Ensemble deep learning of CNN vs vision transformers for brain lesion classification on MRI images*

Norelhouda. Laribi[1,†], Djamel. Gaceb[1,*,†], Fayçal. Touazi[1,†], Abdellah. Rezoug[1,†], Abdelmoumen. Sahad[1,†] and Massine Omar. Reggai[1,†]

[1] *LIMOSE laboratory, Computer science department, University M'hamed Bougara, Independence Avenue, 35000 Boumerdes, Algeria*

Abstract

Deep learning models, such as Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), have achieved state-of-the-art performance in the classification of brain lesions on MRI images. However, the complexity of this type of images requires CNNs to use deeper architectures with more parameters to effectively capture their high-dimensional features and subtle variations. On one side, ViTs provide a different approach to tackle this challenge, but they require larger datasets and more computational costs. On the other side, ensemble deep learning techniques, such as Bagging, Stacking, and Boosting, can help mitigate these limitations by combining multiple CNN models. This study explores these methods and compares them to evaluate their accuracy and efficiency using three approaches: CNN-based transfer learning, ViT-based transfer learning, and ensemble deep learning techniques, such as Bagging, Stacking, and Boosting based on XGBoost, AdaBoost methods. The experiments, carried out on four MRI image datasets with different levels of complexity and types of brain lesions, show that the combination of CNNs with ensemble techniques offers very competitive performances to those of individual CNNs and ViTs, with interesting improvements compared to already existing approaches.

Keywords

Brain Lesions, MRI images, Transfer Learning, Ensemble deep learning, Bagging, Stacking, Boosting, AdaBoost, XGBoost, AdaBoost, Convolutional Neural Networks, Vision Transformers

1. Introduction

Brain lesions, including those caused by strokes, tumors, and ischemic injuries, pose significant diagnostic challenges due to their complex nature. These complexities involve including varying shapes, sizes, locations, textures and appearances, as well as variability in presentation in medical imaging. Traditional imaging methods often require specialized expertise, advanced techniques, and a deep understanding of subtle patterns within images, which can lead to variability in interpretation and be limited by the available data. Deep learning has significantly improved the detection, classification, and segmentation of complex medical conditions, including brain lesions. Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) have emerged as leading solutions in medical imaging tasks due to their ability to automatically extract and learn from complex features.

Transfer learning involves adapting models that are pre-trained on large datasets, such as ImageNet, for use in medical imaging, however, its effectiveness depends on model's architecture and the size of the dataset. For example, when comparing CNNs to ViTs, CNNs often face challenges in generalizing across specific datasets, particularly when dealing with the deep features of brain lesions or the limited availability of labeled medical data.

To address the performance gap between CNNs and ViTs in classification tasks, this study proposes a comprehensive approach to enhance the performance of CNN models (in feature extraction mode) to match that of ViTs. This is achieved by utilizing transfer learning to extract features from CNNs

IDDM'24: 7th International Conference on Informatics Data-Driven Medicine, November 14 - 16, 2024, Birmingham, UK

[*] Corresponding author.

[†] These authors contributed equally.

n.laribi@univ-boumerdes.dz (Norelhouda. Laribi); d.gaceb@univ-boumerdes.dz (Djamel. Gaceb); f.touazi@univ-boumerdes.dz (Fayçal. Touazi); a.rezoug@univ-boumerdes.dz (Abdellah. Rezoug); sahad1.abdelmoumen@gmail.com (Abdelmoumen. Sahad); reggai.massine@gmail.com (M. Omar. Reggai)

0000-0002-6178-0608 (Djamel. Gaceb); 0000-0001-5949-5421 (Fayçal. Touazi); 0000-0001-7326-9164 (Abdellah. Rezoug)

and employing ensemble methods such as Bagging, Stacking, and Boosting (e.g., XGBoost, AdaBoost). These techniques allow CNNs to generalize better by detecting more global features and capturing complex relationships from diverse extracted features. As a result, this approach achieves performance comparable to ViTs and significantly improves generalization for brain lesions classification, such as tumor and stroke predictions, thereby closing the gap between CNNs and ViTs in both feature extraction. The rest of the paper is organized as follows: Section 2 reviews the related work in brain lesion classification using machine and deep learning methods, Section 3 describes the proposed approaches, Section 4 covers the experimental results with the different datasets used. A discussion of the results and a comparison with existing work have been provided at the end of this section.

2. Related work

In recent years, several machine learning and deep learning approaches have been explored to improve accuracy in predictive tasks on Brain Stroke CT Image dataset. Sailasya et al. [1] conducted a comparative study using Logistic Regression (LR), Decision Trees (DR), and Random Forest (RF), achieving accuracies of 78%, 66%, and 73%, respectively. Similarly, Aniwat et al. [2] employed a Deep Neural Network (DNN), achieving a significant improvement with an accuracy of 96.21%. Devet et al. [3] extended this line of research by experimenting with Support Vector Machines (SVM), RF, and CNN, obtaining accuracies of 68%, 74%, and 74%, respectively. Further advancements were observed by Santwana et al. [4] who achieved 87.22% accuracy with RF, and Akter et al. [5], whose RF model yielded an accuracy of 95.30%. Tursynova et al. [6] employed CNN with an accuracy of 81%, while Yeo et al. [7] combined RNN and CNN architectures, achieving 93%. Gupta et al. [8] demonstrated the effectiveness of DenseNet-121, reporting an accuracy of 96%. Luis et al. [9] also reported high performance using a DNN, with 92.70% accuracy. Moreover, a hybrid approach combining Genetic Algorithm (GA) with Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) achieved accuracies of 95.35% and 96.45%, respectively [10]. The proposed method, combining XGBoost with VGG16, outperforms the previous approaches with an accuracy of 98.6%, indicating a significant improvement over traditional and deep learning methods. Another study [11] utilized a CNN-based transfer learning approach for automated stroke lesion classification, achieving 95.06% precision. Additionally, a recent integrated approach [12] combining ResNet50, Vision Transformer, and AutoML, enhancing both slice-level and patient-wise predictions on CT volume for real brain Stroke Dataset, with 87% and 92% accuracy, respectively, outperforming standalone models like VGG-16, VGG-19, ResNet50, and ViT by 9%, highlighting the advantage of leveraging transformers within hybrid architectures.

For Acute Ischemic Stroke (AIS) lesions, classification is less common compared to segmentation tasks, due to the complexity of these images and the need for large, diverse datasets with labeled outcomes. However, classification remains relevant for tasks such as stroke detection, subtype identification (ischemic vs. hemorrhagic), severity grading, and patient recovery prediction. Recent studies have explored the use of advanced models for AIS classification. One study [13] utilized ResNet101 and DenseNet201 models on an MRI dataset [14], achieving an impressive accuracy of 97.5% using a 10-fold cross-validation approach. Another study [15] evaluated the performance of transfer learning-based CNN models (Inception-v3, EfficientNet-b0, and a modified LeNet), finding that Inception-v3 had the best overall accuracy (86.3%) and performance metrics. Additionally, another study [16] highlighted the benefits of combining clinical and imaging data to predict 3-month functional outcomes in AIS patients. This study used a multi-center dataset of 4,147 patients and developed three models: Model A (clinical data), Model B (imaging data), and Model C (integrated model). Model C, which combined clinical and imaging data, demonstrated the highest performance with an AUC of 78.6%, outperforming both Model A (AUC: 75.7%) and Model B (AUC: 72.5%).

Building on advancements in Vision Transformers (ViTs), recent studies have investigated their application in brain imaging to address challenges in prediction accuracy and data efficiency. One study [17] leveraged a transformer architecture instead of traditional CNNs, where the proposed ViT model achieved 80.03% accuracy on training and 81.20% on testing. Other approaches [18, 19] improved the prediction of brain stroke lesions on CT imaging data by combining non-contrast CT images with patient reports through a multi-modal fusion framework based on the Transformer architecture. These approaches leveraged complementary information from both sources, demonstrating the potential of Transformer-based models in multi-modal medical applications, achieving state-of-the-art performance. Additionally, to address the challenge of data imbalance in stroke diagnosis, a recent study [20] introduced a ViT model designed to improve stroke classification accuracy by employing data augmentation. The model's performance significantly increased from 97.25% accuracy to 98.75%, demonstrating how data augmentation effectively enhances predictive accuracy.

For brain lesions, various studies used MRI images from patients with Glioma, Meningioma, and Pituitary tumors on the Brain Tumor Dataset [21] based on ensemble learning approaches. Research [22] utilized ensemble deep learning models to improve diagnostic prediction for brain tumors, integrating CNNs. The stacked ensemble of three models, VGG19, Inception v3, and ResNet10, achieved a high accuracy of 96.6% for binary classification. A novel brain tumor classification model [23], using the XGBoost algorithm to address the challenge of limited data, employed image augmentation through a conditional Generative Adversarial Network (cGAN). It introduced a hybrid feature fusion approach, combining deep features, 2D Fractional Fourier Transform (2D-FrFT) features, and geometric features to capture both global and local information from MRI scans. The model achieved a high classification accuracy of 98.79% and sensitivity of 98.77%, outperforming state-of-the-art algorithms on the Brain Tumor Dataset. Another paper [24] introduced bagging to improve CNN predictions, where EfficientNetB0's accuracy increased from 97.71% to 99.64% when features extracted from it were fed into a bagging tree classifier, leading to a significant improvement in overall performance on the Brain Tumor Dataset. Another research [25] highlighted that AdaBoost, which uses a boosting technique to convert weak models into stronger ones, achieved a higher accuracy of 95% compared to 89% for Random Forest on the Brain Tumor Dataset.

Another area of research in brain tumors focuses on detecting tumors and predicting the presence of MGMT promoter methylation, a crucial biomarker for determining glioma treatment response using MRI scans. The MICCAI initiative has introduced a comprehensive and complex dataset in this field, known as the RSNA MICCAI-MICCAI dataset [26], which is widely used for advancing research in brain tumor detection and MGMT status prediction. In recent years, the application of ensemble learning methods has significantly impacted the analysis of the RSNA MICCAI, Brain Tumor Radio Genomic Classification Challenge dataset. Researchers have utilized ensemble deep learning frameworks to enhance classification accuracy, especially for predicting MGMT promoter methylation status, a key biomarker in glioma treatment. One study [27] compared three different deep learning approaches—whole-brain, slice-wise, and voxel-wise—for predicting MGMT promoter methylation status using fivefold cross-validation, leveraging YOLOv5x, pre-trained on the COCO dataset, and a 3D-DenseNet169 model from the MONAI models package with transferred weights. Among the approaches, the whole-brain method consistently outperformed the others, achieving the highest accuracy (55.26% to 66.37%) and AUC-ROC values (up to 0.6597), likely due to its ability to capture the 3D spatial relationships in the brain. The voxel-wise approach followed, with better performance (accuracy: 59.23% to 63.71%, AUC-ROC: up to 0.6608), while the slice-wise approach had the lowest accuracy (50.00% to 60.34%) and AUC-ROC values (0.5023 to 0.6236), indicating it missed important spatial information. This demonstrates the advantage of 3D models in predicting MGMT methylation status from MRI data. Another study [28] highlighted the significant challenges in using current deep learning models, which struggle with MGMT promoter methylation status prediction from MRI data. Specifically, the results showed low accuracy, with performance close to random guessing, in which the ResNet10 models achieved AUC values between 0.54 and 0.59, indicating the need for optimization.

This paper introduces an ensemble approach that leverages advanced deep learning techniques, incorporating transfer learning and ensemble methods combining the strengths of CNNs, to improve brain lesion classification.

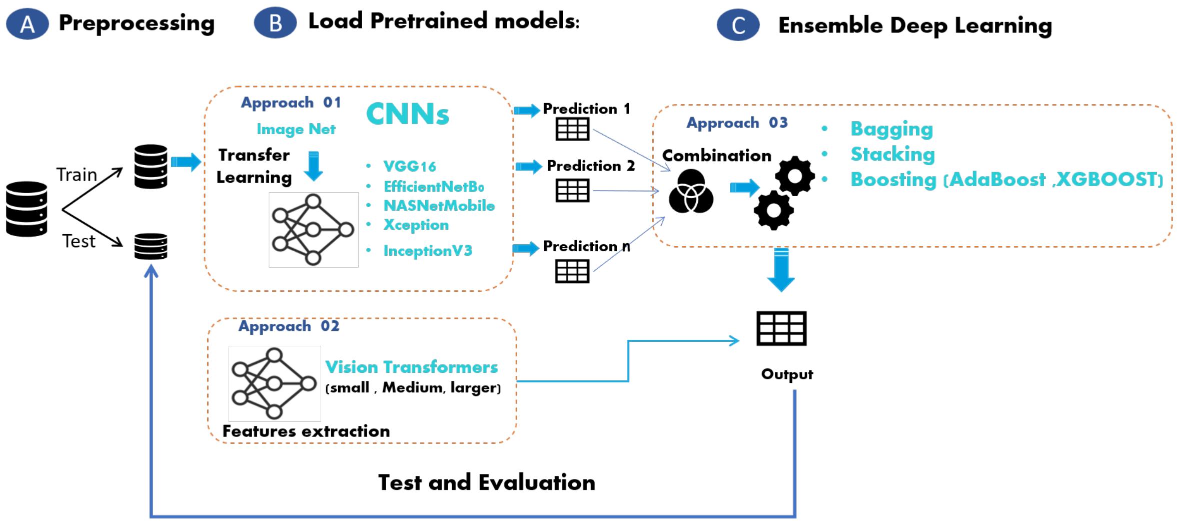

Figure 1: Proposed approaches for brain lesion classification: combining transfer learning of CNNs, ViTs and ensemble deep learning techniques.

3. Proposed approaches

In this section, the proposed approaches are presented, aimed at enhancing the accuracy of brain lesion classification. Building on the strengths of various CNN and ViT architectures with advanced deep learning techniques, including transfer learning and ensemble learning methods (see Figure 1). The study tested these approaches using four distinct public datasets, each with its own complexities: The Brain Stroke CT Image Dataset [21], Brain Tumor MRI Database [21], Acute Ischemic Stroke MRI [14], and RSNA MICCAI Dataset [26] (see more details in Section ??). The goal was to explore and compare the performance and convergence of the approaches based on the varying complexities of both models and medical datasets.

3.1. Approach 1: Transfer Learning of CNNs

This approach allows us to compare different CNNs for extracting features of brain lesions using pre-trained architectures with varying computational complexities and parameter counts, including lightweight models like EfficientNetB0 and NASNetMobile, medium models like InceptionV3, and heavyweight models like VGG16. It uses two Transfer Learning modes: feature extractor and fine-tuning, with different levels of fine-tuning (see Figure 2, right).

3.2. Approach 2: Transfer Learning of ViTs

This approach allows us to explore the effectiveness of ViTs for brain lesion classification through transfer learning (see Figure 2, left). By leveraging pre-trained ViTs [29] on ImageNet, the research evaluates their capability as feature extractors across different datasets. Various ViT architectures of different complexities are tested, including SwinTransformerV2Tiny [30] (28.3M parameters), MobileViT_V2_100 [31] (4.9M parameters), and larger models like ViT L/32 and ViT L/16, with 307M parameters. Smaller ViT models, such as ViT B/32 and ViT B/16 (86M parameters), are also evaluated, as these models have demonstrated high accuracy for brain tumor detection tasks.

3.3. Approach 3: Ensemble Deep Learning

The approach aims to enhance the accuracy and generalization of CNN-based models in medical imaging tasks by implementing three key ensemble deep learning strategies: parallel methods (Bagging and Stacking) and sequential methods (Boosting: XGBoost and AdaBoost). Bagging employs bootstrapping

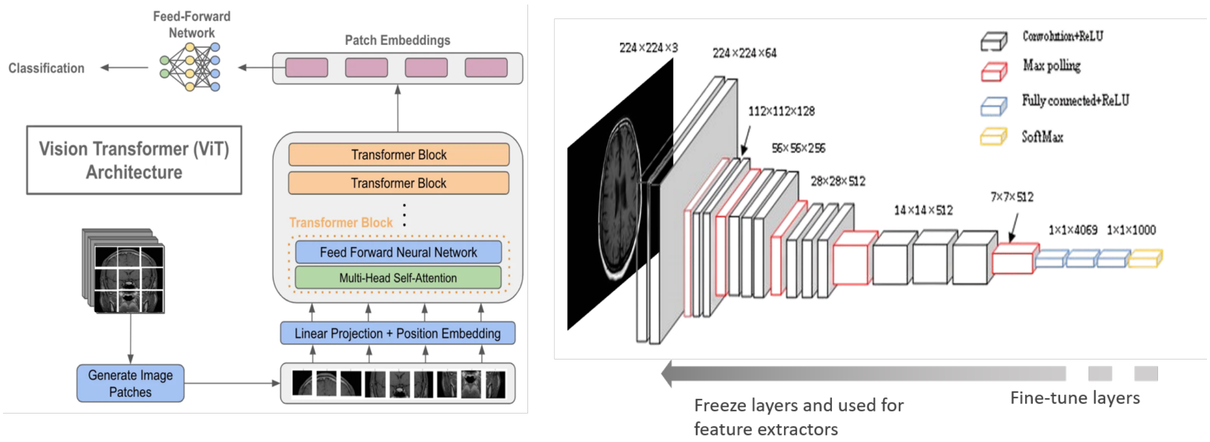

Figure 2: An example of deep learning architectures for classification: VGG-based CNN (right) and ViT (left).

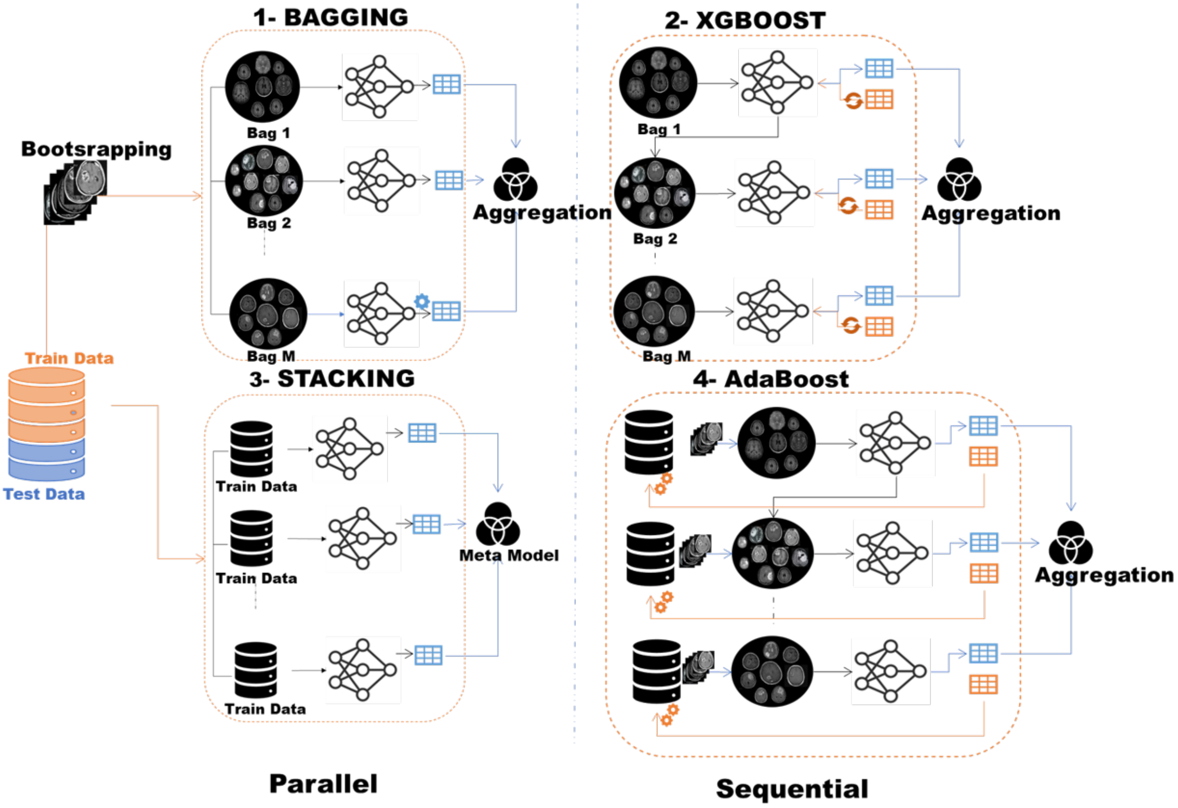

Figure 3: General diagram of ensemble deep learning approaches, proposed for brain lesion classification: Parallel (Bagging, Stacking) and sequential (XGBoost, and AdaBoost).

to generate diverse subsets from the original training data, enabling multiple CNN models to train independently, each focusing on different aspects of feature extraction.

Stacking aims to increase model diversity and maximizing classification accuracy in CNN-based medical imaging applications. For this, regression techniques will be used to improve models' generalization across complex datasets by refining predictions and capturing continuous relationships that are not fully detected in Brain Lesions classification tasks, enabling them to match the performance of ViTs. Finally, Boosting, with both XGBoost and AdaBoost, sequentially integrates CNN feature outputs, progressively refining the learning process and further enhancing classification accuracy. AdaBoost, in particular, adjusts model weights to focus on misclassified instances, complementing XGBoost's

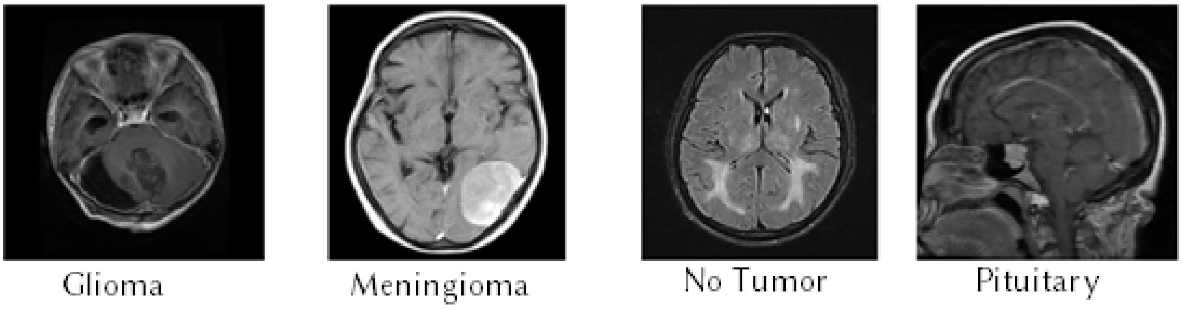

Figure 4: Image examples from the Brain Tumor MRI Dataset [21].

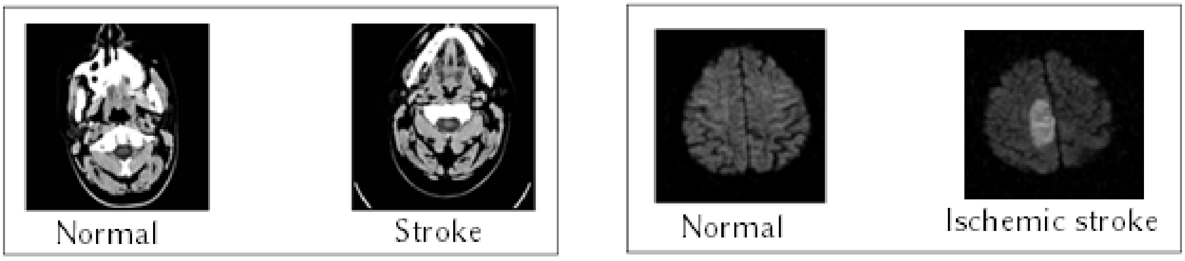

Figure 5: Image examples from stroke datasets (Left: Brain Stroke CT Images [32], Right: Acute Stroke MRI Dataset [14]).

gradient-based approach to improve model performance.

4. Experimentation and results

In our experiments, a focus is on retraining various CNN architectures, originally trained on ImageNet, to evaluate their performance on brain lesion classification using two transfer learning modes: feature extraction and fine-tuning. Models of different complexities are tested, including VGG16 (138M parameters), EfficientNetB0 (5.3M parameters), NASNetMobile (5.3M parameters), Xception (22.9M parameters), and InceptionV3 (23.9M parameters). By comparing their performance across these transfer learning modes, we gained valuable insights into how model complexity affects results in medical imaging tasks. In parallel, ViTs of different complexities were also retrained in feature extraction mode, including the 86M-parameter ViT-B32, the 28M-parameter Swin TransformerV2, and the 5.6M-parameter MobileViTV2. Additionally, ensemble learning techniques are applied on CNNs and compared to ViT models. The goal was to evaluate the trade-off between leveraging model complexity and enhancing prediction diversity through ensemble methods for CNNs to match Vision Transformers' generalization.

To conduct our experiments, various brain lesion datasets are used: Brain Tumor MRI dataset [21] (see Figure 4), Brain Stroke CT dataset [32] (see Figure 5, left), Acute Ischemic Stroke MRI dataset [14] (see Figure 5, right), and RSNA MICCAI dataset [26] (see Figure 6).

In Figures 4, 5, and 6, sample images from the datasets are presented to showcase the types of lesions and the data quality used for training and evaluation, followed by a table summarizing the size of each dataset and the number of classes (see Table **??**). Each dataset is divided into 80% for training and 20% for testing. For the Brain Tumor MRI dataset, a multi-class classification approach is employed, where the model is trained to classify the data into four distinct tumor types: Glioma, Meningioma, No Tumor, and Pituitary. In contrast, for the remaining datasets—Brain Stroke CT, Acute Ischemic Stroke MRI, and the RSNA MICCAI dataset—a binary classification approach is applied, categorizing instances into two classes: lesion present or lesion absent. For the RSNA MICCAI dataset, only the public training dataset is used to perform binary classification tasks (presence or absence of MGMT value), as the public test dataset was initially used for a regression task.

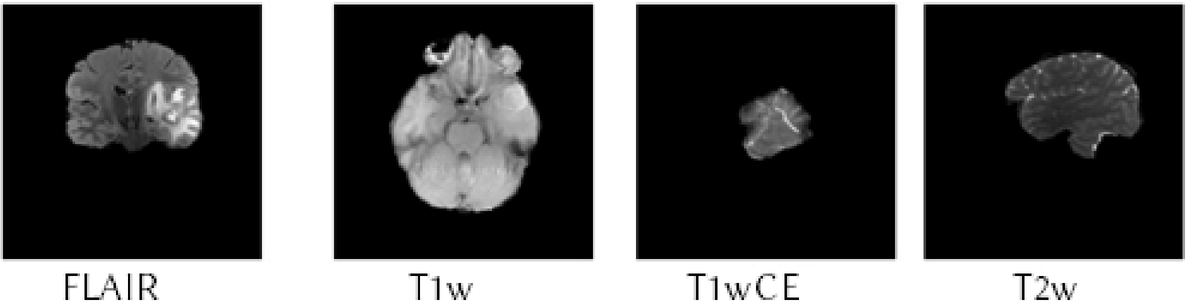

Figure 6: Image examples from the RSNA MICCAI Dataset: FLAIR, T1w, T1wCE, and T2w modalities [26].

Table 1
Overview of Dataset used for Brain Lesion Classification

Datasets	Size	Number of Classes
Acute Ischemic Stroke MRI Dataset	2,000	`2 (Acute Ischemic Stroke, Non-Stroke)`
Brain Stroke CT Images Dataset	2,501	`2 (Stroke, Non-Stroke)`
Brain Tumor MRI Dataset	7,023	`4 (Glioma, Meningioma, No Tumor, Pituitary)`
RSNA MICCAI Dataset	290,923	`2 (MGMT Presence, No MGMT Presence)`

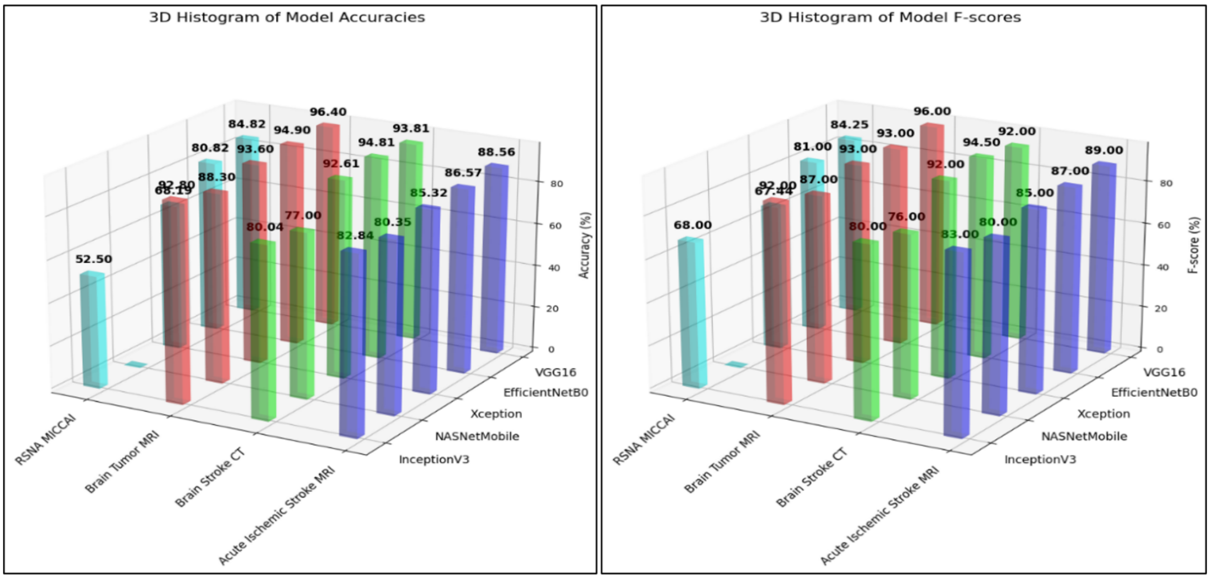

Figure 7: Performance comparison of different CNN models across multiple brain lesion datasets using Transfer Learning in feature extraction mode.

4.1. Results of approach 01: Transfer Learning of CNNs

This approach is based on comparing the performance of several CNN models (VGG16, EfficientNetB0, InceptionV3, NASNetMobile, and Xception) using two fine-tuning modes: feature extraction (see Figure 7) and fine-tuning (see Figure 8). These models are evaluated on different datasets, aiming to identify challenges related to the complexity of both the datasets and the CNN models.

In feature extraction mode, VGG16 consistently performs well, achieving a high accuracy of 96.4% on the Brain Tumor MRI dataset without extensive fine-tuning. EfficientNetB0 also demonstrates strong performance, achieving 94.81% accuracy on the Brain Stroke CT dataset and 94.9% accuracy on the Brain Tumor MRI dataset, though it performs lower on the Acute Ischemic Stroke MRI dataset with 86.57% accuracy. InceptionV3 shows unstable results, performing well on the Brain Tumor MRI dataset (92.8% accuracy) but struggling on the RSNA MICCAI dataset (52.50% accuracy), which is more complex.

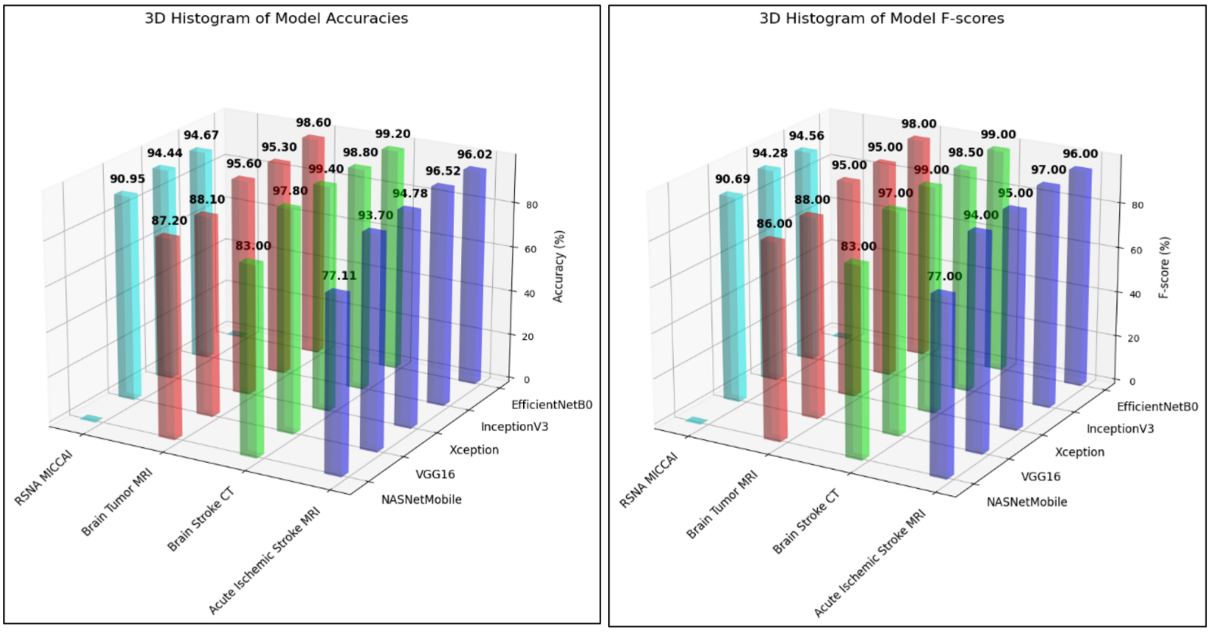

Figure 8: Performance comparison of different CNN models across multiple brain lesion datasets using Transfer Learning in fine-tuning mode.

NASNetMobile performs decently on the Brain Tumor MRI dataset but exhibits weaker results on the Brain Stroke CT dataset. Xception also faces difficulties, particularly on the RSNA MICCAI dataset, where it achieves only 68.19% accuracy. These results indicate that the feature extractor mode faces limitations on datasets like Acute Ischemic Stroke MRI and RSNA MICCAI, where robust models such as InceptionV3, VGG16, and Xception struggle, necessitating deeper fine-tuning. In fine-tuning mode, EfficientNetB0 and InceptionV3 consistently outperform the others, especially on datasets like Acute Ischemic Stroke MRI. Xception also shows improvement for this dataset, achieving 94.78% accuracy and a 95% F-score. VGG16 performs well, reaching 97.80% accuracy and a 97% F-measure on the Brain Stroke CT dataset. However, its performance decreases with more complex datasets like RSNA MICCAI, where it achieves 90.95% accuracy and a 90.69% F-measure. Even with fine-tuning, Xception reaches 94.44% accuracy and a 94.28% F-measure on the RSNA MICCAI dataset. By contrast, NASNetMobile struggles with fine-tuning, as demonstrated on the Brain Stroke CT dataset, where it only achieves 83% across all metrics.

The study highlights the importance of considering dataset complexity when choosing feature extraction models. While InceptionV3 performs poorly on complex datasets initially, fine-tuning improves its performance to 94.67%. EfficientNetB0 and InceptionV3 show potential for medical imaging tasks but face challenges with larger datasets.

Building on these observations, the next section explores the results of our second approach: Application of Transfer Learning on ViTs to compare their performance to that of the previously used CNN models on brain lesion imaging datasets, highlighting their effectiveness and adaptability across diverse dataset complexities.

4.2. Results of approach 2: Transfer Learning on ViTs

In this approach, we analyze the effectiveness of ViTs on the previously mentioned datasets in feature extractor mode. For this, we employed various configurations of ViTs (B/16, L/16, B/32, and L/32) with different patch sizes, each pre-trained on large-scale datasets such as ImageNet-21k and ImageNet-1k. These models were then adapted for the brain lesion classification task, with images resized to 224×224 pixels to ensure optimal input size for the transformers.

ViTs have demonstrated high accuracy in feature extractor mode, consistently outperforming CNNs across various datasets. For instance, ViT L/16 achieved 97.9% accuracy on the Brain Tumor MRI dataset

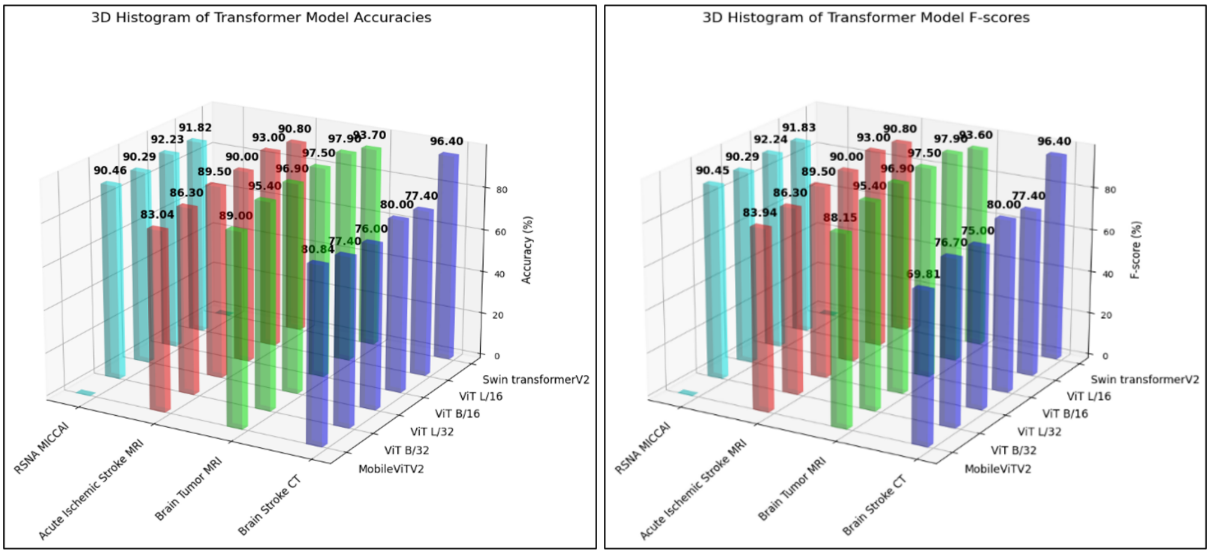

Figure 9: Performance comparison of different ViTs models across multiple brain lesion datasets using transfer learning in feature extractor mode

and 92.23% on complex datasets like RSNA MICCAI. However, on the Brain Stroke CT dataset, ViT models often fail to outperform, with performance typically ranging between 75% and 80% accuracy. Notably, smaller ViT models, such as ViT B/16, achieve higher performance at 80% across all metrics.

In contrast, Swin Transformers V2 and MobileViT V2 have shown difficulties converging on complex datasets like RSNA MICCAI. On the Brain Stroke CT dataset, Swin Transformers achieved a strong accuracy of 96.4%, whereas MobileViT V2 exhibited lower performance, with an accuracy of 80.84%.

To further enhance classification robustness and accuracy across brain lesions datasets, the following sub-section investigates ensemble deep learning with CNNs as an alternative strategy to address the performance limitations observed with individual CNNs and ViTs.

4.3. Results of approach 03: Ensemble Deep learning

The study aims to enhance the performance of CNNs through an ensemble approach and compare them with powerful models like ViTs. The ensemble method leverages model diversity to address CNNs' limitations in extracting local features, thereby improving their ability to capture global dependencies and enabling them to match the performance of ViTs. The experiment begins by implementing parallel ensemble methods, such as Bagging and Stacking. Subsequently, a sequential boosting ensemble approach is applied using both AdaBoost and XGBoost. The objective is to evaluate the performance of these ensemble models and understand how they handle feature diversity across various datasets, with a focus on matching the feature extraction capabilities of ViTs.

4.3.1. Bagging

In this experiment, different bootstrap samples (bags) are created, and a model in feature extraction mode is retrained on each of these bags. The models are then combined by aggregating their outputs using either voting or averaging their predictions. During experimentation, the size of the bags is varied between 40% and 60% to optimize the performance of each model. The sub-models are retrained without data augmentation, using the Adam optimizer with a learning rate of 0.01, except for the RSNA MICCAI dataset, where a learning rate of 0.001 is used. The loss function is set to cross-entropy (categorical for the Brain MRI dataset and binary for the other datasets), with performance metrics including accuracy, precision, recall, and F-score.

Bagging enhances the performance of models that underperform in feature extraction mode, particularly for InceptionV3 and NASNetMobile. On the Acute Ischemic Stroke MRI dataset, Bagging improves

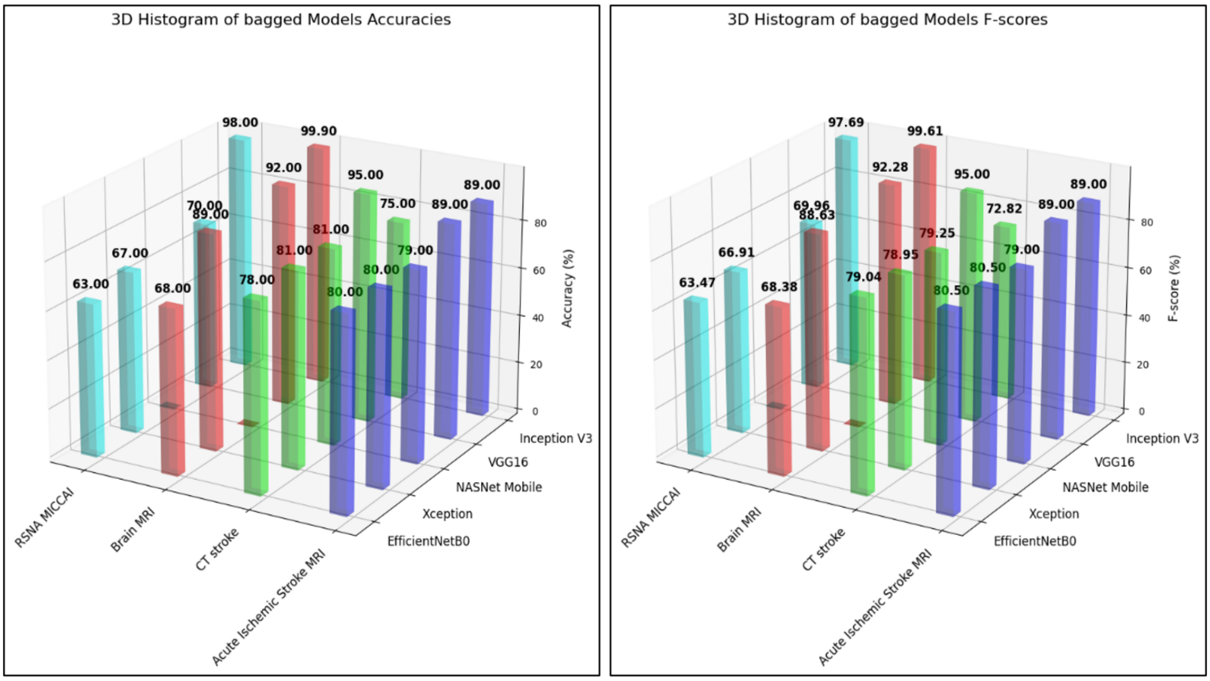

Figure 10: Performance comparison of different CNNs across multiple brain lesion datasets using ensemble deep learning based on the bagging approach.

InceptionV3's accuracy from 82.84% to 89.1%, and NASNetMobile's from 77.11% to 79.4%. On the RSNA MICCAI dataset, InceptionV3's accuracy increases significantly, jumping from 52% to 97.71%, and on the Brain Tumor MRI dataset, it improves from 92% to 99%. Similarly, NASNetMobile shows improvement on the Brain Stroke CT dataset, with accuracy rising from 77% to 81%.

For well-performing models like VGG16, Bagging slightly improves accuracy, increasing it from 88.56% to 89.3%. However, for EfficientNetB0 and Xception, Bagging decreases performance, with EfficientNetB0 dropping from 86.57% to 80.3%. These results indicate that Bagging improves feature extraction performance in models such as InceptionV3 and NASNetMobile but can introduce noise and reduce performance in well-optimized models like EfficientNetB0 and Xception, likely due to the homogeneity of models.

To address this issue, we introduce more heterogeneity in model performance. The next sub-section explores a stacking approach, combining multiple model types to leverage their diversity and create a more balanced, adaptable ensemble.

4.3.2. Stacking

In this experiment, the effectiveness of regression techniques is explored to derive continuous predictions from data that was originally structured for classification tasks, for this end, stacking is applied on two types of lesions (tumors and strokes) using MRI datasets: The Brain Tumor MRI dataset and the Acute Ischemic Stroke MRI dataset. We aimed to uncover latent relationships between outputs predictions that might be hidden in classification, moving beyond simple class labels to capture more complex patterns within the data.

The experiment demonstrates that stacking CNN models significantly improves performance in both classification and regression tasks. On the Brain Tumor MRI dataset, for classification, the stacked CNN achieved the highest accuracy (96.9%) compared to individual models like VGG16 and EfficientNetB0. When applied to regression, the stacking approach outperformed all other models, achieving near-perfect 99% accuracy, precision, recall, and F1-score. This suggests that stacking, particularly when using regression, enhances model generalization and predictive capability by uncovering latent relationships that single models may miss.

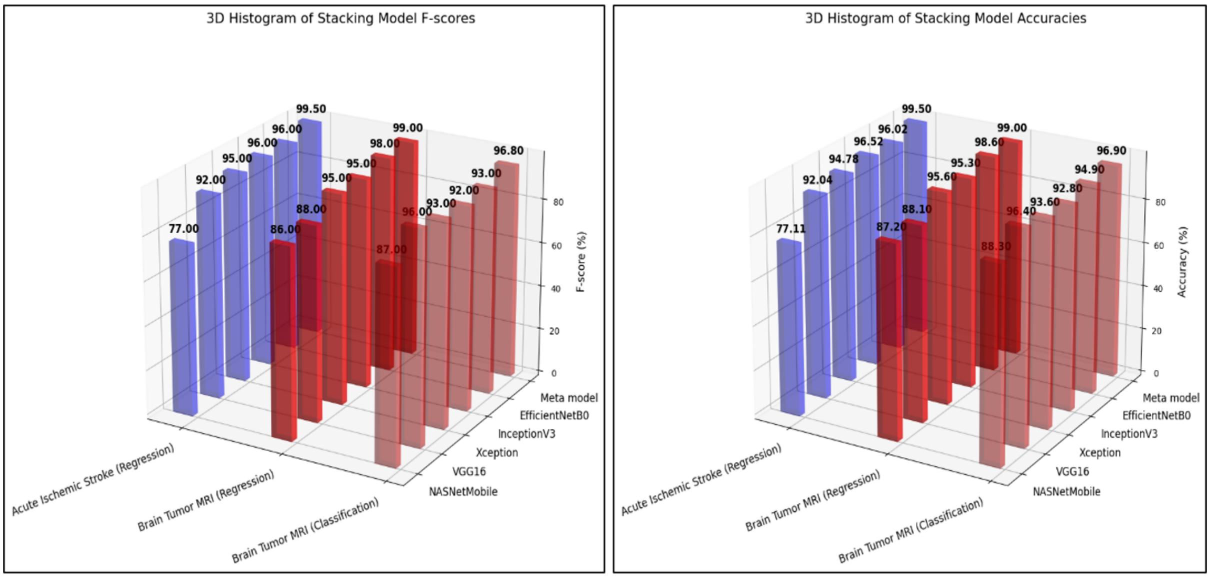

Figure 11: Performance comparison of stacking approach on Brain tumor MRI dataset and Acute Ischemic Stroke dataset.

On the Acute Ischemic Stroke MRI dataset, the performance of different models highlights the effectiveness of both classification and regression approaches. These results indicate that stacking CNNs with regression techniques offers superior capabilities in capturing complex, continuous patterns in both datasets. While ViTs and Swin Transformers perform well, especially on the Brain Tumor MRI dataset, they fall short compared to the stacking CNN with regression, which consistently delivers near-perfect results across both datasets. This suggests that stacking CNNs using regression techniques may offer superior capabilities for capturing complex, continuous relationships in medical imaging datasets, achieving the highest predictive performance compared to vision transformers.

Up to now, we evaluated parallel approaches, such as bagging and stacking, and demonstrated their effectiveness across brain lesion datasets. In the next sub-section, we will explore sequential approaches (Boosting: AdaBoost and XGBoost), focusing on how they leverage a sequence of models ‘instances to iteratively correct predictions and adapt to data variations for improved accuracy.

4.3.3. Boosting

- AdaBoost

VGG16 and Xception are the top-performing models, achieving high accuracy and F-score values (up to 99%) across both datasets, particularly excelling on the Brain Stroke CT Image dataset. EfficientNetB0 also performs well, especially on the Acute Ischemic Stroke MRI dataset, where it achieves 98%, though it shows a slight decline in performance on CT images. InceptionV3 maintains consistent and high performance, ranging between 95.8% and 98.26%. In contrast, NASNetMobile underperforms compared to the other models, with metrics around 83–84%, indicating that it may not be well-suited for these tasks.

- XGBoosT

To enhance model diversity for smaller datasets and improve overall performance of the ensemble model, alternatives like XGBoost are explored with low learning rate, particularly for Acute stroke MRI, Brain Stroke CT, and Brain Tumor MRI datasets.

4.3.4. Performance of top-models using our ensemble approaches

The results presented in Table **??** demonstrate that ensemble deep learning techniques consistently achieve top performance across all brain lesion datasets, particularly when using VGG16 and InceptionV3

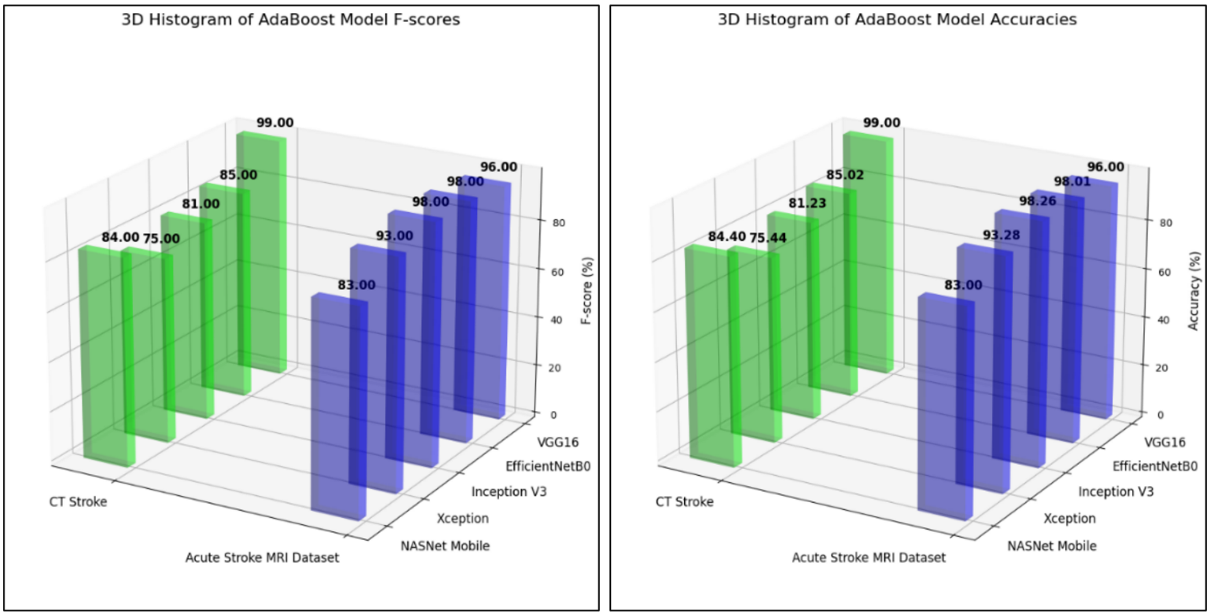

Figure 12: Performance comparison of different CNNs across multiple brain lesion datasets utilizing ensemble deep learning with AdaBoost approach.

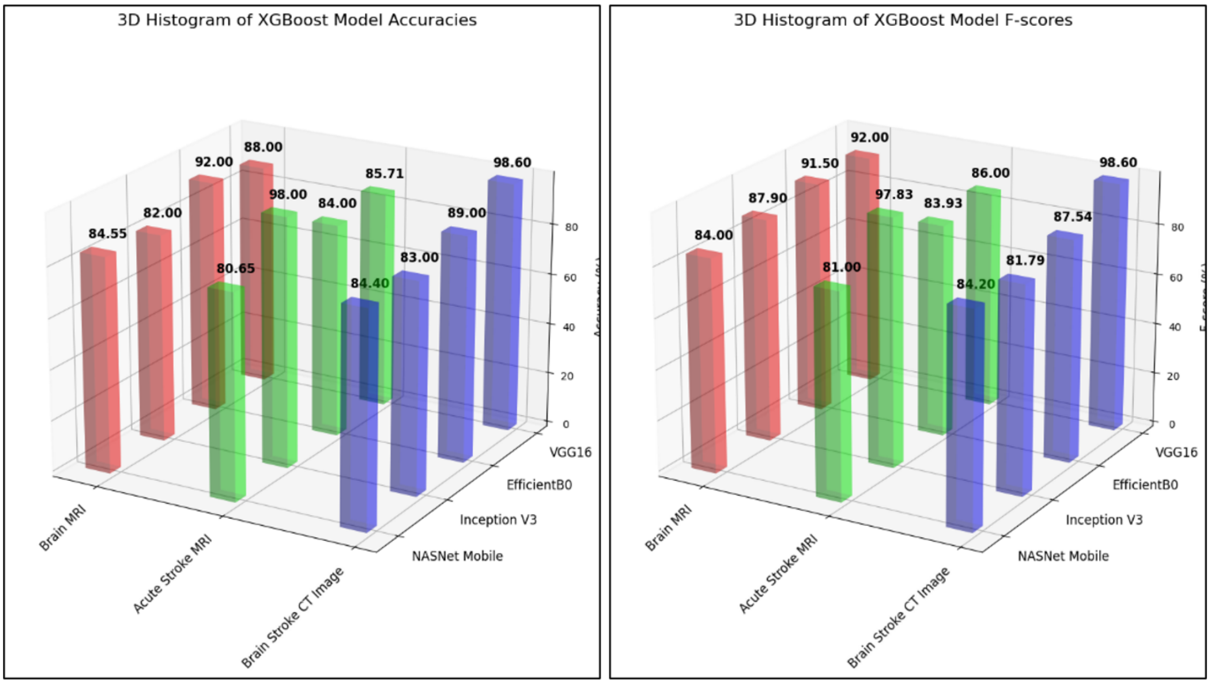

Figure 13: Performance comparison of different CNNs across multiple brain lesion datasets utilizing ensemble deep learning with XGBoost approach.

models. VGG16, when combined with methods such as AdaBoost, Bagging, and XGBoost, excels on the Brain Stroke CT dataset, achieving accuracy scores of 99%, 95%, and 98%, respectively. Similarly, InceptionV3 shows significant results with different ensemble methods: it achieves 98.26% accuracy with AdaBoost on the Acute Stroke MRI dataset, a near-perfect 99.9% accuracy with Bagging on the Brain Tumor MRI dataset, and a strong 98% accuracy on the RSNA MICCAI dataset. These findings highlight that the combination of ensemble techniques with VGG16 and InceptionV3 is particularly effective for various brain lesion classification tasks.

Table 2
Performance Comparison of Top Models Using Ensemble Approaches

Approach	Model	Accuracy	F-Score	Dataset
Bagging	InceptionV3	98%	97.69%	RSNA MICCAI
	InceptionV3	99.9%	99.61%	**Brain Tumor MRI**
	VGG16	95%	95%	**Brain Stroke CT**
Stacking	CNN (Reg)	94.20%	95%	Acute Stroke MRI
	CNN (Reg)	96.9%	96.8%	Brain Tumor MRI
AdaBoost	EfficientNetB0	98.01%	98%	Acute Stroke MRI
	InceptionV3	98.26%	98%	Acute Stroke MRI
	VGG16	99%	99%	**Brain Stroke CT**
XGBoost	VGG16	98.6%	98.6%	**Brain Stroke CT**
	InceptionV3	98%	97.83%	Acute Stroke MRI

4.4. Comparison and discussion

The study demonstrates that while ViTs consistently outperform traditional CNNs in feature extraction tasks across multiple brain lesion datasets, CNNs can significantly close this performance gap through the use of ensemble deep learning techniques such as Bagging, Stacking, and regression-based ensemble learning. ViTs excel at capturing global features in an image, whereas CNNs traditionally focus on local features. However, by incorporating ensemble learning—specifically Stacking and Bagging—CNNs can diversify their feature extraction capabilities, effectively introducing global feature understanding that is more characteristic of ViTs.

Stacking CNN models, enhanced with regression techniques, improves their ability to generalize across complex datasets. Regression helps by refining predictions and capturing continuous relationships in the data that are not fully exploited in classification tasks. As a result, stacked CNNs with regression achieved near-perfect performance, surpassing ViTs, with accuracies of 99% on the Brain Tumor MRI dataset and 99.5% on the Acute Ischemic Stroke MRI dataset. This level of performance exceeds the best results from ViTs, such as ViT L/16, which achieved 97.9% accuracy on the Brain Tumor MRI dataset and 93% on the Stroke dataset.

Moreover, the study shows that Bagging boosts the performance of underperforming models like InceptionV3 and NASNetMobile, allowing them to match or even surpass the performance of ViTs on certain datasets, such as the Brain Stroke CT dataset. For instance, Bagging helped InceptionV3 improve from 82.84% to 89.1% on the Acute Ischemic Stroke MRI dataset and from 52% to 97.71% on the RSNA MICCAI dataset, matching the performance of ViT models. However, while Bagging benefits these underperforming models, it may introduce noise in highly optimized models such as EfficientNetB0 and Xception, leading to a performance drop.

Further enhancements with AdaBoost have been particularly effective for CNN models such as VGG16 and InceptionV3, enabling them to outperform ViTs in some brain lesion classification tasks. For example, VGG16 achieves 99% accuracy on the Brain Stroke CT dataset, while InceptionV3 reaches 98.26% accuracy on the Acute Ischemic Stroke MRI dataset, both surpassing the top-performing ViTs. However, even with AdaBoost, models like NASNetMobile continue to fall short compared to ViTs.

Furthermore, the study reveals that boosting techniques like AdaBoost and XGBoost outperform ViTs on brain stroke lesion datasets such as Acute Ischemic Stroke MRI and Brain Stroke CT. Specifically, AdaBoost with VGG16 achieves the highest performance on Brain Stroke CT with 99% accuracy and F-score, while XGBoost with VGG16 performs strongly with 98.6% accuracy on the same dataset. Similarly, on the Acute Stroke MRI dataset, XGBoost with InceptionV3 (98% accuracy) and AdaBoost with VGG16 (96% accuracy) outperform vision transformers, which achieve only 93% accuracy with ViT L/16 on Acute Stroke MRI and 96.4% accuracy with SwinV2 on Brain Stroke CT.

In addition, Bagging with InceptionV3 outperforms other approaches for brain tumor lesion classifi-

Table 3
Performance Comparison of Our Top-Model Approaches on Different Datasets

Dataset	Top-Model	Accuracy	Approach
Acute Ischemic Stroke MRI	ViT L/16	93.00%	TL of ViTs
	InceptionV3 / VGG16	89.00%	Bagging CNNs
	Meta Model	99.50%	Stacking CNNs
	InceptionV3	98.00%	XGBoost CNN
	VGG16	96.00%	AdaBoost CNN
Brain Stroke CT	SwinV2	96.40%	TL of ViTs
	InceptionV3	95.00%	Bagging CNNs
	VGG16	98.60%	XGBoost CNN
	VGG16	**99.00%**	**AdaBoost CNN**
Brain Tumor MRI	ViT L/16	97.90%	TL of ViTs
	InceptionV3	**99.90%**	**Bagging CNNs**
	EfficientNetB0	92.00%	XGBoost CNN
	Classification Meta Model	99.00%	Stacking CNNs
	Regression Meta Model	97.00%	Stacking CNNs
RSNA MICCAI	ViT B/16	92.23%	TL of ViTs
	InceptionV3	**98.00%**	**Bagging CNNs**

cation, demonstrating strong generalization compared to ViTs. Specifically, Bagging with InceptionV3 achieves a remarkable 99.9% accuracy and F-score on the Brain Tumor MRI dataset and 98% accuracy with a 97.69% F-score on the RSNA MICCAI dataset, indicating near-perfect predictive performance. In contrast, ViT L/16 achieves 97.9% accuracy and a 98% F-score on the Brain Tumor MRI dataset and 92.23% accuracy with a 92.24% F-score on RSNA MICCAI, suggesting that while ViTs are reliable, they do not reach the high generalization and stability levels observed with Bagging.

This highlights Bagging's capability to stabilize and enhance model accuracy, making it highly effective for complex medical imaging tasks where Transformers fall short in achieving comparable accuracy and robustness.

Moreover, our experiments demonstrate that the proposed approaches outperform previous state-of-the-art methods across multiple datasets (see Table **??**). For the Brain Tumor MRI dataset, the XGBoost + VGG16 model achieved an accuracy of 98.6%, surpassing the performance of DNN (96.21%), Random Forest (95.30%), and GA + BiLSTM (96.45%). Similarly, for the Acute Ischemic Stroke MRI dataset, our XGBoost + InceptionV3 model achieved 98% accuracy, exceeding the performance of ResNet101 and DenseNet201 (97.5%).

In conclusion, while ViTs set a high benchmark for performance in medical imaging, combining CNNs with ensemble techniques such as stacking-based regression, bagging, or boosting allows CNNs to match or even exceed the performance of ViTs. This finding is particularly significant as these techniques improve the generalization and feature extraction capabilities of CNNs, enabling them to compete effectively with state-of-the-art ViTs on challenging datasets.

5. Conclusion

This study highlights the effectiveness of ensemble deep learning techniques in enhancing the performance of CNNs to match or even surpass ViTs in medical imaging tasks. By employing methods such as stacking, bagging, and boosting, CNNs significantly improved their feature extraction capabilities, achieving near-perfect accuracy across various brain lesion datasets. These findings demonstrate that, when ensemble approaches are carefully applied, CNNs can serve as a powerful and competitive alternative to state-of-the-art ViTs in complex medical imaging classification challenges. However, ViTs maintain an advantage on more complex datasets, such as RSNA MICCAI, where their ability to capture

Table 4
Performance Comparison of Our Approaches Against State-of-the-Art Classification Techniques Across Multiple Datasets

Ref	Year	Method	Accuracy	Dataset
[2]	2021	DNN	96.21%	Brain Tumor MRI
[5]	2021	RF	95.30%	Brain Tumor MRI
[10]	2023	GA + BiLSTM	96.45%	Brain Tumor MRI
Ours	2024	**XGBoost + VGG16**	**98.60%**	Brain Tumor MRI
Ours	2024	ViT L/16	97.90%	Brain Tumor MRI
[13]	2022	ResNet101, DenseNet201	97.50%	Acute Ischemic Stroke MRI
Ours	2024	XGBoost + InceptionV3	98.00%	Acute Ischemic Stroke MRI
Ours	2024	**AdaBoost + Fine-Tuning VGG16**	**99.00%**	Acute Ischemic Stroke MRI
Ours	2024	ViT L/16	93.00%	Acute Ischemic Stroke MRI
[33]	2023	Stacking CNN	92.00%	RSNA MICCAI
Ours	2024	Bagging (InceptionV3)	98.00%	RSNA MICCAI
Ours	2024	ViT B/16	92.23%	RSNA MICCAI

intricate features demonstrates their superior performance.

6. Declaration on Generative AI

This paper is written entirely by the authors without any use of GenAI tools and services.

References

[1] G. Sailasya, G. L. A. Kumari, Analyzing the performance of stroke prediction using ml classification algorithms, International Journal of Advanced Computer Science and Applications 12 (2021).

[2] A. Phaphuangwittayakul, Y. Guo, F. Ying, A. Y. Dawod, S. Angkurawaranon, C. Angkurawaranon, An optimal deep learning framework for multi-type hemorrhagic lesions detection and quantification in head ct images for traumatic brain injury, Applied Intelligence (2022) 1–19.

[3] S. Dev, H. Wang, C. S. Nwosu, N. Jain, B. Veeravalli, D. John, A predictive analytics approach for stroke prediction using machine learning and neural networks, Healthcare Analytics 2 (2022) 100032.

[4] S. Gudadhe, A. Thakare, A. M. Anter, A novel machine learning-based feature extraction method for classifying intracranial hemorrhage computed tomography images, Healthcare Analytics 3 (2023) 100196.

[5] B. Akter, A. Rajbongshi, S. Sazzad, R. Shakil, J. Biswas, U. Sara, A machine learning approach to detect the brain stroke disease, in: 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2022, pp. 897–901.

[6] A. Tursynova, et al., Deep learning-enabled brain stroke classification on computed tomography images, Comput. Mater. Contin 75 (2023) 1431–1446.

[7] M. Yeo, et al., Evaluation of techniques to improve a deep learning algorithm for the automatic detection of intracranial haemorrhage on ct head imaging, European Radiology Experimental 7 (2023) 17.

[8] M. Gupta, P. Meghana, K. H. Reddy, P. Supraja, Predicting brain stroke using iot-enabled deep learning and machine learning: Advancing sustainable healthcare, in: International Conference on Sustainable Development through Machine Learning, AI and IoT, 2023, pp. 113–122.

[9] L. Cortés-Ferre, M. A. Gutiérrez-Naranjo, J. J. Egea-Guerrero, S. Pérez-Sánchez, M. Balcerzyk, Deep learning applied to intracranial hemorrhage detection, Journal of Imaging 9 (2023) 37.

[10] M. A. Saleem, et al., Innovations in stroke identification: A machine learning-based diagnostic

model using neuroimages, IEEE Access 12 (2024) 35754–35764. doi:10.1109/ACCESS.2024.3369673.

[11] B. Kaya, M. Önal, A cnn transfer learning-based approach for segmentation and classification of brain stroke from noncontrast ct images, International Journal of Imaging Systems and Technology 33 (2023) 1335–1352.

[12] R. Raj, J. Mathew, S. K. Kannath, J. Rajan, Strokevit with automl for brain stroke classification, Engineering Applications of Artificial Intelligence 119 (2023) 105772.

[13] B. Tasci, I. Tasci, Deep feature extraction based brain image classification model using preprocessed images: Pdrnet, Biomedical Signal Processing and Control 78 (2022) 103948.

[14] M. N. Faiz, T. Badriyah, S. F. Kusuma, Classification of intracranial hemorrhage based on ct-scan image with vision transformer (vit) method, in: 2024 International Electronics Symposium (IES), 2024, pp. 454–459.

[15] K.-Y. Lee, C.-C. Liu, D. Y.-T. Chen, C.-L. Weng, H.-W. Chiu, C.-H. Chiang, Automatic detection and vascular territory classification of hyperacute staged ischemic stroke on diffusion weighted image using convolutional neural networks, Scientific Reports 13 (2023) 404.

[16] H. Jo, et al., Combining clinical and imaging data for predicting functional outcomes after acute ischemic stroke: an automated machine learning approach, Scientific Reports 13 (2023) 16926.

[17] M. N. Faiz, T. Badriyah, S. F. Kusuma, Classification of intracranial hemorrhage based on ct-scan image with vision transformer (vit) method, in: 2024 International Electronics Symposium (IES), 2024, pp. 454–459.

[18] D. Ma, M. Wang, A. Xiang, Z. Qi, Q. Yang, Transformer-based classification outcome prediction for multimodal stroke treatment, arXiv preprint arXiv:2404.12634 (2024).

[19] Z. A. Samak, P. Clatworthy, M. Mirmehdi, Transop: transformer-based multimodal classification for stroke treatment outcome prediction, in: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), 2023, pp. 1–5.

[20] O. Katar, O. Yildirim, Y. Eroglu, Vision transformer model for efficient stroke detection in neuroimaging, in: 2023 4th International Informatics and Software Engineering Conference (IISEC), 2023, pp. 1–6.

[21] M. Nickparvar, Brain tumor mri dataset, Kaggle, 2021. URL: https://doi.org/10.34740/KAGGLE/DSV/2645886 [Accessed: 30-Sep-2024].

[22] Z. H. N. Al-Azzwi, A. N. Nazarov, Brain tumor classification based on improved stacked ensemble deep learning methods, Asian Pacific Journal of Cancer Prevention: APJCP 24 (2023) 2141.

[23] M. Jha, R. Gupta, R. Saxena, A framework for in-vivo human brain tumor detection using image augmentation and hybrid features, Health Information Science and Systems 10 (2022) 23.

[24] Y. Modaresnia, F. A. Torghabeh, S. A. Hosseini, Efficientnetb0's hybrid approach for brain tumor classification from mri images using deep learning and bagging trees, in: 2023 13th International Conference on Computer and Knowledge Engineering (ICCKE), Iran, 2023, pp. 234–239.

[25] M. F. Khan, P. Khatri, S. Lenka, D. Anuhya, A. Sanyal, Detection of brain tumor from the mri images using deep hybrid boosted based on ensemble techniques, in: 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1464–1467.

[26] U. Baid, et al., The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification, arXiv 2107.02314 (2021). URL: https://www.kaggle.com/datasets/jonathanbesomi/rsna-miccai-png/data [Accessed: 30-Sep-2024].

[27] S. Faghani, B. Khosravi, M. Moassefi, G. M. Conte, B. J. Erickson, A comparison of three different deep learning-based models to predict the mgmt promoter methylation status in glioblastoma using brain mri, Journal of Digital Imaging 36 (2023) 837–846.

[28] N. Saeed, S. Hardan, K. Abutalip, M. Yaqub, Is it possible to predict mgmt promoter methylation from brain tumor mri scans using deep learning models?, in: International Conference on Medical Imaging with Deep Learning, Zürich, Switzerland, 2022, pp. 1005–1018.

[29] D. Alexey, An image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint 2010.11929 (2020).

[30] Z. Liu, et al., Swin transformer v2: Scaling up capacity and resolution, in: Proceedings of the

IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 12009–12019.

[31] S. Mehta, M. Rastegari, Separable self-attention for mobile vision transformers, arXiv preprint 2206.02680 (2022).

[32] Brain stroke ct image dataset, https://www.kaggle.com/datasets/afridirahman/brain-stroke-ct-image-dataset, ???? Accessed: 30-Sep-2024.

[33] M. Khaled, D. Gaceb, F. Touazi, C. A. Aouchiche, Y. Bellouche, A. Titoun, New cnn stacking model for classification of medical imaging modalities and anatomical organs on medical images, in: 6th International Conference on Informatics Data-Driven Medicine, Bratislava, Slovakia, 2023, pp. 174–188.