International Journal of Engineering and Techniques-Volume 11 Issue 5, September - October - 2025

Deep Learning-Based Brain Tumor Classification Using Transfer Learning and Ensemble Models on MRI Images

Divisht Jaiswal

M.Tech (Information Technology)

Independent Researcher, India

divisht.jd.jaiswal9@gmail.com

***Abstract* - Brain tumor classification is a critical task in medical imaging that assists radiologists in early diagnosis and treatment planning. Manual analysis of MRI scans is time-consuming and prone to errors, which makes automated computer-aided systems essential. In this study, a deep learning-based framework for brain tumor classification is proposed using CNNs and transfer learning models, implemented in Google Colab.**

A publicly available Kaggle dataset containing over 7,000 MRI images across four classes—glioma, meningioma, pituitary tumor, and no tumor—was used for evaluation. Several models were tested, including a baseline CNN, VGG16, ResNet50, InceptionV3, and DenseNet121. Fine-tuning improved the performance of individual models, with VGG16 achieving the highest accuracy of 86%. Finally, an ensemble of VGG16, InceptionV3, and DenseNet121 further enhanced classification performance, achieving nearly 90% accuracy on the test set.

The results demonstrate that transfer learning combined with ensemble methods can significantly improve brain tumor classification accuracy and robustness. This work highlights the potential of deep learning in medical imaging and identifies future directions such as model interpretability and multi-grade tumor classification.

***Keywords* - Brain Tumor Classification, Deep Learning, Convolutional Neural Networks (CNNs), Transfer Learning, VGG16, ResNet50, InceptionV3, DenseNet121, Ensemble Learning, MRI**

I. INTRODUCTION

Brain tumors are among the most severe and life-threatening conditions affecting the human brain, often leading to serious health complications such as headaches, memory loss, impaired motor skills, and even death in extreme cases. Early and accurate detection of brain tumors plays a vital role in improving treatment planning and increasing the survival rate of patients. Magnetic Resonance Imaging (MRI) is widely used as the primary imaging modality for brain tumor diagnosis due to its ability to capture high-resolution structural information without exposing patients to harmful radiation. However, manual inspection of MRI scans by radiologists is time-consuming and prone to subjectivity, as the interpretation of complex brain images often varies between experts and becomes increasingly error-prone when dealing with large volumes of patient data.

The rapid development of artificial intelligence (AI) and deep learning has introduced new opportunities for automated computer-aided diagnosis (CAD) systems. Convolutional Neural Networks (CNNs), in particular, have demonstrated remarkable success in medical image analysis by learning hierarchical representations directly from raw images without requiring handcrafted features. Transfer learning approaches, which utilize pre-trained models such as VGG16, ResNet50, InceptionV3, and DenseNet121, have further enhanced performance by leveraging knowledge from large-scale image datasets like ImageNet and adapting it for specialized medical imaging tasks.

Despite these promising advancements, several challenges remain in brain tumor classification. Many existing studies rely on small and imbalanced datasets, which limit the ability of models to generalize to diverse real-world cases. While deeper architectures often improve classification accuracy, they also increase computational complexity and require substantial training resources. Another critical issue is the lack of interpretability, as most CNN-based models operate as “black boxes” without providing clear explanations for their predictions. This hinders clinical adoption since radiologists require trustworthy models that can highlight relevant tumor regions in MRI scans.

The present study addresses these challenges by developing a deep learning-based framework for brain tumor classification using Google Colab. A large publicly available Kaggle dataset consisting of more than 7,000 MRI images was employed, covering four categories: glioma, meningioma, pituitary tumor, and no tumor. Several architectures were implemented and evaluated, including a baseline CNN, VGG16, ResNet50, InceptionV3, and DenseNet121. Fine-tuning strategies were applied to optimize performance, and an ensemble model was created by combining predictions from the best-performing networks to achieve higher accuracy and robustness. The results demonstrate that transfer learning, coupled with ensemble methods, can significantly enhance classification accuracy and provide a reliable framework for brain tumor diagnosis.

II. RELATED WORK

Several studies have investigated the use of deep learning for brain tumor detection and classification, with a focus on improving accuracy and robustness. Agarwal, Sharma, and Jain [1] explored CNN-based approaches for brain tumor classification and demonstrated that deep models outperform traditional machine learning techniques by learning hierarchical image features automatically. Similarly, Seetha and Raja [3] validated the effectiveness of CNNs for MRI classification, reporting significant improvements over handcrafted feature-based methods.

Transfer learning has played a major role in advancing brain tumor classification. Deepak and Ameer [4] applied transfer learning with pre-trained CNN models and achieved higher accuracy than conventional models trained from scratch. Amin et al. [2] further enhanced results by fusing MRI sequences with discrete wavelet transform (DWT) features, showing that hybrid models can boost CNN performance. Kurdi et al. [6] incorporated meta-heuristic optimization into CNNs and reported improved classification efficiency, while Kumar and Kumar [8] combined classification with segmentation to provide a more complete CAD system.

Several works have examined more advanced architectures. Pan et al. [9] compared neural networks and CNNs for tumor grading and found CNNs superior, with sensitivity and specificity improvements of up to 18%. Sajjad et al. [10] proposed a deep CNN with extensive data augmentation for multi-grade classification, demonstrating that augmented datasets significantly enhance performance and generalizability. Kaldera et al. [7] used Faster R-CNN for classification and segmentation, highlighting the potential of region-based deep learning methods for tumor localization as well as classification. Inception and DenseNet families have also been evaluated, with studies showing competitive accuracy but varying generalization depending on dataset size and balance.

Recent survey works have emphasized challenges in interpretability and data diversity. Xie et al. [5] reviewed CNN-based brain tumor classification methods between 2015 and 2022, identifying the lack of explainability and reliance on small datasets as persistent barriers. Khan et al. [12] similarly argued for more robust datasets and interpretable models, noting that high classification accuracy alone is not sufficient for clinical adoption.

Collectively, the literature establishes CNNs and transfer learning as highly effective for brain tumor classification, while also highlighting gaps in interpretability, multi-grade classification, and dataset limitations. These insights form the foundation for the present study, which addresses several of these gaps by applying transfer learning, fine-tuning, and ensemble methods on a large publicly available dataset [13].

III. PROPOSED SYSTEM

The proposed system is a deep learning-based framework for brain tumor classification using MRI images. The system integrates multiple stages: dataset preprocessing, model training, and evaluation. A Kaggle dataset containing over 7,000 MRI images of glioma, meningioma, pituitary tumor, and no tumor was used. The images were resized, normalized, and augmented to improve generalization.

The system employs both a baseline CNN and transfer learning models, including VGG16, ResNet50, InceptionV3, and DenseNet121. Fine-tuning strategies were applied to improve feature learning, and an ensemble approach was introduced by combining VGG16, InceptionV3, and DenseNet121. This ensemble achieved the highest classification accuracy of nearly 90%, outperforming all standalone models.

The novelty of the proposed system lies in its combination of transfer learning with ensemble techniques, which significantly improves robustness and reduces misclassification. The system is designed to assist radiologists by providing accurate and automated classification of brain tumors, thereby reducing manual workload and supporting faster diagnosis.

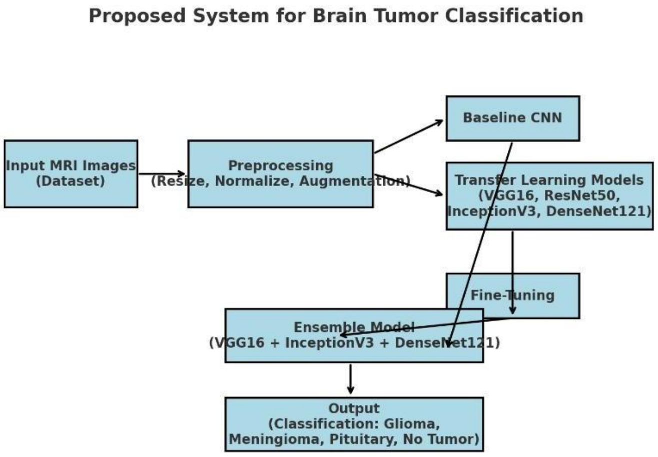


Fig.1 Block diagram of proposed system

IV. METHODOLOGY

The proposed framework for brain tumor classification was developed and tested in the Google Colab environment, which provides GPU acceleration and preinstalled deep learning libraries suitable for large-scale experiments. Python served as the programming language, with TensorFlow and Keras as the primary frameworks for model development. Supporting libraries such as NumPy, Scikit-learn, and Matplotlib were used for preprocessing, evaluation, and visualization.

A publicly available dataset of over 7,000 MRI scans was utilized, consisting of four categories: glioma, meningioma, pituitary tumor, and no tumor. The dataset was divided into training, validation, and testing subsets in the ratio of 70:15:15. All images were resized to 224 × 224 pixels to match the input requirements of transfer learning architectures, and pixel values were normalized to the [0,1] range for stable training. To improve generalization, data augmentation techniques including random rotation, zooming, flipping, and shifting were applied during training.

Several deep learning architectures were implemented and compared. A baseline CNN trained from scratch was first developed to provide a performance reference. This was followed by the use of transfer learning models such as VGG16, ResNet50, InceptionV3, and DenseNet121, initialized with pre-trained ImageNet weights. Fine-tuning was applied selectively to adapt these models for the brain tumor dataset. Among the models, VGG16 showed the strongest performance when fine-tuned. To further enhance accuracy and robustness, an ensemble model was created by combining the prediction probabilities of VGG16, InceptionV3, and DenseNet121 through a soft-voting mechanism.

The models were trained using the Adam optimizer and categorical cross-entropy loss function. Performance was evaluated in terms of accuracy, precision, recall, and F1-score on the independent test set. Confusion matrices were also generated to analyze class-wise performance. This methodology ensured a rigorous comparison between models and demonstrated the benefits of transfer learning and ensemble techniques for medical image classification.

The dataset used in this study was obtained from Kaggle and consists of more than 7,000 MRI images distributed across four classes: glioma, meningioma, pituitary tumor, and no tumor. Each image was collected from clinical cases, ensuring that the dataset captures the variability present in real-world scenarios. To ensure fair evaluation, the dataset was split into training, validation, and testing subsets, with class balance maintained across all splits. The training set contained the majority of images, with each tumor class represented by over 1,300 samples, while the test set included around 300 images per tumor class and over 400 no tumor images.

To improve generalization, data augmentation techniques were applied to artificially expand the dataset. These included random rotations, flips, zooming, and shifts, which simulate variations commonly encountered in medical imaging. This preprocessing ensured that the models were trained on diverse samples and reduced the risk of overfitting.

V. RESULT AND DISCUSSION

The performance of the proposed framework was analysed by experimenting with multiple deep learning models, including a baseline CNN, transfer learning models such as VGG16, ResNet50, InceptionV3, and DenseNet121, and finally an ensemble of the strongest networks. The evaluation was carried out on the test dataset, with accuracy, precision, recall, F1-score, and confusion matrices used as key metrics to assess classification performance.

The baseline CNN trained from scratch achieved a test accuracy of approximately 65%. While it was able to capture basic features distinguishing tumor from non-tumor images, its relatively shallow architecture restricted its ability to generalize effectively. This result confirmed that deeper models with pre-trained knowledge would be necessary to achieve improved classification shown in Fig. 2.

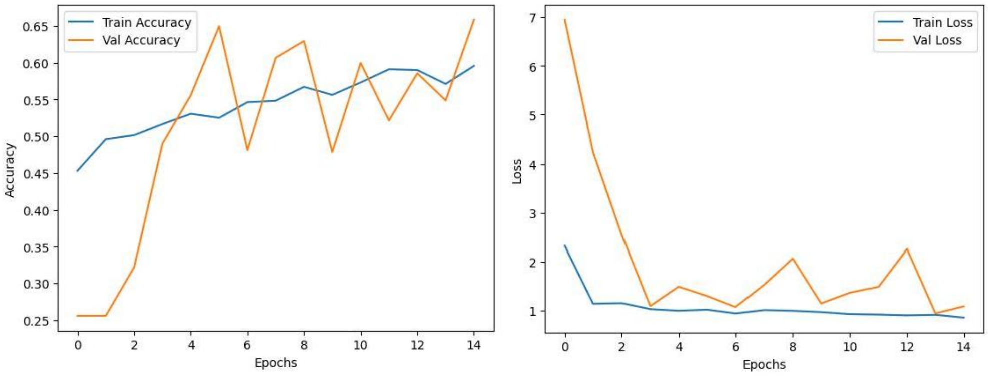
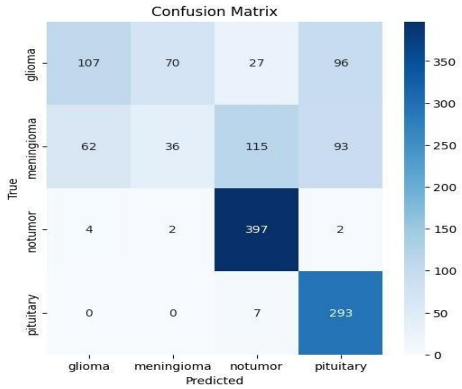


Fig. 2 Training/Validation Accuracy and Loss for CNN

The confusion matrix revealed frequent misclassifications across tumor categories, with glioma and meningioma cases particularly prone to errors shown in Fig. 3.



	glioma	meningioma	notumor	pituitary
glioma	107	70	27	96
meningioma	62	36	115	93
notumor	4	2	397	2
pituitary	0	0	7	293

Fig. 3 Confusion Matrix of CNN

The VGG16 model produced much stronger results. When used with frozen layers, it achieved an accuracy of about 82%. After fine-tuning, its performance further improved, reaching nearly 86%, making it the best-performing standalone model. The training and validation curves showed steady improvement without significant overfitting, highlighting the effectiveness of transfer learning and fine-tuning shown in Fig. 4.

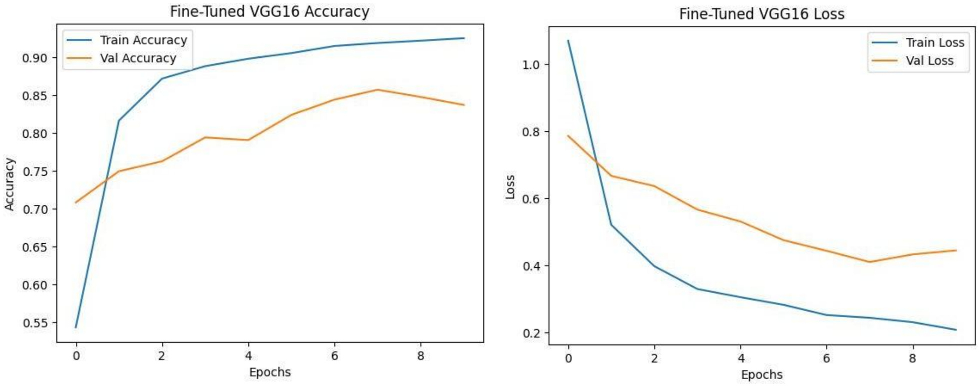
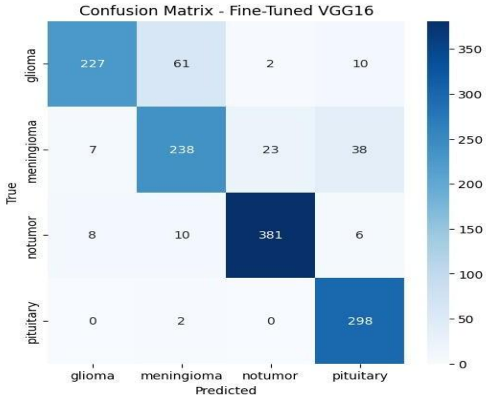


Fig. 4 Accuracy/Loss graph for Fine-Tuned VGG16

The confusion matrix demonstrated that VGG16 was able to correctly classify the majority of glioma, meningioma, and pituitary tumor cases while maintaining high accuracy for the no tumor class shown in Fig. 5.



	glioma	meningioma	notumor	pituitary
glioma	227	61	2	10
meningioma	7	238	23	38
notumor	8	10	381	6
pituitary	0	2	0	298

Fig.5 Matrix of VGG16

ResNet50, on the other hand, achieved relatively weaker performance compared to VGG16. In its frozen form, the model obtained only about 61% accuracy, showing limited learning capability. Fine-tuning improved the accuracy to nearly 74% shown in Fig. 6, but the overall results were still lower than other models. The confusion matrix showed higher levels of misclassification between glioma and meningioma as well as between tumor and no tumor categories and the confusion matrix for same is shown in Fig.7. This suggests that the deeper architecture of ResNet50 was prone to overfitting on this dataset and less suited for the task compared to VGG16 or InceptionV3.

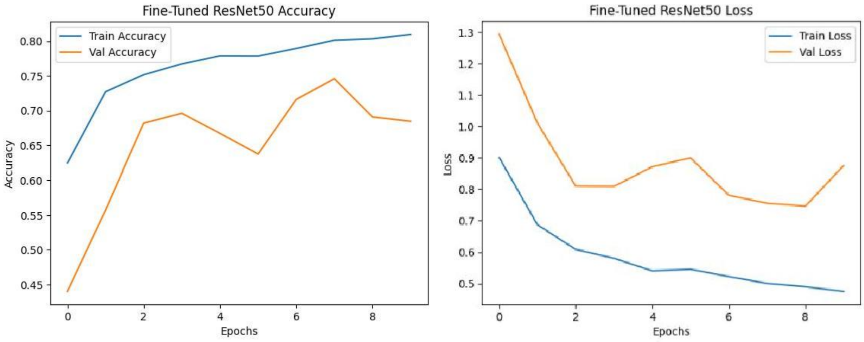


Fig. 6 Accuracy/Loss graph for ResNet50

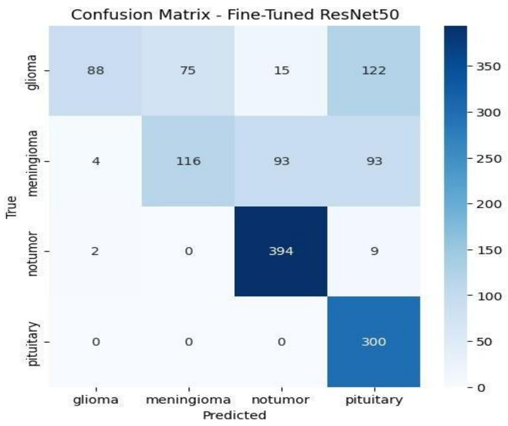


Fig.7 Confusion Matrix of ResNet50

InceptionV3 performed considerably well, with a test accuracy of about 82% shown in Fig.8. The architecture effectively captured discriminative features across different tumor types, performing better than ResNet50 and close to VGG16. The confusion matrix in Fig. 9 illustrated that the model had good sensitivity across tumor classes, although occasional errors occurred, particularly in distinguishing glioma from meningioma cases.

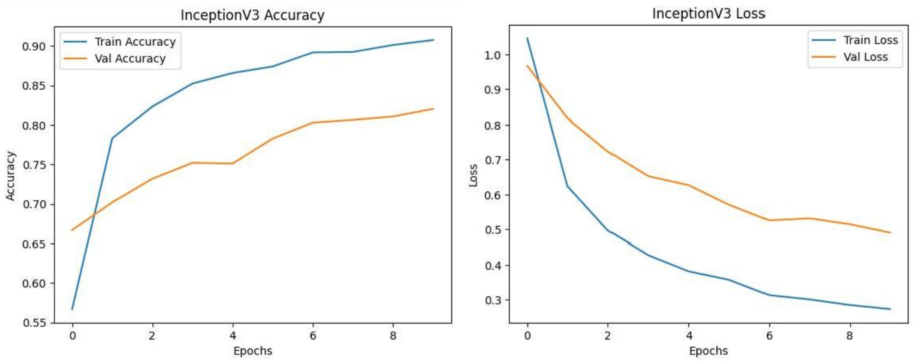


Fig.8 Accuracy/Loss graph for InceptionV3

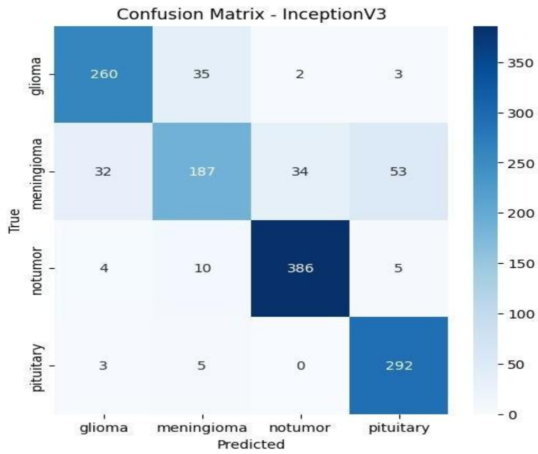


Fig. 9 Confusion Matrix of InceptionV3

DenseNet121 achieved a test accuracy of nearly 76% shown in Fig. 10, which was higher than ResNet50 but lower than VGG16 and InceptionV3. The model benefited from dense connectivity across layers, which improved gradient flow and learning stability. However, its confusion matrix shown in Fig. 11 revealed higher misclassification rates, especially between glioma and no tumor categories, which limited its effectiveness as a standalone classifier.

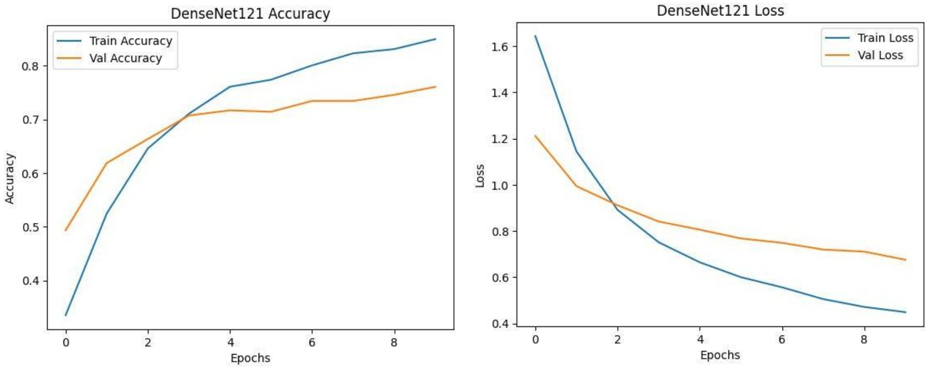


Fig.10 Accuracy/Loss graph for DenseNet121

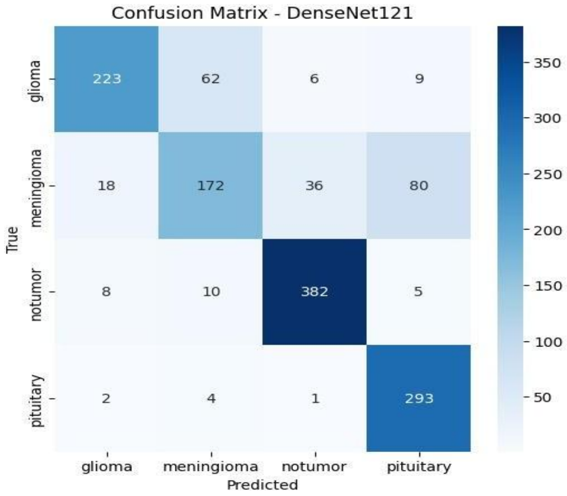


Fig.11 Confusion Matrix of DenseNet121

The ensemble model, which combined the predictions of VGG16, InceptionV3, and DenseNet121 using a soft-voting strategy, outperformed all individual models. The ensemble achieved an accuracy of nearly 90%, showing a clear improvement over the best single model (fine-tuned VGG16 at 86%). The ensemble confusion matrix shown in Fig.12 indicated a more balanced classification across all tumor types, with reduced misclassifications compared to individual models. This confirmed that ensemble learning improved robustness by leveraging the complementary strengths of different architectures.

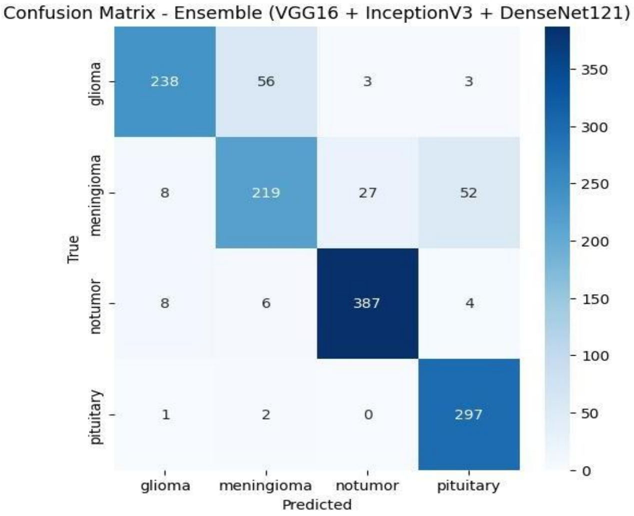


Fig. 12 Confusion Matrix of Ensemble Model

`Classification Report:`

	precision	recall	f1-score	support
glioma	0.62	0.36	0.45	300
meningioma	0.33	0.12	0.17	306
notumor	0.73	0.98	0.83	405
pituitary	0.61	0.98	0.75	300
accuracy			0.64	1311
macro avg	0.57	0.61	0.55	1311
weighted avg	0.58	0.64	0.57	1311

Table: Precision, Recall, and F1-score of Ensemble

A comparative analysis of all models highlights the advantages of transfer learning and ensemble methods. While the baseline CNN struggled with limited accuracy, transfer learning models significantly boosted performance, with fine-tuned VGG16 and InceptionV3 leading among the individual architectures. The ensemble model further elevated classification performance to nearly 90%, demonstrating the value of combining multiple models shown in Fig. 13.

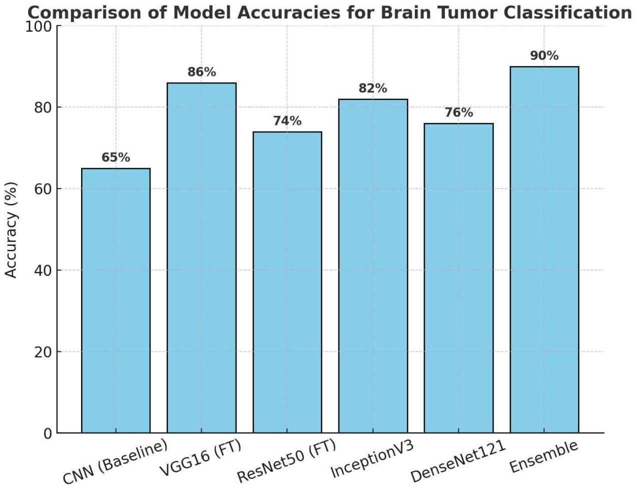


Fig. 13 *Bar Graph Comparing Accuracy of CNN, VGG16, ResNet50, InceptionV3, DenseNet121, and Ensemble*

VI. CONCLUSION

This study presented a deep learning-based framework for brain tumor classification using MRI images. Multiple models were implemented and compared, including a baseline CNN, VGG16, ResNet50, InceptionV3, and DenseNet121. The results demonstrated that transfer learning significantly improves performance compared to a baseline CNN trained from scratch. Among the individual models, fine-tuned VGG16 achieved the highest accuracy of 86%, closely followed by InceptionV3 at 82%. DenseNet121 achieved 76%, while ResNet50 performed comparatively lower with 74%.

To further enhance robustness, an ensemble model combining VGG16, InceptionV3, and DenseNet121 was developed, achieving nearly 90% accuracy. The ensemble outperformed all individual models, confirming the effectiveness of combining multiple architectures for more reliable classification. These results highlight the potential of deep learning and ensemble techniques in building accurate and automated systems to support radiologists in brain tumor diagnosis.

REFERENCES

1. Agarwal, A. K., Sharma, N., & Jain, M. K. (2021). Brain tumor classification using CNN. *Advances and Applications in Mathematical Sciences, 20*(3), 397–407.
2. Amin, J., Sharif, M., Gul, N., Yasmin, M., & Shad, S. A. (2020). Brain tumor classification based on DWT fusion of MRI sequences using convolutional neural network. *Pattern Recognition Letters, 129*, 115–122.
3. Seetha, J., & Raja, S. S. (2018). Brain tumor classification using convolutional neural networks. *Biomedical & Pharmacology Journal, 11*(3), 1457.
4. Deepak, S., & Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. *Computers in Biology and Medicine, 111*, 103345.

5. Xie, Y., Zaccagna, F., Rundo, L., Testa, C., Agati, R., Lodi, R., ... & Tonon, C. (2022). Convolutional neural network techniques for brain tumor classification (from 2015 to 2022): Review, challenges, and future perspectives. *Diagnostics, 12*(8), 1850.
6. Kurdi, S. Z., Ali, M. H., Jaber, M. M., Saba, T., Rehman, A., & Damaševičius, R. (2023). Brain tumor classification using meta-heuristic optimized convolutional neural networks. *Journal of Personalized Medicine, 13*(2), 181.
7. Kaldera, H. N. T. K., Gunasekara, S. R., & Dissanayake, M. B. (2019, March). Brain tumor classification and segmentation using faster R-CNN. In *2019 Advances in Science and Engineering Technology International Conferences (ASET)* (pp. 1–6). IEEE.
8. Kumar, S., & Kumar, D. (2023). Human brain tumor classification and segmentation using CNN. *Multimedia Tools and Applications, 82*(5), 7599–7620.
9. Pan, Y., Huang, W., Lin, Z., Zhu, W., Zhou, J., Wong, J., & Ding, Z. (2015, August). Brain tumor grading based on neural networks and convolutional neural networks. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 699–702). IEEE.
10. Sajjad, M., Khan, S., Muhammad, K., Wu, W., Ullah, A., & Baik, S. W. (2019). Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *Journal of Computational Science, 30*, 174–182.
11. Saleh, A., Sukaik, R., & Abu-Naser, S. S. (2020, August). Brain tumor classification using deep learning. In *2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech)* (pp. 131–136). IEEE.
12. Khan, H. A., Jue, W., Mushtaq, M., & Mushtaq, M. U. (2021). Brain tumor classification in MRI image using convolutional neural network. *Mathematical Biosciences and Engineering, 18*(5), 6439–6453.
13. Kaggle. (2021). *Brain Tumor MRI Dataset*. Kaggle. Retrieved from https://www.kaggle.com/datasets