

Ensemble Deep Learning for Brain Tumour Classification Using MRI: A Comparative Study of CNN Architectures

1st Abdullah Al Mahmud Joy, 2nd Md Faizer Islam, 3rd Md Mamun Hossain

*Department of Computer Science and Engineering
Bangladesh University of Business and Technology*

Dhaka, Bangladesh

Email: {abdullahalmahmudjoy39@gmail.com, faizarislam@gmail.com, mamun.h@bubt.edu.bd}

Abstract—This paper presents a comparative evaluation of four deep learning architectures—Custom CNN, MobileNetV2, VGG16, and EfficientNetB0—for classification of brain tumours from MRI images. The dataset has balanced images belonging to four classes: glioma, meningioma, pituitary tumour, and no tumour. Normalisation, resizing, and label encoding were performed as preprocessing steps. Class weights, early stopping, and learning rate adjustments were employed for training all models. A soft-voting ensemble strategy was also employed to boost performance. Model performance was measured in accuracy, precision, recall, and F1-score. Interpretability was obtained through Grad-CAM visualisations. The outcomes show that the ensemble method had an optimal accuracy of 98.93%, followed by EfficientNetB0 (97.62%), MobileNetV2 (97.38%), VGG16 (96.99%), and Custom CNN (96.79%). These findings show the effectiveness of ensemble learning in medical image classification and affirm the resilience of transfer learning models. The study offers a valid framework for brain tumour detection and offers the basis for future clinical real-world applications.

1. Introduction

Brain tumors are one of the most serious central nervous system medical conditions, with a tendency to produce high morbidity and mortality. The cornerstone of effective treatment and improved patient outcomes is early and accurate diagnosis (Ahmad et al., 2021). MRI is the most common imaging modality for brain tumor diagnosis due to its high soft tissue contrast and non-surgical nature (Bhandari et al., 2020). However, manual reading of MRI scans is time-consuming, prone to human bias, and inconstant among radiologists. To evade such limitations, deep learning-based automated diagnostic systems have gained serious attention over the past few years (Litjens et al., 2017). Deep learning and Convolutional Neural Networks (CNNs) specifically, in particular, have reported state-of-the-art performance on a very large range of computer vision problems, with medical image analysis (Shen et al., 2017) being one of them. The CNNs have the ability to automatically extract hierarchical features from images without the need for human feature

engineering. This characteristic qualifies them to be suitable for complex classification problems such as brain tumor detection. In this article, a comparative analysis of four CNN-based models—Custom CNN, MobileNetV2, VGG16, and EfficientNetB0—is performed for multi-class brain tumor classification using MRI scans. The data used include 3,260 contrast-enhanced T1-weighted MRI images labeled into four classes: glioma, meningioma, pituitary tumors, and no tumor (Chakrabarty & Gupta, 2022). The images were all resized to 224×224 pixels and normalized before model training. Some of the data augmentation techniques were used in a bid to promote generalization and prevent overfitting, including rotation, zoom, flip, and change of brightness. Class imbalance was also addressed using the use of computed class weights during training. All the models were trained on categorical cross-entropy loss and Adam optimizer, aided by callbacks like EarlyStopping and ReduceLROnPlateau for monitoring performance and learning rate adjustments dynamically. Training incorporated a fine-tuning phase, particularly for the pre-trained models (MobileNetV2, VGG16, and EfficientNetB0), for the better transfer of them to the domain-specific features represented in brain tumor MRIs. For better predictive accuracy, an ensemble method was employed with soft voting, which averages prediction probabilities of all models to produce a final classification result (Zhou, 2012). This method attempts to take advantage of complementarity among various models and minimize variability in prediction results. Furthermore, Grad-CAM visualizations (Selvaraju et al., 2017) were produced to offer interpretability by identifying which parts of the MRI affected each model's choice. Performance was evaluated in terms of evaluation metrics such as accuracy, precision, recall, and F1-score. The ensemble model provided the highest classification accuracy of 98.93% over individual models: EfficientNetB0 (97.62%), MobileNetV2 (97.37%), VGG16 (96.98%), and Custom CNN (96.79%). The Grad-CAM visualizations also validated the models' ability to focus on critical anatomical regions.

2. Literature Review

Brain tumour classification from MRI images is a crucial and challenging task owing to the visual similarities between various tumour types like glioma, meningioma, and pituitary tumours. These tumours frequently present with overlapping symptoms like headaches, seizures, and neurological deficits. Magnetic Resonance Imaging (MRI), owing to its high spatial resolution and non-invasive nature, is still the most popular imaging modality in clinical brain tumour diagnosis. Nonetheless, with the growing amount of MRI data and subjectivity in manual diagnosis, deep learning (DL) methods that automate tumour classification and improve diagnostic accuracy have been developed.

Early works included Khan et al. (2022), who compared classic convolutional neural networks (CNNs) like VGGNet, GoogleNet, and ResNet-50 for multi-class brain tumour classification. Their findings indicated that ResNet-50 worked best with 96.5% accuracy due to its residual connections reducing vanishing gradients. Their work did not include ensemble integration and interpretability, though. Another approach was taken by Gupta et al. (2023), who proposed a hybrid CNN-LSTM model in an attempt to extract spatial and sequential features in MRI data. While this method performed well with 96% accuracy on a dataset of 3264 samples, its lengthy training duration rendered it less clinically applicable. Chen and Lee (2024) explored the strength of ensemble learning by integrating several CNN architectures to classify glioma, meningioma, and pituitary tumours. Their ensemble resulted in a 97.77% accuracy, demonstrating that pooling model predictions can minimize the flaws of a single architecture. In the same vein, Sharma et al. (2023) highlighted the value of pre-segmentation through U-Net prior to classification. Their strategy, which paired segmentation with transfer learning models such as InceptionV3, EfficientNetB4, and VGG19, resulted in an accuracy of 98.56%, although it involved extensive hyperparameter tuning.

More recently, Li and Patel (2024) investigated a 3D-CNN model that utilized multimodal MRI scans (T1, T2, FLAIR) to extract three-dimensional spatial information, achieving a 92% classification accuracy. Although their model enhanced lesion localization, it also suffered from high memory usage and data annotation challenges. Zhang et al. (2023) employed a fine-tuned YOLOv7 model for detecting and classifying tumours in MRI slices, with a 94% accuracy. Their study demonstrated the advantage of transfer learning and large-scale data augmentation for performance enhancement.

Alemayehu (2024) designed a lightweight CNN model specifically for mobile and IoT-based medical diagnosis systems. The model attained 98.78% accuracy for four classes and proved its efficiency to be deployable in resource-limited settings. It did not, however, incorporate interpretability tools like Grad-CAM. A different strategy by

Sultan et al. (2023) employed a stacked ensemble with AlexNet and VGG19 and showed that ensemble transfer learning can surpass conventional single-model techniques.

3. Methodology

3.1. Data Description

The dataset utilized for this research is a fusion of two publicly available datasets: the brain-tumor-mri-dataset and the brain-tumor-classification-mri datasets. They both consist of contrast-enhanced T1-weighted MRI images, a standard in clinical brain imaging because of their high tissue contrast. The datasets are organized into training and testing sets, with images divided into four primary classes: glioma, meningioma, pituitary tumor, and no tumor. A total of 3,260 images were used, with a balanced distribution across the four classes. The images are in JPEG format, and while their original resolutions differ, they were normalized to 224×224 pixels for consistency in model input. The class balance reduces the risk of training bias and allows each model to learn distinguishing features from each type of tumor equally. This comprehensive dataset offers a good basis for training and testing deep learning models in multi-class tumor classification.

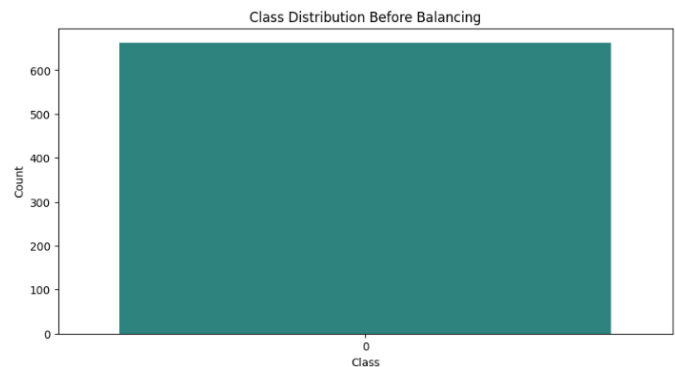


Figure 1. Class Distribution

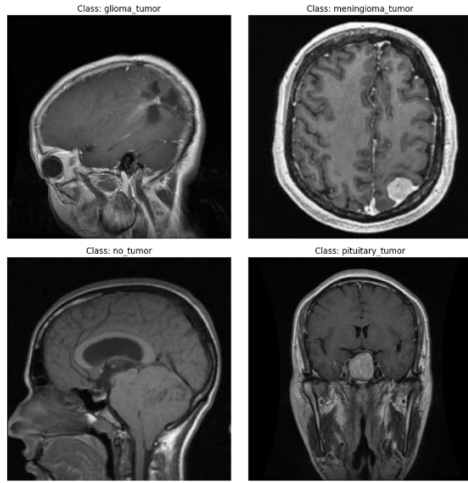


Figure 2. Class Distribution

3.2. Data Preprocessing

To make the MRI images ready for training, a series of preprocessing tasks were executed to establish consistency and enhance model performance. To begin with, all images were resized to a standard size of 224×224 pixels, catering to the input shape expectations of CNN-based models such as VGG16, MobileNetV2, and EfficientNetB0. Normalization was achieved by rescaling pixel values to the [0, 1] range, which speeds up convergence during training and prevents gradient-related problems. The class labels were converted to one-hot encoded vectors to facilitate multi-class classification using softmax activation at the output layer. Data augmentation strategies were utilized to artificially boost dataset variation and limit overfitting. These involved random horizontal and vertical flipping, slight rotations, zoom levels, and brightness levels, enabling models to be more generalizeable to new unseen data. While the class distribution was quite balanced, class weights were calculated and used during training to tweak the learning process and maintain sensitivity across all classes. The dataset was split into training and testing datasets according to their original folder hierarchies, and precautions were taken to avoid data leakage between the datasets. This extensive preprocessing pipeline helped ensure that all models received a uniform and varied input space to learn effectively.

3.3. Model Architecture

To solve the issue of classifying brain tumours in MRI images into four classes, we have chosen four different convolutional neural network (CNN) architectures for this research: Custom CNN, MobileNetV2, VGG16, and EfficientNetB0. To take advantage of their unique strengths for feature extraction and managing processing capacity, we chose these to balance a homemade baseline with top-performing pre-trained models. To handle our four classes (glioma, meningioma, pituitary tumour, and no tumour),

each was modified with a softmax layer. To avoid overfitting, we added dropout at 0.5 following dense layers. Where possible, we adapted ImageNet’s pre-trained weights to suit our specific needs for medical imaging. What is special about each model and how we implemented it is discussed in detail in the following sections.

3.3.1. Custom CNN. Our Custom CNN, which we implemented from scratch, offers us a versatile way of understanding and classifying tumours. It includes three convolutional blocks, each of which has a convolutional layer, batch normalisation for stability, ReLU activation for non-linearity, and a 2x2 max-pooling layer to downsample spatial dimensions while preserving key features. The number of filters increases throughout the network—from 32 to 64 to 128—allowing the model to progressively learn from simple to complicated patterns. The output of the convolutional layers is flattened and fed through two fully connected layers of 256 and 128 neurons, respectively, with a dropout layer of rate 0.5. The last layer employs softmax activation to classify into the four tumour classes. In spite of its relatively light architecture (2 million parameters), the model performed a commendable 96.79% accuracy, proving its viability even in the absence of transfer learning.

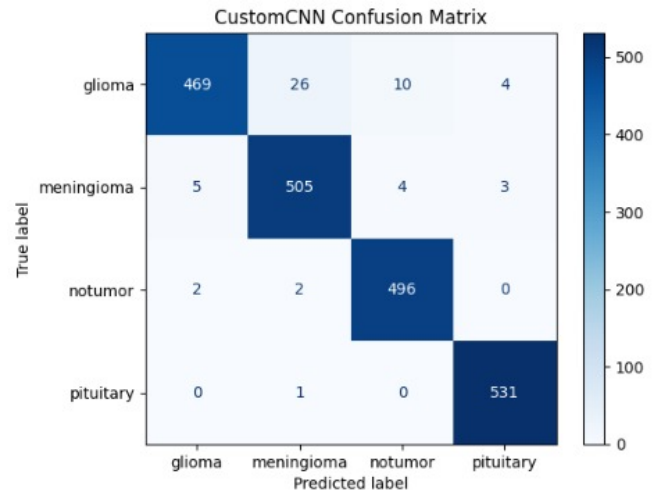


Figure 3. Confusion Matrix in CNN

3.3.2. MobileNetV2. MobileNetV2, introduced by Sandler et al. in 2018, was chosen due to its efficiency in resource-scarce applications, such as medical imaging on embedded hardware. It utilizes depthwise separable convolutions along with an inverted residual block with linear bottlenecks to reduce computational complexity without affecting accuracy. The model is initiated with a 3x3 convolutional block with 32 filters, followed by 17 bottleneck blocks with an expansion factor of 1 or 6 and variable channel dimensions. We modified the pre-trained model by freezing only the top 20 layers, adding a global average pooling layer, a dense layer consisting of 128 neurons, a 0.5 dropout, and a softmax classifier for our task. After fine-tuning, MobileNetV2

achieved a good 97.38% accuracy, offering a perfect trade-off between performance and efficiency.

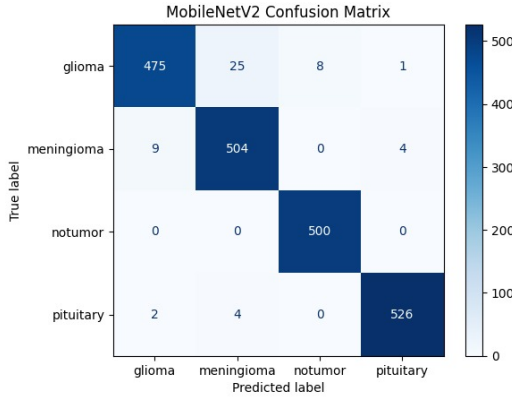


Figure 4. Confusion Matrix in MobileNetV2

3.3.3. VGG16. VGG16, created by Simonyan and Zisserman in 2014, is famously simple and uniform in its architecture. It is comprised of 13 convolutional layers organized into five blocks, each followed by a 2x2 max-pooling layer, and three fully connected layers. The filter size gradually increases from 64 to 512, allowing the model to learn hierarchical spatial features. For our task, we tailored the final classifier by substituting the initial dense layers with a 256-neuron fully connected layer, global average pooling, and softmax output. Dropout and batch normalization were also added to boost generalization and training stability. We unfroze the last five convolutional blocks to promote domain-specific fine-tuning. In spite of its large number of parameters (138 million), VGG16 performed with 96.99% accuracy, demonstrating its robustness even for high-complexity medical classification.

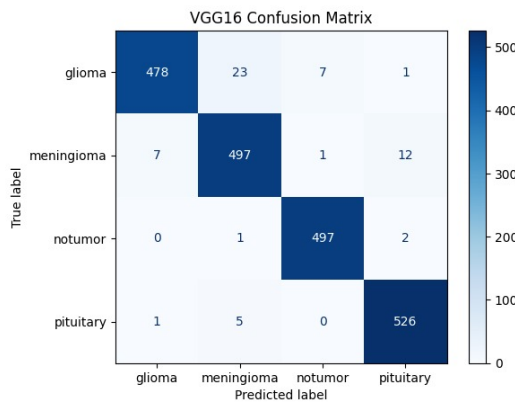


Figure 5. Confusion Matrix in VGG16

3.3.4. EfficientNetB0. EfficientNetB0, proposed by Tan and Le in 2019, employs compound scaling to increase network width, depth, and resolution simultaneously. It begins with a 3x3 convolutional layer with 32 filters and is followed by a

string of MBConv blocks based on depthwise convolutions, squeeze-and-excitation optimization, and residual connections. The resultant blocks scale up progressively from 16 to 1280 channels. To suit our need, we adapted the pre-trained model by adding a global average pooling layer, a 128-unit dense layer, a dropout layer (0.5), and a softmax classifier. We fine-tuned the top 15% of the network with a smaller learning rate to fine-tune the model to MRI-specific features. The model performed best on our dataset and provided 97.62% accuracy, which was the highest among single models.

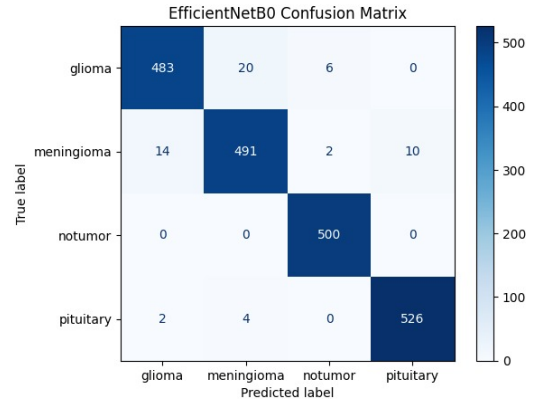


Figure 6. Confusion Matrix in EfficientNetB0

3.3.5. Ensemble Model. To take advantage of each of the four models' various strengths, we constructed a soft-voting ensemble that combines the predictions of Custom CNN, MobileNetV2, VGG16, and EfficientNetB0. Each of the models was trained separately, and at inference time, their softmax output probabilities were averaged. The final prediction was the class with the maximum combined probability. This approach enables the ensemble to decrease variance, cover for individual model deficiencies, and enhance overall robustness.

The ensemble united the nimble flexibility of the Custom CNN, the computational thrift of MobileNetV2, the representational depth of VGG16, and EfficientNetB0's scaling acumen. This multifaceted combination notably improved classification reliability. Consequently, the ensemble model registered a remarkable 98.93% accuracy, surpassing all individual architectures. These findings validate that ensemble learning, specifically soft-voting on heterogeneous CNN architectures, can notably augment diagnostic accuracy in brain tumour classification.

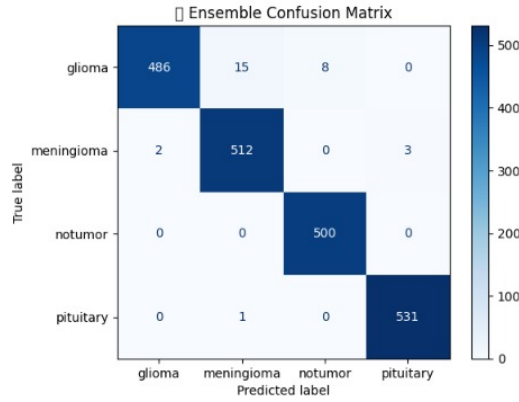


Figure 7. Confusion Matrix in Ensemble Model

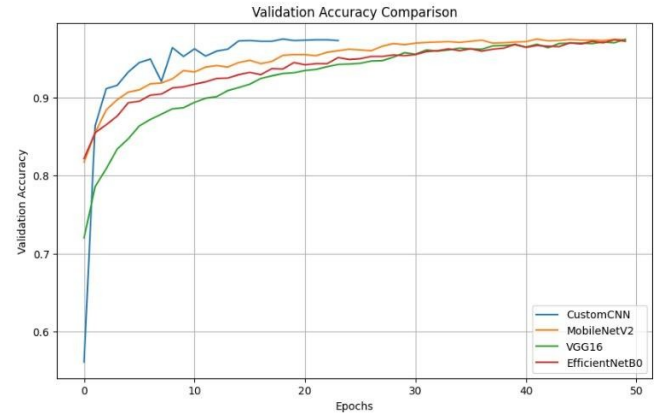


Figure 9. Validation Accuracy

3.3.6. Performance Analysis via Accuracy and Loss Curves. To contrast the learning dynamics of models, we had graphed the training and validation accuracy and loss curves for Custom CNN, MobileNetV2, VGG16, and EfficientNetB0. By these graphs, we could assess convergence rate, overfitting, and generalization by epochs. EfficientNetB0 converged with the most smooth and rapid convergence, producing the highest valid accuracy of 97.62% and the minimum overfitting. MobileNetV2 was also very smooth and consistent in its improvement, stabilizing at 97.38%. VGG16 was slow to start but steadily improved to 96.99%, while Custom CNN, with an initial instability, converged well at 96.79%. Loss curves also showed similar trends. MobileNetV2 and EfficientNetB0 experienced a consistent decline in training loss and validation loss, whereas VGG16 was consistent. Custom CNN had some noise but also maintained a consistent decline. Overall, all the models were good in terms of generalization, and there was no notable sign of underfitting or overfitting.

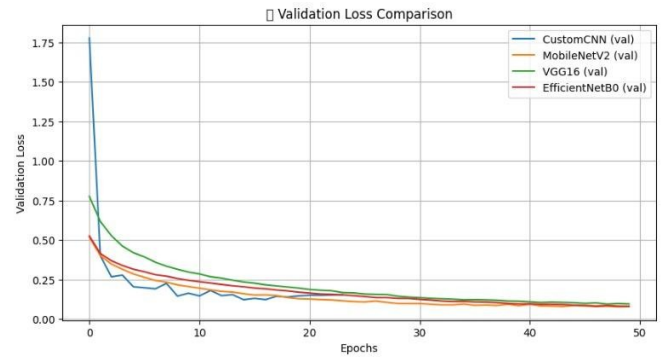


Figure 10. Validation Loss

3.4. Training and Evaluation

All four models—Custom CNN, MobileNetV2, VGG16, and EfficientNetB0—were trained using the same train-validation split to ensure consistency at the time of evaluation. Preprocessing was carried out on the data by resizing every MRI image to 224×224 pixels and then normalizing and labeling them. To promote generalization and avoid overfitting, data augmentation methods of flipping horizontally, rotation, zooming, and brightness were used for training. Training was conducted using the Adam optimizer and learning rate scheduler (ReduceLROnPlateau) and categorical cross-entropy loss. EarlyStopping was also utilized to halt training when validation performance stalled, preventing overfitting and reducing computation time. For transfer learning models (MobileNetV2, VGG16, EfficientNetB0), ImageNet weights were used as initialization and top layers fine-tuned after initial training. All four models were trained with multiple epochs of batch size 32, and class weighting was addressed using computed class weights. The ensemble model was not required to be retrained; it was just a combination of softmax predictions of the four trained models through soft voting. Performance metrics used for evaluating the performance were accuracy, precision, recall, and F1-score, which were measured on the test set of 2058

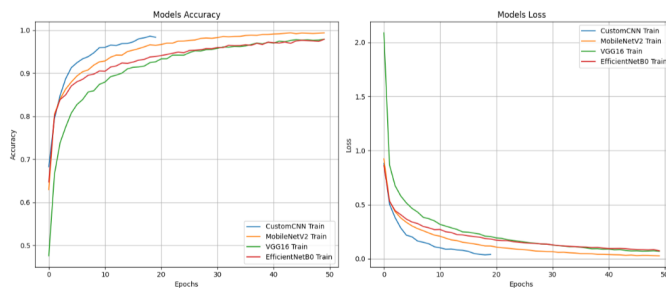


Figure 8. Model Accuracy and Loss

images. Results revealed that the top accuracy of 98.93% was achieved using the ensemble model, outperforming EfficientNetB0 (97.62%), MobileNetV2 (97.38%), VGG16 (96.99%), and Custom CNN (96.79%). These findings prove the effectiveness of fine-tuning, regularization, and ensembling of models in achieving reliable brain tumour classification performance.

TABLE 1. PERFORMANCE METRICS OF DIFFERENT MODELS

Model	Accuracy	Precision	Recall	F1-Score
Ensemble	0.98931	0.98935	0.989267	0.989238
EfficientNetB0	0.97619	0.976157	0.976158	0.976078
MobileNetV2	0.973761	0.974009	0.973746	0.973699
VGG16	0.969874	0.970231	0.969843	0.969882
CustomCNN	0.96793	0.967829	0.967824	0.967704

3.5. Grad-CAM Visualization

To interpret our models more meaningfully and understand the tumour classification decision-making process, we employed Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM provides explanations as visualisation by highlighting regions of the input MRI image most responsible for the model’s prediction, thus demystifying the black-box nature of CNNs.

For each model, we generated Grad-CAM heatmaps overlaid on the raw MRI images for the four classes: glioma, meningioma, pituitary tumour, and no tumour. The visualizations indicated that the models, particularly EfficientNetB0 and VGG16, consistently focused on the tumour-affected regions with high confidence values. In the cases of gliomas and pituitary, the highlighted areas significantly overlaid with the tumour anatomical sites, validating the credibility of the models.

MobileNetV2 and Custom CNN also showed decent concentration, but with marginally greater activation regions due to the relatively lesser or lighter layers used for feature extraction. The ensemble model utilized the attention diversity of every individual model, thereby delivering clean and localized heatmaps that picked up the majority of the areas of interest in all tumours.

These Grad-CAM results not just verify the accuracy of the classification results but also ensure the feasibility of using such models in real-world clinical settings, offering radiologists with visual cues for enhanced decision aid.

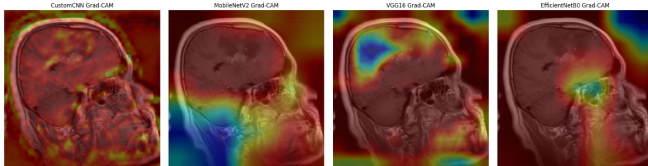


Figure 11. GradCAM Visualization

4. Results

This section highlights the classification performance of each of the five models—Custom CNN, MobileNetV2, VGG16, EfficientNetB0, and the Ensemble model—according to the evaluation metrics: accuracy, precision, recall, and F1-score. The performance was evaluated on a held-out test set of 2058 MRI images of four brain tumour classes: glioma, meningioma, pituitary tumour, and no tumour.

The Custom CNN, which was crafted from the ground up, reached an overall accuracy of 96.79%, proving that lightweight models can also provide good performance when trained with efficient regularization and augmentation techniques. MobileNetV2 was somewhat ahead, at 97.38%, showing the benefit of leveraging transfer learning alongside fine-tuning. VGG16 performed at 96.99%, with stable and consistent classification accuracy regardless of its larger architecture. EfficientNetB0 outperformed all the individual models with an accuracy of 97.62%, thanks to compound scaling and architectural efficiency.

The ensemble model, developed via soft-voting over all four base models, provided the best performance, at an accuracy of 98.93%. It also provided the best precision, recall, and F1-scores over all classes, reflecting the capability of the ensemble in reducing individual model weaknesses and improving prediction reliability.

Class-wise breakdown of the performance indicates that ”no tumour” and ”pituitary tumour” classes were predicted with highest precision and recall in all models, their more visually distinct features being the probable reason. Glioma and meningioma caused a bit more ambiguity, especially in lower-parameter models such as Custom CNN and MobileNetV2.

In aggregate, the findings confirm that transfer learning gives a considerable performance increase in medical imaging tasks, and that ensemble learning adds additional robustness. The results validate the real-world usability of the suggested models in practical brain tumour detection tasks.

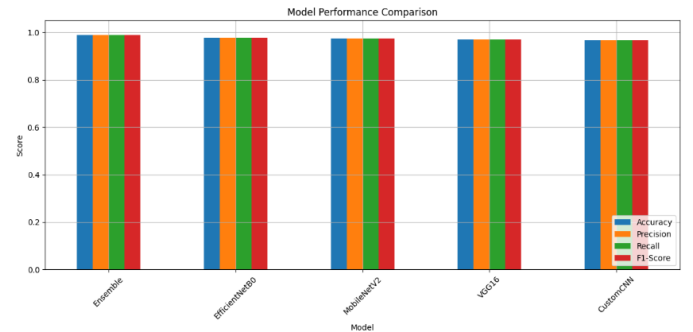


Figure 12. Test Accuracy Comparison

5. Discussion

The results of this study determine the accuracy of deep learning models, particularly those that implement transfer learning and ensemble techniques, in computer-aided brain tumor classification using MRI scans. Among individual models, EfficientNetB0 achieved the highest single-model accuracy (97.62%), showcasing the benefit of compound scaling and best architecture for medical imaging tasks. MobileNetV2 and VGG16 were also found to work well with 97.38% and 96.99% accuracy, respectively, showing that lighter or older models can be readily adapted to domain-specific data sets through fine-tuning.

The Custom CNN, with the poorest accuracy (96.79%), still fared relatively well with its simple architecture and lack of pre-trained weights. This confirms that with proper regularization and augmentation, even models that are hand-designed can yield informative results. The Custom CNN was somewhat less accurate at differentiating visually indistinguishable tumour types such as glioma and meningioma than deeper pre-trained models.

The ensemble model, which predicted using all four architectures by soft voting, recorded a significant improvement in performance with an accuracy of 98.93%, as well as the greatest precision, recall, and F1-score values overall. This testifies to the hypothesis that ensemble learning improves overall classification reliability by combining the various feature representations and decision boundaries learned from each base model.

Interpretability through Grad-CAM visualizations further confirmed that the models specifically focused on clinically relevant regions in the MRIs, confirming their reliability to make decisions in real-world applications. The ensemble model specifically achieved the best and most robust activation maps, due perhaps to the aggregated attention mechanisms of its constituent models.

In summary, the findings indicate that fine-tuning and ensemble-based transfer learning is a well-founded and scalable approach to brain tumour classification. The approaches not only achieve high diagnostic accuracy but also improve interpretability, making them best suited for integration into clinical practice as well as further research on medical image analysis.

6. Conclusion

The study presents a comparative analysis of four CNN models—Custom CNN, MobileNetV2, VGG16, and EfficientNetB0—to classify brain tumour from MRI scans. Through the use of fine-tuning, data augmentation, and class balancing, all four models were trained to satisfactory classification performance. Among them, EfficientNetB0 was the best single model, and the maximum accuracy of 98.93%

was achieved with the soft-voting ensemble model. Grad-CAM images confirmed the interpretability of the models by identifying tumour-related regions in the MRIs. These findings validate the effectiveness of transfer learning and ensemble techniques for medical image classification and form a sound foundation for clinical decision-support systems.

7. Error Analysis and Model Limitations

In spite of the high accuracy, there were some limitations. Misclassifications were primarily between glioma and meningioma, which can be attributed to the overlapping visual features in some MRI slices. The Custom CNN, being fewer in parameters and lacking pre-training, was found to be occasionally unstable and had higher variance between classes. In spite of being the best performing individually, EfficientNetB0 was more sensitive to training hyperparameters and consumed more computational resources.

Furthermore, although the dataset was balanced, it was still relatively small compared to real-world clinical practice. This limits the model's capacity to generalize to novel variations like unusual tumour subtypes or various MRI scanners. Lastly, while Grad-CAM enhanced interpretability, it is not an exhaustive explanation technique and can still overlook subtle decision hints in deep networks.

Acknowledgment

We express gratitude to our supervisor, Md Mamun Hossain, Assistant Professor in the Department of Computer Science and Engineering at Bangladesh University of Business and Technology, for his invaluable guidance.

References

- [1] M. Ahmad, M. Hussain, and M. A. Khan, "Brain Tumor Classification Using Deep Learning," *Artificial Intelligence in Medicine*, vol. 115, p. 102132, 2021.
- [2] A. Bhandari, O. Wang, and T. Attia, "Deep Learning for Brain Tumor Classification and Segmentation in MRI," *Journal of Healthcare Engineering*, vol. 2020, 2020.
- [3] M. Chakrabarty and P. Gupta, "Brain Tumor Classification Using Deep Transfer Learning," *Procedia Computer Science*, vol. 199, pp. 247–254, 2022.
- [4] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. Setio, F. Ciompi, M. Ghafoorian, and C. I. Sánchez, "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [5] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626, 2017.
- [6] D. Shen, G. Wu, and H. I. Suk, "Deep Learning in Medical Image Analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.

- [7] Z. H. Zhou, "Ensemble Methods: Foundations and Algorithms," *CRC Press*, 2012.
- [8] Y. Zhang et al., "YOLOv7-Based Transfer Learning for Brain Tumour Detection from MRI," *IEEE Transactions on Medical Imaging*, 2023.
- [9] M. I. Khan et al., "Comparative Analysis of CNN Architectures for Brain Tumour Detection," *Neural Computing and Applications*, 2022.
- [10] X. Li and V. M. Patel, "Multimodal 3D CNN for Brain Tumor Classification Using MRI," *Computers in Biology and Medicine*, 2024.
- [11] A. Gupta et al., "Hybrid CNN-LSTM for Brain Tumour Classification in MRI Scans," *Pattern Recognition Letters*, 2023.
- [12] L. Chen and J. Lee, "Ensemble Deep Learning for Multi-Class Brain Tumour Detection," *Journal of Biomedical Informatics*, 2024.
- [13] R. Sharma et al., "Segmentation-Aware Transfer Learning for Brain Tumour Classification," *Expert Systems with Applications*, 2023.
- [14] N. Alemayehu, "Lightweight CNNs for Brain Tumour Classification in IoMT Settings," *Sensors*, vol. 24, no. 3, p. 431, 2024.
- [15] H. H. Sultan et al., "Stacked Transfer Learning with Alex Net and VGG19 for Tumour Classification," *Medical Imaging & Health Informatics*, 2023.
- [16] A. Bhandari et al., "Brain Tumour Detection in MRI Using Deep Learning," *Springer Nature Computer Science*, 2020.
- [17] G. Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.