Jurnal Ilmiah KURSOR
Menuju Solusi Teknologi Informasi
Vol. 13, No. 3, Juli 2026
ISSN 0216 – 0544
e-ISSN 2301– 6914

Optimization EfficientNetV2 model variant using Grad-CAM for multiple MRI brain tumor classification

Denisa Septalian Alhamda [a], Wahyu Syaifullah J [b], Prasetyaning Estu Pratiwi [c], Surjo Hadi[d], Wan Suryani Wan Awang [e], I Gede Susrama Mas Diyasa [f]

[a,b]Department of Data Science, Faculty of Computer Science, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia
[c] Department of Medical Education, Faculty of Medicine, University of Pembangunan Nasional Veteran Jawa Timur Indonesia
[d] Department of Industrial Engineering, Faculty of Engineering, University of Yos Soedarso Surabaya, Indonesia
[e] Faculty of Informatics and Computing, University Sultan Zainal Abidin Besut Campus, 22200 Besut, Terengganu, Malaysia
[f] Department Master of Information Technology, Faculty of Computer Science, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya,Indonesia
E-mail: 21083010113@student.upnjamtim.ac.id, wahyu.s.j.saputra.if@upnjatim.ac.id, prasetyaning.estu.fk@upnjatim.ac.id, bestsuryo@yahoo.com, suryani@unisza.edu.my, igsusrama.if@upnjatim.ac.id

Abstract

Fast and accurate diagnosis plays a critical role in effectively treating brain tumors. This study optimized and evaluated the EfficientNetV2 architecture through transfer learning, fine-tuning, and data augmentation, using three variants Small, Medium, and Large to classify MRI images into four categories: glioma, meningioma, pituitary tumors, and no tumor. Grad-CAM visualization was employed to enhance interpretability, providing a clear view of the critical regions in the MRI images that influenced the model's decisions. Grad-CAM was tested across all model variants, and the best results were observed with EfficientNetV2-Large, where the model successfully highlighted the key areas associated with brain tumors. Among the variants, EfficientNetV2-Large achieved the best performance, with 99.85% accuracy, 99.60% precision, 99.65% recall, and 99.50% F1-score. However, this model required the longest computation time of 288 seconds per step, which may not be feasible in resource- constrained environments. Overall, this study underscores the potential of EfficientNetV2 models in revolutionizing brain tumor diagnosis by balancing accuracy, efficiency, and interpretability through advanced optimization techniques.

Key words: Brain tumors, MRI classification, EfficientNetV2, Grad-CAM, Deep learning.

INTRODUCTION

In the world of health, there are various types of diseases that can threaten human life, one of which is a brain tumor. A brain tumor is a condition in which cells in or around the brain grow abnormally and uncontrollably.

This type of tumor is very dangerous, accounting for about 1.35% of all malignant tumor cases and contributing to 29.5% of cancer deaths globally [1]. Based on Alther's research in 2020 [2], Symptoms of brain tumors can differ based on the tumor's size and location, often resulting in increased intracranial pressure over time. Brain tumors are classified into two main categories: primary brain tumors, which develop from the brain's own cells, and secondary brain tumors, which arise from cancer that has spread from other parts of the body [3].

The number of brain tumor cases worldwide continues to rise significantly. Data from the Global Cancer Observatory (2020) shows that there are 308,102 cases of brain tumors worldwide, with the highest distribution in Asia at 54.2% and a high global mortality rate, reaching 71% for patients under the age of 60 [4]. Unfortunately, research efforts and early detection of brain tumors are still limited. Conventional techniques such as biopsy and manual examination are time-consuming and usually require more than one specialist to confirm the diagnosis [5]. This increases the need for more efficient approaches to aid the diagnosis process.

Artificial intelligence has demonstrated significant potential in solving diverse challenges within the medical field, particularly in medical image classification. In the diagnosis of brain tumors, MRI imaging is essential due to its ability to show a clear distinction between soft and hard tissues in the brain [6]. Machine learning and deep learning-based approaches enable the analysis of MRI images with high accuracy, which is not always easy to achieve with conventional methods. An algorithm frequently utilized in medical image classification is the Convolutional Neural Network (CNN), which excels at identifying intricate patterns within medical image data [7][8].

EfficientNet has emerged as a favored CNN architecture due to its capability to balance computational efficiency and accuracy effectively. The model is designed to balance input depth, width, and resolution, enabling efficient use in a wide range of model sizes, from EfficientNet-B0 to EfficientNet-B7 [9]. In 2021, Google launched EfficientNetV2 which offers improvements in training and inference speed and accuracy. Recent research has shown that EfficientNetV2 in combination with global attention mechanisms can achieve very high accuracy in MRI-based brain tumor classification [10].

EfficientNetV2 is a model developed by Google in 2021 as an advanced version of EfficientNet with significant improvements in training efficiency and speed. The model is designed using a balanced scaling approach, where the width, depth, and resolution of the input are optimized simultaneously to achieve better results in image classification. EfficientNetV2 offers three main variants Small (S), Medium (M), and Large (L), each optimized for different computational requirements. Unlike the previous versions, EfficientNetV2 uses a technique called Fused-MBConv, which combines depthwise separable block convolution with standard convolution to speed up inference. In research conducted by Huang in 2022 EfficientNetV2 showed improved accuracy of up to 98.84% on medical datasets with complex structures, such as X-rays and CT-Scans, compared to other models with comparable parameters [11]. It is also reported to be faster in training than other deep learning architectures, which makes it ideal for medical applications that require high accuracy and computational efficiency. These advantages make EfficientNetV2 a good choice in brain tumor classification studies, where precision and speed are crucial in supporting the medical diagnosis process.

Previous research has shown the great potential of EfficientNet architecture in improving the accuracy and efficiency of medical diagnosis. Anagun and colleagues in 2023 introduced a CNN-based brain tumor diagnosis system using EfficientNetV2-S. The system demonstrated optimal performance with 99.85% accuracy and 98.21% F1-score in brain tumor classification, surpassing various other CNN architectures such as ResNet18 and InceptionV4 [12]. In addition, Al Moteri's research in 2024 showed that the modification of EfficientNetV2-S with a cyclic learning rate strategy successfully increased the accuracy of breast cancer classification to 99.68%. This research also emphasizes the importance of accessibility in medical diagnosis, focusing on the application of this system for patients with disabilities [13]. These two studies confirm that the utilization of EfficientNetV2, especially the optimized variant, is able to improve precision and efficiency in medical image classification, making it a superior choice in the development of intelligent diagnosis systems.

Previous studies have emphasized the significance of combining preprocessing techniques with deep learning models to enhance the accuracy of brain tumor classification. Abedini and colleagues in 2024 developed a method using Segment Anything Model (SAM) followed by CNN- based classification, which successfully focused the analysis on the lesion by removing the background from the MRI image. This method has substantially improved the accuracy of tumor detection and classification during the initial stages of diagnosis [12]. Ji-hyeon Lee in 2023 also reported that the application of a Gaussian filter was able to reduce noise in MRI images, while data augmentation with AutoAugment enriched the variety of training data, thus improving the performance of the EfficientNetV2L model by 2.87% compared to the original data without augmentation [13]. These findings emphasize that the combination of preprocessing methods and advanced deep learning architectures is an essential component to support faster and more accurate diagnosis in the medical context.

In a literature review conducted in 2024, Pacal and colleagues introduced a novel approach leveraging the EfficientNetv2 architecture integrated with a Global Attention Mechanism (GAM) and Efficient Channel Attention (ECA) for accurate classification of brain tumors using MRI images. This technique is designed to enhance the model's ability to focus on essential features within intricate MRI scans, thereby increasing both the accuracy and efficiency of tumor detection. The method achieved an impressive accuracy rate of 99.76% on the utilized public dataset. Additionally, the Grad-CAM technique was employed to provide visual insights, offering a transparent and interpretable explanation of the model's decision-making process, which is crucial for clinical diagnostics. This study underscores the potential of deep learning architectures in medical imaging, particularly for brain tumor classification, and highlights the critical role of attention mechanisms in improving the precision and interpretability of AI-driven diagnoses [14].

This study optimizes and evaluate EfficientNetV2 architecture for brain tumor classification by experimenting with three model variants: Small, Medium, and Large. A significant challenge in this domain lies in analyzing MRI images captured from multiple perspectives, such as top, back, and side views, which can substantially influence model performance. The researchers adopted a transfer learning approach to address this, leveraging pre-trained weights and biases from EfficientNetV2 models trained on large- scale datasets like ImageNet. further optimization was conducted by fine-tuning hyperparameters and utilizing the EfficientNetV2 balanced scaling approach, which adjusts input resolution, depth, and width to achieve superior performance. Data augmentation techniques were also employed to create diverse variations of the MRI dataset, enhancing the model's ability to generalize from limited multi-sectional data. Grad-CAM (Gradient-weighted Class Activation Mapping) was integrated to improve interpretability, enabling the visualization of key regions in MRI images that contributed most significantly to classification decisions. This integration enhances model transparency and provides valuable insights for clinical validation. EfficientNetV2 was selected for its superior training efficiency and parameter optimization compared to earlier architectures. It is an ideal candidate for delivering faster, more accurate, and resource-efficient solutions to support clinical diagnoses. The combination of transfer learning, hyperparameter fine-tuning, data augmentation, and interpretability strategies enables the model to extract features effectively and improve the accuracy and efficiency of brain tumor classification.

MATERIAL AND METHODS

Figure 1 is the research flow in this study designed systematically to optimize the variant of the EfficientNetV2 model for brain tumor classification using multi-sectional MRI images [15]. These images are organized into four distinct categories: Glioma, Meningioma, Absence of Tumor, and Pituitary. The next step is MRI image preprocessing, where noisy images are enhanced using a denoising autoencoder, resulting in reconstructed, cleaner images. In addition, To enhance the dataset's diversity and reduce the risk of overfitting during training, data augmentation techniques like rotation and flipping are employed.

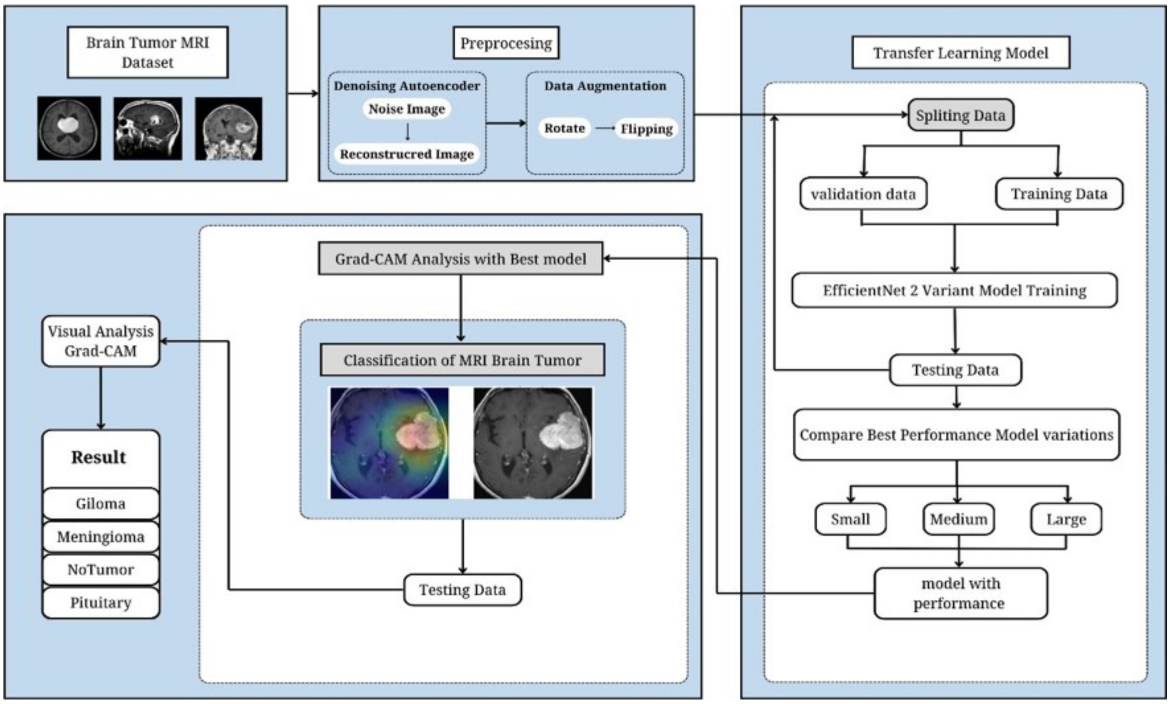

Fig. 1. Research workflow MRI classification

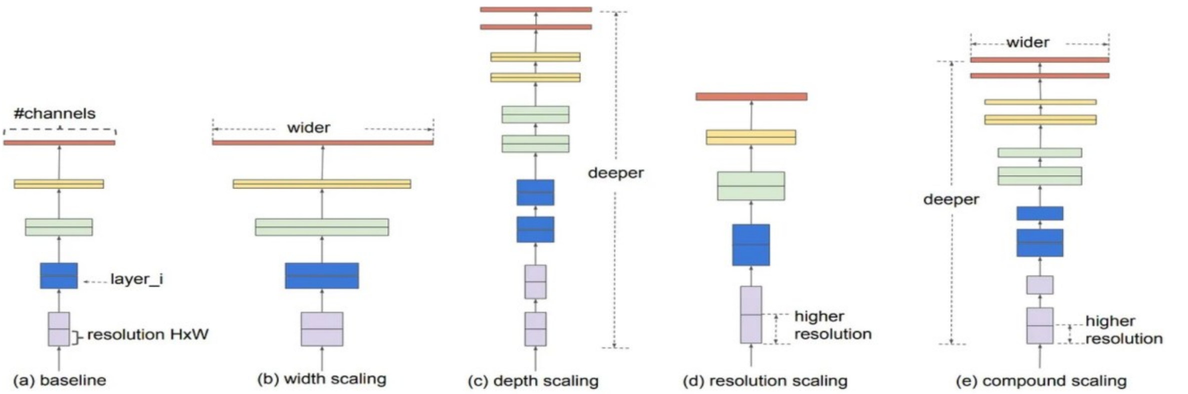

Fig. 6. Scaling the depth, width, and image resolution to create different variations of the EfficientNet model

Following preprocessing, a transfer learning approach is applied. The dataset is divided into training, validation, and test subsets, and EfficientNetV2 variants (Small, Medium, and Large) are trained using pre-trained weights from ImageNet to exploit the model's feature extraction capabilities. Each model variant is evaluated for performance using accuracy and training time efficiency metrics to assess computational feasibility for clinical use. Gradient-weighted Class Activation Mapping (Grad-CAM) is applied to the best- performing model to improve interpretability. Grad-CAM visualizes critical regions within the MRI images that contribute most to the classification decisions, providing insights that meet clinical validation requirements. Finally, the classification results are presented, categorizing the MRI images into one of the four tumour classes. The integration of Grad-CAM ensures transparency and interpretability, overcoming key challenges in multi-slice image analysis and improving the accuracy and reliability of brain tumour classification in real-time clinical settings.

Data Acquisition

At the initial stage, MRI image data was The dataset, sourced from Kaggle, includes MRI scans of various brain tumor types, glioma, meningioma, and pituitary tumors along with non-tumorous conditions. It comprises 7,023 grayscale images in ".jpg" format (512x512 pixels), distributed as 1,621 glioma, 1,645 meningioma, 2,000 no-tumor, and 1,757 pituitary tumor images. These images, captured from multiple perspectives and brain slices, allow the model to learn diverse patterns for tumor classification effectively.

Meningioma tumors develop from the protective layers encasing the central nervous system, whereas pituitary tumors arise near the gland situated between the brain and the nasal cavity. Both are benign, meaning they do not spread to other tissues, unlike gliomas, which are malignant and can metastasize to other parts of the body [16]. With its comprehensive representation of benign and malignant tumor types, this dataset provides a strong basis for training and testing classification models. Figure 2 visually depicts the tumor types and their respective brain locations, helping to highlight the structural and positional distinctions critical for accurate classification. For the example datsaet is shown in Figure 2.

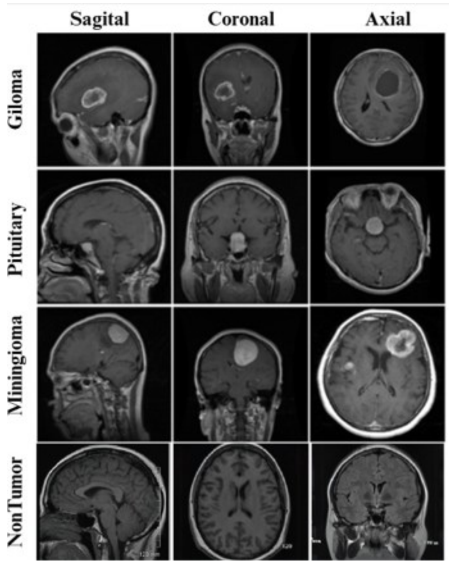

Fig. 2. Dataset images for each class

Denoising Data and Risizes MRI image

The Denoising Autoencoder improves the quality of MRI images by eliminating noise from the image data [17][18]. This model processes noisy MRI images as input and reconstructs clearer, higher- quality versions through an encoding and decoding mechanism. The primary objective of this approach is to retain essential information in the MRI images while minimizing visual distortions that could hinder the accuracy and performance of the classification model [19].

Next step aims to reduce the dimension of the image data without removing important features. The main purpose of this resizing is to improve computational efficiency during the data training process. output result of the MRI image rescaling process using the open cv library show in Figure 4. Specifically, the image size was changed from 512 x 512 pixels to 150 x 150 pixels. Smaller images allow training to take place faster and can accelerate convergence [20]. The use of smaller images also helps the network to generalize better as less data is used, resulting in a lower risk of overfitting.

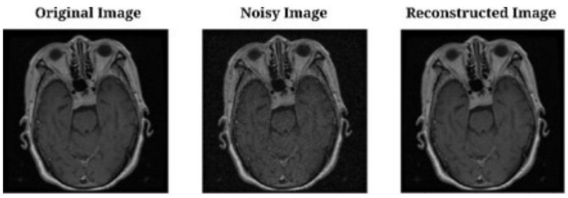

Fig. 3. Processing denoising MRI image

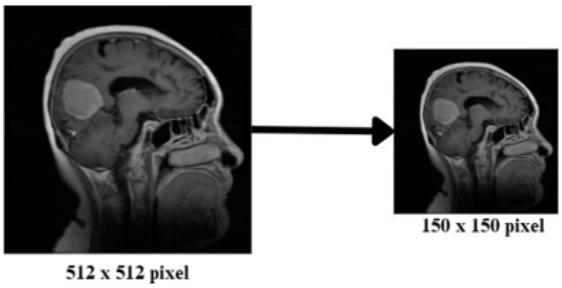

Fig. 4. Resizing the MRI image

Data Augmentation

The process of data augmentation seeks to enhance the diversity of the image dataset through multiple transformations, contributing to better classification performance by the model. In this study, the augmentation techniques applied include horizontal and vertical image flipping, brightness adjustment of 0.2, zoom with a range of 0.2 for x and y axes, translation or image shift of 0.2, and random rotation of up to 20% of the original image [21]. Figure 5 is the output of the results of the data augmentation process used.

This augmentation technique is designed to introduce variations in the data more dynamically, thus supporting the EfficientNetV2 Meidun and Large model in learning from more varied data and strengthening the model's capacity to generalize to unseen data. With this approach, classification performance is expected to be optimized as the risk of overfitting is lowered.

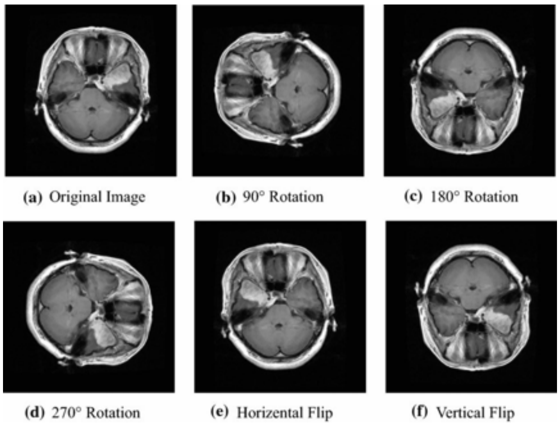

Fig. 5. Data augmentation output

EfficienetV2 Model Architecture

EfficientNetV2 models use a combined approach with Fused-MBConv and MBConv. Fused-MBConv is a combination of depthwise and standard convolutions that improves training efficiency in the initial layer, while MBConv is predominant in the final layer to expand the depth of the model [16][22]. Each model (S, M, L) follows a compound scaling rule, which balances depth, width, and image size for optimal performance, similar to EfficientNetV1 but with additional adjustments. Scaling efficiency and accuracy adjustments usually depend on the architecture of the model.

A. EfficientNetV2-S (Small): Designed with a lower number of parameters and smaller image size, making it more computationally efficient and suitable for small-scale tasks [23].
B. EfficientNetV2-M (Medium): This model adds additional layers and channels, making it more suitable for applications with medium requirements [24].
C. EfficientNetV2-L (Large): With greater depth and width, this model is geared towards tasks that require more extensive feature extraction while maintaining efficient training speed[25].

Figure 6 represents the scaling up of deep learning models, especially on the EfficienrNetV2 architecture.It is important to understand the difference between the two types of convolution blocks often used in convolutional neural networks for computational efficiency and performance, especially on devices with limited computing power. Mobile Bottleneck Convolution (MBConv) and Fused-MBConv blocks are popular architectures in modern deep learning models such as EfficientNet. They are designed to reduce the number of parameters and speed up the computational process while maintaining accuracy. MBConv uses a depthwise convolution layer to reduce the number of operations, while Fused-MBConv removes this layer and combines it with regular convolution for more specific applications. Figure 7, Mobile Bottleneck Convolution (MBConv) and Fused-MBConv.

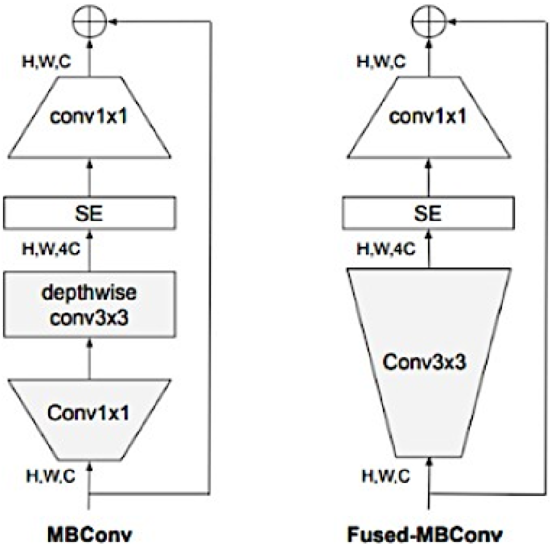

Fig. 7. Structure of MBConv and Fused-MBConv blocks

The EfficientNetV2 architecture adopts a progressive learning strategy, which dynamically adjusts key training parameters such as image size and regularization settings (e.g., dropout rate) as training progresses [26]. This innovative approach ensures that the model learns efficiently by starting with smaller images and less stringent regularization, which accelerates initial convergence. As training continues, the image size and regularization levels gradually increase, enabling the model to refine its feature extraction and improve generalization. This staged progression maintains a delicate balance between overfitting and underfitting at each phase of training, allowing the model to achieve high accuracy while optimizing computational efficiency. Architecture model EfficieNetV2 show in Figure 8.

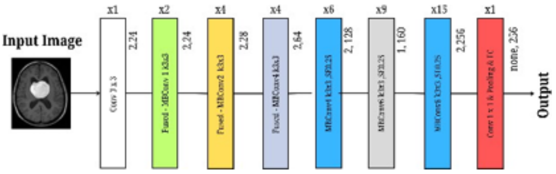

Fig. 8. Architecture of EfficientNetV2 model

The EfficientNetV2 family of models leverages a sophisticated scaling method known as compound scaling, which ensures a balanced growth of the model's dimensions depth, width, and input resolution. This scaling approach is mathematically expressed by the formula in Equation (1).

$$F(\alpha,\beta,\gamma) = \alpha^d x \beta^w \; x \; \gamma^r \tag{1}$$

α = scale factor for network depth
γ = scale factor for resolution
β = scale factor for width

with alpha, beta, and gamma as constants determined through architectural search to optimize the balance between accuracy and computational efficiency EfficientNetV2 architectural approach and scaling principles provide significant performance improvements on datasets such as ImageNet, CIFAR-10, and CIFAR-100, surpassing previous ConvNet and Transformer models in both speed and accuracy.

Grad-CAM Visualization

Gradient-weighted Class Activation Mapping (Grad-CAM) technique is a visualization method that identifies and highlights key areas in MRI images, such as tumor regions, which are crucial in the classification process [27]. This technique allows the model to focus on critical features by utilizing spatial attention generated from Grad-CAM activation maps [28]. The structure and implementation of Grad-CAM are illustrated in Figure 9.

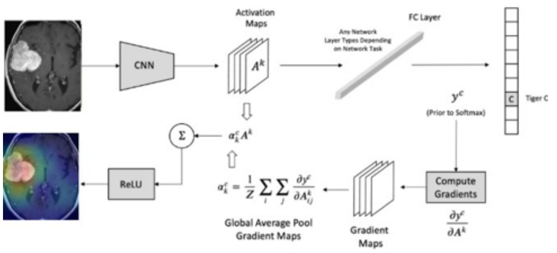

Fig. 9. Grad-CAM architecture CNN

Metrics Evaluating Model Performance

In assessing the predictive performance of a classification system, evaluation methods are needed, one of which is the confusion matrix. Confusion matrix displays the prediction results by comparing the classification produced by the system with the actual classification. A confusion matrix assesses a model's performance by providing essential metrics, including accuracy, precision, recall, and the F1-score, all based on the actual observed values [29]. the values within the measurement matrix consist of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Each event probability that should be positive (P) or negative (N) is represented in these values. These values are then used to calculate evalution metrics of Equation (2),(3),(4),(5).

$$Accuracy = \frac{TP+FN}{TP+FP+FN+TN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

RESULT AND DISCUSSION

The training settings for the model experiments ensured consistency in parameter configuration across all architectures. Experiments were conducted with a batch size of 32, an initial learning rate of 0.1, a minimum learning rate of 1e-6, and a maximum learning rate of 4e-5. The training process was set to multiple epochs but used early stopping to stop training when performance reached a plateau or started to degrade, followed by fine tuning training using ImageNet transfer lerning to ensure optimal efficiency. For evaluation, accuracy and validation loss are monitored, with cross- entropy used as a loss function to guide the optimization process.

Best Variant Model Result

Advantage of this model Table 1 shows the number of parameters in various model variations: Small, Medium, and Large. As can be seen, the number of parameters increases significantly with model size, from Small to Large. The increase in parameters leads to higher model complexity, necessitating additional computational resources for both training and execution [30]. As a result, this increase in parameters leads to extended training times, higher memory demands, and reduced inference speed.

Table 1. Model parameter EfficientNetV2

Variants	Total Parameters	Trainable Parameters	Non-Trainable Parameters
Small	20,336,484	20,331,360	5,124
Medium	53,155,512	53,150,388	5,124
Large	117,751,972	117,746,848	5,124

After identifying the parameters, the The data will be trained using the EfficientV2 model. This training is performed on both the regular model and a model that has been fine-tuned with ImageNet data. The graphs illustrate the accuracy of each dataset, with the blue line indicating the performance of the training data and the orange line showing the performance of the validation data. The results are as follows Figures 10, 11 and 12 show two graphs each illustrating the output process of the training data and the tuned data.

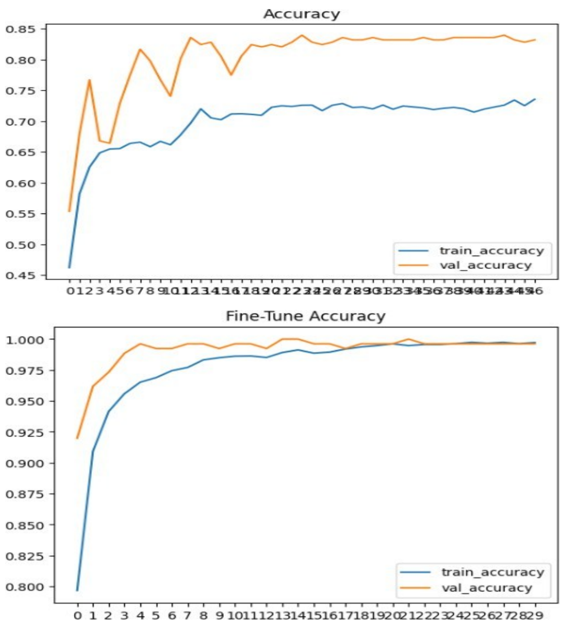

Fig. 10. Accuracy chart EfficientNetV2-S

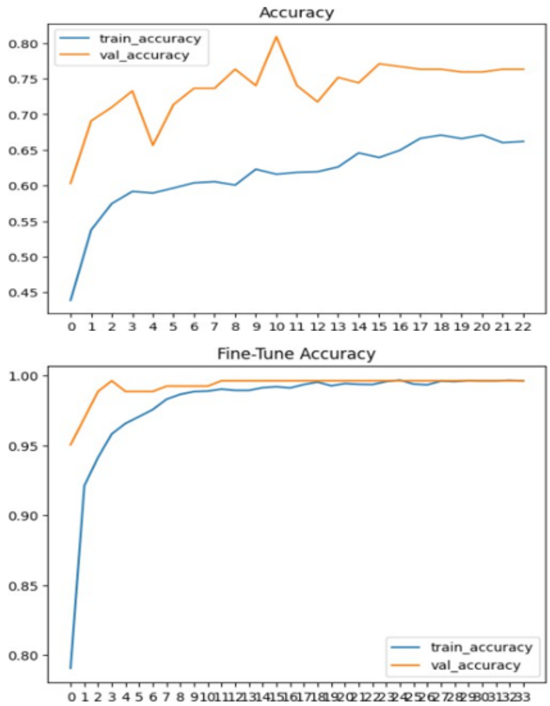

Fig. 11. Accuracy chart EfficientNetV2-M

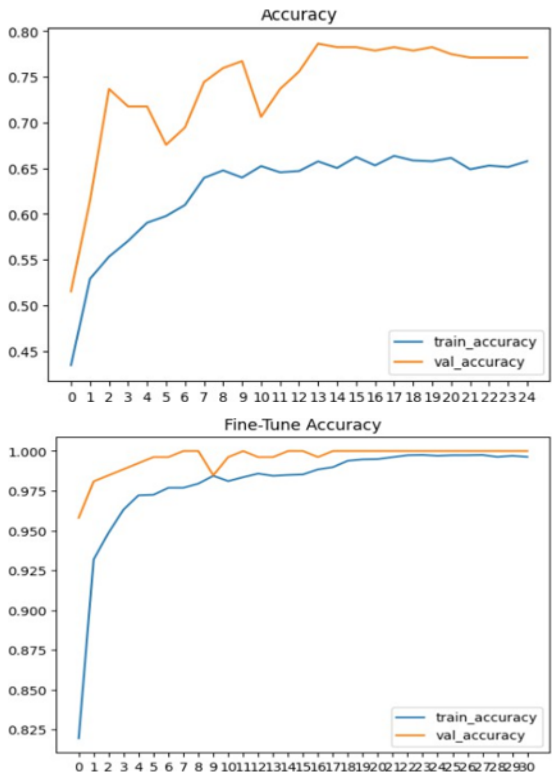

Fig. 12. Accuracy chart EfficientNetV2-L

Figures 10, 11, and 12 present the accuracy charts for EfficientNetV2-S, EfficientNetV2-M, and EfficientNetV2-L during the initial training and fine-tuning phases, demonstrating the effectiveness of progressive training and fine-tuning strategies. During the initial training phase, EfficientNetV2-S (Figure 10) shows a steady improvement in training accuracy, with validation accuracy increasing gradually, indicating effective generalization. EfficientNetV2-M (Figure 11) exhibits a similar pattern, although validation accuracy fluctuates slightly due to the model's increased complexity. EfficientNetV2-L (Figure 12), the most significant variant, achieves the most robust performance, consistently improving validation accuracy despite minor noise during training. In the fine-tuning phase, all models demonstrate rapid convergence, with EfficientNetV2-L achieving near-perfect validation accuracy close to 100%, outperforming the more minor variants.

The fine-tuning process is critical in optimizing these models' performance. This phase involves unfreezing the top layers of the pre-trained model while keeping earlier layers frozen to retain their general feature extraction capabilities. Specifically, layers before the 100th layer remain non-trainable, while the top layers are unfrozen and trained further, allowing the model to specialize in brain tumor classification. The model mitigates overly confident predictions and stabilizes training by introducing label smoothing in the categorical cross-entropy loss function. Additionally, the learning rate is reduced by a factor of 10 (BASE_LR/10) to ensure precise weight updates, and class weighting addresses any class imbalance in the dataset.

The fine-tuning phase is extended for an additional 100 epochs starting from the last epoch of the initial training phase. This extended training period and adaptive callbacks allow the model to converge to its optimal state. EfficientNetV2-S reaches a validation accuracy exceeding 98%, while EfficientNetV2-M stabilizes above 99%. The EfficientNetV2-L variant outshines the others, achieving near-perfect validation accuracy and highlighting its capacity to handle complex classification tasks at the cost of increased computational requirements.

The combined use of progressive learning, compound scaling, and fine-tuning ensures that the EfficientNetV2 models achieve state-of-the-art performance. While smaller models like EfficientNetV2-S are more efficient and suitable for resource- constrained environments, more significant variants like EfficientNetV2-L provide unparalleled accuracy, making them ideal for high-precision applications. These results underscore the importance of leveraging fine- tuning techniques, adaptive learning strategies, and scalable architectures for achieving robust performance in brain tumor classification. After looking at the accuracy of each model, the next step is to look at the confusion matrix value of each model, which can be seen in Figure 13,14 and 15.

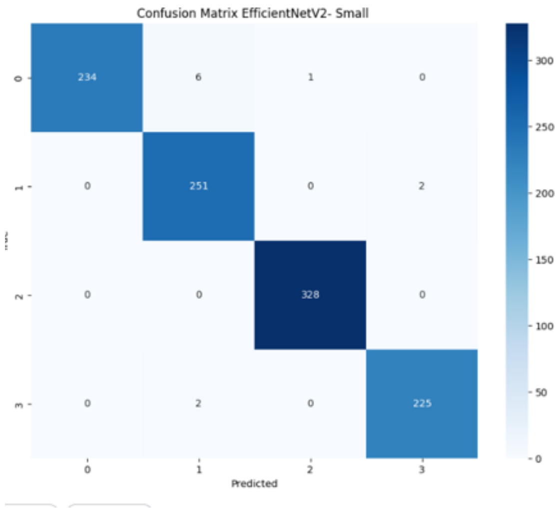

Fig. 13. Confusion matrix EfficientNetV2 - S

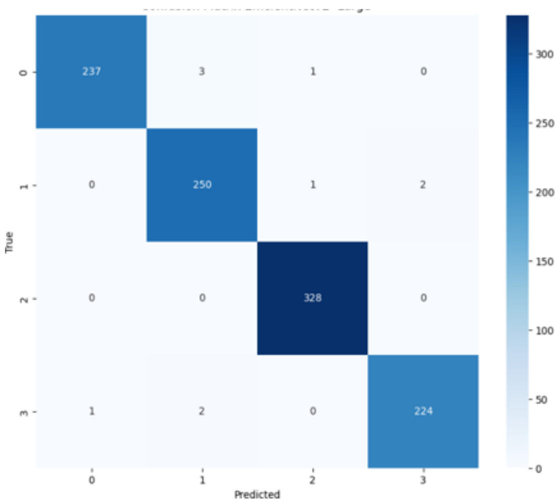

Fig. 14. Confusion matrix EfficientNetV2 - M

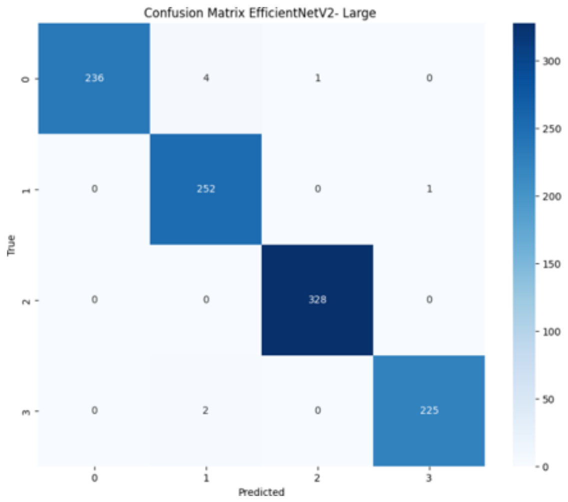

Fig. 15. Confusion matrix EfficientNetV2 - L

The confusion matrix analysis confirms that classification performance improves with increasing model complexity across the EfficientNetV2 variants. For the EfficientNetV2-Small (S) model, the results are promising, with high values concentrated along the main diagonal, demonstrating the model's capability to classify most instances correctly. However, some misclassifications occur, as indicated by the non-zero values outside the diagonal. While these errors are relatively few, they highlight the limitations of the smaller model in handling more complex or ambiguous cases.

The EfficientNetV2-Medium (M) model shows a noticeable improvement compared to the Small variant. The main diagonal contains a higher proportion of correct predictions, and the off-diagonal values are significantly reduced, indicating fewer misclassifications. This suggests that the Medium model better captures and recognizes data patterns, enhancing overall performance.

The EfficientNetV2-Large (L) model performs best among the three variants. Its confusion matrix displays a perfect concentration of values along the main diagonal, signifying the highest number of correct classifications. Misclassifications are minimal, with negligible off-diagonal values, making this model highly reliable for tasks requiring precise and accurate predictions.

Based on the confusion matrix analysis, the EfficientNetV2-Large model is the most effective variant, achieving the highest accuracy with the fewest misclassifications. This makes it an ideal choice for applications demanding high precision and reliability, such as brain tumor classification, where the cost of errors can be significant. The performance progression from Small to Large variants illustrates the trade-off between computational complexity and classification accuracy, with the Large model offering superior results for scenarios where computational resources are not a primary constraint. Table 2 displays the Table 2 displays the accuracy results of each model results of each model.

Table 2 provides a comprehensive summary of the evaluation metrics across the three variants, detailing key performance indicators and enabling a comparative analysis of their effectiveness. Experimental results show that the small and medium size models exhibit a better balance between training, validation, and testing performance, with consistently high accuracy.

Table 2. Evaluation metrics for the three EfficentNetV2 model variants fine-tune

Variants	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Small	99.72	99.46	99.50	99.48
Medium	99.61	99.45	99.00	99.43
Large	99.85	99.60	99.65	99.50

The Large model, while excelling in training and validation accuracy, showed a drop in performance on test data, indicating potential overfitting and slightly lower generalization ability. In addition to the accuracy evaluation results, we can look at the computational side to consider which model suits our needs. The graphical visualization of the performance of the three models in Figure 16.

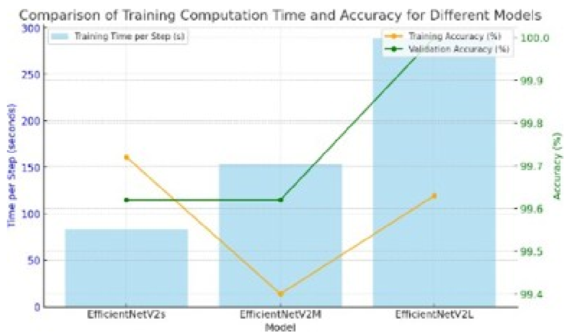

Fig. 16. Comparison of training model computation

After looking at the graph, we can see the values in Table 3 below, which shows the computational comparison results of the three models. Based on the comparison graph in Figure 16, the performance data of the EfficientNetV2-Small (S), Medium (M), and Large (L) models show significant differences in terms of training time, loss, and accuracy during the fine-tuning phase. The EfficientNetV2-Small (S) model showed outstanding efficiency, with the shortest training time of 83 seconds per step.

Table 3. Evaluation metrics for the three EfficentNetV2 model variants fine-tune

Variants	Train Time Perstep (s)	Loss	Train Accuracy (%)	Val Loss	Val Accuracy (%)
Small	83s	0.3605	99.72	0.3564	99.62
Medium	153s	0.3674	99.61	0.3625	99.62
Large	288s	0.3585	99.75	0.3501	100.00

The model achieved a training loss of 0.3605 and a training accuracy of 99.72%. The validation loss was slightly lower at 0.3564, with a validation accuracy of 99.62%. These results highlight that the small variant provides high accuracy while maintaining computational efficiency, making it an ideal choice for applications with limited computational resources and tight time constraints.

EfficientNetV2-Medium (M) model exhibits a longer training time of 153 seconds per step, almost double that of the small variant. The model achieved a training loss of 0.3674 and a training accuracy of 99.61%, with a validation loss of 0.3625 and a validation accuracy of 99.62%. Although the performance metrics are comparable to the small model, the significant increase in training time does not offer substantial improvements in accuracy or loss, making this model less efficient for tasks where computational resources or time are limited.

EfficientNetV2-Large (L) model achieved the best overall performance, with a training loss of 0.3585 and a training accuracy of 99.75%. The validation loss is the lowest among all variants, at 0.3501, and the validation accuracy reaches 100%. However, this model takes the longest training time, at 288 seconds per step, and thus requires significant resources. This makes the large variant particularly suitable for applications that want to achieve the highest possible accuracy, regardless of time or resource constraints.

The EfficientNetV2-Large model offers the highest accuracy and reliability, ideal for tasks that demand maximum precision. The refinement results clearly illustrate the trade-off between efficiency and performance, with the optimal choice varying based on application needs.

Evaluation and Visualization Grad-CAM

Grad-CAM (Gradient-weighted Class Activation Mapping) was implemented across all EfficientNetV2 variants (Small, Medium, and Large) to interpret and visualize the regions in MRI images that significantly influence the model's classification decisions. This method provides crucial insights into the models' decision-making processes by generating heatmaps highlighting critical regions in the input images.

EfficientNetV2-Large demonstrated the most consistent and accurate Grad-CAM visualizations among the three models, which aligns with its superior performance metrics across all evaluations. As depicted in the figure, the Grad-CAM visualizations for the EfficientNetV2-Large model show a clear and precise focus on relevant regions of the brain MRI scans for each class: meningioma, pituitary, glioma, and no tumor. The results of the Grad-CAM visualization using the EfficintNetV2-Large model can be shown in Figure 17.

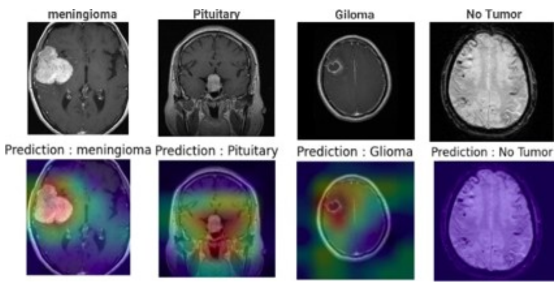

Fig. 17. Visualization classification model EfficientNetV2-Large with Grad-CAM

The heatmap powerfully highlights the tumor area for the meningioma class, confirming the model's ability to localize and classify the abnormal mass accurately. Similarly, for the pituitary class, the activations are concentrated precisely on the pituitary gland, reflecting the model's capability to recognize anatomical structures effectively.

For the glioma class, the visualization indicates a high-intensity focus on the tumor area, demonstrating the model's robustness in distinguishing this type of tumor. In the case of no tumor, the model generates a more diffuse heatmap, appropriately reflecting the absence of significant focal abnormalities, and correctly classifies the image as "no tumor."

Compared to the Grad-CAM results of the Small and Medium variants, EfficientNetV2-Large consistently produced the most interpretable and accurate heatmaps. The Small and Medium models occasionally exhibited less precise focus, leading to misclassifications in some instances, as reflected in their slightly lower evaluation metrics.

Table 4. Evaluation metrics for the three EfficentNetV2 model variants fine-tune

Model	Year	Author	Result
Proposed model EfficientNet+ Grand-CAM	2024	Pacal [14]	99.77%
EfficientNetV2 B1-B7	2024	Singh [31]	93.80%
Optimation EfficientNet B1-B7	2024	Ilsam [32]	99.69%
Proposed Method	2026	Denisa	99.85%

Overall, EfficientNetV2-Large stands out for its high accuracy, minimal misclassifications, and ability to provide interpretable visual evidence for its predictions. This makes it a highly reliable and explainable model, particularly suitable for critical medical applications like brain tumor classification, where transparency and trust in the decision-making process are essential. After seeing the evaluation of the model in this study, a comparison will then be made to previous research shown in Table 4.

CONCLUSION

This study optimized the EfficientNetV2 architecture variants (Small, Medium, and Large) for MRI-based brain tumour classification using transfer learning, fine-tuning, and data augmentation techniques. The optimized model excellently classified MRI images into four categories: glioma, meningioma, pituitary tumour, and no tumour. Grad-CAM improved interpretability by visualizing important areas in the MRI images that affect the model's decision. results showed that the EfficientNetV2-Large model provided the best performance, with 99.85% accuracy, 99.60% precision, 99.65% recall, and 99.50% F1-score. However, this model requires the longest computation time of 288 seconds per step, which may not be suitable fo resource-constrained environments. In contrast, the EfficientNetV2-Small model offers the best computational efficiency with still high accuracy (99.72%) and much shorter training time. Grad-CAM demonstrated the model's ability to accurately identify tumour areas on MRI, providing consistent and interpretative visualizations, especially in the Large variant. This confirms the potential of this model to support faster and more accurate medical diagnosis. Overall, this study confirms that the EfficientNetV2 architecture, especially the Large variant, is well suited for MRI- based brain tumour diagnosis applications, with an optimal combination of high accuracy, interpretability, and efficiency. However, the choice of model used should be tailored to specific needs, such as the availability of computing resources and the desired level of precision.

REFERENCES

[1] P. Bansal et al., “We are IntechOpen , the world ’ s leading publisher of Open Access books Built by scientists , for scientists TOP 1 %,” Intech, vol. i, no. tourism, p. 15, 2016, [Online]. Available: https://www.intechopen.com/books/advanced-biometric- technologies/liveness-detection-in- biometrics

[2] B. Alther, V. Mylius, M. Weller, and Gantenbein, “From first symptoms to diagnosis: Initial clinical presentation of primary brain tumors,” Clin. Transl. Neurosci., vol. 4, no. 2, p.

2514183X2096836, 2020, doi: 10.1177/2514183x20968368.

[3] Laurent, "Brain Tumor Management: One Day Symposium and Workshop," no. December, pp. 1–13, 2017.

[4] H. Sung et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA. Cancer J. Clin., vol. 71, no. 3, pp. 209–249, 2021, doi: 10.3322/caac.21660.

[5] T. Singh, R. R. Nair, T. Babu, A. Wagh, A. Bhosalea, and P. Duraisamy, "BrainNet: A Deep Learning Approach for Brain Tumor Classification," Procedia Comput. Sci., vol. 235, pp. 3283–3292, 2024, doi: 10.1016/j.procs.2024.04.310.

[6] X. Guo, T. Liu, and Q. Chi, "Brain tumor diagnosis in MRI scans images using Residual/Shuffle Network optimized by augmented Falcon Finch optimization," Sci. Rep., vol. 14, no. 1, pp. 1–21, 2024, doi: 10.1038/s41598-024-77523-2.

[7] L. Alzubaidi et al., Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.

[8] A. Iqbal, M. A. Jaffar, and R. Jahangir, "Enhancing Brain Tumour Multi-Classification Using Efficient-Net B0-Based Intelligent Diagnosis for Internet of Medical Things (IoMT) Applications," Inf., vol. 15, no. 8, 2024, doi: 10.3390/info15080489.

[9] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," Proc. Mach. Learn. Res., vol. 139, pp. 10096–10106, 2021.

[10] Y. Anagun, "Smart brain tumor diagnosis system utilizing deep convolutional neural networks," Multimed. Tools Appl., vol. 82, no. 28, pp. 44527–44553, 2023, doi: 10.1007/s11042-023-15422-w.

[11] M. Al Moteri, T. R. Mahesh, A. Thakur, V. Vinoth Kumar, S. B. Khan, and M. Alojail, "Enhancing accessibility for improved diagnosis with modified EfficientNetV2-S and cyclic learning rate strategy in women with disabilities and breast cancer," Front. Med., vol. 11, no. March, pp. 1– 15, 2024, doi: 10.3389/fmed.2024.1373244.

[12] Mani Abedini, "Classification of MRI Brain Tumor Images using Deep Learning Segment Anything Model for segmentation and Deep Convolution Neural Network," World J. Adv. Res. Rev., vol. 23, no. 2, pp. 1153–1161, 2024, doi: 10.30574/wjarr.2024.23.2.2469.

[13] J. H. Lee, J. W. Chae, and H. C. Cho, "Improved Classification of Different Brain Tumors in MRI Scans Using Patterned-GridMask," IEEE Access, vol. 12, no. March, pp. 40204–40212, 2024, doi: 10.1109/ACCESS.2024.3377105.

[14] I. Pacal, O. Celik, B. Bayram, and A. Cunha, "Enhancing EfficientNetv2 with global and efficient channel attention mechanisms for accurate MRI-Based brain tumor classification," Cluster Comput., vol. 1, 2024, doi: 10.1007/s10586-024-04532-1.

[15] L. I. Kesuma, Ermatita, and Erwin, "ELREI: Ensemble Learning of ResNet, EfficientNet, and Inception- v3 for Lung Disease Classification based on Chest X-Ray Image," Int. J. Intell. Eng. Syst., vol. 16, no. 5, pp. 149–161, 2023, doi: 10.22266/ijies2023.1031.14.

[16] O. A. Montesinos López, A. Montesinos López, and J. Crossa, Multivariate Statistical Machine Learning Methods for Genomic Prediction, no. January. 2022. doi: 10.1007/978-3-030-89010-0.

[17] Dr. T Subburaj and Bhavana. S, "Image Noise Reduction with Auto- encoders using TensorFlow," Int. J. Adv. Res. Sci. Commun. Technol., pp. 86–91, 2024, doi: 10.48175/ijarsct-19016.

[18] L. Pfaff et al., "Self-supervised MRI denoising: leveraging Stein's unbiased risk estimator and spatially resolved noise maps," Sci. Rep., vol. 13, no. 1, pp. 1–13, 2023, doi: 10.1038/s41598-023-49023-2.

[19] K. L. Radke et al., "Deep Learning- Based Denoising of CEST MR Data: A Feasibility Study on Applying Synthetic Phantoms in Medical Imaging," Diagnostics, vol. 13, no. 21, 2023, doi: 10.3390/diagnostics13213326.

[20] N. Pudjihartono, T. Fadason, A. W. Kempa-Liehr, and J. M. O'Sullivan, "A Review of Feature Selection Methods for

Machine Learning-Based Disease Risk Prediction," Front. Bioinforma., vol. 2, no. June, pp. 1–17, 2022, doi: 10.3389/fbinf.2022.927312.

[21] Y. Xu and R. Goodacre, "On Splitting Training and Validation Set: A Comparative Study of Cross- Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning," J. Anal. Test., vol. 2, no. 3, pp. 249–262, 2018, doi: 10.1007/s41664-018-0068-2.

[22] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 36th Int. Conf. Mach. Learn. ICML 2019, vol. 2019–June, pp. 10691–10700, 2019.

[23] H. Song, "FST-EfficientNetV2: Exceptional Image Classification for Remote Sensing," Comput. Syst. Sci. Eng., vol. 46, no. 3, pp. 3959–3978, 2023, doi: 10.32604/csse.2023.038429.

[24] J. hyeon Lee, J. woo Chae, and H. chong Cho, "Improved Classification of Brain-Tumor MRI Images Through Data Augmentation and Filter Application," J. Electr. Eng. Technol., vol. 18, no. 4, pp. 3135–3142, 2023, doi: 10.1007/s42835-023-01542-8.

[25] S. Tummala, V. S. G. Thadikemalla, S. Kadry, M. Sharaf, and H. T. Rauf, "EfficientNetV2 Based Ensemble Model for Quality Estimation of Diabetic Retinopathy Images from DeepDRiD," Diagnostics, vol. 13, no. 4, 2023, doi: 10.3390/diagnostics13040622.

[26] Y. Sun, L. Ning, B. Zhao, and J. Yan, "Tomato Leaf Disease Classification by Combining EfficientNetv2 and a Swin Transformer," Appl. Sci., vol. 14, no. 17, 2024, doi: 10.3390/app14177472.

[27] M. M. M, M. T. R, V. K. V, and S. Guluwadi, "Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50," BMC Med. Imaging, vol. 24, no. 1, pp. 1–19, 2024, doi: 10.1186/s12880-024-01292-7.

[28] D. Nova, K. Hardani, and I. Ardiyanto, "Decoding brain tumor insights : Evaluating CAM variants with 3D U- Net for segmentation," vol. 9, no. 2, pp. 262–273, 2024. https://doi.org/10.21924/cst.9.2.2024.1477

[29] I. G. S. M. Diyasa, W. S. J. Saputra, A. N. Gunawan, D. Herawati, S. Munir, and S. Humairah, "Abnormality Determination of Spermatozoa Motility Using Gaussian Mixture Model and Matching-based Algorithm," J. Robot. Control, vol. 5, no. 1, pp. 103–116, 2024, doi: 10.18196/jrc.v5i1.20686.

[30] M. M. Taye, "Understanding of Machine Learning with Deep Learning :," Comput. MDPI, vol. 12, no. 91, pp. 1–26, 2023. https://doi.org/10.3390/computers12050091.

[31] R. Singh, C. Prabha, M. Malik, and A. Goyal, "A Robust Deep Learning Model for Brain Tumor Detection and Classification Using Efficient Net: A Brief Meta-Analysis," J. Adv. Res. Appl. Sci. Eng. Technol., vol. 49, no. 2, pp. 26–51, 2025, doi: 10.37934/araset.49.2.2651.

[32] M. M. Islam, M. A. Talukder, M. A. Uddin, A. Akhter, and M. Khalid, "BrainNet: Precision Brain Tumor Classification with Optimized EfficientNet Architecture," Int. J. Intell. Syst., vol. 2024, 2024, doi: 10.1155/2024/3583612.