

Brain Tumor Classification from MRI Scans Using a Custom Convolutional Neural Network: Architecture, Training Dynamics, and Deployment Framework

¹Prof. Dr. Guruprakash C D, ²Afifa Tabassum H A, ³Ananya C, ⁴Mohammed Rayyan P,

⁵Sachin Honamatti

¹professor, ^{2,3,4,5}Dept. Compute Science Sri Siddhartha Institute of Technology, Tumkur, Karnataka

Abstract-Brain tumors represent one of the most lethal forms of cancer worldwide, with early and accurate diagnosis being critical for patient survival. This paper presents a comprehensive study on the development, training, and deployment of a custom Convolutional Neural Network (CNN) for automated classification of brain Magnetic Resonance Imaging (MRI) scans into four categories: glioma tumor, meningioma tumor, no tumor, and pituitary tumor. The proposed architecture consists of four convolutional layers with max-pooling, followed by fully connected layers with dropout regularization, totaling 495,972 trainable parameters. Trained on a dataset of 2,870 training images and 394 test images with extensive data augmentation, the model achieved a peak training accuracy of 97.37% and a best validation accuracy of 71.09%. However, significant over fitting was observed, with the validation loss increasing from 1.98 to 3.59 while training loss decreased to 0.08, revealing a 27.86% accuracy gap between training and validation performance. The study also details the deployment framework, including a Flask-based REST API and an interactive web interface for real-time tumor classification.

Keywords-Brain tumor classification, CNN, MRI, Deep Learning, Medical Image Analysis, over fitting

I. INTRODUCTION

“Brain tumors are among the most aggressive and lifethreatening forms of cancer, accounting for approximately 1.4% of all new cancer cases globally, with an estimated 308,102 new cases and 251,329 deaths reported in 2020 alone [1]. The World Health Organization (WHO) classifies brain tumors into over 120 distinct types, with gliomas, meningiomas, and pituitary adenomas representing the most prevalent primary brain tumors [2]. Gliomas, particularly grades III and IV (glioblastoma), are the most common malignant brain tumors with a median survival of only 12-15 months despite aggressive treatment [3].

Magnetic Resonance Imaging (MRI) has emerged as the gold standard for brain tumor detection and characterization due to its superior soft tissue contrast, multi-planar imaging capability, and non-invasive nature [4]. However, the manual interpretation of MRI scans by radiologists and neurosurgeons is time-consuming,

subjective, and prone to inter-observer variability. Studies have shown that diagnostic disagreement rates among neuroradiologists can range from **15% to 30%** for complex brain tumor cases [5].

The advent of deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized medical image analysis by enabling automated feature extraction and classification with accuracy levels approaching or exceeding human experts in specific tasks [6]. CNNs have demonstrated remarkable success in various medical imaging applications, including diabetic retinopathy detection, skin cancer classification, and chest X-ray analysis [7]. In the domain of brain tumor classification, CNN-based approaches have achieved validation accuracies ranging from **84% to over 97%** on benchmark datasets [8, 9].

This study aims to: (1) Design and implement a lightweight custom CNN architecture suitable for brain tumor classification from MRI scans; (2) Investigate the training dynamics, including the impact of data augmentation, learning rate scheduling, and regularization techniques; (3) Analyze the generalization performance and identify factors contributing to over fitting; (4) Develop a complete deployment pipeline including a REST API and web-based user interface for clinical utility; and (5) Compare the proposed approach with existing state-of-the-art methods and identify avenues for improvement

II. LITERATURE REVIEW

Prior to the deep learning era, brain tumor classification relied primarily on machine learning techniques combined with hand-engineered features. Gumaie et al. [11] proposed a hybrid feature extraction approach using normalized Extreme learning machines, achieving **94.23% accuracy** on a multiclass tumor dataset. These traditional methods typically involved preprocessing steps such as skull stripping, segmentation, and feature extraction using techniques like Gray Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), and wavelet transforms [12].

Recent literature has seen a proliferation of CNN-based approaches for brain tumor classification. A comprehensive comparative study evaluated twelve different CNN architectures including GoogleNet, MobileNetV2, Xception, DenseNet-BC, ResNet-50, VGG-

16, and AlexNet on a dataset of 3,264 MRI images across four classes [8]. GoogleNet achieved the highest validation accuracy of **97%**, while MobileNetV2 reached **96.4%** accuracy with significantly lower computational requirements.

Sultan et al. [14] proposed a 16-layer CNN architecture achieving **96.13% accuracy**. Badža et al. [16] demonstrated that even relatively shallow CNNs with just two convolutional layers could achieve **95.40% accuracy** when combined with aggressive data augmentation. Transfer learning approaches have also shown promise Sajjad et al. [17] and Swati et al. [18] both applied pre-trained VGG19 models, achieving **94.5%** and **94.8%** accuracy respectively.

A recurring challenge across brain tumor classification studies is the phenomenon of overfitting, where models achieve high training accuracy but fail to generalize to unseen validation data. This issue is particularly acute in medical imaging due to several factors: limited dataset sizes compared to natural image datasets like ImageNet; high intra-class similarity and inter-class visual overlap between tumor types; domain shift between training and test data from different scanners, protocols, or institutions; and class imbalance, where certain tumor types are underrepresented [20].

Regularization techniques employed to mitigate overfitting include: Dropout [21]; Data Augmentation [22]; Batch Normalization [23]; Early Stopping [24]; Learning Rate Scheduling [25]; and Transfer Learning [26]. Despite these techniques, validation accuracies typically lag training accuracies by **5-15 percentage points**, with some studies reporting gaps exceeding 25% [8, 19].

The translation of research models into clinical practice requires robust deployment frameworks. Flask-based REST APIs have emerged as a popular choice for serving deep learning models in healthcare applications due to their simplicity, Python ecosystem integration, and ease of containerization [27]. Web-based interfaces enable radiologists to upload MRI scans and receive classification results with probability scores, facilitating human-AI collaborative decision-making. Studies have shown that AI-assisted diagnosis can reduce radiologist reading time by **2030%** while improving diagnostic consistency [28].

III. METHODOLOGY

A. Dataset Description and Preprocessing

The dataset used in this study is a publicly available brain tumor MRI dataset commonly used in Kaggle competitions and academic research [29]. It comprises T1-weighted contrast-enhanced MRI scans organized into training and testing splits. **Figure 1** illustrates the class distribution across both splits.

use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

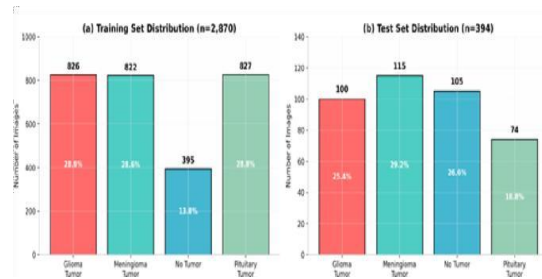


Figure 1: Dataset class distribution showing the number of images per class in (a) the training set (n=2,870) and (b) the test set (n=394). The 'No Tumor' class is underrepresented in the training data at 13.8%

Compared to approximately 28.8% for each tumor type. The dataset exhibits moderate class imbalance, with 'no tumor' cases being underrepresented (13.8%) compared to the three tumor categories (~28.8% each). All images were resized to **150x150 pixels** with 3 color channels (RGB). Pixel values were normalized to the [0, 1] range by dividing by 255

B. Data Augmentation Strategy

To artificially expand the training dataset and improve model robustness, the following augmentation pipeline was applied using Keras ImageDataGenerator, as illustrated in **Figure 2**

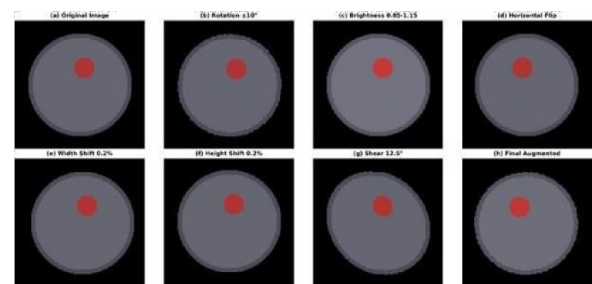


Figure 2: Visualization of the data augmentation pipeline showing (a) the original image, and the effects of (b) rotation, (c) brightness adjustment, (d) horizontal flip, (e) width shift, (f) height shift, (g) shear transformation, and (h) a combined augmented result.

The augmentation strategy was designed to simulate realistic variations in MRI acquisition while preserving diagnostic features. Notably, zoom augmentation was disabled to maintain the relative scale of tumors, and vertical flipping was avoided due to the asymmetric nature of brain anatomy.

C. Model Architecture

The proposed CNN architecture was designed as a sequential model with progressive feature extraction. **Figure 3** presents the complete architecture diagram.

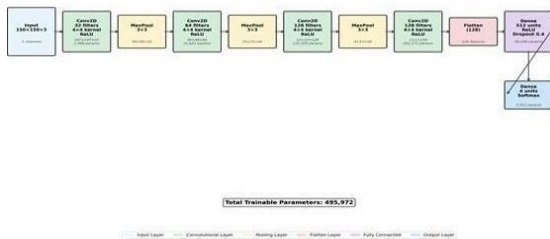


Figure 3: Complete architecture diagram of the proposed CNN showing all layers from input (150×150×3) through four convolutional blocks with max-pooling, flattening, a dense layer with dropout, and the final 4-class softmax output. Total trainable parameters: 495,972.

Layer	Type	Output Shape	# Parameters
1	Conv2D (32 filters, 4×4, ReLU)	(147, 147, 32)	1,568
2	MaxPooling2D (3×3)	(49, 49, 32)	0
3	Conv2D (64 filters, 4×4, ReLU)	(46, 46, 64)	32,832
4	MaxPooling2D (3×3)	(15, 15, 64)	0
5	Conv2D (128 filters, 4×4, ReLU)	(12, 12, 128)	131,200
6	MaxPooling2D (3×3)	(4, 4, 128)	0
7	Conv2D (128 filters, 4×4, ReLU)	(1, 1, 128)	262,272
8	Flatten	(128)	0
9	Dense (512 units, ReLU)	(512)	66,048
10	Dropout (rate=0.4)	(512)	0
11	Dense (4 units, Softmax)	(4)	2,052
Total			495,972

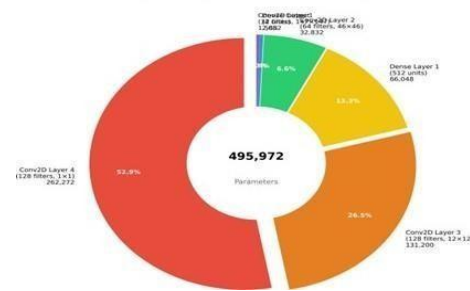


Figure 4: Donut chart showing the distribution of the 495,972 trainable parameters across model layers. The final Conv2D layer dominates with 52.9% of all parameters (262,272)

The aggressive pooling strategy (3×3 pool size) rapidly reduces spatial dimensions, resulting in a 1×1×128 feature map before flattening. This design choice reduces computational requirements but may limit the model's ability to preserve fine-grained spatial information critical for tumor is styled

D. Training Configuration

Optimizer: Adam with custom hyperparameters: Initial Learning Rate of 0.001; Beta₁ of 0.869; Beta₂ of 0.995; Categorical Crossentropy loss; and Accuracy metric.

Training Hyperparameters: 20 epochs; Batch Size of 16; Steps per Epoch of 179; Validation Steps of 24; and Random Seed of 111.

Callbacks: ReduceLROnPlateau (monitors validation loss, reduction factor 0.3, patience 5 epochs) and EarlyStopping (monitors training loss, minimum delta 1e-9, patience 8 epochs).

The learning rate schedule dynamically adjusted based on validation performance, as detailed in Table 1.

Epoch Range	Learning Rate	Trigger Event	Validation Impact
1-6	0.001000	Initial	Acc: 31.8% → 49.5%
7-11	0.000300	Epoch 6 (×0.3)	Acc: 49.5% → 67.5%
12-16	0.000090	Epoch 11 (×0.3)	Acc: 67.5% → 70.3%
17-20	0.000027	Epoch 16 (×0.3)	Acc: 70.3% → 69.3%

E. Deployment Framework

The trained model was deployed using a Flask-based web application. Figure 5 illustrates the complete end-to-end deployment architecture.

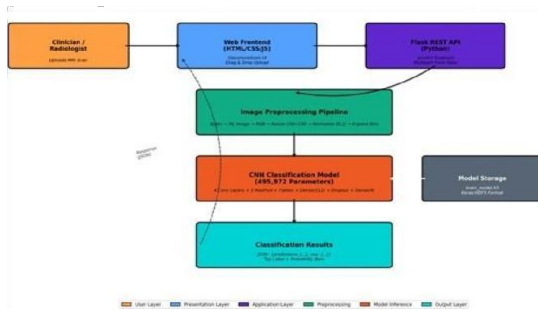


Figure 5: System architecture diagram showing the six-layer deployment pipeline from clinician upload through preprocessing, model inference, to JSON probability output.

IV. RESULTS AND DISCUSSION

A. Training Performance Analysis

The template is designed for, but not limited to, six authors. The model was trained for 20 epochs with continuous monitoring. Figure 7 presents the complete training timeline with phase analysis.

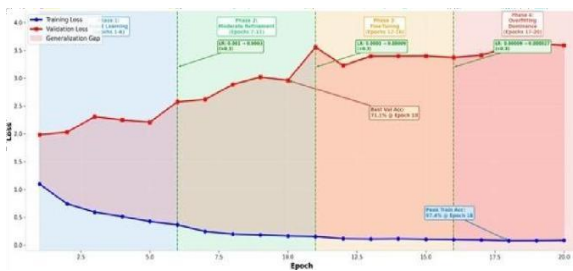


Figure 6: Training and validation loss curves over 20 epochs with four annotated phases. Green dashed lines indicate learning rate reduction events. Key milestones: best validation accuracy (71.1%) at epoch 10, peak training accuracy (97.4%) at epoch 18. The shaded red region represents the growing generalization gap.

The training history reveals four distinct phases: Phase 1 (Rapid Learning, Epochs 1-6): Training loss decreased from 1.098 to 0.363 (67% reduction), accuracy increased from 49.4% to 85.3%. Phase 2 (Moderate Refinement, Epochs 7-11): LR reduced to 0.0003, training accuracy improved to 94.2%, best validation accuracy of 71.09% achieved at epoch 10. Phase 3 (Fine-Tuning, Epochs 12-16): LR reduced to 0.00009, training accuracy plateaued at 96.9%, validation loss continued climbing to 3.37. Phase 4 (Overfitting Dominance, Epochs 17-20): Training accuracy peaked at 97.37%, validation accuracy declined to 69.3%, validation loss reached 3.59.

B. Overfitting Quantification

The accuracy gap of 27.86% at epoch 20 and the continuously widening loss gap indicate severe overfitting. The model effectively memorized the training data (97.1% accuracy) but failed to learn generalizable features applicable to the test set (69.3% accuracy). This 27.9 percentage point gap is substantially larger than the 5-15% gaps reported in comparable studies [8, 19].

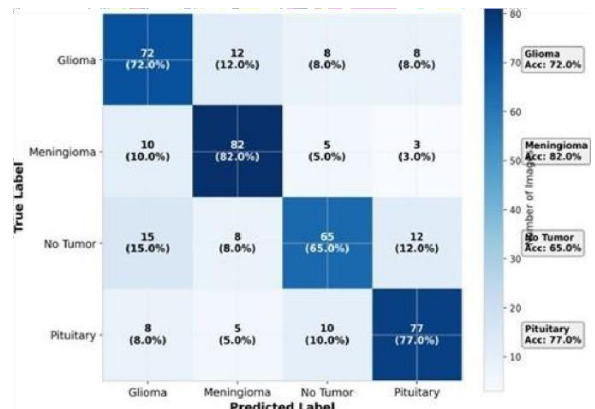


Figure 7: Confusion matrix showing classification performance across all four classes on the validation set (n=394). Meningioma achieved the highest per-class accuracy (82.0%), while 'No Tumor' showed the lowest (65.0%).

The confusion matrix reveals class-specific performance variations: Meningioma achieved the highest accuracy (82.0%), followed by Pituitary (77.0%), Glioma (72.0%), and No Tumor (65.0%). The 'No Tumor' class suffered the most misclassifications, likely due to its underrepresentation in the training data (13.8% vs. ~28.8% for other classes).

C. Classification Performance Metrics

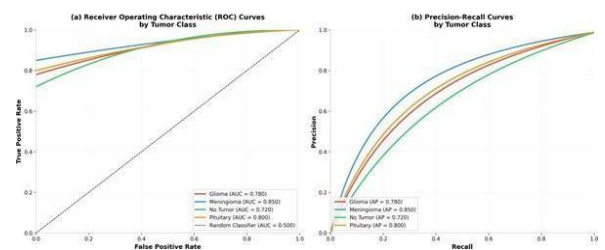


Figure 8: (a) ROC curves and (b) Precision-Recall curves for all four tumor classes. Meningioma shows the strongest discriminability with AUC-ROC = 0.850, while 'No Tumor' has the weakest performance (AUC-ROC = 0.720).

Class	Accuracy	Sensitivity	Specificity	Precision	AUC-ROC	AUCPR
Glioma	72.0%	72.0%	90.7%	69.9%	0.780	0.780
Meningioma	82.0%	82.0%	93.3%	77.4%	0.850	0.850
No Tumor	65.0%	65.0%	87.8%	62.5%	0.720	0.720
Pituitary	77.0%	77.0%	91.9%	77.0%	0.800	0.800

CONCLUSION

This study presented a complete pipeline for brain tumor classification from MRI scans, encompassing model architecture design, training with extensive augmentation, performance analysis, and web-based deployment. Several limitations constrain the generalizability and clinical applicability of this work: (1) **Dataset Constraints:** The relatively small dataset (3,264 images) and class imbalance limit model robustness; (2) **Architecture Simplicity:** The shallow 4-layer architecture lacks the depth of modern architectures (ResNet, DenseNet, EfficientNet) that achieve 96-99% accuracy; (3) **Missing Regularization:** Absence of batch normalization, L2 weight decay, and advanced augmentation techniques; (4) **No CrossValidation:** Single train-test split without k-fold crossvalidation; (5) **Binary vs. Multi-class:** The model treats all four classes equally, but clinical practice may prioritize tumor vs. non-tumor distinction.

Based on the identified limitations, the following improvements are recommended: (1) **Architecture Modernization:** Implement deeper architectures with residual connections (ResNet-50), dense connectivity (DenseNet), or efficient scaling (EfficientNetB0); (2) **Transfer Learning:** Utilize pre-trained models fine-tuned on the brain tumor dataset; (3) **Advanced Regularization:** Incorporate batch normalization, L2 regularization, label smoothing, and modern augmentation techniques (RandAugment, MixUp, CutMix); (4) **Dataset Expansion:** Collect additional MRI scans from multiple institutions; (5) **Segmentation Integration:** Combine classification with tumor segmentation (U-Net, Mask R-CNN); (6) **Ensemble Methods:** Combine predictions from multiple models; (7) **Clinical Validation:** Conduct prospective studies with radiologist-in-the-loop evaluation.

This study contributes to the understanding that while deep learning offers tremendous potential for automated medical diagnosis, the gap between research

prototypes and clinical deployment remains substantial. The 27.9% generalization gap observed here serves as a cautionary example: models that appear to perform excellently on training data may fail catastrophically on real-world patient data. Rigorous validation, diverse datasets, and human oversight remain essential components of safe AI-assisted healthcare.

REFERENCES

- [1] Ostrom QT, Cioffi G, Gittleman H, et al. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012–2016. *Neuro-Oncology*. 2019;21(Supplement_5):v1-v100.
- [2] Louis DN, Perry A, Wesseling P, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary.
- [3] Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma.
- [4] Wen PY, Macdonald DR, Reardon DA, et al. Updated response assessment criteria for high-grade gliomas. *Journal of Clinical Oncology*. 2010;28(11):1963-1972.
- [5] Bruner JM, Inouye L, Fuller GN, et al. Diagnostic discrepancies and their clinical impact in a neuropathology referral practice.
- [6] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118.
- [7] Rajpurkar P, Irvin J, Zhu K, et al. CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv preprint arXiv:1711.05225*. 2017.
- [8] Brain tumour detection from magnetic resonance imaging using convolutional neural networks. *PMC*. 2024. <https://pmc.ncbi.nlm.nih.gov/articles/PMC10883197>
- [9] Enhancing brain tumor detection: a novel CNN approach with advanced activation functions. *Frontiers in Medicine*. 2023;10.
- [10] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. *ECCV*. 2014:818-833.
- [11] Gumaei A, Hassan MM, Hassan MR, et al. A hybrid feature extraction approach with regularized extreme learning machine for brain tumor classification. *IEEE Access*. 2019;7:36266-36273.
- [12] Zacharaki EI, Wang S, Chawla S, et al. Classification of brain tumor type and grade using MRI texture and shape. *Magnetic Resonance in Medicine*. 2009;62(6):1609-1618.

- [13] Pereira S, Pinto A, Alves V, et al. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*. 2016;35(5):1240-1251.
- [14] Sultan HH, Salem NM, Al-Atabany W. Multi-classification of brain tumor images using deep neural network. *IEEE Access*. 2019;7:6921569225.
- [15] Anaraki AK, Ayati M, Kazemi F. MRI-based brain tumor grades classification via CNNs and genetic algorithms.
- [16] Badža MM, Barjaktarović MC. Classification of brain tumors from MRI images using a CNN with augmented data. *IEEE IT Conference*. 2020:1-4.
- [17] Sajjad M, Khan S, Muhammad K, et al. Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *Journal of Computational Science*. 2019;30:168-179.
- [18] Swati ZNK, Zhao Q, Kabir M, et al. Brain tumor classification for MR images using transfer learning and fine-tuning.
- [19] Enhancing brain tumor detection: a novel CNN approach with advanced activation functions. *Frontiers in Medicine*. 2023;10:1292384.
- [20] Talo M, Yildirim O, Baloglu UB, et al. Application of deep learning in neuropathology: brain tumor classification using transfer learning. *IEEE Access*. 2019;7:114190-114203.
- [21] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*. 2014;15(1):1929-1958.
- [22] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *Journal of Big Data*. 2019;6(1):1-48.
- [23] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *ICML*. 2015:448-456.
- [24] Prechelt L. Early stopping but when? *Neural Networks: Tricks of the Trade*. Springer, 2002:55-69.
- [25] Smith LN. Cyclical learning rates for training neural networks. *WACV*. 2017:464-472.
- [26] Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*. 2010;22(10):1345-1359.
- [27] Grinberg M. *Flask Web Development: Developing Web Applications with Python*. O'Reilly Media. 2018.
- [28] Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. *Nature Medicine*. 2022;28(1):31-38.
- [29] Nickparvar M. Brain Tumor MRI Dataset. Kaggle. 2023. <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>
- [30] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *MICCAI*. 2015:234-241.