Brain Tumor Classification from MRI Images Using a Custom CNN, XceptionNet, EfficientNet-B0, and Vision Transformer with Explainable AI Analysis

Yujin Jeon
Software Engineering M.Sc.
Univ. of Europe for Applied Sciences
Potsdam 14469, Germany
yujin.jeon.developer@gmail.com

Raja Hashim Ali
Department of Business
Univ. of Europe for Applied Sciences
Potsdam 14469, Germany
hashim.ali@ue-germany.de

***Abstract*—Brain tumors are one of the fatal diseases in which early diagnosis and accurate classification greatly influence a patient's prognosis and survival rate. Recently, deep learning based image analysis technology has attracted attention as a key research field that dramatically improves the automatic classification and diagnosis accuracy of brain tumors in the medical field and simultaneously increases clinical reliability and efficiency. Existing brain tumor detection studies have focused on limited scale MRI datasets and single model performance improvements, lacking generalization, explainability, and in-depth analysis of ensemble effects of different models in real world clinical settings. Therefore, this study systematically compares and analyzes ViT, XceptionNet, EfficientNet-B0, and basic CNN with various evaluation indicators and XAI based visualizations to present practical grounds and guidelines for building a reliable automatic brain tumor diagnosis system. Using Kaggle's Brain Tumor MRI dataset, this study divided the data into training and test sets after systematic preprocessing processes such as deduplication, resize, grayscale transformation, Z-score normalization, and data augmentation. After that, the four basic CNN, XceptionNet, EfficientNet-B0, and ViT models were trained under the same conditions, and the classification performance and explainability of each model were quantitatively compared and analyzed through major evaluation indicators and Grad-CAM based visual interpretation. Comparing the brain tumor classification performance of the four models in this study, XceptionNet and EfficientNet-B0 achieved excellent accuracy of 0.98 or higher on all key indicators, and ViT also showed high performance but somewhat lower levels. EfficientNet-B0 achieves high accuracy with relatively few parameters and computational amount, demonstrating both practical efficiency and practicality. In particular, XceptionNet and EfficientNet-B0 had little misclassification in most classes, and the results of explainability analysis using Grad-CAM can increase clinical reliability, such as paying attention to actual tumor sites. This study systematically compares the performance and efficiency of various deep learning models (CNN, XceptionNet, EfficientNet-B0, ViT) and presents practical grounds and guidelines for the development of reliable automatic brain tumor diagnosis systems through Grad-CAM based explainability analysis.**

***Index Terms*—Brain Tumor, Deep Learning, Convolutional Neural Networks (CNN), XceptionNet, EfficientNet, Vision Transformers (ViT), Explainable AI (XAI), Magnetic Resonance Imaging (MRI)**

I. Introduction

Brain tumors are malignant or benign tumors formed by the rapid growth of abnormal cells in normal tissues of the brain [1]. If it is not accurately diagnosed early, it can be a fatal threat to life, and the patient's quality of life is also rapidly deteriorated. Currently, brain tumor diagnosis is mainly done through magnetic resonance imaging (MRI), and MRI plays an important role in clinical practice by providing high resolution brain tissue images without the use of radiation [2]. However, existing MRI image analysis often relies on subjective judgment from specialists, so there is a limit to the accuracy or speed of reading [3]. To overcome these limitations, research on automatic detection of brain tumors using deep learning technology has been actively conducted recently [4]. In particular, deep learning based medical image analysis technology shows high accuracy beyond human reading ability and greatly contributes to diagnostic assistance and increased efficiency of medical staff [5]. Deep learning models can identify microscopic tumor patterns in MRI images, which significantly help in early diagnosis and treatment planning. In addition, the ability to automatically classify the location, size, and type of tumors provides an important basis for establishing personalized treatment strategies. These technologies have the potential to significantly improve access to healthcare, enabling remote diagnosis and treatment planning even in areas where healthcare is scarce. Therefore, MRI image based deep learning brain tumor detection research has become an essential and urgent research challenge in the medical field [6].

Recently, deep learning technology is rapidly spreading beyond traditional image processing techniques in the field of medical image analysis [7]. Deep learning is a high performance machine learning approach based on artificial neural networks to self learn data and automatically extract complex patterns. In particular, Convolutional Neural Networks (CNNs) demonstrate excellent performance in extracting and classifying features from images effectively. These CNN based deep learning technologies can detect the location and size of tumors in brain MRI images with high accuracy, which

is highly utilized in clinical use. Furthermore, CNN models can quickly process vast amounts of medical imaging data, minimizing the likelihood of errors due to subjective judgments by specialists. Recently, new deep learning models such as Vision Transformer (ViT) have emerged to compensate for the structural limitations of CNNs, further improving medical image analysis performance [8]. Transformer based models can more effectively capture microscopic differences between tumors and normal tissues within MRI images using a self attention mechanism. Advances in these deep learning techniques not only increase the accuracy of brain tumor diagnosis, but also significantly reduce the diagnosis time, which greatly contributes to improving the early treatment possibility of patients [9]. Furthermore, deep learning models provide accuracy beyond human reading capabilities, helping medical staff make more reliable judgments. In conclusion, deep learning is positioning itself as a technology that presents the most important and practical solutions in the field of MRI image-based brain tumor detection [10].

A. Gap Analysis

Deep learning research on brain tumor detection has been actively conducted to date, but there are still several unresolved technical limitations [11]. First of all, most studies are based on relatively limited scale MRI imaging datasets, resulting in insufficient performance validation in various real world clinical settings. Due to this, the generalization ability of the model to overcome differences in image quality, which depends on the hospital or the imaging device, has not been sufficiently evaluated. In particular, research using explainable artificial intelligence (XAI) technology to help medical staff trust the diagnosis results in clinical settings is still in its infancy [12]. Explainable models support medical staff to make more accurate and faster decisions by visually providing the basis for diagnosis, but in-depth research using them has been insufficient to date. In addition, most studies focus on improving the performance of a single deep learning model, so there has not been enough research on ensemble techniques that combine the advantages between different models. While these ensemble techniques can complement the limitations of individual models, specific application cases and performance evaluations in the medical imaging field are still lacking. In addition, studies to evaluate and improve the efficiency and speed of deep learning models in real world clinical settings are also insufficient, requiring practicality evaluation for clinical application. After all, future studies need to focus on more extensive datasets and performance validation in real-world clinical settings, improving the explainability of models, and developing efficient ensemble techniques. This is expected to further increase the clinical utility of deep learning-based brain tumor detection technology.

B. Research Questions

1) RQ1: How do the ViT, XceptionNet, EfficientNet-B0, and a basic CNN differ in terms of accuracy, precision, recall, and other key performance metrics when detecting brain tumors from MRI images?
2) RQ2: How does model complexity (such as number of parameters and computational cost) or architectural choice (CNN vs Transformer) influence the performance and efficiency of these models in brain tumor detection?
3) RQ3: Which tumor types are most frequently misclassified by each model, and what patterns or trends can be observed from the confusion matrices and class-wise metrics?
4) RQ4: How does the accuracy of the basic CNN model change as the number of training epochs increases (in steps of 10 up to 100), and what can we learn from this trend about the training needs of a simple CNN for brain tumor detection?
5) RQ5: What insights can be gained from Grad-CAM visualizations about the decision-making process of each CNN-based model for brain tumor detection?

C. Problem Statement

This work addresses the problem of accurate and efficient detection of brain tumors using deep learning models in brain MRI images. Specifically, this study analyzes and compares the performance differences of Vision Transformer (ViT), XceptionNet, EfficientNet-B0, and the CNN model using key evaluation metrics such as accuracy, precision, recall, and f1-score. Furthermore, we systematically evaluate the complexity of the model which includes the impact of the number of parameters and computational cost of the model on brain tumor detection performance and efficiency. Through this, we want to clarify what advantages and limitations CNN based and Transformer based models have in the actual brain tumor diagnosis process, respectively. In addition, the frequency and cause of each model's incorrect classification of specific tumor types (glioma, meningioma, pituitary tumor) are analyzed through confusion matrices and class specific metrics to deeply identify misclassification patterns. By observing the accuracy changes according to the number of training sessions (epoch) of the CNN model step by step, we also check the appropriate training conditions for the simple CNN structure to achieve sufficient performance in brain tumor detection. Furthermore, we apply explainable artificial intelligence (XAI) techniques such as Grad-CAM to visually analyze how each CNN based model's decision making process takes place, and evaluate how reliable the insights derived from them are in real diagnostic environments. By solving these problems, we intend to provide practical grounds and guidelines for building a reliable and explainable automatic brain tumor diagnosis system in the medical field. In addition, the analysis methods and results presented through this study are expected to be important reference materials for conducting similar medical imaging analysis studies in the future. As a result, this study seeks to complement the limitations of existing studies in the field of MRI imaging based brain tumor detection and contribute to creating practical and clinical value.

D. Novelty of this study

This study presents an original approach to overcome the limitations of existing brain tumor detection studies and to increase practical clinical applicability. In particular, the integrated analysis of the performance, structure, and interpretability of various deep learning models differentiates them from previous studies. The Novelty of this study can be summarized in five ways below.

- Comprehensive model performance evaluation: The performance of various deep learning models (ViT, XceptionNet, EfficientNet-B0, basic CNN) was systematically compared and evaluated in MRI based brain tumor detection problems.
- Impact of model complexity and architecture: The effect of structural complexity, such as the number of parameters and the amount of computation of the model, on actual diagnostic efficiency and performance was quantitatively analyzed.
- Tumor type misclassification analysis: It was specifically identified using a confusion matrix and class specific metrics to see what type of misclassification each model tends to show by tumor type.
- Optimal training conditions for CNN: By analyzing the performance change according to the learning epoch of the basic CNN model, the optimal learning conditions that can be used in actual clinical practice were derived.
- Explanability via Grad-CAM visualization: Through visual interpretation using Grad-CAM, the judgment basis of CNN based models is presented intuitively, increasing the reliability and explainability of the diagnosis.

E. Significance of Our Work

In this study, we compared and analyzed various deep learning models such as ViT, XceptionNet, EfficientNet-B0, and basic CNN for MRI based brain tumor detection. Data preprocessing, learning, and evaluation processes were consistently applied to all models to ensure fair comparisons. Key performance indicators such as accuracy, precision, recall, and f1-score were calculated for each model, and ViT, XceptionNet, and EfficientNet-B0 showed good performance overall. In particular, EfficientNet-B0 has been shown to achieve high accuracy and efficiency simultaneously despite its lightweight structure. XceptionNet also recorded very high accuracy with its deep network structure and effective feature extraction ability. Comparing model complexity with computational cost, we find that lightweight models such as EfficientNet-B0 can be a practical alternative in environments with limited computational resources. Through the analysis of misclassification patterns by tumor type, we found that certain models tended to cause confusion in some classes. As a result of observing the performance change according to the increase in the learning epoch of the basic CNN model, it was found that the performance improvement stagnated after a certain epoch. Visual explanations using Grad-CAM confirmed the model's tendency to focus on the site where the tumor is actually located, suggesting the possibility that medical staff can trust the model's judgment basis. As a result of comparing various models and methods overall, we presented practical criteria for selecting optimal models and learning conditions based on clinical environment and resources. As a result, this study provides a systematic analysis and discussion that considers both practicality and reliability in the field of MRI image based brain tumor detection.

II. Literature Review

Various deep learning based image classification techniques are being used in medical image analysis fields such as brain tumor classification. Convolutional neural networks (CNNs), the most basic structure, are widely used to automatically extract key features of medical images as they can effectively capture spatial patterns of images. The Xception network is an extension of the existing CNN structure, and by applying the depthwise separable convolution method, efficient and powerful feature extraction is possible with fewer parameters. These Xception are frequently chosen in recent studies that simultaneously pursue high accuracy and computational efficiency. EfficientNet-B0 is a lightweight model that introduces compound scaling techniques to optimize the size and accuracy of the model in a balanced way, and it is characterized by significantly reducing the amount of computation and the number of parameters while maintaining high performance. This allows EfficientNet-B0 to be applied practically even in resource constrained situations, such as in real world clinical settings. Vision Transformer (ViT) is a recent transformer based image classification model that divides images into patch units to learn the correlation of each patch through the attention mechanism. Unlike CNN, ViT can effectively utilize global contextual information, showing excellent classification performance even in medical images with complex patterns. These various models have their respective structural advantages and limitations, and in practical applications, the optimal combination is selected in consideration of data characteristics, computational resources, and interpretability. Recent studies have applied these models alone or in ensemble form to simultaneously increase the accuracy and reliability of brain tumor classification.

A. Convolutional Neural Networks

Yoon *et al.* (2025) [13] demonstrated that a hybrid ensemble combining Xception and parallel deep CNNs achieved a 99.09% accuracy for four-class brain tumor classification, surpassing other state-of-the-art models such as VGG19 and ResNet152V2. This highlights the significant impact of CNN architectures and their ensembles in enhancing classification performance for complex multi-class medical imaging tasks. Bairagi *et al.* (2023) [14] compared various CNN architectures, including AlexNet, VGG-16, and GoogLeNet, and achieved a high accuracy of 98.67% with AlexNet. The study emphasized the superior efficiency and accuracy of automated CNN-based classification over manual MRI interpretation. Saeedi *et al.* (2023) [15] performed multiclass brain tumor classification

using a 2D CNN with eight convolutional and four pooling layers, reporting 96.47% accuracy and an AUC close to 1. Additionally, they showed that the 2D CNN offered both simplicity and high performance compared to auto-encoder and other machine learning techniques. Vankdothu *et al.* (2022) [16] proposed a hybrid CNN-LSTM model, achieving 92% accuracy and outperforming standalone CNN and RNN models. Their work indicated that a CNN combined with LSTM can be applied to real-time, IoT-based diagnostic systems.

B. XceptionNet

Verma *et al.* (2024) [17] reported that a fine-tuned Xception model using the Kaggle dataset achieved approximately 98% training accuracy along with high precision, recall, and F1-score. This study demonstrated that the Xception model can effectively classify multiple types of brain tumors, offering the potential to improve clinical diagnostic accuracy. Hossain *et al.* (2023) [18] utilized the Xception architecture as one of several transfer learning backbones for brain tumor classification and reported an accuracy of 94.5% on a dataset of 3,264 MRI images. Their study showed that while Xception achieved strong performance, ensemble models such as IVX16 outperformed individual backbones, indicating that model combination can further enhance classification accuracy in this domain. Kushwaha *et al.* (2022) [19] developed an ensemble CNN model combining VGG16, AlexNet, Inception Net, and Xception Net, demonstrating consistently high accuracy across multiple MRI datasets. In this study, Xception excelled in feature extraction and classification, particularly for Schwannoma and other disease types. Yoon *et al.* (2025) [13] showed that a hybrid ensemble model combining Xception with a parallel deep CNN outperformed state-of-the-art models like VGG19 and ResNet152V2, achieving 99.09% classification accuracy. The results indicate that Xception, when integrated with other networks, provides outstanding classification and generalization performance for four-class brain tumor classification tasks.

C. EfficientNet

Shah *et al.* (2022) [20] demonstrated that a fine-tuned EfficientNet-B0 model, when combined with image enhancement and data augmentation, achieved an overall accuracy of 98.87%, outperforming VGG16, InceptionV3, Xception, ResNet50, and InceptionResNetV2. This highlights the model's superior classification and detection capabilities for brain tumors, although further details on generalization were not discussed. Zulfiqar *et al.* (2023) [21] showed that fine-tuning pre-trained EfficientNetB2, along with data augmentation and Grad-CAM visualization, resulted in a test accuracy of 98.86% for three-class brain tumor classification. The study also provided visual evidence of the model's attention through Grad-CAM but did not fully explore comparisons with non-EfficientNet models or external datasets. Mahesh *et al.* (2024) [22] reported that integrating EfficientNetB0 with explainable AI techniques like Grad-CAM yielded an accuracy of 98.72%, with precision and recall exceeding 97% for all classes. This approach improved both the classification reliability and interpretability, offering enhanced trust for clinical use, though real-world validation is still needed. Tariq *et al.* (2025) [23] showed that EfficientNetV2 achieved a 95% accuracy with F1, precision, and recall all reaching 0.96 for brain tumor classification on a large-scale MRI dataset. Their results indicate that EfficientNetV2 provides robust and reliable multi-class classification performance, outperforming ViT in this context and further improving diagnostic accuracy when used as part of an ensemble, though real-world clinical validation remains necessary.

D. Vision Transformer (ViT)

Hossain *et al.* (2023) [18] compared several transfer learning and deep learning models for brain tumor classification, including ViT, and found that the proposed IVX16 ensemble achieved the highest accuracy of 96.94% among all evaluated methods. This work highlights the potential of ViT models in multi-class brain tumor detection when combined with ensemble strategies and explainable AI. Poornam *et al.* (2024) [24] proposed the VITALT system, which incorporates ViT with attention, S-BiFPN, linear transformation module, and soft-quantization, achieving classification accuracy between 98.82% and 99.15% across four benchmark datasets. Their results demonstrate that ViT-based systems, when enhanced with multi-scale feature fusion and attention mechanisms, can deliver highly reliable and consistent results in medical imaging applications. Wang *et al.* (2024) [25] introduced RanMerFormer, a method based on a pre-trained Vision Transformer backbone with token merging and a randomized vector functional-link head, reaching an impressive accuracy of 99.77% on their dataset. This approach illustrates that optimizing the efficiency and expressiveness of ViT architectures enables rapid and precise brain tumor classification. Tariq *et al.* (2025) [23] utilized both EfficientNetV2 and ViT for multi-class brain tumor classification, reporting that ViT alone achieved 90% accuracy and that a geometric mean ensemble of the two models further increased accuracy to 96%. These findings suggest that integrating ViT with convolutional models can enhance diagnostic performance and support comprehensive, automated brain tumor detection.

Please refer to the literature review summaries in the table I.

III. Methodology

The entire methodology of this study was conducted in the order of data collection, preprocessing and refining, design and learning of various deep learning models, and model evaluation and comparison. After collecting images from the Brain Tumor MRI Dataset, duplicate images were removed and preprocessed such as Z-score normalization and resize. After that, four models, CNN, XceptionNet, EfficientNet-B0, and ViT, were trained under the same conditions, and the performance of each model was evaluated by various indicators such as accuracy, precision, recall, and f1-score to select the optimal brain tumor classification model. (shown in Fig 1).

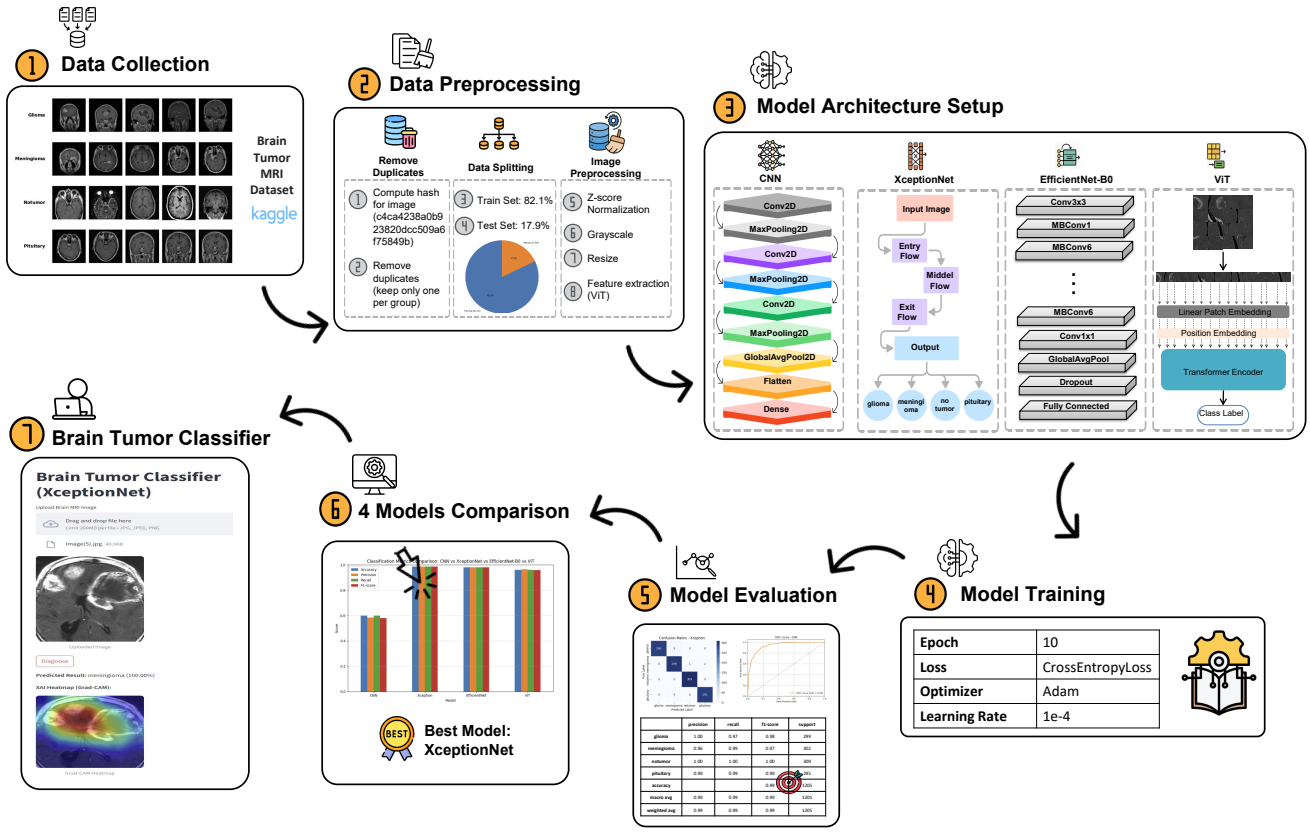

Fig. 1. The entire workflow of this study. After data collection, it was divided into data for learning and testing through preprocessing processes such as removing duplicate images and normalizing. After that, various deep learning models (CNN, XceptionNet, EfficientNet-B0, and ViT) were designed and learned, and the performance of each model was evaluated through various indicators.

A. Dataset

The dataset used in this study is the Brain Tumor MRI Dataset provided by Kaggle, a public dataset contributed by Masoud Nickparvar. The original data consists of a total of 7,023 brain MRI images, but duplicate images were identified and removed during the preprocessing process, and finally trained and verified using 6,726 images. This dataset consists of four classes: ”glioma”, ”meningioma”, ”no tumor” and ”pituitary”. Each image was a two dimensional scan image, and all images were preprocessed to a 224×224 resolution for consistency of analysis. In addition, each pixel value was normalized between 0 and 1 to improve the learning stability and performance of the model. The data were divided into training, validation, and testing so that the generalization performance of the model could be objectively evaluated. Data augmentation techniques such as rotation and flip were applied to secure data diversity. Patient personal information is provided as images of completely removed anonymization and is used only for research purposes. By consistently applying data preprocessing, normalization, and augmentation processes, the performance of several deep learning models could be compared and evaluated fairly. In conclusion, this study contributed to the development of a reliable automatic brain tumor diagnostic model by systematically preprocessing and utilizing the Kaggle public dataset. (shown in Fig 2)

B. Detailed Methodology

The first step in this study is the data collection and preprocessing process. First of all, using the Brain Tumor MRI Dataset provided by Kaggle, duplicate images were identified and removed from 7,023 original images, and only 6,726 images were finally used for analysis. After that, each image was resized to a size of 224×224 and converted to grayscale to remove unnecessary color information. For the stability and consistency of model learning, the pixel values were standardized by applying the Z-score normalization technique. By applying rotation and inversion as a data augmentation technique, the model can learn various types of brain tumors. The preprocessed and augmented images were divided into learning (82.1%) and testing (17.9%). Deduplication was done by calculating the image hash value, leaving only one image in each group. This thorough preprocessing and data management process played an important role in enhancing the reliability of model learning and evaluation. This step is a foundational task to ensure the quality of the data, and to provide optimized

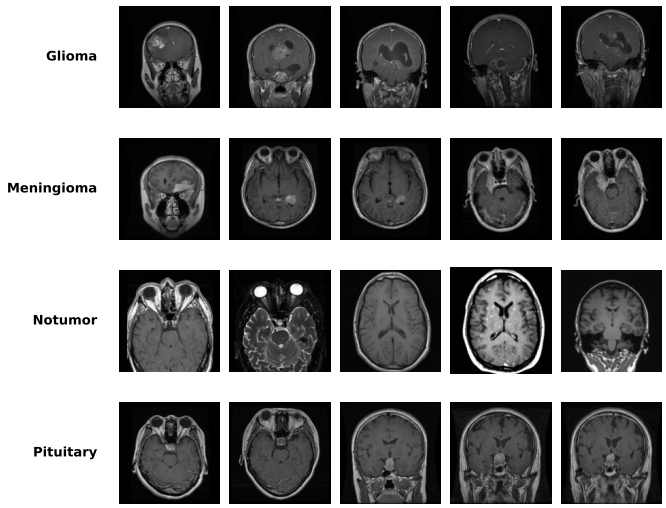

Fig. 2. Sample image of Brain Tumor MRI Dataset released by Kaggle. Each image is based on a brain MRI scan and is classified into four classes: glioma, meningioma, no-tumor, and pituitary. In this study, the brain tumor detection performance of various deep learning models was evaluated using the corresponding dataset. (Dataset source: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset/data)

inputs for learning subsequent deep learning models. As a result, it is designed to minimize unnecessary variables in the data preprocessing process and to make model performance comparison fair.

In the second step, the process of designing and learning the structure of various deep learning models took place. In this study, a total of four models were selected: basic CNN, XceptionNet, EfficientNet-B0, and Vision Transformer (ViT). Each model is designed to automatically extract key features from input images and classify the presence and type of brain tumors. CNN, XceptionNet, EfficientNet-B0 and others use convolutional and pooling layers repeatedly to effectively identify spatial patterns within images. The ViT model divides the image into patch units and then inputs each patch information into the transformer encoder to learn global features. All models set CrossEntropyLoss as a loss function and Adam as an optimizer, and the learning rate was unified to 1e-4. Learning was conducted for 10 epochs, and the accuracy of learning and verification of the model was recorded for each epoch. In the model structure design, the role and hyperparameters of each layer were standardized to unify the experimental conditions. This step allows a quantitative analysis of how structural differences in multiple models affect actual brain tumor detection performance.

The final step is model evaluation, performance comparison, and explainability analysis. For all models, a test dataset was used to calculate key performance indicators such as accuracy, precision, recall, and f1-score. In addition, class specific performance evaluation and confusion matrix analysis were used to closely observe which type of brain tumor each model was misclassified. In this process, XceptionNet showed the highest overall performance, and EfficientNet-B0 and ViT also showed excellent results. Grad-CAM was used to visually analyze which image area the CNN based models actually focuses on and determines. As a result of the explainability analysis, it was confirmed that the model tended to respond strongly to the site where the tumor was actually present. The overall workflow was systematically conducted step by step, from data preparation to model learning, evaluation, and visual interpretation. These procedures allowed us to provide not only simple performance figures of the model, but also reliable diagnostic evidence in the clinical field. In the end, the pipeline of this study presents a systematic methodology that considers both the practicality and reliability of automatic diagnosis of brain tumors. This approach can be used as a standardized framework for various medical imaging analysis studies in the future.

C. Evaluation Metrics

In this study, various evaluation indicators were used to reflect the characteristics of MRI based brain tumor classification problems. First, Accuracy is defined as the proportion of images that match the actual label and the predictive label among all test images, and in this study, all four classes(glioma, meningioma, pituitary, and notumor) were calculated. Second, Precision, Recall, and F1-score were calculated for each class and averaged in two ways (macro and weighted). The weighted average is a sum of the overall performance by reflecting the ratio of the number of samples in each class, and performance evaluation most similar to the actual data distribution is possible. For example, the weighted average precision of XceptionNet was 0.9871, showing high reliability in all classes of prediction. The Macro mean is a simple arithmetic mean that equally reflects the evaluation metrics of each class, evaluating the overall equilibrium performance of the model, including minority classes, in real world environments with severe data imbalances. The macro F1-score of EfficientNet-B0 was 0.98, showing balanced performance across all classes. In addition, for a comprehensive comparison of model complexity and efficiency, the number of parameters (Params(M)) and floating-point operations (FLOPs(G)) were also measured for each model. Params(M) was calculated as the total number of trainable parameters divided by one million, and FLOPs(G) was calculated as the sum of floating-point operations across all layers divided by one billion, as shown in Equations 1 and 2. As such, each model used several indicators to evaluate not only overall performance but also reliability and balance for a specific class. The performance evaluation method in this study is focused on reflecting various situations that are important when applied in practice in clinical settings, not just the percentage of correct answers.

$$\text{Params(M)} = \frac{\sum_{l=1}^{L} P_l}{10^6} \tag{1}$$

where P_l denotes the number of trainable parameters in layer l, and L is the total number of layers in the model.

$$\text{FLOPs(G)} = \frac{\sum_{l=1}^{L} F_l}{10^9} \quad (2)$$

where F_l denotes the number of floating point operations required for layer l, and L is the total number of layers in the model.

D. Experimental settings

All experiments in this study were conducted using the Brain Tumor MRI Dataset by Kaggle. After removing duplicate images, the entire dataset was divided into training (82.1%) and testing (17.9%). Each image was resized to a size of 224×224 and subjected to grayscale transformation and Z-score normalization. Data augmentation techniques such as rotation and inversion were applied to secure data diversity. The model compared in this study consists of basic CNN, XceptionNet, EfficientNet-B0, and Vision Transformer (ViT). All models were trained and evaluated on the same dataset that underwent the same preprocessing and augmentation process. The experiment was conducted in a PyTorch based framework environment, and NVIDIA Tesla T4 GPU was used. Reproducibility was secured by using the same random seed so that randomness was not involved in both learning and testing processes. Performance evaluation was conducted using various indicators such as accuracy, precision, recall, and F1-score. For fairness in comparison, the experiment was conducted by applying the same data division, preprocessing, and augmentation conditions to all models.

The hyper-parameter settings (Table II) used in this study were applied equally to all models for consistency and reproducibility of the experiment. Learning was conducted for a total of 10 epochs, and the learning rate was fixed at 0.0001. The batch size was set to 32, 64, and 64 for learning, validation, and testing, respectively, to promote efficient memory utilization and performance improvement. The optimization algorithm adopted Adam optimizer, and the loss function used CrossEntropyLoss suitable for the multi class classification problem. The number of classes to be classified is four in total, consisting of glioma, meningioma, pituitary, and notumor. The network structure was implemented in a total of four models. The basic CNN (Fig 3) extracts features by repeatedly stacking 2D convolutional layers and pooling layers, followed by a final classification through a fully connected layer. XceptionNet (Fig 4) and EfficientNet-B0 (Fig 5) leverage deep convolutional neural network structures and efficient computational blocks to learn more complex and rich features. Vision Transformer (ViT) (Fig 6) divides an image into several patches and then inputs information about each patch into a transformer encoder to learn global features. Such hyperparameter and network architecture settings enable fair performance comparisons of all models.

IV. Results

We compared the brain tumor classification performance of ViT, XceptionNet, EfficientNet-B0, and basic CNN models. For all models, four major performance indicators were calculated: accuracy, precision, recall, and F1-score. As a result, XceptionNet and EfficientNet-B0 performed very well over 0.98 on all metrics. ViT models also recorded high overall performance, but showed somewhat lower values than XceptionNet and EfficientNet-B0. The basic CNN model showed lower classification accuracy than other models, with around 0.6 in all performance indicators. The difference in performance by model was consistent across all indicators. In particular, XceptionNet maintained overall classification accuracy and consistency through deep network and efficient convolutional structure. All experiments were conducted on the same dataset and under the conditions, and the performance evaluation was limited to numerical based objective results. Quantitative performance by model can be found in comparison chart 7 and table III.

The CNN model showed a parameter of about 0.09M, FLOPs of 0.52G, and a training time of 6.01 minutes. XceptionNet recorded a parameter of 20.82M, FLOPs of 4.6G, and a learning time of 16 minutes. EfficientNet-B0 had a parameter of 4.01M, FLOPs of 0.39G, and a training time of 6.62 minutes. ViT took 85.8M of parameter, 16.87G of FLOPs, and 33.51 minutes of learning time. As the model complexity (number of parameters) increased, the overall accuracy tended to increase. The least complex CNN had an accuracy of 0.6 and the most complex ViT had an accuracy of about 0.96. EfficientNet-B0 achieved a high accuracy of about 0.98, despite its small parameters and computational amount. XceptionNet also showed a high accuracy of 0.98 or more and showed excellent performance compared to the amount of computation. ViT requires the most parameters and amount of computation, but the learning time was also the longest. The relationship between complexity and accuracy for each model can be found in Figure 8 and the table IV.

The CNN model is marked by misclassification in the meningioma and notumor classes. In particular, meningioma was often incorrectly classified as a notumor and meningioma. Additionally, glioma tended to be partially mispredicted as pituitary and meningioma. XceptionNet and EfficientNet-B0 recorded very high accuracy overall. These two models had very few misclasses in almost all classes, with virtually perfect classification of notumor and pituitary. The ViT model performed well overall, but misclassification between glioma and meningioma occurred relatively frequently. Specifically, glioma was misclassified 46 times as meningioma, showing a limitation in which the boundaries of these two classes were not clear. Overall, simple CNN were often misclassified due to their lack of clear class-to-class separation, XceptionNet and EfficientNet-B0 performed well in all classes, and ViTs were limited in some class separations. Please refer to Fig 9.

The results of analyzing how the accuracy of the CNN model changes with an increase in the number of epochs (up to 100), and what this trend means for the learning needs of simple CNNs for brain tumor detection are as follows. The validation accuracy at the initial epoch started at about 0.70, reaching the level of 0.73 at 50 epochs, and finally rose to about

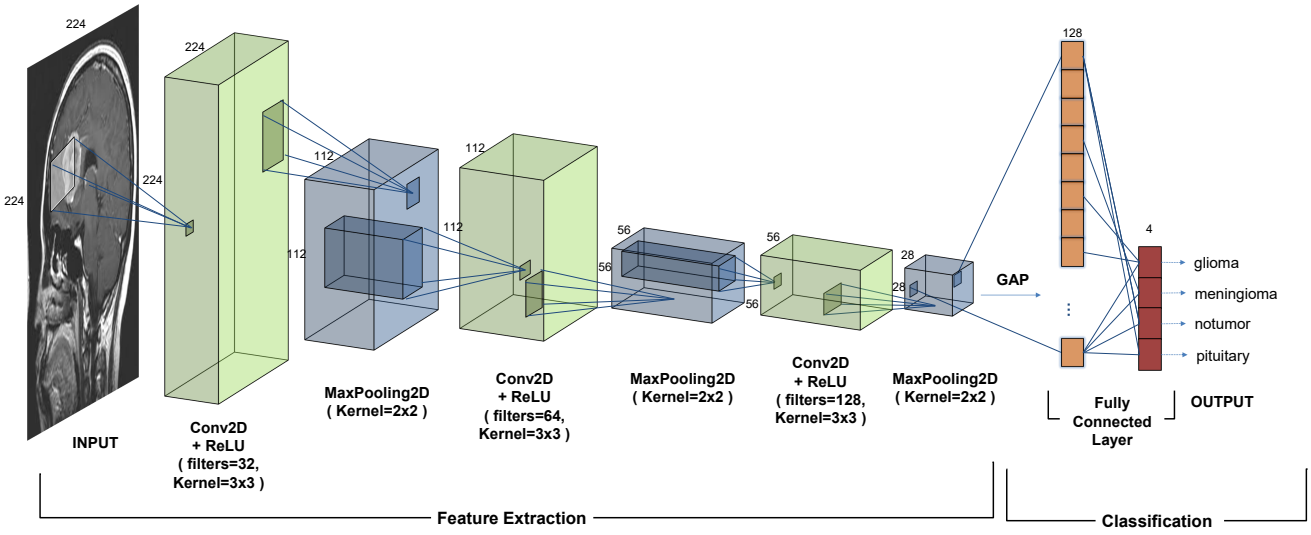

Fig. 3. CNN Architecture. It is a basic convolutional neural network structure that repeatedly stacks 2D convolution and pooling layers to extract key features in the image, and performs final classification through a complete connection layer.

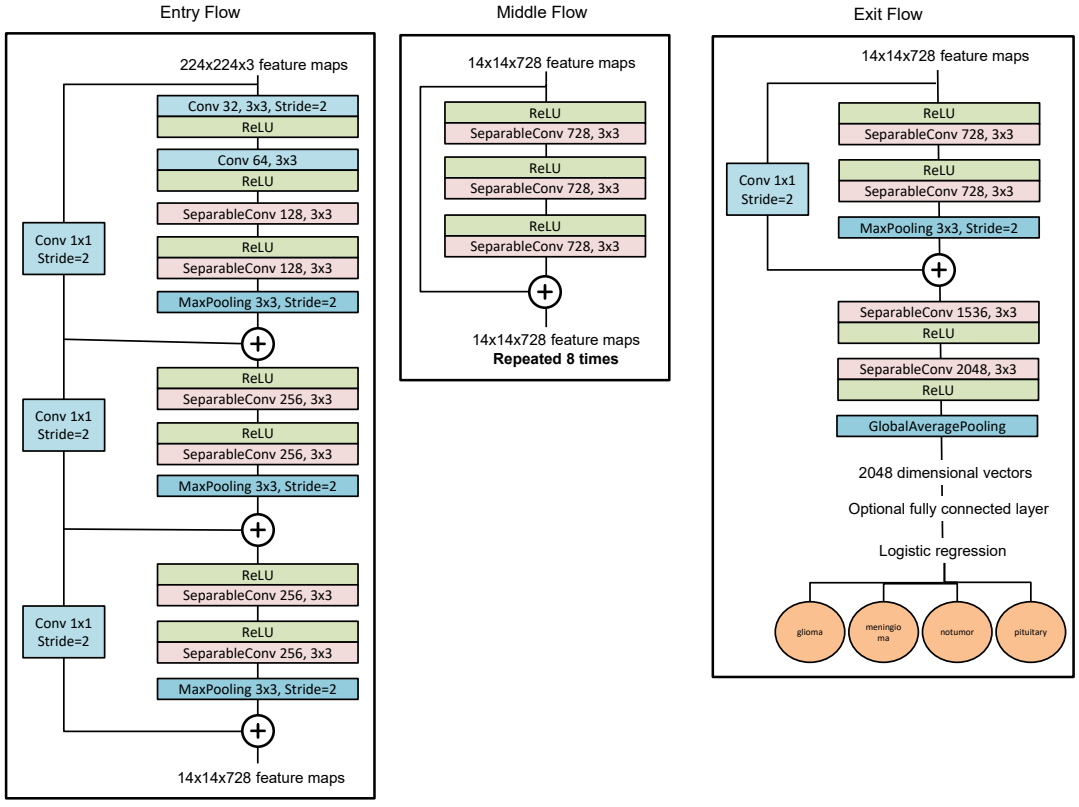

Fig. 4. XceptionNet Architecture. It is a deep neural network structure based on Depthwise Separable Convolution, designed to effectively learn complex image patterns while increasing computational efficiency compared to existing CNNs.

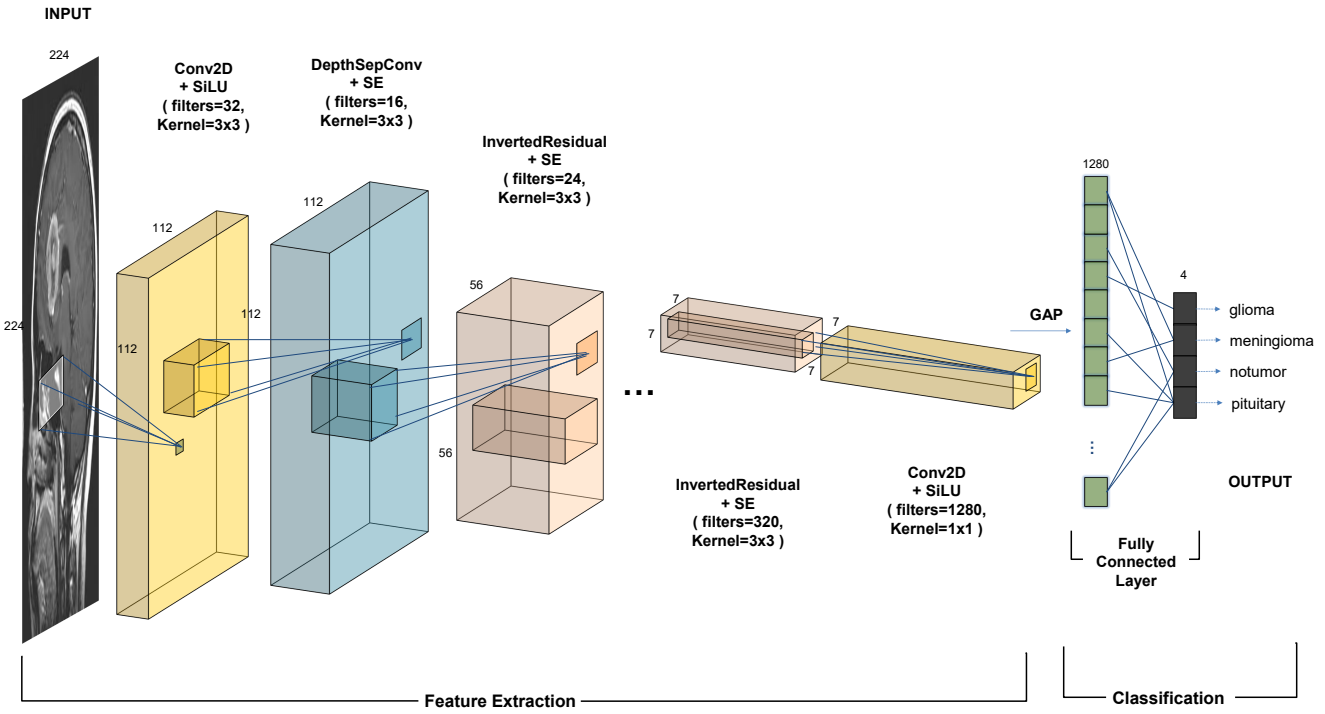

Fig. 5. EfficientNet-B0 Architecture. A lightweight convolutional neural network with a compound scaling strategy that scales model size, depth, and width in balance, simultaneously achieves high classification performance and computational efficiency with few parameters.

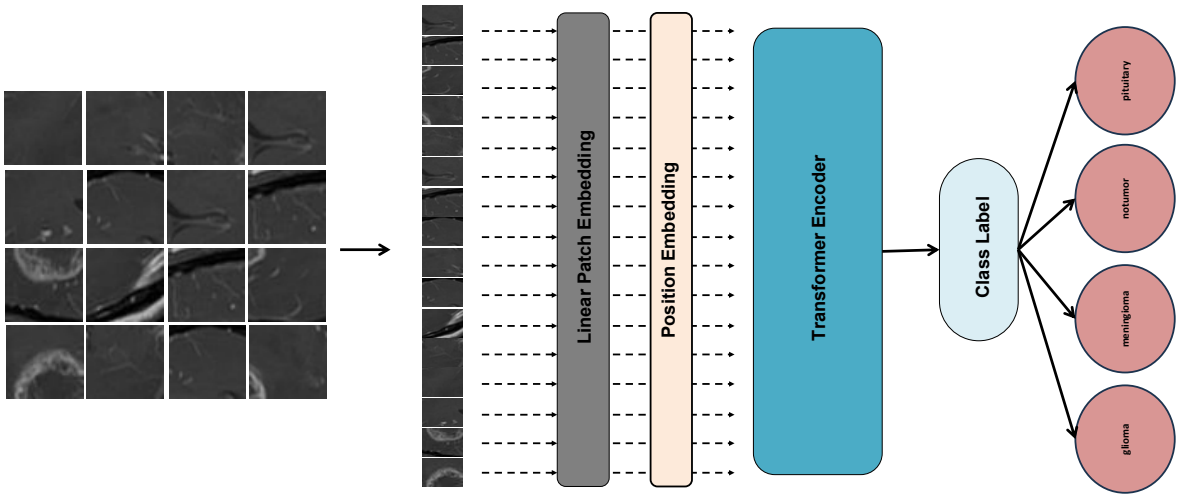

Fig. 6. ViT (Vision Transformer) Architecture. It is a state-of-the-art vision transformer-based model that effectively learns global properties and long term dependencies by segmenting images in patch units and then entering each patch information into a transformer encoder.

0.77 at 100 epochs. Training accuracy also started at about 0.73, steadily increasing as learning progressed, approaching 0.79 at the end. The difference between training accuracy and validation accuracy was not significant. In addition, the training accuracy showed almost no further increase during the final 90-100 epochs. The validation accuracy increased overall, but a phenomenon in which the value fell intermittently was also repeatedly observed. As a result, the CNN model reaches maximum performance in the approximately 80-100 epoch range, after which further learning has little practical benefit. Please refer to Fig 10.

The results of analyzing the decision making process for

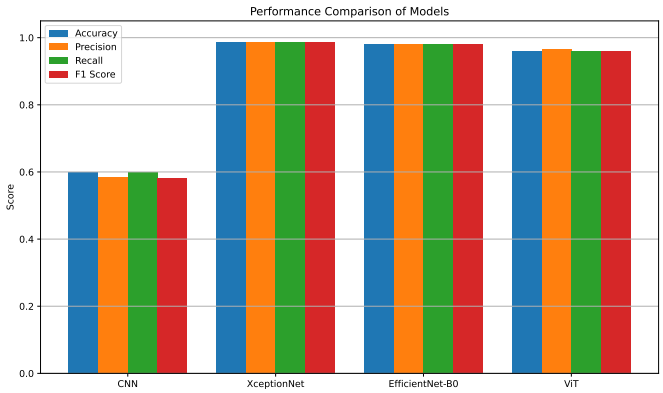

Fig. 7. A graph that visualizes the comparison results of Accuracy, Precision, Recall, and F1-score performance of each model (CNN, XceptionNet, EfficientNet-B0, ViT). In all indicators, XceptionNet and EfficientNet-B0 showed the best values, and CNN recorded relatively low performance.

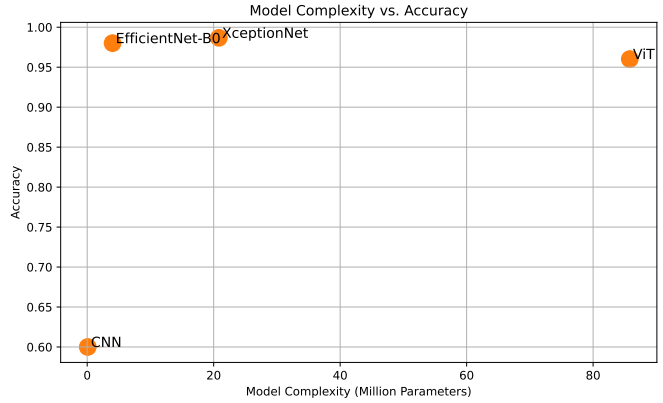

Fig. 8. It is a graph that visually compares the number and accuracy of parameters of each model (CNN, XceptionNet, EfficientNet-B0, ViT). ViT shows the highest complexity and EfficientNet-B0 shows excellent efficiency with few parameters and amount of computation. XceptionNet balances complexity and efficiency, and CNN has the simplest structure overall.

detecting brain tumors of each CNN based model through Grad-CAM visualization results are as follows. In the case of basic CNN, heat maps often tended to be widely emphasized to areas unrelated to the lesion or even unnecessary areas such as the outer skull. This suggests that attention may be focused on areas other than actual lesions, leading to incorrect classification. On the other hand, XceptionNet and EfficientNet-B0 showed that the heat map was more accurately concentrated at the lesion site or the internal area of the brain where the tumor was located. By strongly highlighting the area with tumors in both models, it can be confirmed that the shape and location information of the lesion is effectively utilized in the decision making process. In the images classified as correct answers, activation concentrated in the tumor site was clearly observed. In misclassified cases, activation tended to be dispersed in areas other than tumors in some models, which was directly related to prediction error. XceptionNet showed relatively good emphasis around the tumor even in misclassification situations, but did not guarantee completely correct predictions. As a result, Grad-CAM visualization clearly shows which brain regions are actually used to make decisions by model, and which patterns are repeated during misclassification. Please refer to Fig 11.

V. Discussion

In this study, we compared the performance of four deep learning models: CNN, XceptionNet, EfficientNet-B0, and ViT on the brain tumor MRI image classification problem. The results showed that XceptionNet performed best on all key metrics (accuracy, precision, recall, and F1-Score), followed by EfficientNet-B0. The ViT model also showed high overall performance, but it fell short of XceptionNet or EfficientNet-B0 in some class-to-class divisions. The basic CNN showed lower results than models with complex structures in all indicators, and in particular, the misclassification rate was relatively high. These results clearly demonstrate that complex network structures contribute to practical performance improvements in the field of recent medical imaging analysis. In particular, it differs from previous studies in that it systematically compared the four models with the same data and indicators. We think this result illustrates well the limitations of traditional simple CNNs in complex problems such as brain tumor detection. The comprehensive comparative analysis of this study provides a practical basis for model selection to increase clinical applicability.

We analyzed the effect of the number of parameters, amount of computation(FLOPs), and architecture type on the actual performance and efficiency of the model. EfficientNet-B0 has a very good balance of efficiency and accuracy, and achieves superior results with relatively low computational cost. XceptionNet has many parameters and FLOPs, and has correspondingly high classification performance. ViT required the highest computational cost and training time, but did not guarantee that much performance improvement in all situations. This result simply suggests that a large model does not necessarily perform well, and that the efficiency of architectural design is important. In particular, EfficientNet-B0 has been shown to be able to satisfy both lightweight and performance, which is an important point when applied to real world clinical settings. In the existing paper, only one or two models were compared, but this study is differentiated in that various models and architectures were compared under the same conditions. These detailed comparisons provide practical implications for model selection in real world system design or resource constrained environments.

Through the confusion matrix, the trend of misclassification of each model was identified. In the basic CNN, misclassification occurred most frequently in the meningioma and notumor classes, and the main reason was the unclear boundary between the two classes. XceptionNet and EfficientNet-B0 had very few misclasses in almost all classes, and notumor and fitness were classified near perfection. The ViT model showed good overall performance, but the frequency of misclassification between glioma and meningioma was high. In particular, it

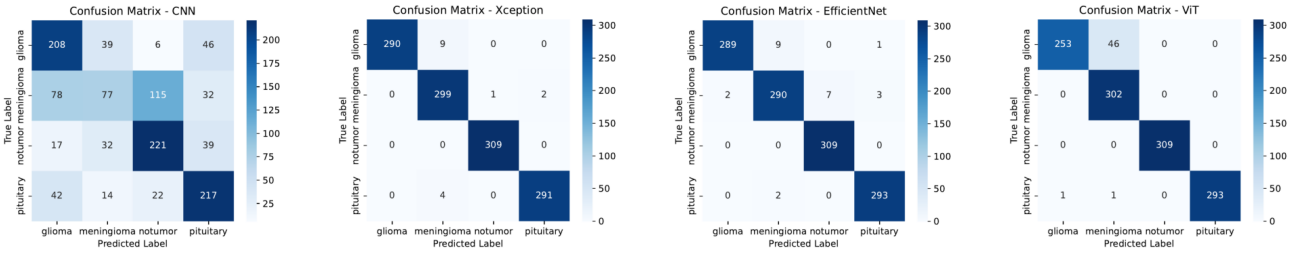

Fig. 9. Confusion matrix representing the class specific prediction results of each model (CNN, XceptionNet, EfficientNet-B0, ViT). Each matrix visualizes the correct prediction and number of misclassification for glioma, meningioma, notumor, and pituitary tumor types, showing differences in classification performance and misclassification patterns by model.

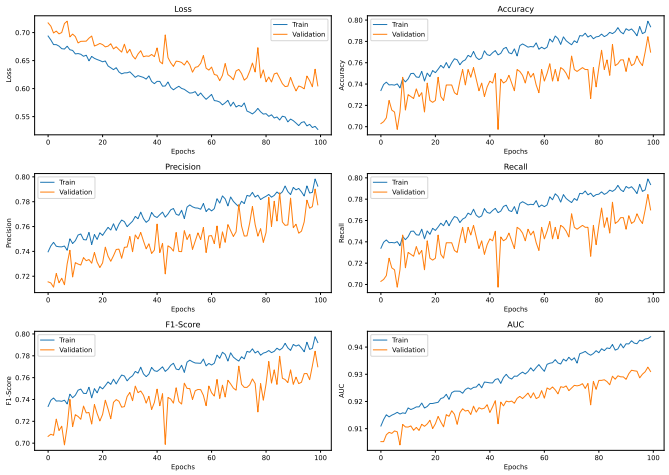

Fig. 10. A graph of changes in Loss, Accuracy, Precision, Recall, F1-Score, and AUC indicators for training and validation data of the basic CNN model for 100 epochs. Each indicator shows an overall gradual improvement over 100 epochs, and the validation indicators show intermittent fluctuations.

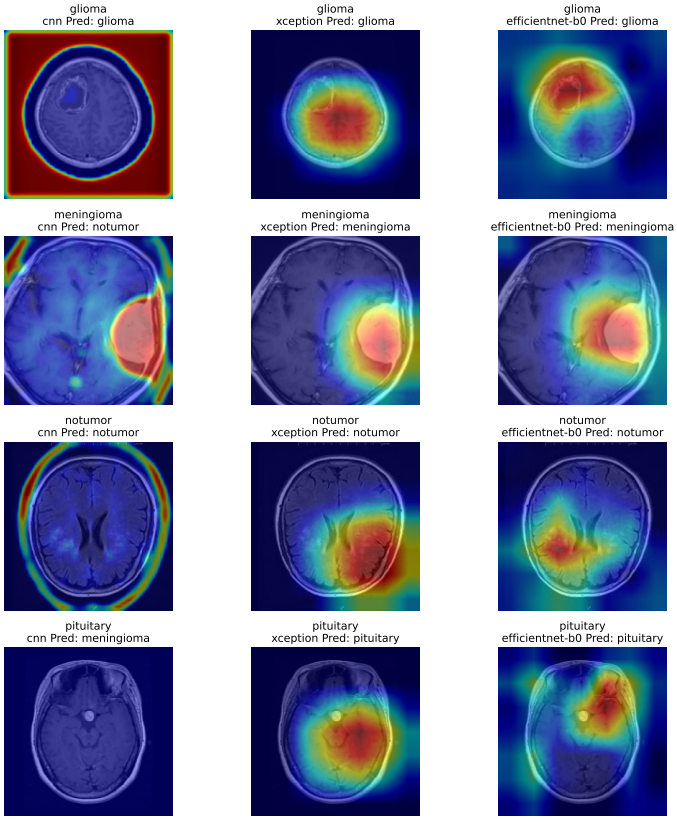

Fig. 11. Grad-CAM visualization results of each CNN based models (CNN, XceptionNet, EfficientNet-B0). Each image shows an activation map noted in actual prediction for each model, and it is possible to compare the concentration of attention to the lesion site and the accuracy of the prediction.

was confirmed that the boundaries between the two classes were not clear, with glioma being incorrectly classified 46 times as meningioma. Overall, simple CNNs were frequently misclassified due to lack of distinction between classes, and the more complex models, the less these problems were minimized. These analyses suggest the need for additional data acquisition or introduction of post-processing techniques focused on clinically important cases.

As a result of analyzing the change in accuracy according to the increase in the epoch of the CNN, both training and validation accuracy steadily increased as the number of epochs increased. The difference between training and validation accuracy was not significant, so overfitting did not occur seriously. However, after 80 epochs, the phenomenon that the increase in accuracy stagnated clearly appeared. In addition, the validation accuracy increased overall, but the phenomenon of intermittent drop was repeatedly observed. These results show that for simple CNNs, additional performance improvements are limited above a certain epoch. The results suggest that proper epoch setting is important for practical performance improvement, even in simple structured models. These learning patterns show that simple CNNs have limitations in data representation and generalization ability compared to more complex models in brain tumor classification problems.

Analysis of the brain tumor detection decision making process of each CNN based model through Grad-CAM visualization revealed distinct differences in the brain regions of interest for each model. The basic CNN model often tended to disperse activation to the brain outskirts or background independent of the lesion. On the other hand, XceptionNet and EfficientNet-B0 formed a concentrated activation map inside the brain where the lesion site or tumor was located. In the images

classified as correct answers, a clear emphasis appeared on the actual tumor site, confirming the basis for rational judgment of the model. In misclassified cases, a pattern of distraction to no-tumor areas was repeatedly observed in some models. XceptionNet tended to emphasize the tumor surroundings relatively better even under misclassification situations. These XAI analyses allow us to visually verify whether the model's decision-making is based on the actual lesion or is biased toward clinically meaningless areas. In this study, the Grad-CAM results of various models were systematically compared, which is an original attempt that has not been sufficiently addressed in previous studies.

A. Limitations

There are several limitations to this study. The MRI dataset used is difficult to generalize to various patient populations and equipment environments. The number of classes of brain tumors is limited, and there are more diverse types and high level lesions in actual clinical practice, so there is a limit to the extended application of the results of this study. Using a predefined hyperparameter and a simple preprocessing scheme, the potential optimal performance of the model was not sufficiently reflected. The performance evaluation of the model is conducted only on fixed validation sets and test sets, which lack further validation on external independent datasets. XAI analysis using Grad-CAM mainly remains in qualitative visualization, so there are limitations in terms of providing reliability and numerical evidence for medical interpretation. In actual clinical settings, more diverse noise, artifacts, and patient specific variables exist, but these complex variables were not sufficiently reflected in this experiment. For Transformer models, sufficient hyperparameter tuning or large scale experiments were difficult due to computational resource limitations. In the process of misclassification analysis and XAI interpretation, some subjective judgments of researchers could be involved. In addition, there is a limit to direct performance comparison because the evaluation index and experimental environment used in this study are not completely the same as other studies.

B. Future Directions

In order to continue to develop this study in the future, several directions can be sought. First of all, research is needed to more broadly evaluate the generalization performance of the model using MRI data collected from various institutions and environments. Further experiments to derive the maximum performance of each model through hyperparameter tuning and optimization of preprocessing pipelines are also important challenges. It is necessary to verify practical applicability in the clinical field through evaluation using external independent test sets or actual clinical data. XAI techniques such as Grad-CAM can be combined with quantitative indicators beyond qualitative analysis to enhance clinical interpretability. In addition to the currently compared networks, it will be meaningful to apply the latest hybrid models. Development of an automated system that can be integrated with actual clinical workflow and interface research that reflects user feedback are also worth considering as follow-up tasks. Based on misclassification case analysis, additional post-processing or ensemble techniques can be introduced to increase accuracy and reliability. Building a personalized AI diagnostic system that integrates patient-specific clinical information and prognostic data can also be an important future challenge. These various research directions will contribute to further increase the reliability and clinical utilization of AI-based diagnostic systems in the field of brain tumor detection.

VI. Conclusion

This study allowed us to systematically analyze the performance and limitations of various deep learning models in MRI-based brain tumor classification. XceptionNet and EfficientNet-B0 showed outstanding results on key metrics such as accuracy, precision, and recall, while ViT also showed competitive performance in some classes. Although the CNN revealed relatively low performance and misclassification trends, it was able to achieve a certain level of classification accuracy even with appropriate training epoxy and hyperparameter settings. Through the model specific confusion matrix and Grad-CAM visualization analysis, it was possible to understand in detail where each network actually pays attention and predicts, and what patterns of misclassification are repeated. As a result of model complexity and computational efficiency analysis, it was confirmed that a model that can secure lightweight structure and performance at the same time like EfficientNet-B0 is practical. Overall, it was found that the higher the complexity of the model, the higher the classification power and the better the clinical reliability. Analysis of interpretability using XAI techniques such as Grad-CAM has contributed to increasing the transparency and explainability of medical artificial intelligence. Based on the above results, it can be concluded that future medical image AI development should consider the actual application environment and clinical interpretation beyond simple performance comparison.

References

[1] A. B. Abdusalomov, M. Mukhiddinov, and T. K. Whangbo, "Brain tumor detection based on deep learning approaches and magnetic resonance imaging," *Cancers*, vol. 15, no. 16, p. 4172, 2023.

[2] V. Satushe, V. Vyas, S. Metkar, and D. P. Singh, "Ai in mri brain tumor diagnosis: A systematic review of machine learning and deep learning advances (2010–2025)," *Chemometrics and Intelligent Laboratory Systems*, vol. 263, p. 105414, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169743925000991

[3] M. F. Ahamed, M. M. Hossain, M. Nahiduzzaman, M. R. Islam, M. R. Islam, M. Ahsan, and J. Haider, "A review on brain tumor segmentation based on deep learning methods with federated learning techniques," *Computerized Medical Imaging and Graphics*, vol. 110, p. 102313, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0895611123001313

[4] B. R. A. A. Ishfaq, Q.U.A., "Automatic smart brain tumor classification and prediction system using deep learning," *Scientific Reports*, vol. 15, p. 14876, 2025. [Online]. Available: https://www.nature.com/articles/s41598-025-95803-3

[5] M. M. S. R. Hosny, K.M., "Explainable ensemble deep learning-based model for brain tumor detection and classification," *Neural Computing and Applications*, vol. 37, pp. 1289–1306, 2025. [Online]. Available: https://doi.org/10.1007/s00521-024-10401-0

[6] N. Musthafa, Q. A. Memon, and M. M. Masud, "Advancing brain tumor analysis: Current trends, key challenges, and perspectives in deep learning-based brain mri tumor diagnosis," *Eng*, vol. 6, no. 5, 2025. [Online]. Available: https://www.mdpi.com/2673-4117/6/5/82

[7] L. Pinto-Coelho, "How artificial intelligence is shaping medical imaging technology: A survey of innovations and applications," *Bioengineering (Basel)*, vol. 10, no. 12, p. 1435, 2023.

[8] A. Parvaiz, M. A. Khalid, R. Zafar, H. Ameer, M. Ali, and M. M. Fraz, "Vision transformers in medical computer vision—a contemplative retrospection," *Engineering Applications of Artificial Intelligence*, vol. 122, p. 106126, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095219762300310X

[9] K. A. Tran, O. Kondrashova, A. Bradley, E. D. Williams, J. V. Pearson, and N. Waddell, "Deep learning in cancer diagnosis, prognosis and treatment selection," *Genome medicine*, vol. 13, pp. 1–17, 2021.

[10] A. Verma and A. K. Yadav, "Brain tumor segmentation with deep learning: Current approaches and future perspectives," *Journal of Neuroscience Methods*, p. 110424, 2025.

[11] S. Krishnapriya and Y. Karuna, "A survey of deep learning for mri brain tumor segmentation methods: Trends, challenges, and future directions," *Health and technology*, vol. 13, no. 2, pp. 181–201, 2023.

[12] B. H. Van der Velden, H. J. Kuijf, K. G. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (xai) in deep learning-based medical image analysis," *Medical Image Analysis*, vol. 79, p. 102470, 2022.

[13] S. Yoon, "Brain tumor classification using a hybrid ensemble of xception and parallel deep cnn models," *Informatics in Medicine Unlocked*, vol. 54, p. 101629, 2025.

[14] V. K. Bairagi, P. P. Gumaste, S. H. Rajput, and K. Chethan, "Automatic brain tumor detection using cnn transfer learning approach," *Medical & Biological Engineering & Computing*, vol. 61, no. 7, pp. 1821–1836, 2023.

[15] S. Saeedi, S. Rezayi, H. Keshavarz, and S. R. Niakan Kalhori, "Mri-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, p. 16, 2023.

[16] R. Vankdothu, M. A. Hameed, and H. Fatima, "A brain tumor identification and classification using deep learning based on cnn-lstm method," *Computers and Electrical Engineering*, vol. 101, p. 107960, 2022.

[17] G. Verma, "Xception-based deep learning model for precise brain tumour classification," in *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*. IEEE, 2024, pp. 1481–1485.

[18] S. Hossain, A. Chakrabarty, T. R. Gadekallu, M. Alazab, and M. J. Piran, "Vision transformers, ensemble model, and transfer learning leveraging explainable ai for brain tumor detection and classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 3, pp. 1261–1272, 2023.

[19] V. Kushwaha and P. Maidamwar, "Btfcnn: Design of a brain tumor classification model using fused convolutional neural networks," in *2022 10th International Conference on Emerging Trends in Engineering and Technology-Signal and Information Processing (ICETET-SIP-22)*. IEEE, 2022, pp. 1–6.

[20] H. A. Shah, F. Saeed, S. Yun, J.-H. Park, A. Paul, and J.-M. Kang, "A robust approach for brain tumor detection in magnetic resonance images using finetuned efficientnet," *Ieee Access*, vol. 10, pp. 65 426–65 438, 2022.

[21] F. Zulfiqar, U. I. Bajwa, and Y. Mehmood, "Multi-class classification of brain tumor types from mr images using efficientnets," *Biomedical Signal Processing and Control*, vol. 84, p. 104777, 2023.

[22] T. Mahesh, M. Gupta, T. Anupama, O. Geman *et al.*, "An xai-enhanced efficientnetb0 framework for precision brain tumor detection in mri imaging," *Journal of Neuroscience Methods*, vol. 410, p. 110227, 2024.

[23] A. Tariq, M. M. Iqbal, M. J. Iqbal, and I. Ahmad, "Transforming brain tumor detection empowering multi-class classification with vision transformers and efficientnetv2," *IEEE Access*, 2025.

[24] S. Poornam and J. J. R. Angelina, "Vitalt: a robust and efficient brain tumor detection system using vision transformer with attention and linear transformation," *Neural Computing and Applications*, vol. 36, no. 12, pp. 6403–6419, 2024.

[25] J. Wang, S.-Y. Lu, S.-H. Wang, and Y.-D. Zhang, "Ranmerformer: Randomized vision transformer with token merging for brain tumor classification," *Neurocomputing*, vol. 573, p. 127216, 2024.

TABLE I

THIS TABLE IS A LITERATURE REVIEW TABLE THAT COMPARING AND ORGANIZING THE DATASETS, METHODS USED, EXPERIMENTAL RESULTS, MAJOR CONTRIBUTIONS, AND LIMITATIONS OF MAJOR PAPERS PUBLISHED IN BRAIN TUMOR DETECTION WITH VARIOUS MACHINE LEARNING TECHNIQUES.

Year Published	Paper Author and Citation	Paper Title	Dataset Used	Methods Used	Results	Contribution(s)	Drawback/ Limitations
2023	Bairagi *et al.* [14]	Automatic brain tumor detection using CNN transfer learning approach	BRATS 2013, BRATS 2015, OPEN I (total 621 images)	AlexNet, VGG-16, GoogLeNet, RNN (mainly AlexNet, VGG-16)	98.67% (AlexNet)	Compared various CNNs; automatic classification	Manual MRI analysis hard; small test size
2023	Saeedi *et al.* [15]	MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques	3,264 MRI images	2D CNN, auto-encoder, 6 ML methods	2D CNN: 96.47%, Auto-encoder: 95.63%	Simple 2D CNN is effective; compared ML/auto-encoder	Auto-encoder complex; single dataset only
2022	Vankdothu *et al.* [16]	A Brain Tumor Identification and Classification Using Deep Learning based on CNN-LSTM Method	Kaggle MRI dataset (3,264 images; 2,870 train, 394 test)	Hybrid: CNN + LSTM (LSTM-CNN)	CNN-LSTM: 92%	IoT-based system, LSTM improves CNN	Long training time
2024	Verma *et al.* [17]	Xception-based Deep Learning Model for Precise Brain Tumour Classification	2,000 MRI images from Kaggle	Fine-tuned Xception, preprocessing	∼98% train accuracy, high precision/recall/F1	Xception effective for 3-class, can aid diagnosis	Needs clinical validation
2022	Kushwaha *et al.* [19]	BTFCNN: Design of a brain tumor classification model using fused convolutional neural networks	Kaggle, Br35H Brain Tumor Dataset, IEEE Data Port (multiple MRI datasets)	Ensemble CNN (VGG16, AlexNet, Inception, Xception), saliency segmentation	97.4∼98.2% (by class)	Ensemble model, high accuracy, consistent	Limited complexity/inference info
2025	Yoon *et al.* [13]	Brain tumor classification using a hybrid ensemble of Xception and parallel deep CNN models	The original Kaggle brain tumor dataset consists of 7023 MRI images	Hybrid: Xception + PDCNN, compared with VGG19/ResNet	Hybrid: 99.09%, Xception: 98.26%	Hybrid outperforms previous SOTA, 4-class	Generalization, inference not discussed
2022	Shah *et al.* [20]	A Robust Approach for Brain Tumor Detection in Magnetic Resonance Images Using Finetuned EfficientNet	3762 MR images(kaggle)	Fine-tuned EfficientNet-B0, TL, augmentation	98.87% accuracy, top among compared	EfficientNet-B0 outperformed other models	Dataset size/generality not discussed
2023	Zulfiqar *et al.* [21]	Multi-class classification of brain tumor types from MR images using EfficientNets	Figshare – Brain Tumor CE-MRI Dataset	TL with EfficientNetB0–B4, Grad-CAM	EfficientNetB2: 98.86%	Fine-tuned EfficientNet, model attention visualized	No non-EffNet/ext. dataset validation
2024	Mahesh *et al.* [22]	An XAI-enhanced efficientNetB0 framework for precision brain tumor detection in MRI imaging	MRI dataset	EfficientNetB0 + Grad-CAM	98.72%, >97% prec/recall	Explainability (Grad-CAM), high accuracy	Needs further real clinical validation
2023	Hossain *et al.* [18]	Vision Transformers, Ensemble Model, and Transfer Learning Leveraging Explainable AI for Brain Tumor Detection and Classification	3,264 MRI images	TL (VGG, Inception, ResNet, Xception, ViT), ensemble (IVX16), XAI	IVX16: 96.94%	Compared TL, ViT, XAI, ensemble best	No binary, no clinical validation
2024	Poornam *et al.* [24]	VITALT: a robust and efficient brain tumor detection system using vision transformer with attention and linear transformation	Four benchmark brain tumor MRI datasets	ViT + attention, S-BiFPN, LTM, soft-quant	98.8∼99.1%	ViT hybrid, robust, consistent across sets	No real-world clinical validation
2024	Wang *et al.* [25]	RanMerFormer: Randomized vision transformer with token merging for brain tumor classification	1426 of glioma, 708 of meningioma, and 930 of the pituitary, collected from 233 patients.	RanMerFormer: ViT + token merge, functional-link head	99.77%	Efficient, fast, high accuracy CAD	No specific metrics/clinical validation
2025	Yujin Jeon	Proposed Work	Brain Tumor MRI Dataset from Kaggle	CNN, XceptionNet, EfficientNet-B0, ViT	CNN: 60%, XceptionNet: 99%, EfficientNet-B0: 98%, ViT: 96%	Systematic, explainable model comparison	Limited to one dataset, reducing generalizability

TABLE II

MAIN NETWORK CONFIGURATION APPLIED IN THIS STUDY. ALL DEEP LEARNING MODELS WERE TRAINED AND EVALUATED BASED ON THE SAME EPOCH, LEARNING RATE, BATCH SIZE, OPTIMIZER, LOSS FUNCTION, AND NUMBER OF CLASSES.

Network Configuration	
Epochs	10
Learning rate	0.0001
batch size(train/val/test)	32/64/64
Optimizer	Adam
Loss	CrossEntropyLoss
num_classes	4

TABLE III

IT IS A TABLE THAT SUMMARIZES THE WEIGHTED AVERAGE AND MACRO AVERAGE VALUES OF EACH MODEL (CNN, XCEPTIONNET, EFFICIENTNET-B0, VIT) FOR ACCURACY, PRECISION, RECALL, AND F1-SCORE. XCEPTIONNET AND EFFICIENTNET-B0 RECORDED VERY HIGH VALUES ABOVE 0.98 IN ALL EVALUATION INDICATORS, AND VIT ALSO PERFORMED WELL OVERALL.

Model	**Accuracy**	**Precision (Weighted)**	**Recall (Weighted)**	**F1-score (Weighted)**	**Precision (Macro)**	**Recall (Macro)**	**F1-score (Macro)**
CNN	0.6	0.5835	0.6	0.5808	0.5838	0.6004	0.5812
XceptionNet	0.9867	0.9871	0.9867	0.9868	0.9871	0.9866	0.9867
EfficientNet-B0	0.9801	0.9802	0.9801	0.98	0.9802	0.98	0.98
ViT	0.9602	0.9653	0.9602	0.96	0.9653	0.9598	0.9599

TABLE IV

IT IS A TABLE THAT SUMMARIZES THE NUMBER OF PARAMETERS (IN MILLIONS), THE AMOUNT OF COMPUTATION (FLOPS, IN BILLION UNITS), AND THE LEARNING TIME (IN MINUTES) OF EACH MODEL (CNN, XCEPTIONNET, EFFICIENTNET-B0, VIT). VIT REQUIRES THE MOST PARAMETERS, AMOUNT OF COMPUTATION, AND LEARNING TIME, AND EFFICIENTNET-B0 SHOWS FAST LEARNING TIME WITH RELATIVELY FEW PARAMETERS AND COMPUTATION.

Model	**Params (M)**	**FLOPs (G)**	**Training Time (min)**
CNN	0.09	0.52	6.01
XceptionNet	20.82	4.6	16
EfficientNet-B0	4.01	0.39	6.62
ViT	85.8	16.87	33.51