JURNAL INFOTEL
VOL. 17, NO. 4, NOVEMBER 2025, PP. 789–809. DOI:10.20895/INFOTEL.V17I4.1290

RESEARCH ARTICLE

An EfficientNet and Dual Path Network Approach for Enhanced Brain Tumor Classification

Andri Agustav Wirabudi[1,2,*], Lia Hafiza[2], Nurwan Reza Fachrurrozi[3], Agus Pratondo[3], and Gagas Ezhar Rahmayadi[4]

[1]Dept. of Intelligence Media Engineering, Hanbat National University, Daejeon, South Korea
[2]Telecommunication Engineering Study Program, School of Electrical Engineering, Telkom University, Jakarta Campus, Jl. Daan Mogot KM 11, West Jakarta 11710, DKI Jakarta, Indonesia
[3]Multimedia Engineering Technology Applied Undergraduate Study Program, School of Applied Science, Telkom University, Main Campus (Bandung Campus), Jl. Telekomunikasi No. 1, Bandung 40257, West Java, Indonesia
[4]Visual Communication Design Study Program, School of Creative Industries, Telkom University, Jakarta Campus, Jl. Daan Mogot KM 11, West Jakarta 11710, DKI Jakarta, Indonesia

*Corresponding email: andriagustaw@telkomuniversity.ac.id

Received: December 20, 2024; Revised: September 05, 2025; Accepted: November 24, 2025.

Abstract: Brain tumor classification is an essential step in medical image analysis, contributing to timely diagnosis and effective treatment planning. This study introduces a brain tumor classification model that integrates EfficientNet with Dual Path Networks (DPN) and a Multi-Head Self-Attention (MHSA) mechanism. The model is applied to classify three major types of brain tumors—glioma, meningioma, and pituitary—using MRI images. The integration of DPN allows the model to leverage both residual and dense connections for enhanced feature representation, while the MHSA module refines global and local contextual information. Experimental evaluation demonstrates that the proposed model achieves an overall accuracy of 97.82%, sensitivity of 97.83%, specificity of 98.41%, precision of 98.34%, and F-score of 98.08%. These results indicate competitive performance compared to widely used architectures such as CNN, ResNet, and DenseNet, suggesting that the combined use of EfficientNet, DPN, and MHSA can provide a robust approach for brain tumor classification.

Keywords: Brain Tumor, Deep Learning, Dual Path Network, EfficientNet, Multi-Head Attention

 ISSN: 2085-3688; e-ISSN: 2460-0997

1 Introduction

The development of digital image processing in the medical field is increasingly being used to facilitate disease diagnosis. One popular application is the use of Magnetic Resonance Imaging (MRI) [1, 2]and X-ray [3] to non-invasively visualize the body's interior, which is highly valuable for early identification of diseases, particularly brain tumors. In practice, medical personnel still have difficulty in classifying the types of tumors that exist. In the current conventional method [4–7] brain tumor classification is carried out manually by radiologists through visual inspection of MRI results. This requires extensive medical expertise to distinguish between tumor types and is prone to inter-observer variability.

Deep learning methods have been widely adopted to improve diagnostic accuracy and reduce human error [8–10]. However, medical imaging research often faces the challenge of relatively small MRI datasets, which can cause overfitting and poor generalization when training deep networks from scratch. To address this, transfer learning (TL) has become essential [8]. By leveraging pretrained weights from large datasets like ImageNet [11], containing over 14 million annotated images across 1,000 categories, models can reuse low- and mid-level features such as edges, textures, and shapes relevant to MRI analysis. TL accelerates convergence, improves robustness, and enhances classification accuracy even with limited medical data [5].

Several recent studies on tumor classification have been conducted, with some considered state-of-the-art (SOTA). Kader et al. [12] developed a CNN-based parotid tumor diagnosis model that showed promise but required large datasets to avoid overfitting, as CNNs mainly capture local information and struggle with complex tumor shapes. Zahra et al. [13] improved accuracy using residual encoder blocks, but this increased complexity and feature redundancy. Wadhah et al. [14] combined feature extraction with SVM, though performance relied heavily on feature quality and scalability was limited for high-dimensional datasets. Zhang et al. [15] enhanced a ResNet-based model [16] with augmentations like Mix-up and rotation, but ResNet remained computationally expensive and excessive augmentation risked unrealistic samples. Ramasamy et al. [17] used a HOG-based machine learning approach, which was limited to low-level features and performed poorly on highly variable data. Fang et al. [18] integrated MHSA with BiGRU to capture long-range dependencies, effectively applying the Transformer concept [19], but the method was computationally and memory intensive and required large datasets for optimal performance.

The adoption of CNN and deep learning has nonetheless significantly improved classification accuracy compared with conventional approaches. Inspired by these studies, this work proposes a brain tumor classification model leveraging EfficientNetB7 [20] enhanced with Dual Path Network (DPN) and Multi-Head Self-Attention (MHSA) modules [21]. DPN integrates residual and dense paths to improve feature diversity and information sharing [22], while MHSA enables the model to attend to multiple regions simultaneously, thus capturing global dependencies and further enhancing classification accuracy.

The discussion in this paper is organized into five sections. Section I introduces the problem, relevant references, and the motivation for developing the proposed approach. Section II presents the methodology, including the mathematical model and architectural details of EfficientNetB7, Dual Path Network (DPN), and Multi-Head Self-Attention (MHSA). Motivated by these observations, this study proposes a brain tumor classification model leveraging EfficientNetB7 integrated with Dual Path Networks (DPN) and Multi-

Head Self-Attention (MHSA) to address both local and global feature dependencies. By combining the representational power of EfficientNet with the feature diversity of DPN and the global context modeling of MHSA, the proposed approach aims to achieve higher accuracy and reliability in classifying glioma, meningioma, and pituitary tumors from MRI images.

2 Related Work

This section reviews related work, covering both conventional and modern classification methods. Early studies laid the foundation for classification techniques and serve as benchmarks for this research. Over time, classification methods have evolved, particularly in medical imaging, aiming to support decision-making and reduce human error.

Conventional methods include the K-Nearest Neighbors (K-NN) algorithm introduced by Stevens et al. [23]. This method determines the class of a sample based on the majority class of its K-neighbors in feature space. Another method, Bayesian classification, was employed by Nazeer et al. [24] for cancer detection using a probabilistic approach known as Bayes' Theorem. This algorithm assumes feature independence and calculates the probability of each class, selecting the one with the highest probability as the final decision. Decision Tree-based approaches, such as those developed by J. Ross Quinlan [25], offer interpretable classification results and have laid the foundation for advanced algorithms like Random Forest [26], which enhances efficiency. However, conventional methods often face challenges in achieving optimal accuracy. Jerome H. et al. [27] addressed this issue by employing Gradient Boosting Machines (GBM), which iteratively reduce errors in previous models, making it highly effective for various classification and regression problems.

In contrast, modern methods, particularly those leveraging Deep Neural Networks (DNNs), have proven more effective than conventional techniques, especially for medical image classification. DNNs also play a pivotal role in detection and segmentation across various fields, such as brain tumor classification using fuzzy clustering combined with support vector machines [6], breast tumor diagnosis via texture characterization from ultrasound images [28], and multi-scale segmentation of thoracic organs for clinical applications [29]. Recent studies utilizing deep learning for brain tumor classification include the work by M. Rizwan et al. [9], which employed a Computer-Aided Diagnosis (CAD) approach to categorize brain tumor radiology images into glioma, meningioma, and pituitary tumor classes. Their process involved pre-processing with Gaussian filters and developing a 16-layer neural network-based architecture.

Another modern approach is You-Only-Look-Once (YOLO), as demonstrated by M.F. Almufareh et al. [8], [30]. This method integrates detection, segmentation, and classification into a unified framework, significantly improving efficiency and accuracy despite increased computational demands and longer training times. Additionally, Asif S. et al. [8] utilized transfer learning (TL) models for MRI-based brain tumor detection, achieving higher accuracy, recall, and precision. Popular classification models applied in other studies include Residual Networks (ResNet) [16], Visual Geometry Group Networks (VGGNet) [30], Efficient Neural Networks (EfficientNet) [20], Spectral-Spatial Bidirectional Transformer-based Convolutional Networks (SSBTCNet) [31], and Densely Connected Convolutional Networks (DenseNet) [32], among others [33], [34].

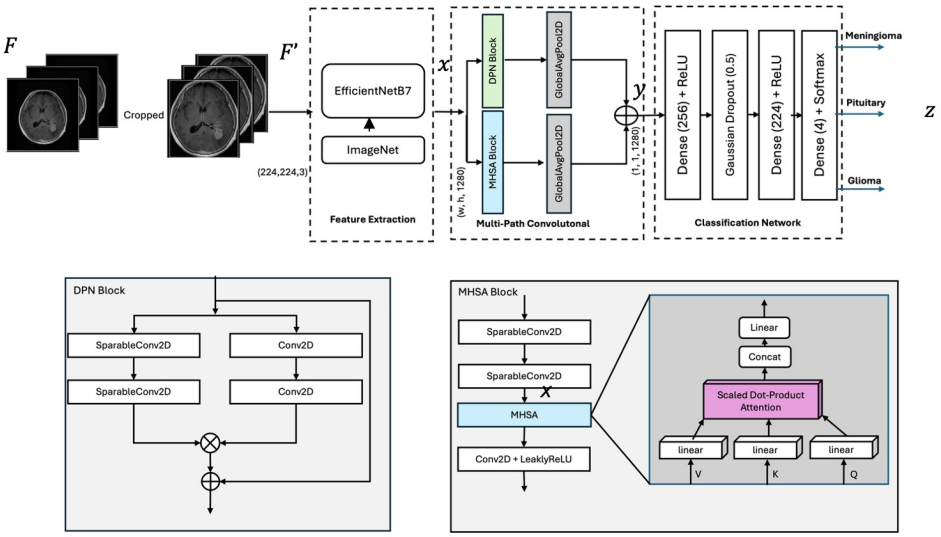


Figure 1: The proposed network architecture model includes MHSA block and DPN block. The model of classification use is EfficientNetB7. The results of the classification in the Classification Network will determine which highest probability value is suitable for determining the type of tumor.

3 Proposed Method

The method we propose in this study is explained in detail; the architecture of the classification model we created is displayed in Figure 1. We adapted the architecture model from the EfficientNet model [20], this DNNs-based architecture is widely used for classification and chose this model because of its ability to achieve high accuracy with less computational resource usage. In addition, to further optimize the final accuracy during the training process, we added two new modules including Multi-Head Self Attention (MHSA) which will divide the model's attention based on three inputs, this allows the model to increase the probability value of the emergence of desired information, thus impacting valid decisions. While the Dual Path Network (DPN) is used to combine the advantages of two main approaches in network convolution such as ResNet and DenseNet, by combining these models into one and implementing them into the EffcientNet model, It will increase the accuracy and efficiency of the existing coding even further.

The complete model utilized is depicted inFigure 1 architecture of the model we suggest. In this model, we divide the model into three stages including Feature Extraction, Multi-path Convolutional, and Classification Network. Each stage has different work and functions during training. The EfficientNetB7 model is used during training the model we created, this model was chosen because based on several previous studies [20] and [32] shows the best performance among its predecessor models B0 to B6, this model uses 64 million parameters with a depth of 813 layers.

During the training process, we standardize the initial input size of Feature F from the image by cropping it to F' with a size of 224×224×3, in addition to standardizing the size,

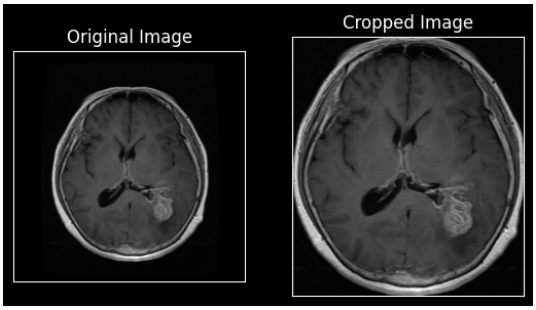


Figure 2: Cropped image.

this crop technique also functions to speed up the training process and focus information on the image during the training process, images with dominant information will often appear rather than backgrounds that do not contain information in them, thus minimizing errors in class classification. Furthermore, the input feature F' is trained into the model and produces output x, this output x is the result of feature extraction during training, these features are then sent to the Multi-path Convolutional stage which contains the MHSA attention block, and the DPN block, in this stage the input x will be trained using both blocks. At the Multi-path Convolutional stage has an important role where the attention block allows the model to focus on certain parts of the relevant input x and ignore the irrelevant parts, in addition, the DPN block is used to reduce overfitting, and improve performance and better accuracy, the results of this stage output y, this output y contains features that have better probability values compared to x, these results are forwarded to the Classification network section to test the accuracy of the previously trained model to produce z. These three stages will be explained in more detail below.

3.1 Cropped Image

The image cropping process, shown in Figure 2, represents the initial stage aimed at focusing on the most informative regions before training and testing. The image processing pipeline begins by converting the input image F to grayscale using weighted contributions from the red, green, and blue channels. Next, a Gaussian blur with a 5×5 kernel and $\sigma = 1.0$ is applied to smooth the image and reduce noise. Thresholding is then performed, setting pixels above 45 to 255 while retaining lower values, thereby preserving important information for classification. Morphological operations follow, with erosion removing small noise and separating objects, and dilation expanding object boundaries to close gaps and connect components, both using a structural element S over two iterations. Finally, contour detection identifies object boundaries from the processed image, allowing cropping to the exact object region without losing information.

3.2 Feature Extraction

At the feature extraction stage, the cropped images were trained using the EfficientNetB7 architecture initialized with pretrained weights from the ImageNet dataset [11]. All models were trained using pretrained weights to ensure fair and consistent comparison results.

During feature extraction, the pretrained layers of EfficientNetB7 were initially frozen, and only the final fully connected layers were trained on the MRI dataset. This allowed the model to leverage previously learned features from ImageNet while adapting to the specific patterns, textures, and structures present in brain MRI images. Subsequently, a fine-tuning step was applied, in which some of the deeper layers of EfficientNetB7 were unfrozen and retrained with a smaller learning rate, enabling the model to further refine its feature representations for the tumor classification task.

ResNet-50 [16], with approximately 25 million parameters and residual connections, is among the most widely adopted architectures in transfer learning for medical image classification. While effective in extracting hierarchical features, this model requires large-scale datasets to achieve stable performance. On limited MRI datasets, ResNet often exhibits performance degradation without the support of transfer learning. In contrast, EfficientNetB7, with over 64 million parameters and 813 layers, employs compound scaling to balance depth, width, and input resolution. When combined with pretrained weights from ImageNet, EfficientNetB7 can extract high-resolution features and fine-grained textures critical in brain MRI analysis, such as glioma boundaries and the complex structures of meningiomas. This explains why, in this study, the integration of EfficientNetB7 with DPN and MHSA not only achieves higher accuracy than ResNet but also utilizes the limited medical dataset more efficiently.

EfficientNetB7 belongs to the EfficientNet family [20], which employs a compound scaling method by simultaneously scaling three key dimensions of the network: depth α^{ϕ}, width β^{ϕ}, and input resolution γ^{ϕ}. Compound scaling involves increasing all network dimensions proportionally at regular scales. For example, by increasing the network depth (number of layers), width (number of filters per layer), and input image resolution simultaneously, the model maintains computational and memory efficiency. In this context, ϕ acts as a global coefficient controlling the scaling level, while α, β, and γ are constants that determine the proportion of each dimension.

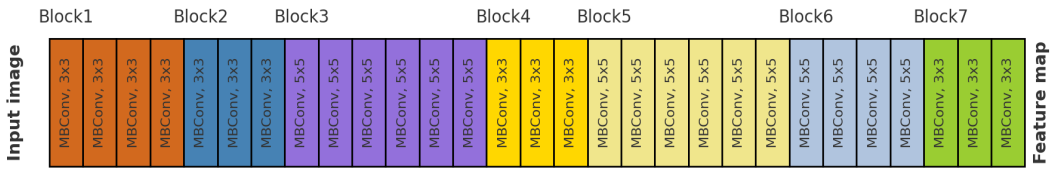


Figure 3: EfficienNetB7 model architecture.

Figure 3 illustrates the architecture of the EfficientNetB7 model used in this study. This model employs a fundamental building block called the Mobile Inverted Bottleneck Convolution (MBConv), also used in MobileNetV2 [35]. Each MBConv block consists of several components, including a depthwise separable convolution, a squeeze-and-excitation (SE) block, and an expansion phase. The expansion phase uses a 1×1 convolution to increase the number of channels, while the depthwise convolution employs a 3×3 or 5×5 convolution, applied individually to each channel. The SE block applies the squeeze-and-excitation mechanism to recalibrate channel-wise feature responses. Finally, the projection phase uses a 1×1 convolution to reduce the number of channels back to the original dimension, completing the MBConv structure.

3.3 Multi-Path Convolutional

In this stage, we added two new blocks including Dual Path Network (DPN) and Multi-Head Self Attention (MHSA), the purpose of using these two blocks is to improve accuracy and provide focus on objects in the image both in the learning process and testing, the objects in question are the parts affected by the tumor and those that are not, so as to minimize errors that occur and suppress overfitting in the model. In addition, the two blocks have different roles and functions during testing and training, the following is a detailed explanation of the two blocks.

3.3.1 Dual Path Network (DPN)

As shown in Figure 1, the Block implementation is carried out before the classification process together with the attention block. We apply this DPN block to improve performance in image recognition tasks, especially in terms of accuracy and computational efficiency. DPN is based on the idea that input information, namely features, can be obtained through two different paths to enrich the feature representation and increase the network capacity, as shown in Figure 4 below.

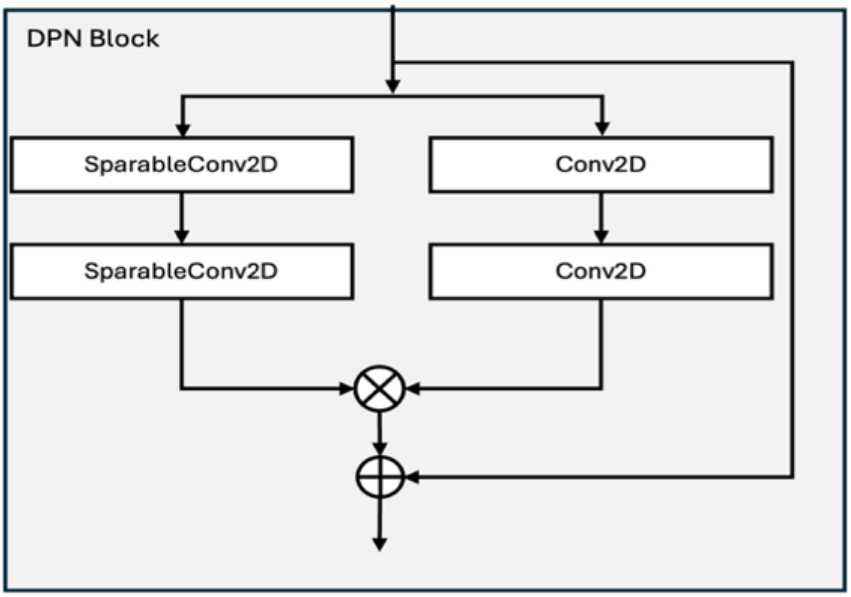


Figure 4: Dual Path DPN Model, this block is divided into three parts, the first part x_1 is SperableConv2d, the second part is Con2D, and the last is the residual which is used to restore to the initial size in the merging process.

In the initial stage, the input feature x is divided into three different parts, namely x_1, x_2, and x_3. To obtain x_1, we use `SeparableConv2D`, a convolution operation designed to reduce the number of parameters required by the model and increase computational efficiency, as shown in Eq. (1).

$$x_1 = \text{ReLU}(\text{Conv2D}(x, \text{Filter}, \text{Kernel\_size}, \text{strides}, \text{padding})) \quad (1)$$

To obtain x_2, we use the standard convolution path `Conv2D`, where x_d represents the output of the depthwise convolution in separable convolution, which is used to combine features from multiple channels without changing the spatial dimensions (width and height) of the input. The computation is shown in Eqs. (2) and (3).

$$x_d = \text{DepthwiseConv2D}(x, \text{Kernel\_size}, \text{strides}, \text{padding}) \quad (2)$$

$$x_2 = \text{ReLU}(\text{Conv2D}(x_d, \text{filters}, 1, 1, \text{padding})) \quad (3)$$

The final output is obtained by combining x_1 and x_2 through element-wise multiplication and adding the residual block x_3 (a 3×3 convolution), as described in Eq. (4).

$$x_{\text{out}} = (x_1 \cdot x_2) + x_3 \quad (4)$$

x_3 is a residual block convolution with a kernel size of 3×3. After obtaining x_1 and x_2, the merging process is carried out as expressed in Eq. (4), where the results of the standard convolution x_1 and the depthwise convolution x_2 are combined through element-by-element multiplication. The input or output from another path, x_3, is then added to produce x_{out}, which represents the DPN output.

3.3.2 Multi-Head Self Attention (MHSA)

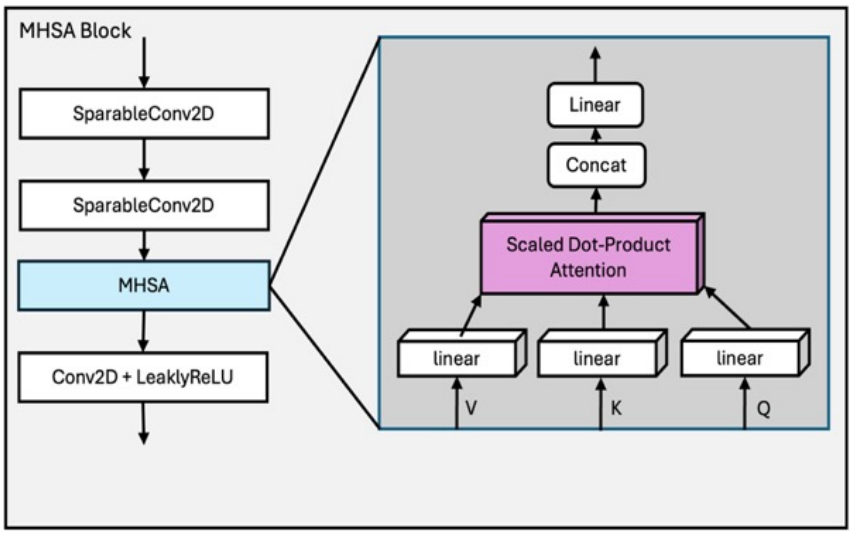


Figure 5: Multi-Head Self Attention (MHSA) has three inputs Query, Key, and Value. Each of the inputs works independently, and each head produces a series of vectors. The outputs of all heads are then concatenate and passed through a linear transformation layer to obtain the final output.

The use of the attention mechanism serves to focus the final output more accurately after being combined with the output from the previous DPN. This block receives input from `SeparableConv2D`, which is then processed through an information-sharing mechanism to refine the final classification results. Multi-Head Self-Attention (MHSA) is one of the key components of the Transformer architecture that significantly improves model performance in producing more accurate predictions.

As shown in Figure 5, the MHSA block consists of three different inputs: Query (Q), Key (K), and Value (V), each operating independently. However, the initial input feature in the model consists of a single tensor, making it unsuitable for direct MHSA processing. Therefore, the input must be transformed using separate convolution operations for each required component. To generate the Q, K, and V values, we apply three separate `SeparableConv2D` layers, one for each part of the input. This ensures that each tensor is computed independently, following the program structure: `query_conv = SeparableConv2D`, `key_conv = SeparableConv2D`, and `value_conv = SeparableConv2D`. Each convolution layer uses the same number of filters and kernel size. After obtaining the MHSA input tensors Q, K, and V, the self-attention mechanism can be performed using standard attention computation steps.

3.3.3 Linear Projections

In the initial stage for each head-i, we perform a linear projection on the input X to obtain three different vectors: query (Q_i), key (K_i), and value (V_i). The process is expressed in Equation (5).

$$Q_i = XW_i^Q, \qquad K_i = XW_i^K, \qquad V_i = XW_i^V \tag{5}$$

Here, X is the input matrix with dimensions $(n \times d_{\text{model}})$, where n denotes the input sequence length and d_{model} represents the model dimension. The matrices W_i^Q, W_i^K, and W_i^V are the learnable weight matrices for the linear projections corresponding to the i-th head, each with size $(d_{\text{model}} \times d_k)$. Meanwhile, Q_i, K_i, and V_i are the resulting projected tensors, each having dimensions $(n \times d_k)$.

3.3.4 Self-Attention Calculation

After getting the values of Q_i, K_i, and V_i, the attention must then be calculated for each head. The process involves several stages as shown in Equation (6).

$$A(Q_i, K_i, V_i) = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \tag{6}$$

In Equation (6), the attention value A is obtained by first computing the attention score $\frac{Q_i K_i^T}{\sqrt{d_k}}$, which is calculated by multiplying the query and key matrices and dividing the result by $\sqrt{d_k}$ to stabilize the gradient. The softmax function is then applied to generate a probability distribution from the attention scores, ensuring that the values sum to 1. The resulting softmax output is used to compute a weighted combination of the value vector V_i. The output dimension of the operation in Eq. (6) is $(n \times d_k)$.

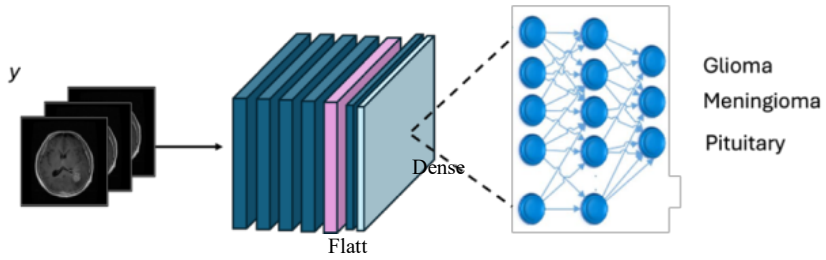


Figure 6: The output from the combination of DPN and MHSA produces output y, which is forwarded to the classification-Network section to obtain the final classification decision results, where there are three types of tumors tested during the study, namely, Glioma, Meningioma, and Pituitary.

3.3.5 Concatenate of Heads

After obtaining the attention value from each head, the output from each head h_n is then combined into a single matrix, where d_k is the dimension of each head, as shown in Eq. (7).

$$\text{Concat}\left(A_{h_1},\, A_{h_2},\, A_{h_3}\right) \in \mathbb{R}^{\,n\times(h\cdot d_k)} \tag{7}$$

The resulting combined matrix has a dimension of $(n \times (h \cdot d_k))$, where h is the number of heads and d_k is the dimensionality of each head.

3.3.6 Final Linear Transformation

The final step, after all components have been processed and the attention results from each head have been obtained, is to combine the outputs from all heads and pass them through the final linear transformation layer to produce the MHSA output, as shown in Eq. (8).

$$\text{MHSA}(Q, K, V) = \text{Concat}\left(A_{h_1},\, A_{h_2},\, A_{h_3}\right) W^O \tag{8}$$

Here, W^O is the final linear projection matrix with dimensions $((h \cdot d_k) \times d_{\text{model}})$. The result of this operation is a matrix with dimensions $(n \times d_{\text{model}})$, which restores the tensor shape to match the original model dimension. The MHSA output is then combined with the DPN output to form the final multi-path convolutional output MPC_y, as shown in Eq. (9).

$$\text{MPC}_y = \text{Concat}(\text{MHSA},\, \text{DPN}) \tag{9}$$

3.4 Classification-Network

In the final stage, tumor type classification is performed by determining the class with the highest probability at the output layer. Figure 6 illustrates the classification network structure, which consists of four MidResBlock units serving as encoder components (shown in green) that extract features from y into the network. This is followed by a flatten layer and a Dense layer used to generate the final classification output.

In the Dense layer, the vector y produced by the multi-path convolution network contains 1792 features. Two Dense layers are used to process these features, each employing

the ReLU activation function. To improve model generalization, a Gaussian dropout rate of 0.01 is applied. The final Dense layer consists of three neurons to produce the categorical output, followed by a Softmax activation function to determine the class with the highest probability.

To evaluate the classification performance, four main criteria are used: false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN). In the multiclass brain tumor classification case, these values are computed separately for each tumor type, where the target class is treated as positive and all other classes as negative. The overall performance of the proposed model is evaluated using Accuracy, Sensitivity, Specificity, Precision, and F-Score, as represented in Eqs. (10)–(14).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}} \tag{10}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{FN} + \text{TP}} \tag{11}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{12}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{13}$$

$$\text{F-Score} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{14}$$

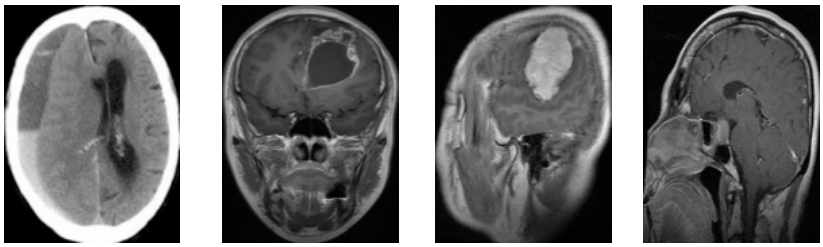
4 Experiment Results

4.1 Cropped Image

In this study, the proposed model was first pretrained on the publicly available ImageNet dataset [11], which contains 14,197,122 images across 1,000 object categories with their corresponding annotations, and then fine-tuned and evaluated on a brain MRI dataset [36] that includes multiple tumor classes in Figure 7 show the cropping results from MRI datasets. To ensure that the model focused on the most relevant regions, all MRI images underwent a preprocessing step involving cropping

4.2 Training Process

All experiments were carried out on an Ubuntu 18.04.6 workstation equipped with an NVIDIA GeForce RTX 3060 GPU (12 GB VRAM) and an AMD Ryzen 5 5600X processor (6 cores, 12 threads) with 32 GB of RAM. The system was configured with CUDA 11.8 for executing the experimental code, while Python 3.8.0 was used within a Docker environment. TensorFlow 2.12 served as the deep learning framework. Model training was performed using the Adam optimization algorithm [37] for 20 epochs with a batch size of 32, a learning rate of 1e-3, and a dropout rate of 0.5. To ensure a fair comparison, all training processes including both the proposed model and the SOTA baseline models [13,14,16,20,30,38]experiments were conducted using models trained from scratch using the ImageNet dataset [11].



a) No Tumor b) Glioma c) Meningioma d) Pituitary

Figure 7: Sample datasets used in the study were: a) no tumor, b) Glioma Tumor, c) Meningioma Tumor, and d) Pituitary.

4.3 Result

The dataset was divided into separate training and testing subsets to ensure an objective evaluation of the proposed model. The model was initialized with pretrained weights, enabling it to learn representative patterns, textures, and structural features from the input images. The training process resulted in a model capable of accurately recognizing and classifying image features. The independent testing subset, which contained images not used during training, was then employed to evaluate the model's performance. This separation prevents bias and ensures that the reported results reflect the model's generalization capability rather than memorization of the training data. Experimental results demonstrate that the proposed approach achieves an accuracy exceeding 90% across all classes, as summarized in Table 1. These findings confirm that the integration of attention blocks and dual-path networks contributes positively to classification performance.

Table 1: Accuracy, Sensitivity, Specificity results obtained in testing

Class Classification	**Menigioma**	**Glioma**	**Pituitary**
Accuracy (%)	97.70	97.52	98.25
Sensitivity (%)	97.81	97.49	98.22
Specificity (%)	98.69	97.31	99.24
Precision (%)	97.58	98.35	99.11
F-Score (%)	97.51	98.04	98.70

Table 1 shows the results obtained from the experiments on each class of tumor that we conducted, these results are the results before being compared with the models from SOTA [13, 14, 16, 20, 30, 38]. From these results, the accuracy value in the meningioma class reached 97.7%, in the Glioma class 97.52%, and in the Pituitary class 98.25% with an average overall value of 97.82%. In addition to the accuracy value obtained, the average overall sensitivity, specificity, precision and F-score obtained were 97.83-98.08%. This proves that the combination of two new blocks can increase the probability value tested to produce better accuracy. We use the Confusion matrix in Table 1 to explain the achievement of the model classification process based on the number of test data samples used.

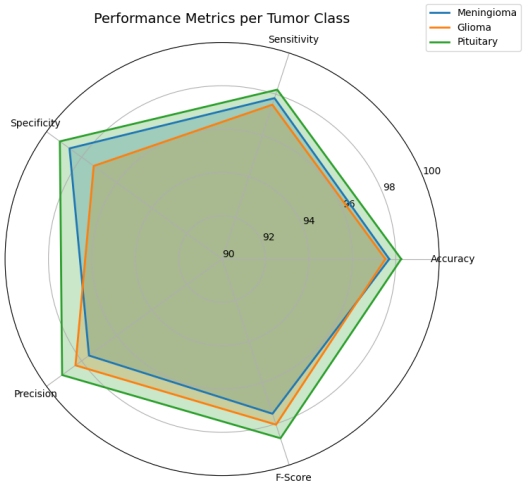


Figure 8: Performance metric results of proposed model.

To provide a more intuitive illustration of the classification performance for each class, the evaluation results from Table 1 are visualized in the form of a Spider Chart, as shown in Figure 8. The diagram presents five key metrics accuracy, sensitivity, specificity, precision, and F-score, each represented on a radial axis. The visualization shows that all three tumor classes meningioma, glioma, and pituitary achieved consistently high performance, with all values exceeding 97%. The pituitary class demonstrates the best performance, particularly in specificity (99.24%) and precision (99.11%), indicating the model's strong capability to detect pituitary tumors accurately with minimal false positives. The glioma class exhibits slightly lower specificity (97.31%), although it still achieves a high precision value (98.35%). Meanwhile, the meningioma class remains stable, with balanced metrics ranging from 97.5% to 98.7%. This visualization confirms that the proposed model provides consistent and well-balanced classification results across all classes, with the most significant advantage observed in pituitary detection.

Table 2: Confusion matrix

Actual Class	**Predicted Class**		
	Meningioma	**Glioma**	**Pituitary**
Meningioma	917	15	5
Glioma	14	902	10
Pituitary	10	5	886

From the results in Table 2, overall, the evaluated classification models showed particularly satisfactory performance in detecting and classifying the three types of tumors. The high sensitivity and specificity values for each type of tumor indicate that our proposed model can recognize several types of classes very accurately. While high specificity shows that the model can accurately identify negative cases, high sensitivity expresses our model's capacity to recognize actual positive cases.

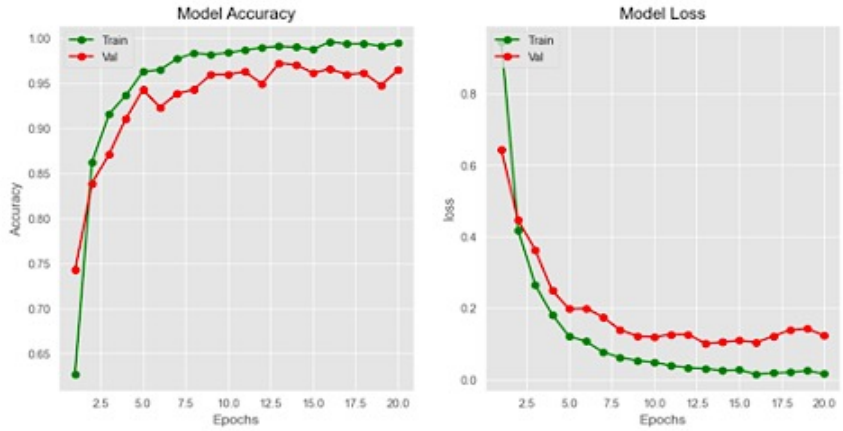


Figure 9: Model accuracy performance against loss from ImageNet dataset.

Figure 9 shows two graphs depicting the model performance of over 20 training epochs. The chart on the left displays the "Model Accuracy" and the chart on the right displays the "Model Loss". In the "Model Accuracy" graph, the vertical axis serves as the accuracy rate, while the horizontal axis speaks for the number of epochs. The accuracy on the training data (Train) is shown by the green curve, and the accuracy on the validation data (Val) is represented by the red curve. From the graph, it can be observed that the training accuracy rises sharply at the beginning and then gradually levels off, stabilizing around 92–97%. The validation accuracy also increases initially but begins to fluctuate and slightly decrease after 15 epochs. To mitigate overfitting, early stopping was applied, and training was halted at epoch 15 despite being scheduled for 20 epochs, ensuring high model accuracy while controlling overfitting. The loss graph illustrates the evolution of the loss values during training, with the horizontal axis representing the number of epochs and the vertical axis representing the loss. The green curve represents the training loss, while the red curve represents the validation loss. Initially, both losses decrease rapidly; however, after approximately 5 epochs, the validation loss begins to rise, suggesting potential overfitting. Overall, Figure 9 demonstrates that the model achieved high accuracy on both training and validation data but exhibited slight overfitting after 15 epochs, as indicated by the increasing validation loss.

Figure 10 shows the Receiver Operating Characteristic (ROC) curves for the classification of brain tumors with three classes: meningioma, glioma, and pituitary based on Table 2. The horizontal axis represents the False Positive Rate (FPR), while the vertical axis represents the True Positive Rate (TPR). Each ROC curve shows the model's performance in distinguishing between the respective tumor class and other classes. Curves closer to the upper left corner indicate better classification performance. Based on the figure, the curves for meningioma and pituitary almost reach the upper left corner with an Area Under the Curve (AUC) value of 0.99 each, while the curve for glioma is slightly lower with an AUC of 0.98. This indicates that the tested model has excellent ability in distinguishing the three types of tumors, with a nearly perfect classification accuracy. The dashed diagonal curve shows random performance as a reference.

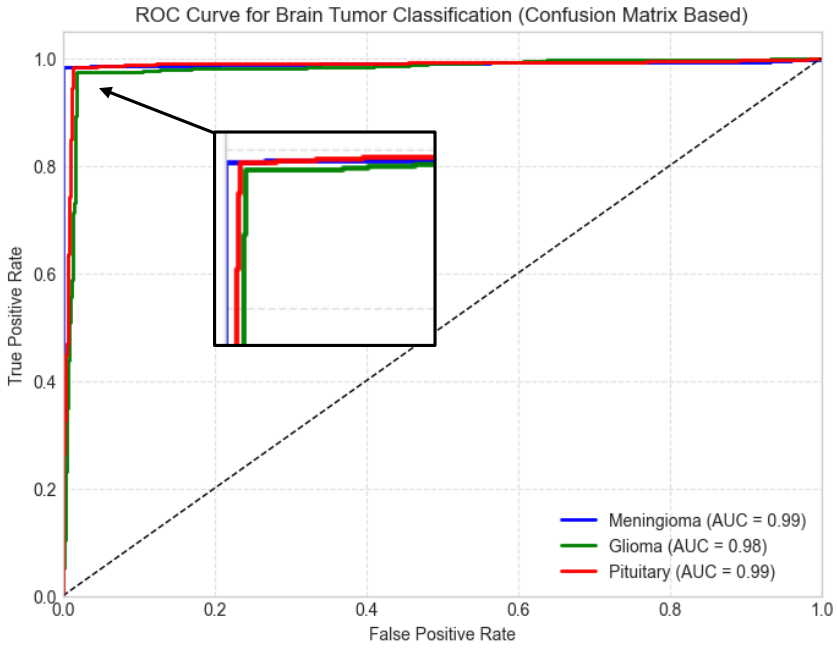


Figure 10: ROC curves of proposed model for brain tumor classification.

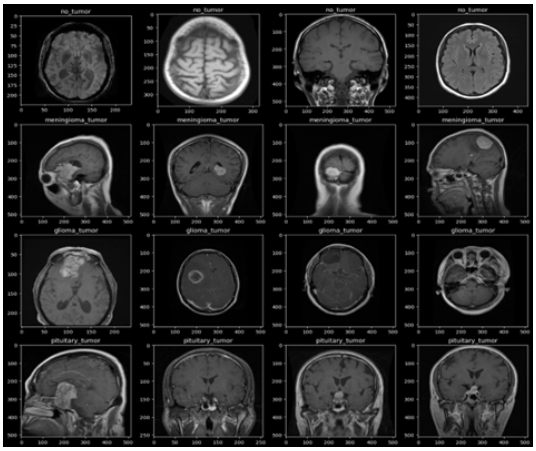


Figure 11: MRI data visualization Coursera dataset [36] classification results of tumor sample testing based on class.

From Figure 11, the visualization results of Magnetic Resonance Imaging MRI data on the human brain are classified into four different categories based on their conditions and classes. There are 16 images divided into a matrix of four rows and four columns, each labelled according to the conditions shown in the image. The first row shows an MRI image labelled "no_tumor" in a normal brain condition without a tumor, the second row shows an

MRI image labelled "meningioma tumor" indicating the presence of a tumor with a meningioma class identified in it, the third row shows an MRI image labelled "glioma_tumor" indicating the presence of a tumor with a glioma class identified, and the last is an MRI image labelled "pituitary_tumor" indicating the presence of a tumor with a Pituitary class identified in it. Each image in the column represents a different view of the brain, such as axial, coronal, and sagittal views. This classification helps in diagnosing different brain conditions with existing conventional methods.

Table 3: Evaluation of the proposed model against EfficientNet and other SOTA models on the MRI dataset

Model	**Accuracy (%)**	**Sensitivity(%)**	**Specificity(%)**	**Precision(%)**	**F-score(%)**
EfficientNetB0 [20]	96.57	96.50	96.60	96.60	96.55
EfficientNetB1 [20]	96.95	96.90	97.00	97.00	96.95
EfficientNetB2 [20]	96.44	96.40	96.50	96.50	96.45
EfficientNetB3 [20]	96.10	96.00	96.20	96.20	96.10
EfficientNetB4 [20]	97.22	97.20	97.30	97.30	97.25
EfficientNetB5 [20]	97.04	97.00	97.10	97.10	97.05
EfficientNetB6 [20]	97.43	97.40	97.50	97.50	97.45
EfficientNetB7 [20]	97.76	97.70	97.80	97.85	97.77
MidResBlock [13]	95.98	X	X	96.08	X
LeNet-based [13]	90.06	X	X	X	X
Hybrid [14]	93.51	90.10	95.28	88.84	89.37
ResNet [16]	90.00	89.90	92.05	92.22	97.78
DenseNet [30]	91.11	90.22	92.05	92.22	91.11
CNN [38]	97.68	96.74	98.16	98.19	97.78
Our Proposed	**97.82**	**97.83**	**98.41**	**98.34**	**98.08**

Table 3 demonstrates that the EfficientNet models (B0-B7) exhibit consistently strong classification performance, with all accuracy values exceeding 90%. Among these models, the EfficientNet B7 achieves the highest performance, accomplishing an accuracy of 97.76% and an F-Score of 97.77%. In contrast, our proposed model, which integrates EfficientNet B7 with Multi-Path Convolutional and Multi-Head Self-Attention (MHSA), achieves the best overall performance over all evaluation metrics. Specifically, the proposed model attains an accuracy of 97.82%, sensitivity of 97.83%, specificity of 98.41%, precision of 98.34%, and an F-Score of 98.08%. These findings show that the proposed model avoids False Positives (FP) and True Negatives (TN) while maintaining a great balance between detecting True Positives (TP) and limiting False Negatives (FN).

Furthermore, the comparison results between the proposed model and other reference models, including LeNet-based, Hybrid, ResNet, DenseNet, and CNN models, using the MRI dataset [36], indicate that the proposed model outperforms the reference models in terms of accuracy and other evaluation metrics [13,14,16,20,30,38]. These results also highlight the key differences between ResNet and EfficientNet when applied through transfer learning. ResNet, although popular for its residual connections and proven effectiveness in feature extraction, typically requires large-scale datasets to achieve stable performance. On limited MRI datasets, ResNet performance often degrades without the support of transfer learning. In contrast, EfficientNetB7, with compound scaling and pretrained weights from ImageNet, is able to maintain high performance even on relatively small datasets. This

efficiency, combined with the integration of DPN and MHSA, explains why the proposed model in this study surpasses the accuracy of ResNet and other architectures in the evaluation table.

However, it is important to note that several factor such as model architecture, dataset type and size, computational resources, and preprocessing techniques (e.g., filter types) can significantly influence the results. Based on the results presented in Table 3, it can be concluded that the integration of Dual Path Network (DPN) blocks and the Multi-Head Self-Attention (MHSA) mechanism significantly improves classification accuracy. The combination of DPN and MHSA within the Multi-Path Convolutional structure enhances the network's ability to extract and capture discriminative features, thereby increasing the probability of correct classification in the network's final stage.

5 Ablation Study

Table 4: Ablation study of the proposed block

Baseline	DPN	MHSA	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F-Score (%)
O	X	X	97.56	97.70	97.80	97.77	O
O	O	X	97.72	97.85	98.02	98.00	O
O	X	O	97.95	98.18	98.08	98.08	O
O	**O**	**O**	**98.15**	**98.34**	**98.15**	**98.34**	**O**

Table 4 presents the results of an ablation study conducted to evaluate the contributions of Dual Path Networks (DPN) and Multi-Head Self-Attention (MHSA) on brain tumor classification using EfficientNetB7 as the baseline model. In this context, "O" indicates that a block is applied, while "X" indicates that a block is not applied. The baseline model alone, without DPN or MHSA, achieved an accuracy of 97.56%, precision of 97.70%, recall of 97.80%, and an F1-score of 97.77%. When DPN was applied, the performance slightly improved, reaching an accuracy of 97.72%, precision of 97.85%, recall of 98.02%, and F-score of 98.00%. Similarly, applying MHSA alone also enhanced the results, particularly in precision (98.18%) and recall (98.08%), yielding an F-score of 98.08%. The combination of both DPN and MHSA produced the best overall performance, with an accuracy of 98.15%, precision of 98.34%, recall of 98.15%, and F-score of 98.34%, indicating that these two modules complement each other by improving feature representation and attention mechanisms for brain tumor classification.

6 Discussion

The proposed model demonstrates strong overall performance, yet class-specific differences are evident. Pituitary tumors achieve the highest scores, particularly in specificity and precision. This advantage can be explained by their consistent anatomical location and clearer structural boundaries in MRI images. In contrast, gliomas often exhibit irregular shapes and diffuse margins, while meningiomas may share visual similarities with surrounding tissues, making them more challenging to classify accurately.

Although the results are encouraging, the ability of the model to generalize to larger and more heterogeneous datasets remains uncertain. In real clinical practice, MRI scans differ widely across institutions, scanners, and acquisition protocols. Such variability may introduce domain shifts that reduce classification reliability. To address this, future studies should evaluate the model using multi-center and multi-modal datasets to ensure robustness and clinical applicability.

Another important consideration is the computational cost of the proposed architecture. The integration of EfficientNetB7, DPN, and MHSA substantially increases the number of parameters and requires significant GPU resources for training and deployment. These requirements may limit real-time use in clinical environments, particularly in resource-constrained settings. Future work should therefore focus on optimizing model efficiency, reducing reliance on high-end hardware, and improving scalability for practical medical applications

7 Conclusion

This study proposes a brain tumor classification model based on an artificial neural network that integrates multi-head self-attention (MHSA), dual-path networks (DPN), and attention blocks. The proposed model demonstrates high accuracy in detecting gliomas, meningiomas, and pituitary tumors, achieving an overall accuracy of 97.82% and outperforming several state-of-the-art approaches [13, 14, 16, 20, 30, 38]. By effectively capturing important features from MRI images, the model significantly improves classification performance. Experimental results show excellent outcomes, with accuracy, sensitivity, specificity, precision, and F-score values of 97.82-98.08%, respectively.

However, this study is limited by the use of a relatively restricted testing dataset, not fully capture the diversity of real-world clinical data. the integration of EfficientNetB7 with DPN and MHSA substantially increases model complexity and computational costs. In addition, although the model achieved high accuracy, further validation on larger and multi-center datasets would be valuable to confirm its robustness across different imaging conditions. Future work may also focus on optimizing the computational efficiency of the proposed architecture and exploring model interpretability to further support its clinical applicability. Nevertheless, the achieved performance indicates that the proposed model holds strong potential for supporting medical detection and diagnosis, improving the accuracy of brain tumor identification, and enabling further research in medical image classification through deep learning and artificial intelligence (AI), which in the future could address data with greater variability and complexity.

Acknowledgments

The authors sincerely thank their colleagues and mentors for their collaborative support and constructive feedback throughout this research. They also gratefully acknowledge the use of the publicly available MRI [36] and the ImageNet [11] datasets pretrained model, which have greatly facilitated the development and evaluation of the proposed model.

References

[1] S. M. Alqhtani, T. A. Soomro, A. A. Shah, A. A. Memon, M. Irfan, S. Rahman, M. Jalalah, A. H. Almawgani, and L. A. B. Eljak, "Improved brain tumor segmentation and classification in brain mri with fcm-svm: a diagnostic approach," *IEEE Access*, vol. 12, pp. 61312–61335, 2024.

[2]

[3] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2097–2106, 2017.

[4] G. Çınarer and B. G. Emiroğlu, "Classificatin of brain tumors by machine learning algorithms," in *2019 3rd international symposium on multidisciplinary studies and innovative technologies (ISMSIT)*, pp. 1–4, IEEE, 2019.

[5] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: a literature review," *BMC medical imaging*, vol. 22, no. 1, p. 69, 2022.

[6] R. S. Ayon, J. R. Mou, S. H. Majed, and R. Rifat, "Brain tumor segmentation and classification using spatial fuzzy c mean and quadratic support vector machine," in *2019 3rd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*, pp. 233–236, IEEE, 2019.

[7] V. Wasule and P. Sonar, "Classification of brain mri using svm and knn classifier," in *2017 third international conference on sensing, signal processing and security (ICSSS)*, pp. 218–223, IEEE, 2017.

[8] S. Asif, W. Yi, Q. U. Ain, J. Hou, T. Yi, and J. Si, "Improving effectiveness of different deep transfer learning-based models for detecting brain tumors from mr images," *IEEE Access*, vol. 10, pp. 34716–34730, 2022.

[9] M. Rizwan, A. Shabbir, A. R. Javed, M. Shabbir, T. Baker, and D. A.-J. Obe, "Brain tumor and glioma grade classification using gaussian convolutional neural network," *IEEE Access*, vol. 10, pp. 29731–29740, 2022.

[10] 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. 65426–65438, 2022.

[11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[12] I. Abd El Kader, G. Xu, Z. Shuai, S. Saminu, I. Javaid, and I. Salim Ahmad, "Differential deep convolutional neural network model for brain tumor classification," *Brain Sciences*, vol. 11, no. 3, p. 352, 2021.

[13] Z. Sobhaninia, N. Karimi, P. Khadivi, and S. Samavi, "Medial residual encoder layers for classification of brain tumors in magnetic resonance images," in *2022 30th International Conference on Electrical Engineering (ICEE)*, pp. 442–445, IEEE, 2022.

[14] Z. Sobhaninia, N. Karimi, P. Khadivi, and S. Samavi, "Medial residual encoder layers for classification of brain tumors in magnetic resonance images," in *2022 30th International Conference on Electrical Engineering (ICEE)*, pp. 442–445, IEEE, 2022.

[15] H. Zhang, H. Lai, Y. Wang, X. Lv, Y. Hong, J. Peng, Z. Zhang, C. Chen, and C. Chen, "Research on the classification of benign and malignant parotid tumors based on transfer learning and a convolutional neural network," *Ieee Access*, vol. 9, pp. 40360–40371, 2021.

[16] M. Zou, M. Ma, A. Wang, Y. Li, and T. T. Toe, "Brain tumor classification model of resnet-50 network based on different data enhancement algorithms," in *2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP)*, pp. 1011–1016, IEEE, 2023.

[17] J. Ramasamy, R. Doshi, and K. K. Hiran, "Detection of brain tumor in medical images based on feature extraction by hog and machine learning algorithms," in *2022 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT)*, pp. 1–5, IEEE, 2022.

[18] F. Fang, X. Hu, J. Shu, P. Wang, T. Shen, and F. Li, "Text classification model based on multi-head self-attention mechanism and bigru," in *2021 IEEE conference on telecommunications, optics and computer science (TOCS)*, pp. 357–361, IEEE, 2021.

[19] V. Ashish, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, p. I, 2017.

[20] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International conference on machine learning*, pp. 10096–10106, PMLR, 2021.

[21] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International conference on machine learning*, pp. 10096–10106, PMLR, 2021.

[22] M. Kumar, P. Samui, D. J. Armaghani, *et al.*, "A novel approach to estimate rock deformation under uniaxial compression using a machine learning technique," *Bulletin of Engineering Geology and the Environment*, vol. 83, no. 7, pp. 1–14, 2024.

[23] N. Nazeer, B. Wajid, I. Nazir, and F. Gohar, "Prediction of malignancy of brain cancer on seer dataset using random forest, svm, and naive bayes classifiers," in *2020 IEEE 23rd International Multitopic Conference (INMIC)*, pp. 1–5, IEEE, 2020.

[24] X. Wang, F. Buontempo, A. Young, and D. Osborn, "Induction of decision trees using genetic programming for modelling ecotoxicity data: adaptive discretization of real-valued endpoints," *SAR and QSAR in Environmental Research*, vol. 17, no. 5, pp. 451–471, 2006.

[25] Z. Jin, J. Shang, Q. Zhu, C. Ling, W. Xie, and B. Qiang, "Rfrsf: Employee turnover prediction based on random forests and survival analysis," in *International Conference on Web Information Systems Engineering*, pp. 503–515, Springer, 2020.

[26] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[27] M. Benaouali, M. Bentoumi, M. Touati, A. T. Ahmed, and M. Mimi, "Segmentation and classification of benign and malignant breast tumors via texture characterization from ultrasound images," in *2022 7th International Conference on Image and Signal Processing and their Applications (ISPA)*, pp. 1–4, IEEE, 2022.

[28] M. I. Khalil, S. Tehsin, M. Humayun, N. Jhanjhi, and M. A. AlZain, "Multi-scale network for thoracic organs segmentation.," *Computers, Materials & Continua*, vol. 70, no. 2, 2022.

[29] M. F. Almufareh, M. Imran, A. Khan, M. Humayun, and M. Asim, "Automated brain tumor segmentation and classification in mri using yolo-based deep learning," *IEEE Access*, vol. 12, pp. 16189–16207, 2024.

[30] Z. Atha and J. Chaki, "Ssbtcnet: semi-supervised brain tumor classification network," *IEEE Access*, vol. 11, pp. 141485–141499, 2023.

[31] A. Abufadel, G. Slabaugh, G. Unal, L. Zhang, and B. Odry, "Interacting active rectangles for estimation of intervertebral disk orientation," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 1, pp. 1013–1016, IEEE, 2006.

[32] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, 2015.

[34] V. Pham, C. Pham, and T. Dang, "Road damage detection and classification with detectron2 and faster r-cnn," in *2020 IEEE International Conference on Big Data (Big Data)*, pp. 5592–5601, IEEE, 2020.

[35] C. Lu, K. Ye, G. Xu, C.-Z. Xu, and T. Bai, "Imbalance in the cloud: An analysis on alibaba cluster trace," in *2017 IEEE International Conference on Big Data (Big Data)*, pp. 2884–2892, IEEE, 2017.

[36] Coursera, "Introduction to neurohacking in R." https://www.coursera.org/, n.d. Online Course.

[37] K. D. B. J. Adam *et al.*, "A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, vol. 1412, no. 6, 2014.

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