.09
3	97.64	92.52	89.17	88.10
4	96.64	92.22	88.48	88.87
5	95.88	90.85	88.18	88.10
Mean	96.63	92.60	88.89	88.29
Std	±0.96	±1.07	±0.48	±0.43

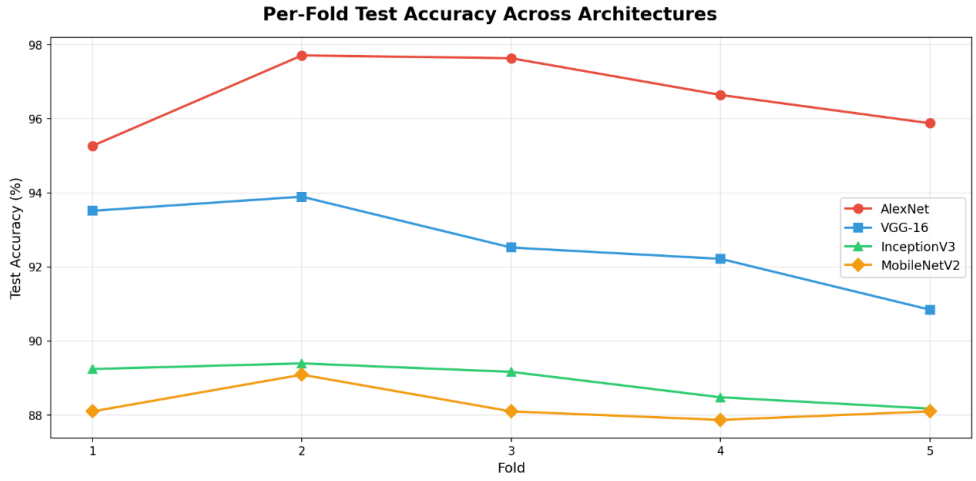

Figure 4. Per-fold test accuracy across 5-fold cross-validation.

Confusion Matrix Analysis

Figure 5 presents the confusion matrices from the best fold model of each architecture. AlexNet exhibited the fewest misclassifications, with primary confusion between glioma and meningioma and between meningioma and no tumor, and correctly identified 404 of 405 no-tumor images. VGG-16 showed increased glioma–meningioma confusion. InceptionV3 and MobileNetV2 exhibited substantially more misclassifications, dominated by meningioma misclassified as pituitary (43 and 60 cases, respectively).

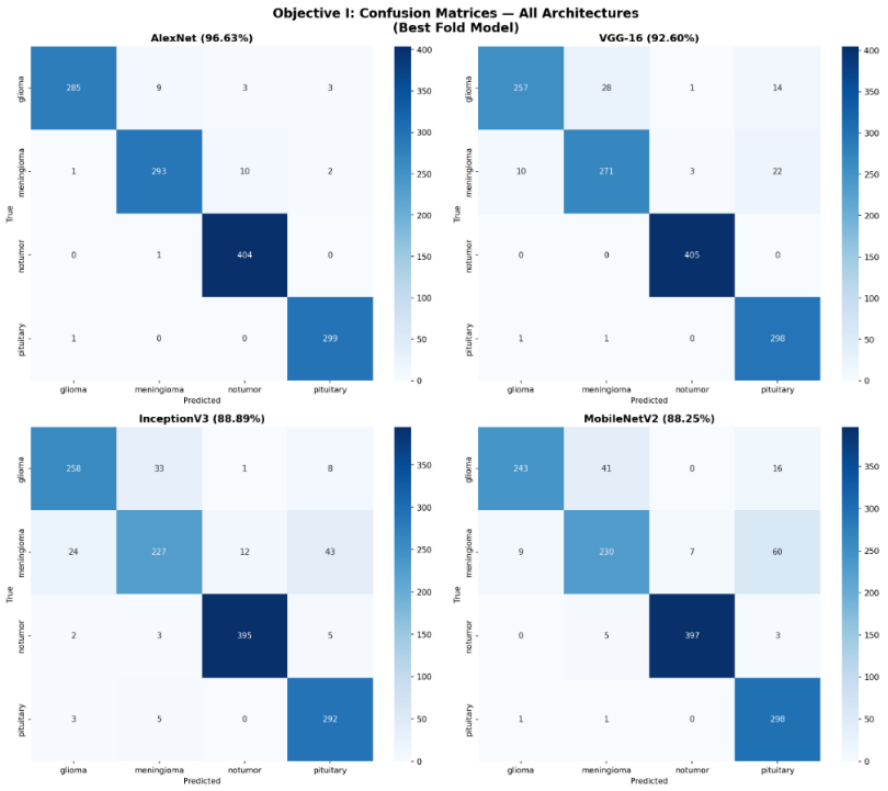

Figure 5. Confusion matrices for the four CNN architectures on the independent test set.

Optimizer Comparison

AlexNet was retrained under two optimizer scenarios using identical cross-validation splits: Adam (learning rate 1×10^{-3}, $\beta_1 = 0.9$, $\beta_2 = 0.999$) and SGDM (learning rate 1×10^{-3}, momentum 0.9), with all other hyperparameters held constant. As shown in Table 4, Adam achieved a marginally higher mean accuracy (97.38% ± 0.36%) than SGDM (96.52% ± 0.86%), a difference of 0.85 percentage points, with lower variance.

Table 4. AlexNet optimizer comparison over 5-fold cross-validation.

Scenario	Test accuracy (Mean ± Std)	Macro F1 (Mean ± Std)	Training time
AlexNet + Adam	97.38% ± 0.36%	0.9735 ± 0.0037	54 min
AlexNet + SGDM	96.52% ± 0.86%	0.9648 ± 0.0089	54 min

Discussion

The most striking finding is that AlexNet, trained from scratch, significantly outperformed three modern architectures used as frozen feature extractors. A plausible explanation lies in the domain gap between natural images and brain MRI. ImageNet pretraining encodes features optimized for natural photographs, where color, texture, and object boundaries differ from the intensity-based contrast and subtle textural cues that distinguish tumor types on T1-weighted MRI. When the backbone is frozen, these features cannot adapt to the medical domain, limiting discriminative power, whereas AlexNet learns features directly from the target distribution. This does not imply that transfer learning is inferior in general; rather, feature extraction without fine-tuning can underperform a model trained from scratch on a sufficiently large medical dataset. This observation is consistent with prior reports that fine-tuning, rather than fixed feature extraction, is often required for strong transfer to medical imaging (Islam et al., 2023; Swati et al., 2019). It also aligns with results on the same dataset, where fine-tuned backbones reached very high accuracy (Disci et al., 2025).

The optimizer comparison reinforces the value of statistical validation. Although Adam achieved a higher mean accuracy than SGDM, the difference was not significant, indicating that prior reports of optimizer superiority based on single-run accuracy should be interpreted with caution (Amin, Sharif, et al., 2022; Chandni et al., 2023). The large Cohen's d alongside

a non-significant P-value also illustrates a known limitation of cross-validation with small k: effect sizes may appear large while the test lacks power to reach significance. Reporting both quantities provides a more complete picture than either alone.

Several limitations should be acknowledged. First, the pretrained backbones were used purely as feature extractors and were not fine-tuned; fine-tuning could substantially change their performance and is an important direction for future work (Mathivanan et al., 2024; Rajput et al., 2023). Second, the statistical tests were based on five folds, limiting statistical power. Third, the evaluation used a single dataset, so generalizability to other institutions and imaging protocols was not assessed here and warrants dedicated cross-dataset investigation.

5. Comparison

A central contribution of this study is that the differences between architectures were subjected to formal statistical validation rather than being asserted from single-run accuracy. A one-way ANOVA on the five-fold test accuracies yielded a significant result ($p < 0.05$), rejecting the null hypothesis of equal mean accuracy and confirming that at least one architecture differs from the others. Tukey HSD post-hoc results are presented in Table 5.

Table 5. Tukey HSD post-hoc pairwise comparisons.

Group 1	Group 2	Mean diff.	Significant
AlexNet	InceptionV3	+7.73%	Yes
AlexNet	MobileNetV2	+8.38%	Yes
AlexNet	VGG-16	+4.03%	Yes
VGG-16	InceptionV3	+3.71%	Yes
VGG-16	MobileNetV2	+4.35%	Yes
InceptionV3	MobileNetV2	+0.64%	No

Five of the six comparisons were significant; only InceptionV3 versus MobileNetV2 was not. Cohen's d effect sizes (Table 6) were very large for all comparisons involving AlexNet, confirming that its advantage is both statistically significant and practically meaningful.

Table 6. Cohen's d effect sizes for pairwise architecture comparisons.

Model A	Model B	Cohen's d	Interpretation
AlexNet	VGG-16	3.5429	Large
AlexNet	InceptionV3	9.1429	Large
AlexNet	MobileNetV2	10.0929	Large
VGG-16	InceptionV3	3.9962	Large
VGG-16	MobileNetV2	4.7624	Large
InceptionV3	MobileNetV2	1.2642	Large

For the optimizer comparison, the Shapiro-Wilk test on the paired differences yielded $W = 0.8777$ ($P = 0.2989$), confirming normality and justifying a paired t-test. The paired t-test produced $t = 1.5292$ ($P = 0.2009$); because P exceeded 0.05, the difference was not statistically significant. Cohen's d was 1.1658 (large). The co-occurrence of a large effect size with a non-significant P-value is attributable to the small sample ($n = 5$ folds), which limits statistical power, and does not invalidate the conclusion of comparable performance. Table 7 summarizes the analysis.

Table 7. Statistical analysis summary of the optimizer comparison.

Statistical measure	Result
Shapiro-Wilk (paired differences)	$W = 0.8777$, $P = 0.2989 \rightarrow$ Normal
Test selected	Paired t-test
Test statistic	$t = 1.5292$, $P = 0.2009$

Significant at 0.05	No
Mean difference (Adam − SGDM)	+0.85 percentage points
Cohen's d	1.1658 (large)

Compared with prior studies on the same dataset that report architecture or optimizer superiority from single-run accuracy (Anwar et al., 2025; Disci et al., 2025; Islam et al., 2023), the statistically validated comparison presented here provides a more measurable and reliable illustration of the research contribution. It shows that, for brain MRI classification, an architecturally simple model trained from scratch can significantly and meaningfully outperform non-fine-tuned transfer learning, and that the apparent advantage of Adam over SGDM does not survive a paired statistical test.

6. Conclusions

This study presented a statistically validated comparison of four CNN architectures and two optimizers for multi-class brain tumor classification on MRI, addressing the objective of evaluating them under identical conditions with formal hypothesis testing. AlexNet trained from scratch achieved the highest mean accuracy of 96.63% and significantly outperformed VGG-16, InceptionV3, and MobileNetV2, as confirmed by one-way ANOVA, Tukey HSD post-hoc testing, and large Cohen's d effect sizes. Adam and SGDM produced statistically equivalent accuracy ($P = 0.2009$), with Adam selected as the final configuration on the basis of marginally higher mean accuracy and lower variance. For brain MRI classification, learning domain-specific features from scratch can surpass non-fine-tuned transfer learning, and the choice between Adam and SGDM does not significantly affect accuracy. Future work will explore fine-tuning of pretrained backbones, larger cross-validation configurations to increase statistical power, and external cross-dataset validation.

Author Contributions: Conceptualization: I.M.D.Y.P. and S.N.M.S.; Methodology: I.M.D.Y.P.; Software: I.M.D.Y.P.; Validation: I.M.D.Y.P. and S.N.M.S.; Formal analysis: I.M.D.Y.P.; Investigation: I.M.D.Y.P.; Data curation: I.M.D.Y.P.; Writing—original draft preparation: I.M.D.Y.P.; Writing—review and editing: S.N.M.S.; Visualization: I.M.D.Y.P.; Supervision: S.N.M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The Brain Tumor MRI dataset used in this study is publicly available on Kaggle (Masoud Nickparvar, 2021). No new data were created in this study.

Acknowledgments: The authors thank Management and Science University for institutional support.

Conflicts of Interest: The authors declare no conflict of interest.

References

Abdusalomov, A. B., Mukhiddinov, M., & Whangbo, T. K. (2023). Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging. *Cancers*, *15*(16), 4172. https://doi.org/10.3390/CANCERS15164172

Amin, J., Anjum, M. A., Sharif, M., Jabeen, S., Kadry, S., & Moreno Ger, P. (2022). A New Model for Brain Tumor Detection Using Ensemble Transfer Learning and Quantum Variational Classifier. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/3236305

Amin, J., Sharif, M., Haldorai, A., Yasmin, M., & Nayak, R. S. (2022). Brain tumor detection and classification using machine learning: a comprehensive survey. *Complex and Intelligent Systems*, *8*(4), 3161–3183. https://doi.org/10.1007/s40747-021-00563-y

Anwar, R. W., Abrar, M., & Ullah, F. (2025). Transformative Transfer Learning for MRI Brain Tumor Precision: Innovative Insights. *IEEE Access*, *13*, 31749–31761. https://doi.org/10.1109/ACCESS.2025.3542451

Booth, T. C., Williams, M., Luis, A., Cardoso, J., Ashkan, K., & Shuaib, H. (2020). *Machine learning and glioma imaging biomarkers*. https://doi.org/10.1016/j.crad.2019.07.001

Bouget, D., Pedersen, A., Jakola, A. S., Kavouridis, V., Emblem, K. E., Eijgelaar, R. S., Kommers, I., Ardon, H., Barkhof, F., Bello, L., Berger, M. S., Conti Nibali, M., Furtner, J., Hervey-Jumper, S., Idema, A. J. S., Kiesel, B., Kloet, A., Mandonnet, E., Müller, D. M. J., … Reinertsen, I. (2022). Preoperative Brain Tumor Imaging: Models and Software for Segmentation and Standardized Reporting. *Frontiers in Neurology*, *13*, 932219. https://doi.org/10.3389/FNEUR.2022.932219

Chandni, Sachdeva, M., & Kushwaha, A. K. S. (2023). The power of deep learning for intelligent tumor classification systems: A review. *Computers and Electrical Engineering*, *106*, 108586. https://doi.org/10.1016/J.COMPELECENG.2023.108586

Deepak, S., & Ameer, P. M. (2023). Brain tumor categorization from imbalanced MRI dataset using weighted loss and deep feature fusion. *Neurocomputing*, *520*, 94–102. https://doi.org/10.1016/J.NEUCOM.2022.11.039

Disci, R., Gurcan, F., Soylu, A., Disci, R., Gurcan, F., & Soylu, A. (2025). Advanced Brain Tumor Classification in MR Images Using Transfer Learning and Pre-Trained Deep CNN Models. *Cancers 2025, Vol. 17, 17*(1). https://doi.org/10.3390/CANCERS17010121

Global Cancer Observatory. (2022). *Cancer site ranking*. (BRAIN CNS).

Gómez-Guzmán, M. A., Jiménez-Beristaín, L., García-Guerrero, E. E., López-Bonilla, O. R., Tamayo-Perez, U. J., Esqueda-Elizondo, J. J., Palomino-Vizcaino, K., Inzunza-González, E., Gómez-Guzmán, M. A., Jiménez-Beristaín, L., García-Guerrero, E. E., López-Bonilla, O. R., Tamayo-Perez, U. J., Esqueda-Elizondo, J. J., Palomino-Vizcaino, K., & Inzunza-González, E. (2023). Classifying Brain Tumors on Magnetic Resonance Imaging by Using Convolutional Neural Networks. *Electronics 2023, Vol. 12, 12*(4). https://doi.org/10.3390/ELECTRONICS12040955

Iliyasu, U., Dima, R. M., & Haruna, L. (2025). Comparative Analysis of Deep Learning Approaches for Brain Tumor Detection: CNN, Improved CNN, and Transfer Learning with VGG-16. *FUOYE Journal of Engineering and Technology*, *10*(1), 90–97. https://doi.org/10.46792/fuoyejet.v10i1.14

Islam, M. M., Barua, P., Rahman, M., Ahammed, T., Akter, L., & Uddin, J. (2023). Transfer learning architectures with fine-tuning for brain tumor classification using magnetic resonance imaging. *Healthcare Analytics*, *4*(8), 100270. https://doi.org/https://doi.org/10.1016/j.health.2023.100270

Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., Hawkins, C., Ng, H. K., Pfister, S. M., Reifenberger, G., Soffietti, R., Von Deimling, A., & Ellison, D. W. (2021). The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro-Oncology*, *23*(8), 1231. https://doi.org/10.1093/NEUONC/NOAB106

Masoud Nickparvar. (2021). *Brain Tumor MRI Dataset*. Kaggle. https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset

Mathivanan, S. K., Sonaimuthu, S., Murugesan, S., Rajadurai, H., Shivahare, B. D., & Shah, M. A. (2024). Employing deep learning and transfer learning for accurate brain tumor detection. *Scientific Reports 2024 14:1*, *14*(1), 7232-. https://doi.org/10.1038/s41598-024-57970-7

Ostrom, Q. T., Patil, N., Cioffi, G., Waite, K., Kruchko, C., & Barnholtz-Sloan, J. S. (2020). CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2013-2017. *Neuro-Oncology*, *22*(12 Suppl 2), IV1–IV96. https://doi.org/10.1093/NEUONC/NOAA200

Prabhas, K. S., Basem, A., Lakshmi, L., Talha, A., Mohammed, S. H., Khan, M. I., & Khedher, N. Ben. (2025). A Deep learning framework for brain tumor detection using CNNs and transfer learning on MRI scans. *Systems and Soft Computing*, *7*, 200389. https://doi.org/10.1016/J.SASC.2025.200389

Rajput, I. S., Gupta, A., Jain, V., & Tyagi, S. (2023). A transfer learning-based brain tumor classification using magnetic resonance images. *Multimedia Tools and Applications 2023 83:7*, *83*(7), 20487–20506. https://doi.org/10.1007/S11042-023-16143-W

Saeedi, S., Rezayi, S., Keshavarz, H., & R. Niakan Kalhori, S. (2023). MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC Medical Informatics and Decision Making*, *23*(1). https://doi.org/10.1186/s12911-023-02114-6

Sarvamangala, D. R., & Kulkarni, R. V. (2022). Convolutional neural networks in medical image understanding: a survey. In *Evolutionary Intelligence* (Vol. 15, Number 1). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s12065-020-00540-3

Srinivas, C., Nandini, N. P., Zakariah, M., Alothaibi, Y. A., Shaukat, K., Partibane, B., & Awal, H. (2022). Deep Transfer Learning Approaches in Performance Analysis of Brain Tumor Classification Using MRI Images. *Journal of Healthcare Engineering*, *2022*. https://doi.org/10.1155/2022/3264367

Swati, Z. N. K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., & Lu, J. (2019). Brain tumor classification for MR images using transfer learning and fine-tuning. *Computerized Medical Imaging and Graphics*, *75*, 34–46. https://doi.org/10.1016/j.compmedimag.2019.05.001