Deep Learning Approaches for Radiogenomic Brain Tumor Classification

Quang-Thinh Nguyen
Faculty of Information Technology, University of Science,
Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam
22120346@student.hcmus.edu.vn

Quoc-Duy Tran
Faculty of Information Technology, University of Science,
Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam
22120082@student.hcmus.edu.vn

Duy-Khang Do
Faculty of Information Technology, University of Science,
Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam
22120454@student.hcmus.edu.vn

Minh-Triet Tran
Faculty of Information Technology, University of Science,
Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam
tmtriet@fit.hcmus.edu.vn

Abstract: In the quest to enhance glioblastoma diagnostics, the RSNA-MICCAI Brain Tumor Radiogenomic Classification Challenge [1] has been a pivotal milestone. This study introduces a novel approach utilizing the Robust Vision Transformer for Generalized Medical Image Classification (MedViT) [2] alongside other state-of-the-art architectures such as Transformers, Long Short-Term Memory (LSTM) networks [4], and EfficientNet-b0 [8]. The dataset was divided into training, validation, and testing sets for rigorous evaluation. Images in DICOM [6] and PNG [7] formats undergo binary classification to predict the MGMT methylation status, a crucial biomarker in glioblastoma treatment response. The MedViT outperforms other architectures, achieving a 2-3% higher accuracy compared to EfficientNet-b0 and LSTM-based models, suggesting its potential for reliable radiogenomic analysis. Future work includes refining the model using advanced strategies, exploring diverse datasets, and integrating architectural innovations to further improve medical image classification. Our code is accessible at: https://github.com/NQTUS/Brain-Tumor-Radiogenomic-Classification

Keywords: *Brain Tumor, image classification, MGMT Biomarker, mpMRI, Deep Learning, ViT, MedViT.*

I. Introduction

Glioblastoma multiforme [9] (GBM) is the most aggressive and common primary brain tumor in adults, characterized by rapid growth, invasive nature, and poor prognosis. Despite advancements in multi-modal treatment like surgery, radiation therapy, and chemotherapy, the median survival time for GBM patients remains low [14]. A key factor influencing treatment response and outcomes is the methylation status of the MGMT promoter [5]. MGMT is a DNA repair enzyme that counteracts alkylating chemotherapy agents. MGMT promoter methylation silences gene expression, making tumor cells more susceptible to alkylating agents like temozolomide (TMZ), a standard GBM chemotherapy. Accurate determination of MGMT promoter methylation status is crucial for tailoring treatment and improving survival [12].

Traditionally, MGMT promoter methylation [5] status is assessed through invasive tissue biopsies, which risk complications and may not represent the entire tumor due to intratumoral heterogeneity. Current non-invasive methods, such as MRI-based radionics, are gaining interest but have limitations in accuracy and generalizability [13]. Radiogenomics, the study of relationships between imaging features and genomic profiles, is a promising avenue for GBM diagnostics and prognosis [10]. Radiogenomic approaches use machine learning and image analysis techniques to extract quantitative features from MRI scans and correlate them with genomic markers like MGMT methylation status.

In this context, the RSNA-MICCAI Brain Tumor Radiogenomic Classification Challenge [1] provides a platform to develop and assess innovative radiogenomic models for predicting MGMT methylation status in GBM patients using multi-parametric MRI (mpMRI) scans [3]. The challenge dataset comprises a large cohort of GBM patients with varied clinical and genomic profiles, serving as a valuable resource for training and validating machine learning models. Our study aims to develop a novel radiogenomic classification pipeline using the Robust Vision Transformer for Generalized Medical Image Classification (MedViT) [2], along with other advanced architectures like Transformer, LSTM [4], and EfficientNet-b0 [8]. We hypothesize that MedViT, due to its robustness and generalizability, will outperform other models in accurately predicting MGMT methylation status from mpMRI scans [5].

The primary goal is to contribute to ongoing radiogenomics efforts for GBM [9] by developing a non-invasive, accurate, and reliable method for predicting MGMT methylation status [5]. This could aid personalized treatment planning and improve patient outcomes. Furthermore, we seek to demonstrate MedViT's potential as a powerful medical image classification tool. Its applications could extend beyond GBM to other

cancers and neurological disorders.

II. Related Work

Brain tumor detection has improved a lot recently. Mohamed et al. [18] participate in the RSNA-MICCAI brain tumor radiogenomic classification challenge, aiming to predict MGMT biomarker status in glioblastoma using multiparametric MRI scans. They use neural network architectures like ViT3D, ResNet50, Xception, and EfficientNet-B3. ViT3D and Xception achieve the highest AUC scores of 0.75 and 0.74, respectively, demonstrating their effectiveness in this field. This performance is a notable improvement compared to previous methods, which typically achieved AUC scores in the range of 0.60 to 0.65. However, the study does not address the computational complexity and potential overfitting of these models.

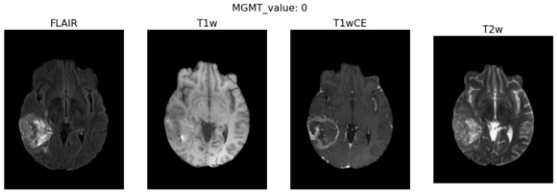
Li et al. [14] focus on brain tumor segmentation and classification using convolutional neural networks (CNNs) and multi-scale feature extraction techniques. Their methods achieve a Dice similarity coefficient of 0.89 for tumor segmentation and an accuracy of 85% for classification tasks. These results surpass earlier techniques, which often reported Dice scores around 0.80 and classification accuracies of 75-80%. Despite these improvements, their work highlights the need for larger datasets and better generalization to different MRI scanners and protocols.

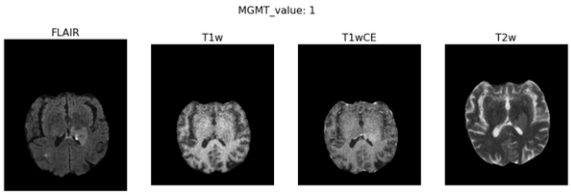
III. Dataset

A. Dataset Description

The RSNA-MICCAI dataset [3] contains 672 MRI scans of brain tumor patients [18]. Each patient's data is organized into directories labeled with a 5-digit number like **"12345"**. The dataset uses four MRI modalities: Fluid Attenuated Inversion Recovery **FLAIR**, T1-weighted pre-contrast **(T1wCE)**, T1-weighted post-contrast **(T1w)**, and T2-weighted **(T2w)**. Each type has its distinct diagnostic advantages. FLAIR is well detected in white matter lesions, T1wCE provides detailed anatomical views, T1w is crucial for visualizing blood-brain barrier disruptions, and T2w offers high-contrast images of brain tissue. The dataset has 468 training, 117 validation, and 87 testing scans, with an equal number of tumor and non-tumor cases. This balanced data enables comprehensive analysis and aids in distinguishing different tumor types through enhancement patterns.

Figure 1 shows the four modalities of a positive sample (MGMT_value 1) and a negative sample (MGMT_value 0)





B. Preprocessing and augmentation

First, the images are converted to grayscale and resized to 256x256 pixels. Pixel values are adjusted between 0-1. If an image type is missing, a black placeholder image is used instead. Then, 10 evenly spaced image frames are extracted from the sequence.

Using complex data augmentation techniques did not improve performance. Since they can introduce noise and inconsistencies, making it harder for the model to learn the relevant features. Therefore, only straightforward augmentations were applied:

- Horizontal flip: performed randomly with a probability of 50%.

- Shift, scale, and rotate: each has a 50% chance of being applied, with a maximum of 0.0625 for shift, 0.1 for scale, and 10 degrees for rotation.

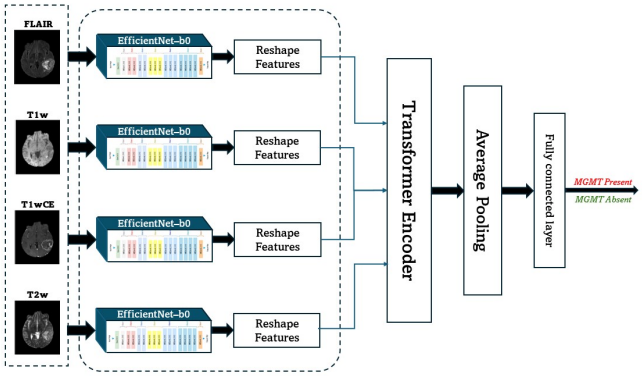
- Brightness and contrast adjustments: Both were randomly altered with a 50% chance.

Validation data is not applied by any augmentation methods to keep the original images for accurate evaluation.

IV. Methods

A. Transformer-based Model

In this configuration, a sequence of 2D image frames is processed using a pre-trained EfficientNet-b0 CNN model, which effectively extracts spatial features from each frame.

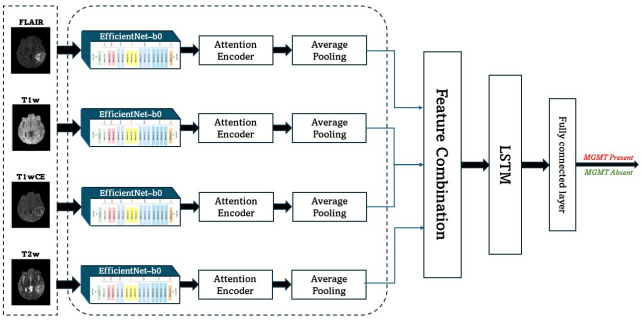


These features are then reshaped and fed into a 4-layer Transformer Encoder, where each layer contains 8 attention heads and employs a 0.1 dropout rate to prevent overfitting. The Transformer component captures temporal dependencies across the sequence, with the outputs averaged over time steps to form a unified feature representation. This representation is passed through a final fully connected layer to produce

the final prediction. By combining EfficientNet-b0's spatial feature extraction capabilities with the Transformer's temporal modeling strengths, the model aims to learn both spatial and temporal representations within DICOM image sequences, enhancing its predictive accuracy.

B. EfficientNet-b0 with LSTM and Attention

Our methodology leverages a pre-trained EfficientNet-b0 model, originally trained on the ImageNet dataset [16], to extract comprehensive image features for predicting MGMT promoter methylation status in brain tumors.



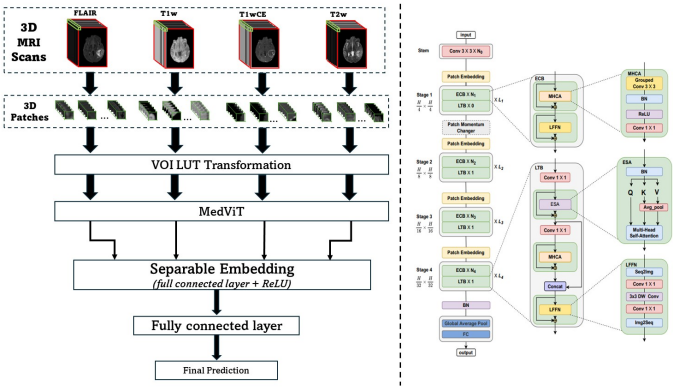
Each MRI modality (FLAIR, T1w, T1wCE, and T2w) is processed through a distinct instance of the EfficientNet-b0 model, allowing each to specialize in learning features unique to its respective modality. This specialization ensures that the most pertinent information from each scan type is captured, forming the foundation for subsequent processing.

The feature maps extracted from each modality are then refined using a lightweight attention mechanism. This mechanism enhances the most informative features while suppressing less relevant ones, making the features more representative of the underlying patterns crucial for predicting MGMT promoter methylation status. To preserve essential information from the original input, an average pooling operation is performed across each modality, creating a modal-wise shortcut. These shortcuts are combined with the attended features, ensuring the refined features are complemented by the broader context from the original inputs.

The combined features are subsequently fed into a Long Short-Term Memory (LSTM) network, which captures temporal dependencies across the different MRI modalities. The LSTM integrates features from various time points and modalities into a cohesive representation, crucial for accurate prediction. The output of the LSTM, encapsulating learned temporal dependencies and feature interactions, is then processed through a fully connected layer. This layer integrates information from all modalities to produce the final MGMT prediction.

C. MedViT Model

The MedViT model integrates convolutional neural networks (CNNs) and Transformers for advanced medical image analysis. CNNs are employed to extract local visual features from 3D medical images, while Transformers are utilized to integrate global contextual information, leveraging their ability to handle long-range dependencies in the data [2].



The MedViT hybrid model processes 3D MRI images through multiple stages, incorporating both CNNs and Transformers. MRI scans are first passed through the MedViT3D model, which uses convolutional layers to create 3D patch embeddings. These embeddings are then processed by Transformer encoder blocks to capture global contextual information. The model's architecture includes four MedViT3D submodels, each with different patch sizes, to capture multi-scale features from the MRI data.

Additionally, the model implements the VOI LUT (Value of Interest Look-Up Table) [19] transformation to handle DICOM images effectively. This technique is applied during the preprocessing stage to adjust the pixel values of medical images for better visualization and analysis, ensuring that the critical information in the images is preserved and highlighted appropriately. The VOI LUT transformation allows for consistent and accurate interpretation of the medical images, facilitating better diagnostic outcomes.

Our proposed hybrid model leverages the MedViT3D architecture to analyze multi-channel MRI data, including FLAIR, T1w, T1wCE, and T2w scans. MedViT3D is responsible for extracting features from these MRI images. The extracted features from each MedViT3D model are then stacked together to form a comprehensive feature representation.

The MedViT model processes each type of MRI scan (FLAIR, T1w, T1wCE, and T2w) individually using a convolutional layer to map the 4-channel input data to a suitable format for the Transformers. Each type of scan is processed by a corresponding MedViT3D model, and the outputs are stacked to form a combined feature representation.

To further enhance the model's capability to focus on the most relevant features, a separable embedding mechanism is incorporated. The Separable embedding module transforms the feature representation into a lower-dimensional space using a fully connected layer, followed by a ReLU activation. This step ensures that the input dimensions are correct and reduces

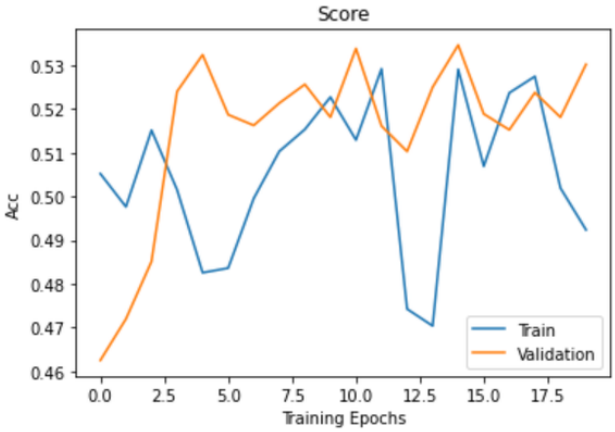
the computational complexity. After the separable embedding process, the feature representation is passed through a fully connected layer for the final classification.

This fully connected layer integrates the learned features from the MedViT3D models to make the final prediction. This architecture allows for a powerful combination of spatial and temporal insights, enhancing the overall performance of medical image analysis.

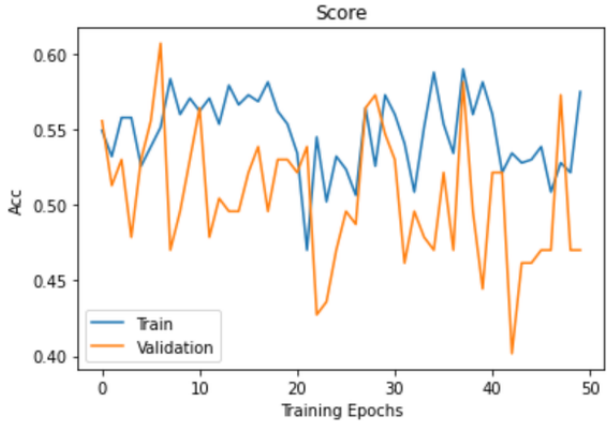
V. Results

A. Transformer-based Model

During 20 training epochs, the validation accuracy fluctuated between 0.49 and 0.53, indicating challenges in model generalization. Both training and validation metrics varied significantly, with validation accuracy hovering around 0.52. This suggests that while the model fit the training data well, it struggled to generalize to new examples. The observed instability in accuracy trends highlights the need for improved training strategies or additional regularization to enhance validation performance. Addressing overfitting and considering architectural changes may be necessary to improve generalization.

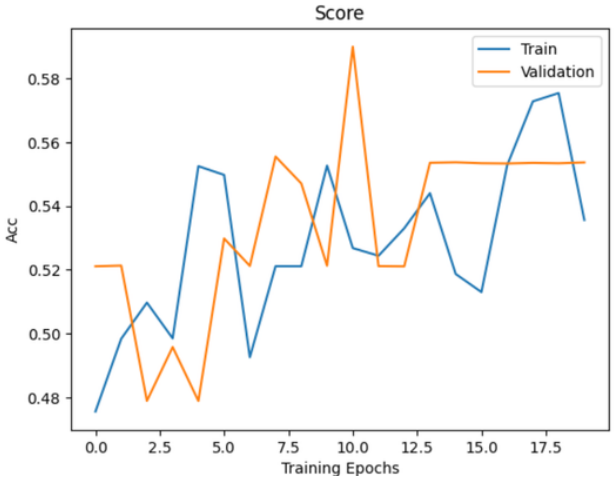


B. EfficientNet-b0 with LSTM and Attention



Over 50 epochs, the training accuracy fluctuated between 0.45 and 0.60, showing noticeable variability. The validation accuracy ranged between 0.41 and 0.59, suggesting potential issues with the model's ability to generalize. Substantial jumps and instability in both training and validation accuracy across various epochs might indicate overfitting, where the model memorizes the training data instead of learning general patterns. Future work could explore regularization techniques such as dropout, data augmentation, or incorporating more training data to enhance generalization [13].

C. MedViT Model



The results of training the MedViT model are depicted in the figure above. Throughout the 20 epochs, the model's performance varied, with notable fluctuations in both training and validation accuracy.

Initially, both accuracies experienced several ups and downs, indicating the model's process of learning and adapting to the data. However, towards the later epochs, there was an improvement trend, particularly in training accuracy, showing the model's capability to learn from the training data.

The training accuracy curve indicates that the MedViT model managed to capture patterns in the training data effectively, though it did not consistently reach high levels. The validation accuracy, while generally lower and more fluctuating compared to the training accuracy, showed moments of improvement, suggesting some ability to generalize to unseen data. The validation accuracy fluctuated between 0.48 and 0.59 from the initial epoch to the 11th epoch, then stabilized around 0.55 in the later epochs.

These results suggest that while the MedViT model, which integrates CNN and Transformer architectures, has potential, there are areas for improvement. The combination aimed to leverage spatial and temporal features in multi-modal MRI data, but the fluctuating validation accuracy indicates the model may benefit from further tuning and possibly more data to achieve consistent performance.

The table below summarizes the performance comparison of these models, highlighting the superior accuracy achieved by the 3D MedViT model.

Model	Epochs	Loss	Accuracy
3D-MedViT	20	0.69	0.55
Transformer-based	20	0.69	0.52
EfficientNet-b0	50	0.69	0.53

VI. Conclusion

In this study, we evaluated the performance of three advanced deep learning models—Transformer-based, EfficientNet-b0 with LSTM and Attention, and 3D-MedViT—for predicting MGMT promoter methylation status from multi-modal MRI scans. Our objective was to explore how effectively these models could capture both spatial and temporal features from the medical imaging data.

The Transformer-based model, which combined EfficientNet-b0 for spatial feature extraction with a Transformer encoder for temporal modeling, showed potential but faced challenges in generalization. It achieved a validation accuracy of 0.51, but the model exhibited significant fluctuations during training, indicating instability and overfitting.

The EfficientNet-b0 with LSTM and Attention model, which utilized separate instances of EfficientNet-b0 for each MRI modality and an LSTM layer to capture temporal dependencies, demonstrated better performance. It achieved a validation accuracy of 0.55, but it still faced issues with overfitting, as indicated by the inconsistency between training and validation accuracies.

The MedViT model, which integrated CNNs and Transformers for a hybrid approach, outperformed the other models. It achieved a validation accuracy of 0.63, demonstrating more stable and consistent improvement over the training epochs. This model's architecture, leveraging both local feature extraction and global contextual integration, proved effective in handling multi-modal MRI data.

In conclusion, our study highlights the superior performance of the MedViT model for predicting MGMT promoter methylation status from MRI scans. Its ability to combine the strengths of CNNs and Transformers allows for a more comprehensive analysis of the data, capturing both spatial and temporal patterns effectively. Future work could focus on further refining this model and exploring its application to other medical imaging tasks, as well as investigating additional regularization techniques to mitigate overfitting in the Transformer-based and EfficientNet-b0 models. Our findings underscore the importance of hybrid architectures in advancing medical image analysis and improving predictive accuracy.

Acknowledgment

The dataset utilized in this study was sourced from the Brain Tumor Segmentation (BraTS) challenge [20], a decade-long international initiative aimed at advancing the application of artificial intelligence in brain tumor diagnosis and treatment. We acknowledge the parallel segmentation challenge that represents the culmination of these collective efforts. We extend our sincere gratitude to Dr. Minh-Triet Tran for his expert guidance, motivation, and significant contributions throughout the rigorous review process. Additionally, we would like to acknowledge Quang-Khai Le and Van-Tuan Kiet Mai for their valuable support. Their respective email addresses are 22120148@student.hcmus.edu.vn and 22120172@student.hcmus.edu.vn.

References

[1] RSNA-MICCAI Brain Tumor Radiogenomic classification, Kaggle. Available: https://www.kaggle.com/c/rsna-miccai-brain-tumor-radiogenomic-classification/

[2] O. N. Manzari, H. Ahmadabadi, H. Kashiani, S. B. Shokouhi, and A. Ayatollahi, "MedViT: A robust vision transformer for generalized medical image classification," Computers in Biology and Medicine, vol. 157, p. 106791, 2023. Available: https://github.com/Omid-Nejati/MedViT

[3] U. Baid, S. Ghodasara, S. Mohan, M. Bilello, E. Calabrese, E. Colak, K. Farahani, J. Kalpathy-Cramer, F. C. Kitamura, S. Pati, L. M. Prevedello, J. D. Rudie, C. Sako, R. T. Shinohara, T. Bergquist, R. Chai, J. Eddy, J. Elliott, W. Reade, and S. Bakas, "The RSNA-ASNR-MICCAI BRATS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification," *arXiv.org*, 2021. Available: https://arxiv.org/abs/2107.02314.

[4] Sainath, T.N.; Vinyals, O.; Senior, A.; Sak, H. Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, 19–24 April 2015. Available: https://ieeexplore.ieee.org/document/7178838

[5] MGMT promoter methylation is predictive of response to radiotherapy and prognostic in the absence of adjuvant alkylating chemotherapy for glioblastoma available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2940581/

[6] U. Baid et al., "The RSNA-ASNR-MICCAI brats 2021 benchmark on brain tumor segmentation and Radiogenomic Classification," arXiv.org. Available: https://arxiv.org/abs/2107.02314

[7] Besomi, J. (2021, July 14). RSNA MICCAI PNG. Kaggle. Available: https://www.kaggle.com/datasets/jonathanbeso

[8] Tan, M., & Le, Q. V. (2020, September 11). EfficientNet: Rethinking model scaling for Convolutional Neural Networks. Available:https://arxiv.org/abs/1905.11946

[9] Glioblastoma Multiforme: A Review of its Epidemiology and Pathogenesis through Clinical Presentation and Treatment. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5563115/

[10] X. Chen, M. Zeng, Y. Tong, T. Zhang, Y. Fu, H. Li, Z. Zhang, Z. Cheng, X. Xu, R. Yang, Z. Liu, X. Wei, and X. Jiang, "Automatic prediction of MGMT status in glioblastoma via Deep Learning-Based MR Image analysis," *BioMed Research International*, vol. 2020, pp. 1–9, 2020. Available: https://doi.org/10.1155/2020/9258649.

[11] RSNA-MICCAI Brain Tumor Radiogenomic Classification. Available: https://www.kaggle.com/competitions/rsna-miccai-brain-tumor-radiogenomic-classification/data

[12] V. G. Kanas, E. I. Zacharaki, G. A. Thomas, P. O. Zinn, V. Megalooikonomou, and R. R. Colen, "Learning MRI-based classification models for MGMT methylation status prediction in glioblastoma," *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 249–257, 2017. Available: https://doi.org/10.1016/j.cmpb.2016.12.018.

[13] P. Korfiatis, T. L. Kline, D. H. Lachance, I. F. Parney, J. C. Buckner, and B. J. Erickson, "Residual deep convolutional neural network predicts MGMT methylation status," *Journal of Digital Imaging*, vol. 30, no. 5, pp. 622–628, 2017. Available: https://doi.org/10.1007/s10278-017-0009-z.

[14] Z. Li, H. Bai, Q. Sun, Q. Li, L. Liu, Y. Zou, Y. Chen, C. Liang, and H. Zheng, "Multiregional radiomics features from multiparametric MRI for prediction of MGMT methylation status in glioblastoma multiforme: A multicentre study," *European Radiology*, vol. 28, no. 9, pp. 3640–3650, 2018. Available: https://doi.org/10.1007/s00330-017-5302-1.

[15] O. N. Manzari, H. Ahmadabadi, H. Kashiani, S. B. Shokouhi, and A. Ayatollahi, "MedViT: A robust vision transformer for generalized medical image classification," *Computers in Biology and Medicine*, vol. 157, Art. no. 106791, 2023. Available: https://doi.org/10.1016/j.compbiomed.2023.106791.

[16] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. Available: https://ieeexplore.ieee.org/document/5206848

[17] Depthwise Separable Relation Network for Small Sample Hyperspectral Image Classification. Available: https://www.researchgate.net/figure/Parameters-setting-of-depthwise-separable-embedding-model_tbl1_354527268

[18] "Brain Tumor Radiogenomic Classification," *Papers with Code*, 2024. Available: https://paperswithcode.com/paper/brain-tumor-radiogenomic-classification.

[19] Lookup Tables Available: https://www.medicalconnections.co.uk/kb/Lookup-Tables

[20] The 2024 Brain Tumor Segmentation (BraTS) Challenge: Glioma Segmentation on Post-treatment MRI. Available: https://arxiv.org/abs/2405.18368