An End-to-End Multi-Head Attention-Based Deep Learning Model for Enhanced Brain Tumor Detection Using MRI

V. Chanemougavel
Department of Computer Science, Arignar Anna Govt. Arts & Science College, Karaikal, Puducherry, India
vchvel@gmail.com (corresponding author)

K. Jayanthi
Department of Computer Application, Government Arts College, C.Mutlur, Chidambaram, Tamil Nadu, India
jayanthirab@gmail.com

Received: 26 August 2025 | Revised: 26 September 2025, 13 October 2025, 14 October 2025, 16 October 2025, and 18 October 2025 | Accepted: 21 October 2025

 DOI: https://doi.org/10.48084/etasr.14326

ABSTRACT

In recent years, the healthcare field has been significantly transformed by advances in technology, with Artificial Intelligence (AI) playing an important role in this process. AI refers to digital systems that emulate human-like intelligence and are widely applied in healthcare. Brain Tumors (BTs), which result from abnormal cell growth in the central nervous system, pose great difficulties in diagnosis and treatment. An early and precise diagnosis is vital for effective treatment. With the help of Magnetic Resonance Imaging (MRI), Deep Learning (DL) models can recognize and classify BTs, aiding in their rapid and easy detection. For accurate detection of BTs, this study presents an Advanced Brain Tumor Classification by integrating DL Models and Optimization Techniques (ABTC-IDLMOT) in biomedical imaging. The objective was to classify the affected BT region using a fine-tuned multi-head attention-based DL model. Initially, preprocessing employs the Wiener Filter (WF) for noise removal and Otsu's threshold for skull removal. In addition, EfficientNetV2M is utilized for feature extraction. Then, a Convolutional Neural Network-based Multi-Head Attention (CNN-MHA) model is used for BT classification. Finally, RMSProp optimization is used to tune the hyperparameters and improve the classification performance of CNN-MHA. The experimental study uses a benchmark MRI dataset. Performance validation of the ABTC-IDLMOT approach showed an accuracy of 96.65%.

Keywords-deep learning; brain tumor classification; RMSProp; preprocessing; biomedical imaging

I. INTRODUCTION

BT is an irregular development of brain cells that might be a visible symptom of cancer. Bening BTs are uniformly combined with inactive cells [1]. Malignant cancers comprise lively cancer cells with a non-uniform architecture. Malignant BTs are divided into two types: (i) primary and (ii) metastatic [2]. BT is one of the deadliest diseases in the world. It is the typical malignancy in elderly people and the third most common among young people [3]. The most common primary BTs are meningiomas, gliomas, and pituitary [4]. The best approach to diagnosing and identifying BTs is a pathologic evaluation of tissue morphology [5]. Among imaging technologies, BT is recognized mainly with MRI without requiring surgical intervention [6]. Classification is conventionally performed physically, which is very time-intensive, subject to error, and possibly not as well-qualified as incorporating refined macroscopic patterns. Therefore, attention is being paid to Computer-Aided Diagnosis (CAD) to increase initial identification. DL is a subset of ML and is proven to have higher efficiency than conventional models. Convolutional Neural Network (CNN) methods, a type of DL, use visual images and typically require less preprocessing [7, 8].

In [9], denoising models and data augmentation were used with two CNN-based methods. In [10], the Sine Cosine Archimedes Optimizer Algorithm (SCAOA) technique was presented. In [11], a DL-based technique utilized a pre-trained VGG16 CNN structure. In [12], the Multiweight New Loss (MWNL) was combined with a Cumulative Learning Strategy (CLS) and a DCNN. In [13], a dual-branch equivalent method integrated a Transformer Module (TM) with a CNN and a Self-Attention Unit (SAU). In [14], a Gray Level Co-occurrence Matrix (GLCM) feature extraction method was used. In [15], two pre-trained DL methods were used, and the subsequent

feature vectors were united to produce a hybrid feature vector utilizing the Partial Least Squares (PLS) model. In [16], EfficientNet was used as the encoder with a U-Net model. In [17], an Orthogonal Channel Shuffle Network (OCSN) was integrated with the Detection Transformer (OS-DETR) model, enforcing internal filter and improving the Attention Mechanisms (AMs). In [18], a model integrated multi-modal MRI data and advanced AMs. The 3D Attention Inception-Residual U-Net [19] integrated advanced feature extraction and attention. Although these studies used diverse DL, optimization, and feature fusion strategies, they mostly lack robustness in handling diverse noise patterns and generalized feature learning. A key research gap is in the limited integration of adaptive preprocessing with dynamic classification models for enhanced BT detection.

This study introduces an Advancing Brain Tumor Classification by integrating DL Models and Optimization Techniques (ABTC-IDLMOT). The major contributions are:

- The preprocessing integrates WF and Otsu's thresholding for accurate skull removal, improving MRI quality, downstreaming processing, and ensuring better feature representation. This also plays a crucial role in the overall classification accuracy.
- EfficientNetV2M is used to extract high-resolution and semantically rich features from multimodal MRI images. This model ensures computational efficiency while preserving significant tumor-related patterns, strengthening the ability to discriminate tumor from non-tumor regions.
- Local and global contextual features are effectively captured by the CNN-MHA model from MRI data, improving the focus on discriminative regions for BT classification and enabling deeper semantic understanding. In addition, RMSProp-based dynamic learning rate tuning ensures stable training, faster convergence, and mitigates overfitting, thus improving overall accuracy and reliability.
- The ABTC-IDLMOT approach is a novel architecture that integrates EfficientNetV2M, CNN-MHA, and RMSProp to improve the precise detection of tumor regions. The novelty is in the synergistic use of these advanced components, which remains largely unexplored in current BT detection.

II. MATERIALS AND METHODS

The proposed ABTC-IDLMOT technique includes various phases, namely feature extraction, classification, preprocessing, and hyperparameter tuning methods. Figure 1 depicts the workflow of the proposed ABTC-IDLMOT method.

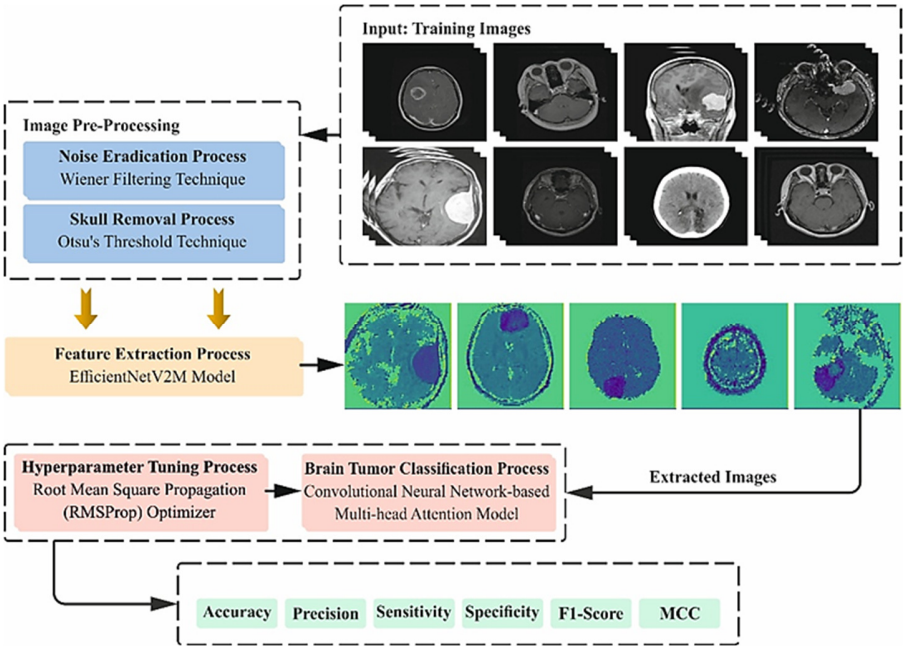

Fig. 1. Workflow of the ABTC-IDLMOT model.

A. Preprocessing Stage

Initially, image preprocessing involves noise removal using WF and skull removal using Otsu's threshold. WF ensures effective noise suppression while preserving crucial image details, outperforming basic filters such as median or Gaussian in medical imaging. Otsu's thresholding is an adaptive, unsupervised approach for accurate skull removal, making it more robust than manual or fixed-threshold methods.

1) Noise Removal Using Weiner Filter (WF)

The WF is a frequently used filter in the frequency field [20]. Blurred images can be restored by Inverse Filtering (IF), but this is highly sensitive to additive noise, often amplifying it during deblurring. The WF minimizes the Mean Squared Error (MSE) by balancing noise reduction and image restoration, providing a linear and more robust approximation of the original image. IF can restore images blurred by low-pass filters, but it is highly sensitive to additive noise, amplifying

both noise and blur. The WF reduces the MSE by balancing noise smoothing and deblurring, acting as a linear approximation of the original image in the Fourier domain:

$$\mathcal{W}(f_1, f_2) = \frac{H^*(f_1,f_2)S_{xx}(f_1,f_2)}{|H(f_1,f_2)|^2 S_{xx}(f_1f_2)+S_{nn}(f_1,f_2)} \quad (1)$$

where $s_{nn}(f_1, f_2)$, and $s_{xx}(f_1, f_2)$ represent the power spectra of the additional noise and the original image, respectively, and $H(f_1, f_2)$ signifies the filter's frequency transfer function.

2) Skull Removal Using Otsu's Threshold

The standard Otsu model is simple, using only the zero and the primary-order collective instants of the gray-level histogram [21].

$$p_i = \frac{n_i}{N}, p_i \geq 0, \sum_{i=1}^{L} p_i = 1 \quad (2)$$

The optimal threshold is decided by a discriminant condition to increase the resulting classes at the gray level. The thresholding approach depends on choosing the lower point amongst dual classes. So, the optimal threshold k^* is described as:

$$\sigma_B^2(k^*) = \max_{1 \leq k \leq L} \sigma_B^2(k) \quad (3)$$

The k range across the maximum is limited to:

$$S^* = \begin{Bmatrix} k; \omega_0\omega_1 = \omega(k)[1-\omega(k)] > 0 \\ \text{or } 0 < \omega(k) < 1 \end{Bmatrix} \quad (4)$$

B. EfficientNetV2M-Based Feature Extraction

EfficientNetV2M is used for feature extraction [22]. This model was selected for its superior accuracy and faster training compared to conventional CNNs, due to its compound scaling and optimized architecture. It effectively captures fine-grained features in medical images while maintaining computational efficiency, making it ideal for MRI-based tumor analysis.

EfficientNet is a form of CNN, and every method in this sequence, from EfficientNet-B0 to EfficientNet-B7, illustrates various scales of the model structure with improved resolution, width, and depth. EfficientNet-B2 improves depth and width for better performance at a slightly higher computational cost, similar to B1. EfficientNet-B3 further enhances these aspects, while B4 and B5 improve accuracy and performance with higher resource demands and computational expenses. B6 and B7 deliver the highest accuracy among the original EfficientNet models but need significantly more computational resources. EfficientNetV2-M and V2-L are advanced versions, integrating architectural improvements aimed at improving both efficiency and accuracy compared to earlier variants.

The resources of EfficientNet-V2M can be calculated as:

$$FLOPs = 0.35 \times (N^2) \times M \times (AF + 1) \quad (5)$$

and

$$Parameters = 0.7 \times (N^2) \times M \times (AF + 1) \quad (6)$$

C. CNN-MHA-Based Classification Process

The CNN-MHA model is employed for BT classification. This model is highly capable of integrating local feature extraction from the CNN with the global context modeling from MHA, which conventional CNNs lack. This integration also improves the focus of the model on tumor-relevant regions, resulting in an enhanced accuracy in BT classification.

The CNN extracts features through layered convolution, pooling, and fully connected layers, capturing increasingly abstract data. Integrating MHA splits the input into heads that focus on diverse regions, improving precision and efficiency. Self-attention runs in parallel across heads, weighting values based on key-query similarity as:.

$$Att = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (7)$$

where K and Q symbolize key and query, respectively, both with the same size d_k, and V represents dimension value d_v. K_i, V_i, and Q_i characterize the i-th subdivisions of key, value, and query, which are gained by multiplying the output u_n from the preceding layer with three weighted matrices.

$$Q_i = u_n^T \cdot W_q^i \quad (8)$$

$$K_i = u_n^T \cdot W_k^i \quad (9)$$

$$V_i = u_n^T \cdot W_v^i \quad (10)$$

where $i \in [1, h]$, h depicts the headcounts, and $\{W_q^h, W_k^h, W_v^h\} \in d_{model} \times d_k$ represent the i-th head trainable weighted matrices. The self-attention is then considered individually for all heads, as:

$$Att_i = soft\max\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) \quad (11)$$

The last output state is gained by linearly transforming and splicing the h-head outputs:

$$O_{mha} = Concat(Att_1, \ldots . Att_h)W^o \quad (12)$$

where $W^o \in d_{model} \times d_{model}$ is the trainable parameter matrix, Att_i is the i-th head attentional output, and O_{mha} signifies the last output condition. Backpropagation is used during pretraining to update weights and biases, with cross-entropy loss measuring the difference between predicted and true labels:

$$\mathcal{L} = -\frac{1}{M}\sum_{m=1}^{M} y_m \log(\hat{y}_m) \quad (13)$$

where M depicts the experiment number, $\hat{y}_m$ refers to the anticipated value, and y_m depicts the true value. Due to the large training set, the Adam optimizer is used to efficiently calculate optimal parameters and prevent gradient oscillations.

D. RMSProp-Based Hyperparameter Tuning Model

Finally, RMSProp hyperparameter tuning is performed [23], chosen for its ability to adapt learning rates for each parameter, ensuring faster and more stable convergence. RMSProp handles non-stationary objectives well, making it ideal for optimizing deep models such as CNN-MHA, unlike standard optimizers. Figure 2 depicts the RMSProp model.

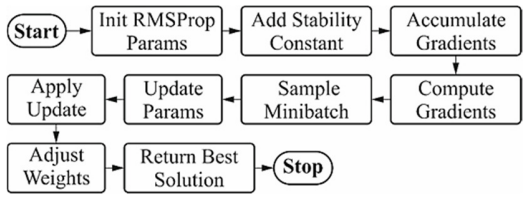

Fig. 2. Flow of the RMSProp model.

RMSprop is a method for decreasing the noise in Neural Networks (NNs) by fine-tuning the errors as they are distributed over the system. RMSProp is an expansion of gradient descent, which utilizes a decaying average of incomplete gradients in the stage dimensions variation for every parameter. Every upgrade is performed based on the expression below, distinctly for every parameter.

$$v_t = \delta * v_{t-1} - (1-\delta) * g_t^2 \quad (14)$$

$$\Delta\omega_t = -\frac{\eta}{\sqrt{y_t+e}} * g_t \quad (15)$$

$$\omega_{t+1} = \omega_t + \Delta\omega_t \quad (16)$$

where η and v_t denote the learning rate and the exponential average of gradient squares, respectively, and g_t signifies the gradient on time t beside w_j. During training, weights are updated with each batch, but some networks overfit, performing well on training data but failing to generalize to unseen data.

III. PERFORMANCE VALIDATION

The proposed ABTC-IDLMOT model was evaluated on the BT MRI dataset [24]. Figure 3 demonstrates samples and preprocessed images. Figure 4 shows samples of extracted features. The proposed method was developed on Python 3.6.5 with an i5-8600k CPU, 4GB GPU, 16GB RAM, 250GB SSD, and 1TB HDD, using a 0.01 learning rate, ReLU, 50 epochs, 0.5 dropout, and a batch size of 5.

TABLE I. DATASET DESCRIPTION

Class	Images
No Tumor	1992
Meningioma	1644
Glioma	1621
Pituitary	1756
Total	7013

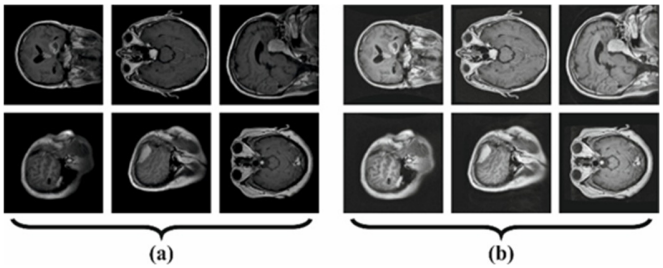

Fig. 3. Images (a) samples and (b) preprocessed.

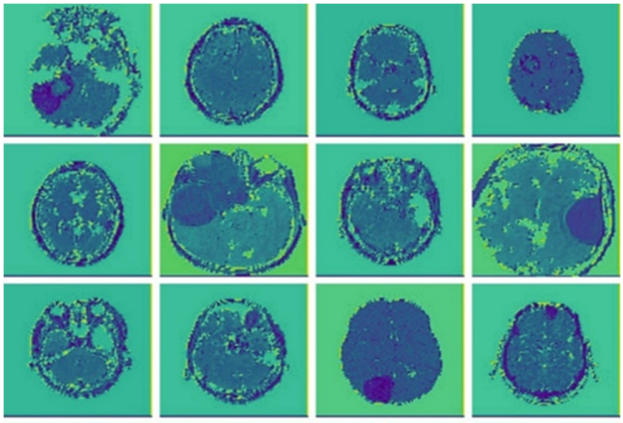

Fig. 4. Samples of extracted features.

Figure 5 shows the classification results of the proposed ABTC-IDLMOT model.

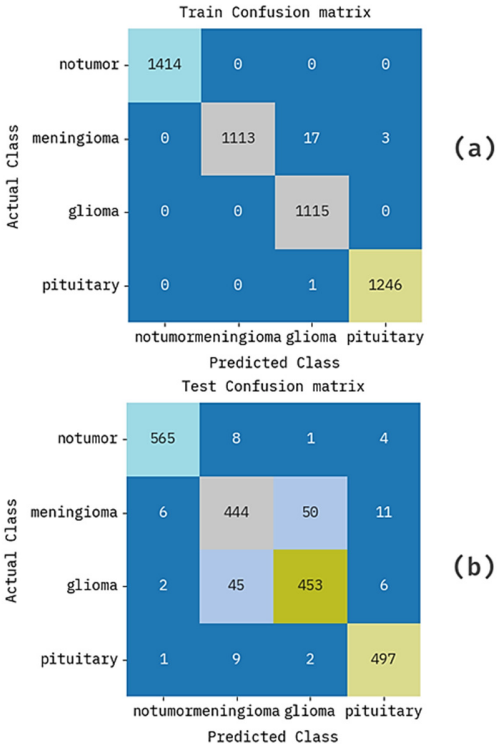

Fig. 5. Confusion matrices for (a) training and (b) testing.

Table II illustrates the classification metrics of the proposed ABTC-IDLMOT method. On the training set, the ABTC-IDLMOT model reached an average $accu_y$ of 99.78%, $prec_n$ of 99.54%, $reca_l$ of 99.54%, $F1_{score}$ of 99.54%, and MCC of 99.40%. On the testing set, the ABTC-IDLMOT achieved an average $accu_y$ of 96.55%, $prec_n$ of 92.92%, $reca_l$ of 92.95%, $F1_{score}$ of 92.93%, and MCC of 90.65%.

Figure 6 depicts the training and validation $accu_y$ of the ABTC-IDLMOT method over 50 epochs.

TABLE II. PERFORMANCE OF ABTC-IDLMOT

Class	$Accu_y$	$Prec_n$	$Reca_l$	$F1_{score}$	MCC
Training set					
notumor	100.00	100.00	100.00	100.00	100.00
meningioma	99.59	100.00	98.23	99.11	98.85
glioma	99.63	98.41	100.00	99.20	98.97
pituitary	99.92	99.76	99.92	99.84	99.79
Average	**99.78**	**99.54**	**99.54**	**99.54**	**99.40**
Testing set					
notumor	98.95	98.43	97.75	98.09	97.37
meningioma	93.87	87.75	86.89	87.32	83.27
glioma	94.96	89.53	89.53	89.53	86.21
pituitary	98.43	95.95	97.64	96.79	95.76
Average	**96.55**	**92.92**	**92.95**	**92.93**	**90.65**

Fig. 6. $Accu_y$ curves of the proposed ABTC-IDLMOT technique.

IV. CONCLUSION

This study presented an ABTC-IDLMOT technique, comprising WF and Otsu's threshold-based image preprocessing, EfficientNetV2M-based feature extraction, CNN-MHA-based BT classification, and RMSProp optimization-based hyperparameter tuning. The experimental study of the proposed technique involved a benchmark image dataset, and the findings were measured using various metrics. The performance validation illustrated a testing accuracy of 96.55%. The limitations of this study include a reliance on a limited dataset, which may affect the generalizability across diverse populations. Also, difficulty can be seen due to highly heterogeneous tumor types and varying MRI quality. Future work may focus on expanding the dataset, integrating multimodal clinical data, and exploring more robust techniques.

REFERENCES

[1] Z. Li and X. Zhou, "A Global-Local Parallel Dual-Branch Deep Learning Model with Attention-Enhanced Feature Fusion for Brain Tumor MRI Classification.," *Computers, Materials & Continua*, vol. 83, no. 1, Apr. 2025, Art. no. 739, https://doi.org/10.32604/cmc.2025.059807.

[2] R. Vankdothu, M. A. Hameed, and H. Fatima, "A Brain Tumor Identification and Classification Using Deep Learning based on CNN-LSTM Method," *Computers and Electrical Engineering*, vol. 101, July 2022, Art. no. 107960, https://doi.org/10.1016/j.compeleceng.2022.107960.

[3] X. Gu, Z. Shen, J. Xue, Y. Fan, and T. Ni, "Brain Tumor MR Image Classification Using Convolutional Dictionary Learning With Local Constraint," *Frontiers in Neuroscience*, vol. 15, May 2021, https://doi.org/10.3389/fnins.2021.679847.

[4] B. R. Sree *et al.*, "Brain Tumor Detection and Classification using Magnetic Resonance Imaging and Machine Learning Approaches," in *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, Mar. 2022, pp. 1729–1734, https://doi.org/10.1109/ICCMC53470.2022.9754077.

[5] R. Sundarasekar and A. Appathurai, "Efficient brain tumor detection and classification using magnetic resonance imaging," *Biomedical Physics & Engineering Express*, vol. 7, no. 5, Apr. 2021, Art. no. 055007, https://doi.org/10.1088/2057-1976/ac0ccc.

[6] L. J. Muhammad, I. Badi, A. A. Haruna, I. A. Mohammed, and O. S. Dada, "Deep Learning Models for Classification of Brain Tumor with Magnetic Resonance Imaging Images Dataset," in *Computational Intelligence in Oncology: Applications in Diagnosis, Prognosis and Therapeutics of Cancers*, K. Raza, Ed. Singapore: Springer, 2022, pp. 159–176.

[7] T. Rohini and P. Praveen, "An Intuitive Approach on Transfer Learning with an IBF+IHP Model for Stroke Classification and Prediction," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19655–19660, Feb. 2025, https://doi.org/10.48084/etasr.9031.

[8] M. Rasool, A. Noorwali, H. Ghandorh, N. A. Ismail, and W. M. S. Yafooz, "Brain Tumor Classification using Deep Learning: A State-of-the-Art Review," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16586–16594, Oct. 2024, https://doi.org/10.48084/etasr.8298.

[9] M. F. Khan, A. Iftikhar, H. Anwar, and S. A. Ramay, "Brain Tumor Segmentation and Classification using Optimized Deep Learning," *Journal of Computing & Biomedical Informatics*, vol. 7, no. 01, pp. 632–640, June 2024.

[10] M. Geetha, V. Srinadh, J. Janet, and S. Sumathi, "Hybrid Archimedes Sine Cosine optimization enabled Deep Learning for multilevel brain tumor classification using MRI images," *Biomedical Signal Processing and Control*, vol. 87, Jan. 2024, Art. no. 105419, https://doi.org/10.1016/j.bspc.2023.105419.

[11] S. Singh and V. Saxena, "Classification and Segmentation of MRI Images of Brain Tumors Using Deep Learning and Hybrid Approach," *International Journal of Electrical and Computer Engineering Systems*, vol. 15, no. 2, pp. 163–172, Feb. 2024, https://doi.org/10.32985/ijeces.15.2.5.

[12] A. Sahu, P. K. Das, I. Paul, and S. Meher, "A Hybrid Deep Learning Framework for Automatic Detection of Brain Tumours Using Different Modalities," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 9, no. 2, pp. 1216–1225, Apr. 2025, https://doi.org/10.1109/TETCI.2024.3442889.

[13] S. Tabatabaei, K. Rezaee, and M. Zhu, "Attention transformer mechanism and fusion-based deep learning architecture for MRI brain tumor classification system," *Biomedical Signal Processing and Control*, vol. 86, Sept. 2023, Art. no. 105119, https://doi.org/10.1016/j.bspc.2023.105119.

[14] S. Rajendran *et al.*, "Automated Segmentation of Brain Tumor MRI Images Using Deep Learning," *IEEE Access*, vol. 11, pp. 64758–64768, 2023, https://doi.org/10.1109/ACCESS.2023.3288017.

[15] M. Aamir *et al.*, "A deep learning approach for brain tumor classification using MRI images," *Computers and Electrical Engineering*, vol. 101, July 2022, Art. no. 108105, https://doi.org/10.1016/j.compeleceng.2022.108105.

[16] P. K. Tiwary, P. Johri, A. Katiyar, and M. K. Chhipa, "Deep Learning-Based MRI Brain Tumor Segmentation With EfficientNet-Enhanced UNet," *IEEE Access*, vol. 13, pp. 54920–54937, 2025, https://doi.org/10.1109/ACCESS.2025.3554405.

[17] K. Deng *et al.*, "OS-DETR: End-to-end brain tumor detection framework based on orthogonal channel shuffle networks," *PLOS ONE*, vol. 20, no. 5, 2025, Art. no. e0320757, https://doi.org/10.1371/journal.pone.0320757.

[18] H. Zhu, J. Huang, K. Chen, X. Ying, and Y. Qian, "multiPI-TransBTS: A Multi-Path Learning Framework for Brain Tumor Image

Segmentation Based on Multi-Physical Information." arXiv, Sept. 18, 2024, https://doi.org/10.48550/arXiv.2409.12167.

[19] V. Sharma, M. Kumar, and A. K. Yadav, "3D AIR-UNet: attention–inception–residual-based U-Net for brain tumor segmentation from multimodal MRI," *Neural Computing and Applications*, vol. 37, no. 16, pp. 9969–9990, June 2025, https://doi.org/10.1007/s00521-025-11105-9.

[20] M. Moshfegh, M. Nikpour, and M. Mobini, "Wavelet-based Denoising of Magnetic Resonance Images Using Optimized Exponential Function Thresholding and Wiener Filter," *International Journal of Engineering*, vol. 37, no. 12, pp. 2560–2569, 2024, https://doi.org/10.5829/ije.2024.37.12c.14.

[21] K. Wisaeng and W. Sa-ngiamvibool, "Brain Tumor Segmentation Using Fuzzy Otsu Threshold Morphological Algorithm." *IAENG International Journal of Applied Mathematics,* vol.53, no. 2, pp .1-12, June 2023.

[22] G. M. S. Himel, M. M. Islam, and M. Rahaman, "Utilizing EfficientNet for sheep breed identification in low-resolution images," *Systems and Soft Computing*, vol. 6, Dec. 2024, Art. no. 200093, https://doi.org/10.1016/j.sasc.2024.200093.

[23] F. Mehmood, S. Ahmad, and T. K. Whangbo, "An Efficient Optimization Technique for Training Deep Neural Networks," *Mathematics*, vol. 11, no. 6, Jan. 2023, Art. no. 1360, https://doi.org/10.3390/math11061360.

[24] M. Nickparvar, "Brain Tumor MRI Dataset." Kaggle, Available: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset.