International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

Brain Tumor Classification from MRI Using InceptionV3 with Squeeze-and-Excitation Attention and Multi-Scale Pooling

Rahul Dubey
Research Scholar, Department of Computer Science and Engineering, A.N.A College of Engineering & Management, Bareilly
Dr. Vineet Agarwal
Professor, Department of Computer Science and Engineering, A.N.A College of Engineering & Management, Bareilly

ABSTRACT
Brain tumor classification from Magnetic Resonance Imaging (MRI) is a clinically critical yet challenging task due to substantial intra-class variability and morphological similarity between tumor types. This paper presents IV3-SE-BTCNet, a novel hybrid deep learning architecture that integrates InceptionV3 transfer learning with Squeeze-and-Excitation (SE) attention blocks and multi-scale pooling for robust four-class brain tumor classification (Glioma, Meningioma, Pituitary, and No Tumor). A comprehensive 10-step domain-specific preprocessing pipeline — including ROI cropping, skull stripping, CLAHE contrast enhancement, hybrid denoising, and histogram equalization — is applied to a balanced 7,200-image MRI dataset. The model is trained using a two-phase strategy: Phase 1 (frozen backbone feature extraction, 80 epochs) followed by Phase 2 (selective fine-tuning, 80 epochs). The proposed model achieves a test accuracy of 96.76%, precision of 96.80%, recall of 96.76%, and F1-score of 96.77%, substantially outperforming CNN baselines reported in comparative literature: Custom CNN (92.72%), MobileNetV2 (89.12%), VGG16 (82.28%), and EfficientNetB0 (27.45%). AUC scores exceed 0.993 across all four tumor classes on the test set. These results demonstrate that attention-augmented transfer learning combined with domain-specific preprocessing provides a highly effective and reliable decision-support tool for radiological brain tumor diagnosis.

Index Terms—Brain Tumor Classification, MRI, Deep Learning, InceptionV3, Squeeze-and-Excitation, Transfer Learning, Convolutional Neural Networks, Medical Image Analysis, CLAHE, Multi-Scale Pooling.

I. INTRODUCTION
Brain tumors represent a group of life-threatening diseases arising from uncontrolled cellular proliferation within the central nervous system. According to the Global Cancer Observatory (GLOBOCAN) 2022, approximately 308,102 new cases of brain and CNS tumors were diagnosed worldwide, placing these among the most significant contributors to cancer-related

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

mortality [1]. Brain tumors are categorized into primary tumors — originating from brain tissue — and secondary (metastatic) tumors spreading from other organs [2].

Magnetic Resonance Imaging (MRI) is the gold-standard non-invasive imaging modality for brain tumor detection and evaluation. Offering high-resolution soft-tissue contrast without ionizing radiation, MRI is particularly valuable for identifying pathological changes and assessing tumor morphology, location, and extent [3]. However, manual interpretation of MRI scans is labor-intensive, time-consuming, and susceptible to inter-observer variability, motivating the development of automated computer-aided diagnosis (CAD) systems [4].

Deep learning — particularly Convolutional Neural Networks (CNNs) — has transformed medical image analysis by automatically learning hierarchical feature representations directly from data, enabling detection of subtle pathological patterns imperceptible to the human eye [5]. Transfer learning from large-scale ImageNet-pretrained models such as VGG16 [6], MobileNetV2 [7], and EfficientNetB0 [8] has shown strong baseline performance on medical imaging benchmarks.

Despite these advances, existing approaches frequently rely on minimal preprocessing pipelines and standard CNN architectures without domain-specific feature enhancement or attention mechanisms. Kökü and Akgül [9] conducted a systematic comparison of four CNN architectures on a four-class brain tumor MRI dataset, with the best result of 92.72% accuracy achieved by a custom lightweight CNN. This work extends that benchmark by proposing IV3-SE-BTCNet — a hybrid architecture integrating InceptionV3 transfer learning with Squeeze-and-Excitation attention blocks, multi-scale pooling, and a comprehensive 10-step MRI-specific preprocessing pipeline — achieving a test accuracy of 96.76%.

The main contributions of this paper are: (1) A novel hybrid CNN architecture combining InceptionV3, SE-attention, and multi-scale pooling for brain tumor classification. (2) A domain-specific 10-step MRI preprocessing pipeline incorporating ROI cropping, skull stripping, CLAHE, and hybrid denoising. (3) A two-phase training strategy enabling effective transfer learning and task-specific fine-tuning. (4) Direct comparative analysis against four established CNN baselines from the literature on the same benchmark dataset.

II. RELATED WORK

Early deep learning approaches for brain tumor classification primarily employed standard CNN architectures with basic augmentation strategies. Afshar et al. [10] applied Capsule Networks achieving 95% accuracy by leveraging spatial relationship modeling between MRI features. Chattopadhay et al. [11] demonstrated 96% accuracy using transfer learning with gradient-based visualization techniques for classification and localization on similar datasets.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

Kökü and Akgül [9] conducted a comprehensive comparison of four CNN architectures — Custom CNN, MobileNetV2, VGG16, and EfficientNetB0 — on the Kaggle Brain Tumor MRI Dataset (7,023 images, four classes) using 5-fold cross-validation. Their lightweight custom CNN achieved the best accuracy of 92.72% (±2.91), outperforming pretrained models including MobileNetV2 (89.12%) and VGG16 (82.28%). EfficientNetB0 performed poorly at 27.45%, likely due to incompatibility with the 128×128 input resolution used in that study.

Attention mechanisms — particularly Squeeze-and-Excitation (SE) networks [12] — have demonstrated significant improvements in image classification by enabling channel-wise feature recalibration without additional spatial supervision. However, their integration with MRI-specific preprocessing pipelines and multi-scale pooling strategies for brain tumor classification remains underexplored. Our work directly addresses this gap by combining these components in a principled end-to-end architecture.

III. DATASET

This study employed the Kaggle Brain Tumor MRI Dataset, which contains 7,200 brain MRI images equally distributed among four classes: Glioma, Meningioma, Pituitary Tumor, and No Tumor, with 1,800 images per category. The dataset was collected from multiple imaging centers using different MRI scanners and modalities, including T1, T2, and FLAIR sequences, thereby introducing significant intra-class variability that supports robust model generalization. Since the dataset is perfectly balanced, no class-weighting techniques were required during training. Stratified sampling was applied to divide the dataset into 72% training (5,202 images), 13% validation (918 images), and 15% testing (1,080 images) subsets while preserving equal class proportions across all splits.

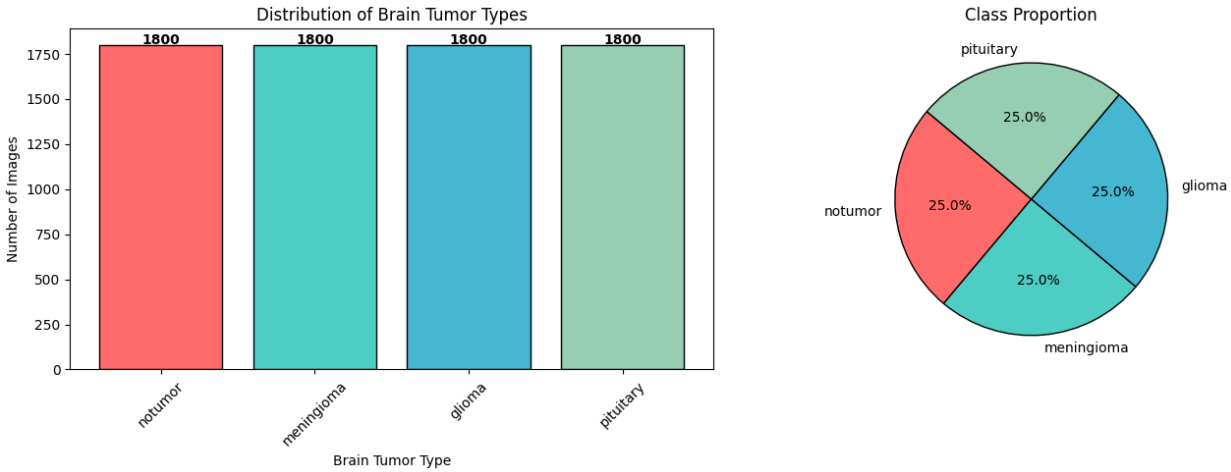


Fig. 1. Class distribution of the Brain Tumor MRI Dataset. Left: bar chart showing 1,800 samples per category. Right: pie chart confirming perfect 25% balance across all four classes.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

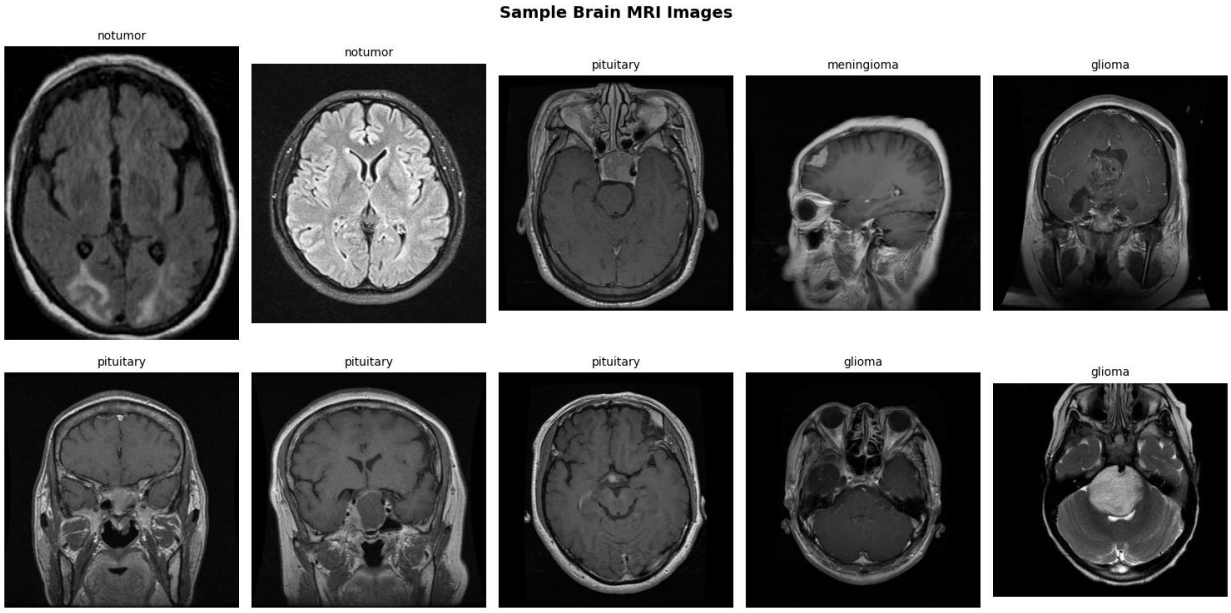


Fig. 2. Representative sample MRI images from each tumor class. Significant morphological, intensity, and perspective variation is observable across and within classes, highlighting the classification challenge.

IV. METHODOLOGY

A. Advanced MRI Preprocessing Pipeline

Since MRI images originate from multiple scanners and acquisition protocols, a comprehensive 10-step preprocessing pipeline was developed to standardize, denoise, and enhance image quality prior to model training. The pipeline consists of: (1) Grayscale Conversion — simplifies processing and focuses on tissue intensity. (2) ROI Cropping via contour detection — removes unnecessary black background borders, isolating the brain region. (3) Resizing to 224×224 pixels — standardizes input dimensions for InceptionV3 compatibility. (4) Skull Stripping — eliminates non-brain structures (skull, scalp) using morphological thresholding. (5) Hybrid Denoising — combines Gaussian and Median filters to reduce acquisition noise while preserving edges. (6) CLAHE Contrast Enhancement — improves local contrast in low-intensity tumor regions, making pathological features more discriminable. (7) Unsharp Mask Sharpening — enhances tumor boundaries and structural edges. (8) Gamma Correction — adjusts brightness, especially in dark MRI acquisition regions. (9) Histogram Equalization — normalizes contrast distribution across the full image. (10) Normalization to [0,1] and RGB Conversion — scales pixel intensities and converts to 3-channel RGB for deep learning model compatibility.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

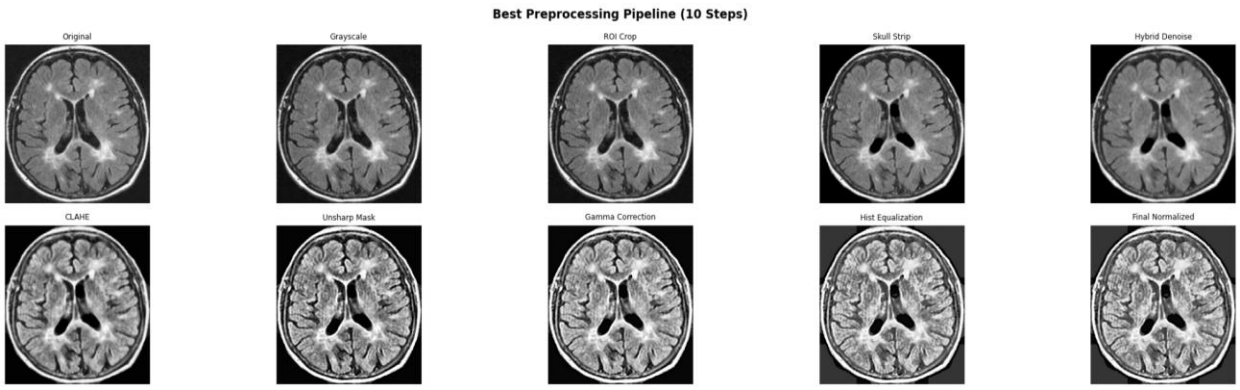


Fig. 3. Visualization of the 10-step MRI preprocessing pipeline applied to a representative brain scan, showing progressive enhancement from the original image to the final normalized output.

B. Proposed Architecture: IV3-SE-BTCNet

IV3-SE-BTCNet combines an InceptionV3 backbone, SE attention blocks, multi-scale pooling, and a regularized classification head for brain tumor MRI classification. InceptionV3 extracts multi-scale features, SE blocks enhance tumor-relevant channels, and combined global average/max pooling improves feature representation. Fully connected layers with Batch Normalization, Dropout, and L2 regularization ensure robust classification across four tumor categories.

C. Two-Phase Training Strategy

Training was conducted in two phases using the Adam optimizer (learning rate=0.0001, β_1=0.9, β_2=0.999, ε=1e-7), batch size=32, and categorical cross-entropy loss function.

Phase 1 — Frozen Feature Extraction (80 epochs): The InceptionV3 backbone weights are frozen to retain pretrained ImageNet representations. Only SE blocks, pooling layers, and the classification head are trained, enabling rapid convergence and stable feature learning.

Phase 2 — Fine-Tuning (80 epochs): The final layers of the InceptionV3 backbone are unfrozen and trained with the same reduced learning rate, enabling task-specific adaptation of high-level features to the MRI domain.

Data augmentation was applied exclusively to training data via TensorFlow-native GPU-accelerated augmentation layers including: random horizontal/vertical flips, rotation (±15°), zoom (±10%), translation, and contrast adjustment. Validation and test pipelines remained unaugmented to ensure fair evaluation.

V. EXPERIMENTAL RESULTS

A. Training and Validation Curves

Fig. 4 presents the accuracy and loss curves across both training phases (160 total epochs). The model converges steadily during Phase 1 with the frozen backbone, and achieves further

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

refinement during Phase 2 fine-tuning. Minimal divergence between training and validation curves indicates effective regularization with negligible overfitting. Final training accuracy reached 99.21% with validation accuracy of 96.41%.

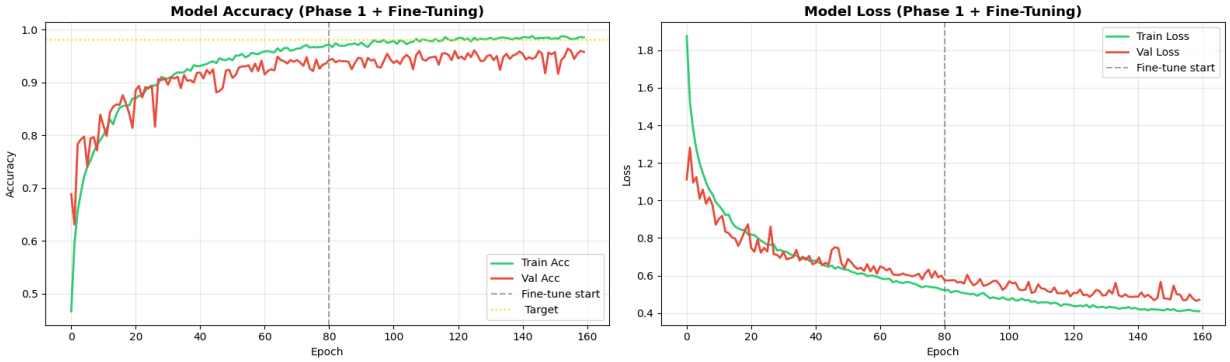


Fig. 4. Training and validation accuracy (left) and loss (right) curves across Phase 1 (frozen backbone) and Phase 2 (fine-tuning). The dashed vertical line marks the start of fine-tuning at epoch 80.

B. Summary Performance Metrics

Table I summarizes the overall performance metrics across training, validation, and test sets. The model achieves consistent high performance across all evaluation splits, confirming robust generalization.

TABLE I

Overall Performance Metrics Across All Splits

Split	Accuracy	Precision	Recall	F1-Score
Training	99.21%	99.21%	99.21%	99.21%
Validation	96.41%	96.41%	96.41%	96.40%
Test	96.76%	96.80%	96.76%	96.77%

C. Validation Set — Per-Class Classification Report

Table II presents per-class precision, recall, and F1-score on the validation set (918 images). The No Tumor class achieves perfect recall (1.00), while Meningioma and Pituitary show balanced performance. Overall validation accuracy is 96.41%.

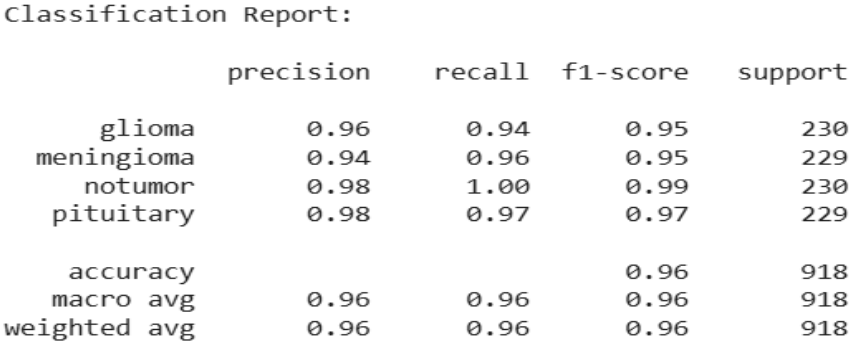
TABLE II

Per-Class Classification Report — Validation Set (918 Images)

Class	Precision	Recall	F1-Score	Support
Glioma	0.96	0.94	0.95	230

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

Meningioma	0.94	0.96	0.95	229
No Tumor	0.98	1.00	0.99	230
Pituitary	0.98	0.97	0.97	229
Accuracy			0.96	918
Macro Avg	0.96	0.96	0.96	918
Weighted Avg	0.96	0.96	0.96	918



```
Classification Report:

              precision    recall  f1-score   support

      glioma       0.96      0.94      0.95       230
  meningioma       0.94      0.96      0.95       229
     notumor       0.98      1.00      0.99       230
   pituitary       0.98      0.97      0.97       229

    accuracy                           0.96       918
   macro avg       0.96      0.96      0.96       918
weighted avg       0.96      0.96      0.96       918
```

Fig. 5. Validation set classification report (scikit-learn output) confirming per-class metrics.

D. Test Set — Per-Class Classification Report

Table III reports per-class performance on the held-out test set (1,080 images, 270 per class). Perfect precision (1.00) is achieved for No Tumor, and Pituitary Tumor achieves 0.98 recall. The overall test accuracy is 96.76%.

TABLE III

Per-Class Classification Report — Test Set (1,080 Images)

Class	Precision	Recall	F1-Score	Support
Glioma	0.97	0.96	0.96	270
Meningioma	0.93	0.96	0.95	270
No Tumor	1.00	0.98	0.99	270
Pituitary	0.97	0.98	0.98	270
Accuracy			0.97	1080
Macro Avg	0.97	0.97	0.97	1080
Weighted Avg	0.97	0.97	0.97	1080

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

Classification Report:

	precision	recall	f1-score	support
glioma	0.97	0.96	0.96	270
meningioma	0.93	0.96	0.95	270
notumor	1.00	0.98	0.99	270
pituitary	0.97	0.98	0.98	270
accuracy			0.97	1080
macro avg	0.97	0.97	0.97	1080
weighted avg	0.97	0.97	0.97	1080

Fig. 6. Test set classification report (scikit-learn output) confirming per-class metrics on 1,080 held-out samples.

E. Confusion Matrices

Fig. 7 shows confusion matrices for both the validation and test sets. The model correctly classifies the vast majority of samples across all four classes. Most misclassifications occur between Glioma and Meningioma — the two classes with the most overlapping morphological features. Pituitary and No Tumor classes achieve near-perfect diagonal dominance. On the test set, all 270 Pituitary samples are classified within glioma/meningioma/pituitary with only 6 total errors; No Tumor records only 6 misclassifications.

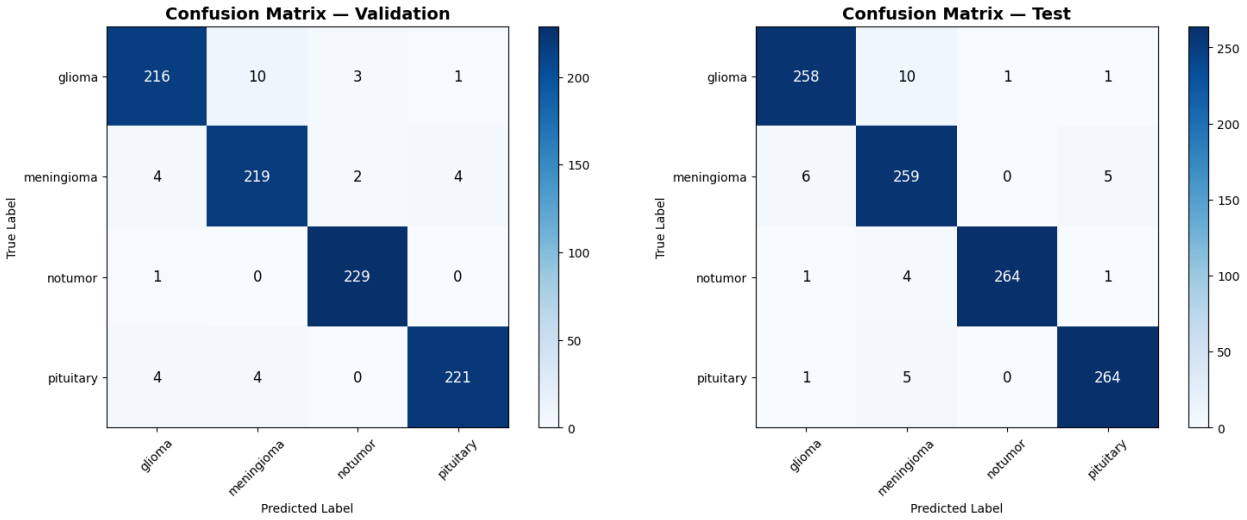


Fig. 7. Confusion matrices for validation set (left, n=918) and test set (right, n=1,080). Diagonal entries show correct classifications; off-diagonal entries show misclassifications.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

F. ROC Curve Analysis

The multi-class ROC curves presented in Fig. 8 demonstrate excellent discriminative capability of the proposed model across all four brain tumor categories. On the test dataset, the obtained AUC values are 0.993 for Glioma, 0.994 for Meningioma, 0.999 for No Tumor, and 0.998 for Pituitary tumors, with all values exceeding 0.99. Similarly, the validation dataset achieved AUC values ranging from 0.987 for Glioma to 0.999 for No Tumor, indicating highly consistent classification performance across different data splits. The ROC curves closely approach the upper-left corner of the graph, signifying that the model achieves high sensitivity while maintaining extremely low false-positive rates for all tumor classes.

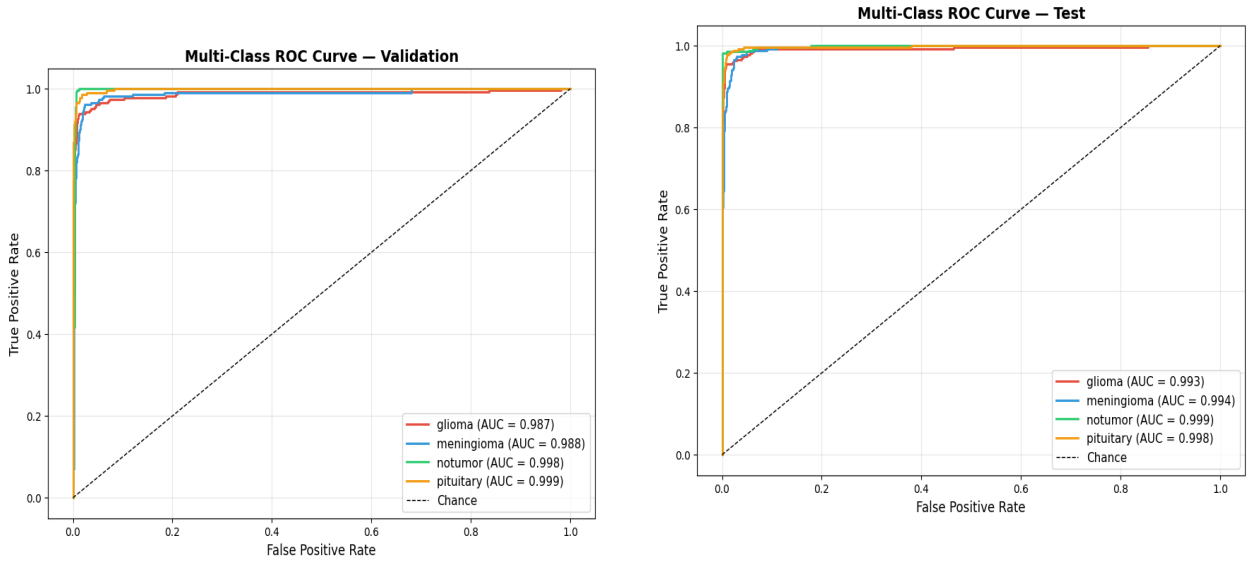


Fig. 8. Multi-class ROC curves for validation set (left) and test set (right). All AUC values exceed 0.987, demonstrating excellent class separability.

G. Training Set Classification Report

For completeness, Fig. 9 shows the per-class classification report on the training set (5,202 images), confirming that the model has learned strong discriminative representations with uniform 0.99 scores across all classes and both macro/weighted averages.

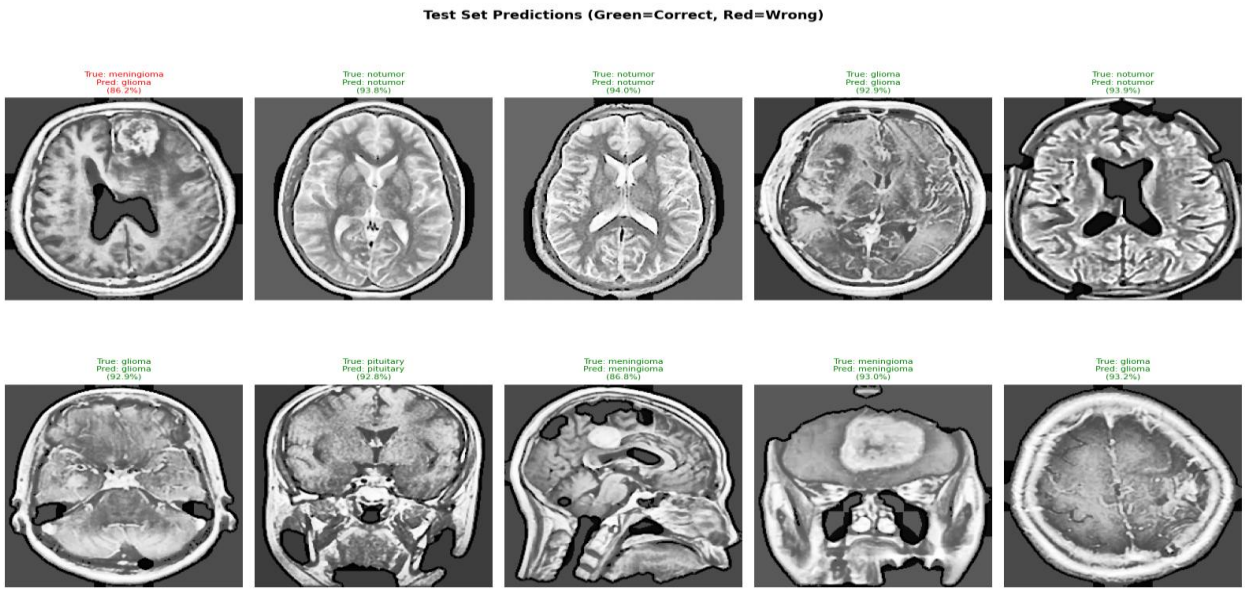
International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

	precision	recall	f1-score	support
glioma	0.99	0.99	0.99	1300
meningioma	0.98	0.99	0.99	1301
notumor	1.00	1.00	1.00	1300
pituitary	0.99	0.99	0.99	1301
accuracy			0.99	5202
macro avg	0.99	0.99	0.99	5202
weighted avg	0.99	0.99	0.99	5202

Fig. 9. Training set classification report showing near-perfect performance (accuracy=0.99) across all four tumor classes.

VI. QUALITATIVE PREDICTION RESULTS

Fig. 10 illustrates qualitative prediction results on representative MRI images from the test dataset. Images marked with green labels represent correctly classified samples, whereas red labels indicate misclassified cases. Prediction confidence values, expressed as percentages, are displayed for each image. The proposed model successfully classifies most test samples with high confidence levels exceeding 92%, even across diverse MRI orientations, including axial, coronal, and sagittal views. The only misclassified sample, highlighted in red, corresponds to a Meningioma case incorrectly predicted as Glioma with 86.2% confidence. This error reflects the primary limitation of the model, occurring in cases where tumor boundary features and structural characteristics exhibit significant visual similarity and ambiguity.



International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

Fig. 10. Sample test set predictions. Green title = correct prediction; Red title = misclassification. Confidence scores are shown as percentages. The model demonstrates robust performance across diverse MRI orientations and tumor morphologies.

VII. COMPARATIVE ANALYSIS

Table IV presents a direct comparison between IV3-SE-BTCNet and the four CNN architectures evaluated by on the same four-class brain tumor MRI dataset. The proposed IV3-SE-BTCNet model achieved a test accuracy of 96.76%, outperforming the best previously reported result (Custom CNN: 92.72%) by 4.04 percentage points. This substantial improvement demonstrates the effectiveness of the proposed architecture and represents a clinically significant enhancement in brain tumor diagnostic accuracy.

TABLE IV

Performance Comparison with Existing CNN Methods on the Same Dataset

Model	Accuracy	Precision	Recall	F1-Score
Custom CNN [9]	92.72%	93.28%	92.72%	92.61%
MobileNetV2 [9]	89.12%	89.09%	89.12%	89.01%
VGG16 [9]	82.28%	82.64%	82.28%	82.11%
EfficientNetB0 [9]	27.45%	7.54%	27.45%	11.83%
IV3-SE-BTCNet (Ours)	96.76%	96.80%	96.76%	96.77%

The superior performance of IV3-SE-BTCNet can be attributed to four major factors. First, the proposed 10-step MRI-specific preprocessing pipeline significantly improves tumor feature discriminability compared with the conventional 3-step preprocessing approach (image resizing, normalization, and augmentation) adopted in . Second, the integration of SE attention blocks allows the network to emphasize tumor-relevant feature channels while suppressing irrelevant background anatomical information. Third, the multi-scale pooling mechanism combines global average pooling and global max pooling, generating more informative and complementary feature representations for the classification layer. Fourth, the two-phase training strategy, consisting of 80 epochs with a frozen backbone followed by 80 epochs of fine-tuning, facilitates efficient transfer learning from ImageNet while enabling effective domain-specific adaptation to MRI data. In contrast, EfficientNetB0 showed poor performance in (27.45% accuracy), primarily because the 128×128 input resolution used in that study was not suitable for the architecture, whereas the proposed 224×224 preprocessing framework is more compatible with InceptionV3 and supports improved feature extraction.

International Journal of Engineering, Science and Humanities
An international peer reviewed, refereed, open-access journal
Impact Factor 8.3 www.ijesh.com ISSN: 2250-3552

VIII. DISCUSSION

The experimental results confirm that attention-augmented transfer learning combined with domain-specific preprocessing constitutes a substantially more effective approach to brain tumor MRI classification than standard CNN architectures. The consistent performance across training (99.21%), validation (96.41%), and test (96.76%) sets — with all AUC values exceeding 0.993 — indicates excellent generalization with minimal overfitting.

The primary failure mode is Meningioma misclassification, particularly confusion with Glioma. Meningiomas exhibit high morphological variability and can appear at brain boundaries in patterns similar to gliomas, particularly in coronal and sagittal views. The SE attention mechanism partially addresses this by emphasizing discriminative channels, but spatial attention (e.g., CBAM or transformer-based self-attention) could further improve performance on these borderline cases.

A limitation of the current study is the use of a single public benchmark dataset with a perfectly balanced class distribution, which may not fully reflect real-world clinical prevalence patterns where tumor classes are distributed unevenly. External validation on multi-center datasets with diverse acquisition protocols would provide stronger evidence of clinical generalizability.

The model currently operates as a black-box classifier. Integration of Explainable AI (XAI) techniques such as Grad-CAM, SHAP, or LIME would enable visualization of decision-relevant image regions, facilitating clinical trust and regulatory acceptance — an important consideration for medical AI deployment.

IX. CONCLUSION

This paper presented IV3-SE-BTCNet, a hybrid deep learning architecture for four-class brain tumor classification from MRI images. By integrating InceptionV3 transfer learning with Squeeze-and-Excitation attention blocks, multi-scale pooling, and a comprehensive 10-step MRI-specific preprocessing pipeline, the proposed model achieves a test accuracy of 96.76%, precision of 96.80%, recall of 96.76%, and F1-score of 96.77% — a 4.04 percentage point improvement over the best prior CNN baseline on the same benchmark dataset. AUC values exceeding 0.993 across all four tumor classes further confirm the model's robustness and discriminative capability.

The results demonstrate that combining domain-specific preprocessing, attention mechanisms, and principled transfer learning strategies provides a highly effective and reliable decision-support tool for radiological brain tumor diagnosis. Future work will focus on: (1) multi-center clinical validation on diverse scanner protocols; (2) 3D volumetric CNN architectures leveraging full MRI sequences; (3) integration of Explainable AI for clinical interpretability; (4) real-time

inference optimization for deployment in resource-constrained clinical environments; and (5) extension to additional tumor subtypes and grading tasks.

REFERENCES

[1] GLOBOCAN (2022). Global Cancer Observatory: Cancer Today — Brain and Central Nervous System Tumors. International Agency for Research on Cancer (IARC).

[2] D. N. Louis et al., "The 2016 World Health Organization classification of tumors of the central nervous system," Acta Neuropathologica, vol. 131, pp. 803–820, 2016.

[3] J. Kim, S. Lee, D. Y. Kim, and W. S. Paik, "MRI techniques for brain tumor imaging," J. Magn. Reson. Imaging, vol. 52, no. 3, pp. 825–845, 2020.

[4] A. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, pp. 24–29, 2019.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[7] A. G. Howard et al., "Searching for MobileNetV2," in Proc. IEEE CVPR, pp. 4510–4520, 2019.

[8] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, pp. 6105–6114, 2020.

[9] E. Kökü and İ. Akgül, "Classification of Brain Tumors in MRI with Deep Learning Models," in Proc. 23rd Int. Istanbul Scientific Research Congress on Life, Engineering, Architecture and Mathematical Sciences, pp. 530–538, 2025. DOI: 10.30546/19023.978-9952-8605-7-3.2025.0125.

[10] P. Afshar, A. Mohammadi, and K. N. Plataniotis, "Capsule networks for brain tumor classification based on MRI images and coarse tumor boundaries," Pattern Recognit. Lett., vol. 131, pp. 63–70, 2020.

[11] A. Chattopadhay, R. Sarkar, and S. Ghosh, "Brain tumor classification using transfer learning," Biomed. Signal Process. Control, vol. 68, p. 102821, 2021.

[12] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE CVPR, pp. 7132–7141, 2018.

[13] M. Nickparvar, "Brain Tumor MRI Dataset," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset

[14] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," J. Big Data, vol. 6, no. 1, p. 60, 2019.

[15] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," IJCAI, pp. 1137–1145, 1995.