Research Paper

Automated Medical Image Segmentation Using Hybrid Transfer Learning Attention Network

Rathiya R[1], Suja GP[2*], Harsha BK[3], Prakash Kanchikere Mathad[4], Sk. Hedayath Basha[5], Kavitha Shanmugakani[6]

1. *Department of Information Technology, Dr. N.G.P. Institute of Technology, Coimbatore, Tamil Nadu, India.*
2. *Department of Computer Science, Muslim Arts College, Thiruvithancode, Kanniyakumari – 629174, Tamil Nadu, India.*
3. *School of Computing and Information Technology, REVA University, Bengaluru, Karnataka, India.*
4. *Department of Electronics and Communication Engineering, Bapuji Institute of Engineering & Technology, Karnataka – 577004, India.*
5. *Department of Electronics and Communication Engineering, R.M.K. College of Engineering and Technology, Puduvoyal, Tamil Nadu, India.*
6. *Department of Computer Science and Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, Tamil Nadu, India.*

Citation Rathiya R, Suja GP, Harsha BK, Mathad PK, Basha SkH, Shanmugakani K. Automated Medical Image Segmentation Using Hybrid Transfer Learning Attention Network. International Journal of Medical Toxicology and Forensic Medicine. 2026; 16:E53261.

https://doi.org/10.22037/ijmtfm.v16.53261

Article info:

Received: 24 August, 2026

Accepted: 01 September, 2026

Published: 19 September, 2026

Keywords:

Brain tumor classification, Magnetic resonance imaging (MRI), Transfer learning, Deep learning, Convolutional neural networks (CNN), EfficientNet, ResNet, DenseNet, Medical image analysis, Computer-aided diagnosis, ROC–AUC, Sensitivity, Specificity

ABSTRACT

Background: Diagnosis of brain tumors using magnetic resonance imaging (MRI) is essential for clinical diagnosis and treatment planning because it is accurate and enables early detection. However, there is a shortage of annotated medical data and inter-class similarity among tumor types, posing a challenge for automated classification systems. The proposed research is a Hybrid Transfer Learning Attention Network (HTLA-Net) for multi-class brain tumor classification, based on pre-trained convolutional neural networks and stage-by-stage fine-tuning.
Methods: The proposed architecture of HTLA-Net is based on a combination of backbone-specific pre-processing, feature extraction, and selective layer modification to expand learning while maintaining consistent generalization. The authors evaluated four state-of-the-art architectures, i.e., EfficientNetB0, ResNet50, MobileNetV2, and DenseNet121, on a publicly available four-class MRI dataset, comprising glioma, meningioma, pituitary macroadenoma, and healthy cases.
Results: The experiment shows that DenseNet121 achieved the best overall performance, with a test accuracy of 85.96, a macro F1-score of 0.8557, a specificity of 0.9532, and an ROC-AUC of 0.9597, attributable to its strong ability to distinguish between classes and its robust diagnostic performance.
Conclusion: The proposed hybrid training plan is an effective way to mitigate overfitting and improve calibration compared to simpler architectures. These findings affirm the suitability and efficiency of deep models that rely on transfer learning for reliable computer-based diagnosis of brain tumors in settings with limited medical practice and data.

*** Corresponding Author:**
Suja G P, PhD
Department of Computer Science, Muslim Arts College, Thiruvithancode, Kanniyakumari – 629174, Tamil Nadu, India.
E-mail: *sujaassistantprofessor@gmail.com*

Introduction

Brain tumors are a severe neurological condition that needs to be diagnosed correctly and in a timely manner to help in treatment and survival. The high soft-tissue contrast and non-invasive nature of magnetic resonance imaging (MRI) make it the preferred imaging modality. But manual interpretation is time-consuming and also prone to radiologist variability, especially in the distinction of glioma, meningioma, and pituitary adenoma. Deep learning has recently been developed to enhance computer-aided medical image analysis by automatically extracting and classifying features. Transfer learning can be applied in medical imaging applications where there are few labeled images. Hence, this paper presents a hybrid transfer learning model for automatic multiclass brain tumor analysis of MRI images, with the objective of improving accuracy, robustness, and computational efficiency. Transfer learning and hybrid feature extraction models are improving segmentation accuracy and clinical decision-making [1]. These systems can minimize variability and interpretation time for MRI, computed tomography and multimodal scans. Cross-dimensional transfer learning can transfer spatial and contextual representations from a pre-trained network to medical imaging domains [2], accelerating convergence, reducing data dependence, and enhancing model stability compared with training from scratch [3]. Encoder–decoder architectures, especially U-Net and its variants, have produced state-of-the-art biomedical segmentation results [4]. However, a model trained on one dataset fails on an unseen dataset. The cross-domain robustness can be improved by learning deep stacked transformations and feature-alignment mechanisms [5]. In scientific fields, CNNs have shown flexibility in biomedical segmentation [7], and deep learning has also shown flexibility in other domains of science [6]. Tumor heterogeneity, indeterminate tumor boundaries, and scarcity of annotated data continue to impact brain tumor segmentation. The current models can also be complex and lack robustness across datasets and optimization for sensitivity, specificity, and accuracy. To overcome these weaknesses, in this paper, we propose HTLA-Net, which integrates a pretrained encoder, multiscale feature fusion, and attention-based refinement for enhancement in boundary localization, contextual understanding, generalization, computational efficiency, and scalable deployment. The key findings are summarized below:

- Proposal of an HTLA-Net with transfer learning and staged fine-tuning for automated multi-class brain tumor classification from MRI images.

- Evaluation of 4 different pre-trained CNN models (EfficientNetB0, ResNet50, MobileNetV2 and DenseNet121).

- Adoption of a two-stage training procedure: First, the backbone's parameters are fixed and then fine-tuned with a lower learning rate.

- Using classification metrics such as accuracy, precision, sensitivity, specificity, macro F1-score, ROC-AUC, PR-AUC and LogLoss for comprehensive evaluation.

The proposed HTLA-Net will address the current shortcomings in medical image segmentation by integrating strength, effectiveness, and clinical reliability within a single deep learning model.

Deep learning technology has led to remarkable improvements in medical image analysis, especially in the detection, classification, and segmentation of tumors. Convolutional neural networks (CNNs) learn hierarchies of features directly from imaging data, as illustrated in Diagrams 1 and 2. Pretrained models such as VGG, ResNet, DenseNet, and EfficientNet have proven to be more effective than traditional machine-learning approaches. Transfer learning is especially useful for tackling small, annotated medical datasets by transferring the knowledge from large natural image repositories. Additionally, recent work has highlighted the need for additional metrics, such as sensitivity, specificity, F1-score, and ROC-AUC, to ensure clinically reliable performance. Applications of medical image segmentation include successful segmentation of radiological images, surgical photos, and organ delineation. Kourounis et al. [9] introduced deep neural networks trained to detect boundaries in an organ-donation image. In a study by Xia et al. [10] of CNNs, encoder-decoder networks, attention models and transformer-based models, U-Net was found to be a basic model. Isensee et al. [11] proposed nnU-Net, which automatically configures the preprocessing, architecture, and training parameters for each dataset. It has been proven effective across all medical segmentation tasks, but it is time-consuming in hardware-limited environments. Minaee et al. [12] classified the different approaches to segmentation into fully convolutional, recurrent, generative adversarial, and transformer-based models. These studies show how handcrafted features have been replaced by data-driven deep architectures. Wang et al. [13] and Jiang et al. [14] pointed out reliability, interpretability, generalization, and clinically meaningful evaluation. Liu et al. [15] also pointed out the challenge of class imbalance, scarce annotations, and inconsistent imaging protocols, as well as the need for multiscale feature fusion and attention mechanisms. Semi-supervised and weakly supervised methods were presented by Wang et al. [16] and required only a few annotations. Deep learning and Transfer learning have proven highly successful in a variety of medical image analysis applications, such as classification, detection, and segmentation. Studies using the encoder-decoder paradigm coupled with attention mechanisms, domain adaptation, and annotation-efficient learning [4, 5, 7–16] have garnered special interest in segmentation-based approaches. The objective of the present investigation differs from the above, as it involves image classification rather than pixel-level segmentation. The above methods have given insights into feature representation and generalization in medical imaging. In this background, this study introduces a method, called Hybrid Transfer Learning Attention Network (HTLA-Net), that is a combination of a feature extraction method, CNN, and a staged fine-tuning method for the classification of Glioma, Meningioma, Pituitary Macroadenoma, and Healthy MRI images. To overcome these drawbacks, we propose the Hybrid Transfer Learning Attention U-Net (HTLA-Net), which combines a pretrained encoder, multiscale feature fusion, and attention-based refinement. The model is designed to enhance the accuracy of tumor-boundary segmentation, improve generalization, and reduce computation time, enabling scalable and clinically reliable medical image segmentation for deployment across diverse imaging scenarios and resource-constrained healthcare environments.

Materials and Methods

Cell Line and Culture Conditions

In this research, a Hybrid Transfer Learning Attention Network (HTLA-Net) is proposed for automated multi-class brain tumor classification from MRI images. The proposed framework is a systematic pipeline comprising dataset preparation, image preprocessing, feature extraction from images using transfer learning, classification, staged fine-tuning, and performance evaluation. Four CNN architectures, namely EfficientNetB0, ResNet50, MobileNetV2, and DenseNet121, are each used as feature-extraction backbones and pre-trained on the ImageNet dataset. In the first phase of training, the pre-trained convolutional backbone is fixed, and the low- and mid-level visual representations are preserved. The extracted feature maps are compactly represented by global average pooling, followed by dropout regularization and a final fully connected classification layer. The softmax activation function outputs probability scores of the 4 target classes. In the second stage of training, the selected higher-level layers of the backbone are unfrozen and fine-tuned with a lower learning rate, enabling fine-tuning and domain-specific adaptation while avoiding overfitting.

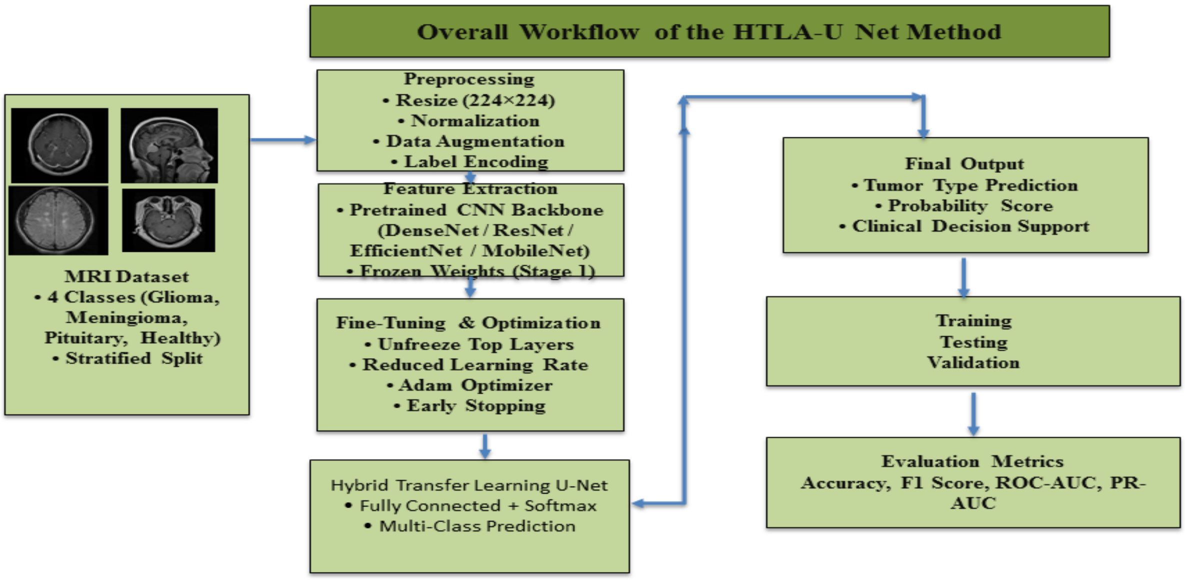
In this paper, we present automated multi-class brain tumor classification using magnetic resonance imaging (MRI). The aim of the suggested framework will be to utilize the pretrained convolutional neural networks to extract robust features and to combine fine-tuning with specific medical data [17]. The architecture is designed to improve discriminative performance across both tumor types, and its computational efficiency is sufficient for successful deployment in clinical settings at scale. The suggested system maintains a systematic pipeline that includes dataset preparation and preprocessing, transfer-learning feature extraction, classification refinement, and performance optimization. Figure 1 presents the block diagram of the overall workflow of the proposed Hybrid Transfer Learning Attention Network (HTLA-Net). The framework incorporates pretrained feature extraction, representation of global pooling, dropout regularization, and classification using a softmax into a staged fine-tuning strategy.

Dataset details

For this study, we used the brain MRI image dataset, which contains 1137 images across four clinically relevant classes: Glioma, Meningioma, Pituitary Macroadenoma, and Healthy. The number of images in the Glioma class is 295, in the Meningioma class is 277, in the Pituitary Macroadenoma class is 274, and in the Healthy class is 291. All images are 2D MRI slices, with categories of each tumor or health label. Stratified data splitting was used to ensure that the training, validation, and testing sets contained the same class distribution approximately. In particular, 70% of images were used for training, 15% for validation, and 15% for testing to preserve class proportions and alleviate sampling bias [18]. Table 1 shows the dataset distribution across classes.

Table 1. Dataset distribution across classes.

Class	Number of Images	Percentage (%)
Glioma	295	25.94%
Meningioma	277	24.36%
Pituitary Macroadenoma	274	24.10%
Healthy	291	25.60%
Total	**1137**	**100%**



International Journal of
Medical Toxicology & Forensic Medicine

Figure 1. Overall workflow of the Hybrid Transfer Learning Attention Network (HTLA-Net).

To ensure the consistency and stability of the data for model training, all MRI images went through a standardized preprocessing pipeline. To meet the input requirements of the pretrained backbone architectures, each image was resized to 224 × 224 pixels. The pixel values were converted to floating-point and preprocessed using the backbone-specific processing functions for EfficientNetB0, ResNet50, MobileNetV2, and DenseNet121. To improve generalization and reduce overfitting, data augmentation was applied during training. Augmentation operations were performed randomly, including horizontal flipping, brightness adjustment, and slight contrast perturbation.

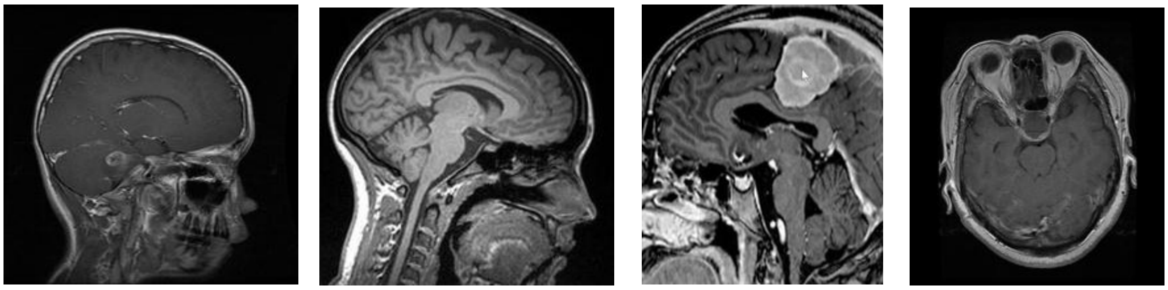
For the multi-class classification, the four class labels were encoded using one-hot encoding [19]. HTLA-Net is built on convolutional backbone models trained on ImageNet. In stage 1, the backbone network is fixed, preserving the low- and mid-level feature representations it has learned. In Stage 2, the higher layers are unfrozen and fine-tuned with a smaller learning rate. The staged training approach enables domain adaptation and mitigates the risk of overfitting to the limited MRI data. The main idea of the proposed HTLA-Net is to use transfer learning with convolutional backbones pre-trained on ImageNet. In the initial phase of training, the convolutional base of each backbone network was frozen to retain the learned low- and mid-level visual representations. The training plan aimed to strike a balance between model complexity and generalization performance, thereby ensuring consistent classification across all tumor classes. The dataset images of 4 classes are provided in Figure 2.

Mathematical *Modeling and Analysis*

Let the labeled brain MRI dataset be defined as

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N} \qquad (1)$$

where $x_i \in \mathbb{R}^{H \times W \times C}$ denotes the i-th MRI image with height H, width W, and C channels; $y_i \in \{1, \dots, K\}$



(a) (b) (c) (d)

International Journal of
Medical Toxicology & Forensic Medicine

Figure 2. Representative MRI images from the four classes (a) Glioma, (b) Healthy, (c) Meningioma, (d) Pituitary Macroadenoma.

is the corresponding class label; N is the total number of samples; and $K = 4$ Represents Glioma, Meningioma, Pituitary Macroadenoma, and Healthy classes.

The objective is to learn a parametric classification function f_θ that maps an input MRI image to a probability distribution over the four classes:

$$f_\theta(x_i) = p_i \quad \textbf{(2)}$$

where $p_i = [p_{i1}, p_{i2}, \dots, p_{iK}]$ is the predicted class-probability vector and θ denotes the trainable model parameters.

Feature Extraction and Classification

The proposed Hybrid Transfer Learning Attention Network (HTLA-Net) consists of a pretrained encoder followed by a classification head. The encoder extracts hierarchical feature representations from the input image. Let the encoder function be denoted as

$$\mathrm{F}_i = \Phi_{\theta_e}(X_i) \ (3)$$

where Φ_{θ_e} represents the pretrained convolutional backbone with parameters θ_e, and $\mathrm{F}_i \in \mathbb{R}^d$ Denotes the global feature vector after pooling.

The classification head transforms the feature vector into logits:

$$\mathrm{F}_i = \Phi_{\theta_e}(X_i) \ (3)$$

where $\mathbf{W} \in \mathbb{R}^{K \times d}$ is the weight matrix and $\mathbf{b} \in \mathbb{R}^K$ is the bias vector.

Probabilistic Output Modeling

The class probabilities are obtained using the softmax function:

$$P(y_i = k|X_i) = \frac{\exp(Z_{i,k})}{\sum_{j=1}^{K} \exp(Z_{i,j})}, k = 1,2, \dots, K \ (5)$$

The predicted class label $\hat{y}_i$ is determined as:

$$\hat{y}_i = \arg\max_k P(y_i = k|X_i) \ \ (6)$$

Consequently, the learned decision function partitions the high-dimensional feature space into K Separable regions corresponding to distinct pathological categories. The probabilistic output defined in (4) provides calibrated confidence estimates, supporting clinically interpretable predictions. This mathematical formulation ensures that the proposed HTLA-Net achieves stable convergence, improved generalization, and robust classification performance on limited medical imaging datasets [21].

Objective Function Formulation

The training objective is to minimize the categorical cross-entropy loss over the dataset. Let $\mathbf{y}_i \in \{0,1\}^K$ Denotes the one-hot encoded label vector. The empirical loss is defined as:

$$\mathcal{L}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k} \log P(y_i = k|X_i) \ (7)$$

Here:

- N denotes the total number of training samples,
- K denotes the number of classes,
- $y_{i,k}$ is the ground-truth indicator for the class k,
- $P(y_i = k|X_i)$ is given by (4).

Regularization and Optimization

To prevent overfitting, L_2 Regularization is applied to the trainable parameters:

$$\mathcal{L}_{reg} = \mathcal{L}(\theta) + \lambda\|\theta\|_2^2 \ \ (8)$$

where $\lambda > 0$ is the regularization coefficient.

Model parameters are optimized using the Adam algorithm. The update rule at iteration t is:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta\frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (9)$$

where:

- η is the learning rate,
- $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected first and second moment estimates,
- ϵ is a small constant for numerical stability.

Training Constraints

The optimization is subject to the following constraints:

$$\theta_e^{(0)} = \theta_{\mathrm{ImageNet}} \quad (10)$$

$$\theta_e \text{ partially frozen during stage 1} \quad (11)$$

$\eta_{\text{fine-tune}} < \eta_{\text{initial}}$ (12)

where:

- $\boldsymbol{\theta}_{\text{ImageNet}}$ denotes pretrained parameters,
- Stage 1 preserves pretrained representations,
- Fine-tuning employs a reduced learning rate.

Performance Metrics Formulation

The classification accuracy is defined as:

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(\hat{y}_i = y_i), \#(13)$$

where $\mathbb{I}(\cdot)$ is the indicator function.

Sensitivity (Recall) and Specificity are defined as:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (14)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (15)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

The macro F1-score is given by:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16)$$

Optimization Strategy and Convergence Analysis

The training process of the proposed HTLA-Net framework is formulated as a constrained nonlinear optimization problem. The objective is to determine the optimal parameter set $\boldsymbol{\theta}^*$ that minimizes the regularized empirical risk defined in (eq:7). Formally, the optimization problem can be expressed as:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \left[-\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k} \log P(y_i \right.$$
$$\left. = k|X_i) + \lambda \|\boldsymbol{\theta}\|_2^2 \right] \quad (17)$$

Here, $\boldsymbol{\theta} = \{\boldsymbol{\theta}_e, \mathbf{W}, \mathbf{b}\}$ represents the complete set of trainable parameters, including encoder parameters $\boldsymbol{\theta}_e$ and classification head parameters $(\mathbf{W}, \mathbf{b})$. The scalar λ Controls the trade-off between empirical loss minimization and parameter regularization.

Optimization is performed using the Adam algorithm, which adaptively updates parameters based on first- and second-order moment estimates of the gradients. At iteration t, the gradient of the objective function with respect to parameters is computed as:

$$g_t = \nabla_\theta \mathcal{L}_{reg}(\theta_t) \quad (18)$$

The biased first and second moment estimates are defined as:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (19)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (20)$$

where β_1 and β_2 are exponential decay rates for moment estimation. Bias-corrected estimates are computed as:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (21)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (22)$$

The parameter update rule is then given by:

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (23)$$

where η denotes the learning rate and ϵ is a small constant introduced for numerical stability.

The optimization is done in two phases. Stage 1: In this step, classification head parameters are updated, but pretrained encoder parameters are not. Stage 2 involves selective fine-tuning, which involves unfreezing the higher layers of the encoder and using a low learning rate. This restricts adaptation to specific domains, allowing domain-specific refinement without disturbing already learned representations. Under mild smoothness assumptions on the loss function and bounded gradients, the Adam optimization scheme will converge to a stationary point of the objective function under non-convex conditions, as found in deep neural networks. The pretrained initialization reduces the dimensionality of the search space and speeds up parameter localization by placing them in a desirable region of the loss landscape. It can be viewed as an overall framework for regularized empirical risk minimization with pretrained constraints. The hierarchical transfer of features, controlled parameter adaptation, and regularized optimization help the model effectively balance bias and variance [22].

Overall, the Hybrid Transfer Learning Attention Network (HTLA-Net) is a novel framework that combines convolutional pretrained feature extraction with fine-tuning and is organized for multi-class brain tumor classification. The architecture aims at balancing accuracy, sensitivity, specificity, and stability, hence offering a dependable base to automated clinical decision support systems in neuro-oncological imaging.

Results

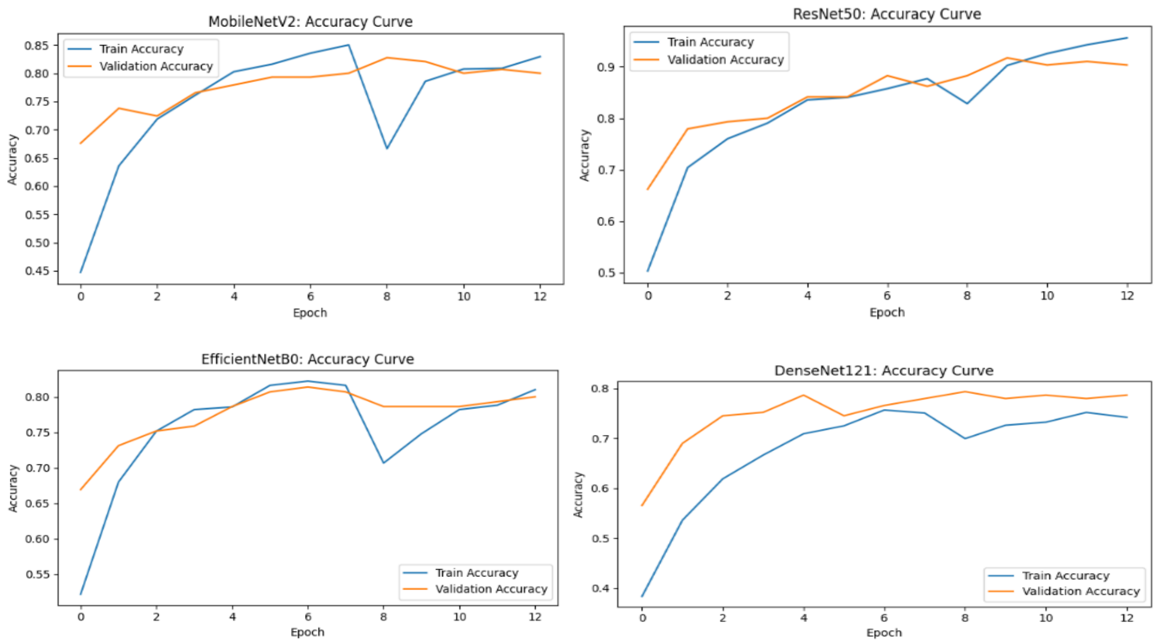
Experimental Setup

The experiments were all conducted using the publicly available Brain cancer MRI dataset, which comprised 1,137 labeled MRI images belonging to four classes: Glioma, Meningioma, Pituitary Macroadenoma, and Healthy cases. Stratified sampling was used to select the data to maintain class balance across the training (70%), validation (15%), and testing (15%) subsets. Each picture was cropped to 224 x 224 pixels and normalized with preprocessing functions specific to the backbone. The proposed Hybrid Transfer Learning was tested with four pretrained convolutional neural network architectures (EfficientNetB0, ResNet50, MobileNetV2, and DenseNet121). The training was done in two phases: first, a frozen-backbone phase, followed by selective fine-tuning with a low learning rate. Categorical cross-entropy loss and the Adam optimizer have been used. Early termination and learning-rate scheduling helped maintain stable convergence and avoid overfitting. Accuracy (ACC), Macro F1-score, Precision, Recall (Sensitivity, SEN), Specificity (SPE), ROC-AUC, PR-AUC, and LogLoss were used as the performance measures. All measures were calculated on the unseen test set.

Training and Validation Behavior

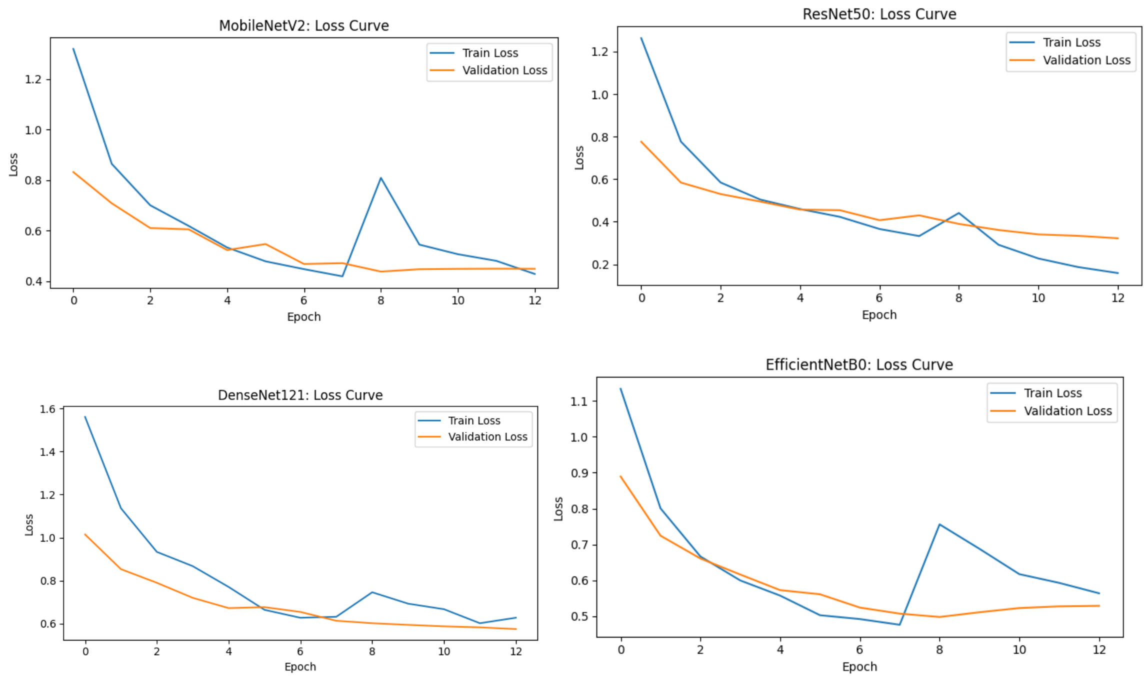
Figure 3 shows the training and validation accuracy curves, while Figure 4 shows the loss curves for the four backbone architectures tested. In the training process, it was seen that progressive learning was achieved in all the models, but the convergence behavior and generalization varied. From the graphs above, it can be seen that DenseNet121 showed relatively stable training and validation performance and achieved the highest test accuracy among the models compared. ResNet50 achieved very high training accuracy, but had a more significant gap between the training and test accuracy, indicating more overfitting. MobileNetV2 and EfficientNetB0 showed comparable convergence and had a lower test accuracy.

DenseNet121 showed smooth convergence with little oscillation between the training and validation curves, indicating stable generalization. Conversely, ResNet50 had a very high training accuracy (98.42) but a moderate generalization gap in test performance, indicating mild overfitting. Both MobileNetV2 and EfficientNetB0 converged equally, but with a relatively lower maximum performance.



International Journal of
Medical Toxicology & Forensic Medicine

Figure 3. Training and validation accuracy curves for EfficientNetB0, ResNet50, MobileNetV2, and DenseNet121.



International Journal of Medical Toxicology & Forensic Medicine

Figure 4. Training and validation loss curves for EfficientNetB0, ResNet50, MobileNetV2, and DenseNet121.

The fact that the two-stage transfer learning approach has a stable validation curve confirms that it is effective at mitigating catastrophic forgetting and enabling domain adaptation.

Quantitative Performance Evaluation

The quantitative results of the four transfer-learning models evaluated are summarized in Table ٢. DenseNet121 attained the highest overall classification accuracy of 85.96%, with the highest macro F1-score (0.8557), precision (0.8557), sensitivity (0.8586), and specificity (0.9532). These results show that DenseNet121 achieved the best overall classification, with a balanced distribution across the four MRI classes. The test accuracy obtained on ResNet50 was 83.63% with a macro F1-score of 0.8303. Notably, ResNet50 outperformed the other models tested, achieving the highest ROC-AUC (0.9737), the highest PR-AUC (0.9285), and the lowest LogLoss (0.3441), indicating strong probabilistic class separability. Its relative training accuracy compared to its test accuracy suggests a moderate gap in generalization. MobileNetV2 achieved a test accuracy of 82.46%, macro F1-score of 0.8005, ROC-AUC of 0.9582, and PR-AUC of 0.8788. These results indicate that a light architecture can still perform competitively in terms of classification performance and may also have an advantage in terms of computational time. EfficientNetB0 achieved a test accuracy of 80.12%, a macro F1-score of 0.7922, a specificity of 0.9336, and a ROC-AUC of 0.9595. Overall, it was the least accurate of the four models, but it had a high ROC-AUC, indicating that the model has class-discrimination capabilities.

The highest ROC-AUC (0.9737) was obtained with ResNet50, which exhibits better class separability in the probabilistic space. Nevertheless, the difference between the training accuracy (98.42) and the test accuracy (83.63) are indicative of moderate overfitting. Nevertheless, it has a value of PR-AUC (0.9285), indicating that it has a high precision-recall trade-off. MobileNetV2 achieved a test accuracy of 82.46% and a ROC-AUC of 0.9582, indicating that lightweight architectures can remain competitive in discriminative

Table 2. Performance comparison of transfer learning models.

Model	Test ACC	Test F1	Precision	Recall (SEN)	Specificity (SPE)	ROC-AUC	PR-AUC	LogLoss
DenseNet121	0.8596	0.8557	0.8557	0.8586	0.9532	0.9597	0.9042	0.4453
ResNet50	0.8363	0.8303	0.8313	0.8359	0.9453	0.9737	0.9285	0.3441
MobileNetV2	0.8246	0.8005	0.8314	0.8225	0.9412	0.9582	0.8788	0.5275
EfficientNetB0	0.8012	0.7922	0.7973	0.8010	0.9336	0.9595	0.8902	0.4937

International Journal of Medical Toxicology & Forensic Medicine

power while reducing computational complexity. The test accuracy of EfficientNetB0 was 80.12, with stable sensitivity and specificity, indicating reliable but relatively weak discriminative power. Across all models, ROC-AUC values were above 0.95 (with slight variation), indicating that the classes in the learned feature space are well separable.

D. ROC and Precision–Recall Analysis

The ROC and precision–recall curves for the evaluated models are shown in Figures 5 and 6. The highest ROC-AUC (0.9737) and PR-AUC (0.9285) were achieved by ResNet50, resulting in the best overall class separability in the probabilistic space. DenseNet121 also performed well in terms of discrimination, with a ROC-AUC of 0.9597 and a PR-AUC of 0.9042. The EfficientNetB0 model had a ROC-AUC of 0.9595 and a PR-AUC of 0.8902, while the MobileNetV2 model had a ROC-AUC of 0.9582 and a PR-AUC of 0.8788. The ROC-AUC values across all four architectures are consistently high, suggesting that the learned representations effectively discriminate between brain MRI classes.

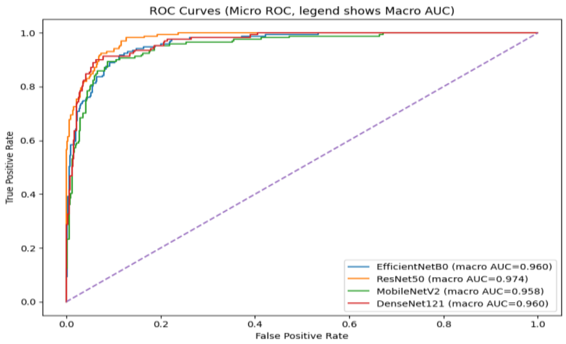


Figure 5. Precision–recall (PR) curves for EfficientNetB0, ResNet50, MobileNetV2, and DenseNet121.

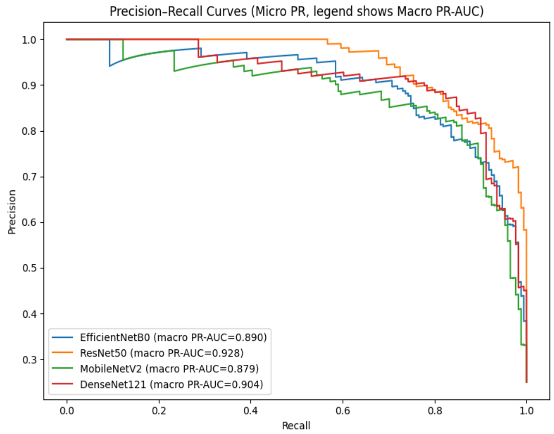


Figure 6. Precision–recall (PR) curves for EfficientNetB0, ResNet50, MobileNetV2, and DenseNet121.

The high PR-AUC values (>0.87 across all models) reflect a strong ability to handle class imbalance and to predict consistently positive values across tumor categories (Figure 7).

Grouped Metric and Baseline Comparison

Figure 8 compares Accuracy, Sensitivity, and Specificity across groups. DenseNet121 was more reliable across all three clinical measures, whereas ResNet50 showed slightly greater variability in sensitivity and specificity. MobileNetV2 had better sensitivity and slightly lower specificity than DenseNet121.

A baseline comparison of Accuracy, ROC-AUC, and LogLoss is shown in Figure 9. Although ResNet50 yielded the best ROC-AUC, DenseNet121 achieved a better balance between accuracy and calibration (LogLoss). A reduced LogLoss indicates greater probabilistic confidence and a lower misclassification penalty.

Discussion

The results of the experiments show that reliable performance can be achieved in multi-class brain tumor classification with transfer learning using controlled fine-tuning, even when the available MRI dataset is not large. The tested DenseNet121 architecture achieved the highest test accuracy (85.96%) and macro F1 score (0.8557), indicating comparatively balanced classification across the four MRI categories. ResNet50, on the other hand, showed the best ROC-AUC (0.9737) and PR-AUC (0.9285), indicating strong class-discriminative performance, but also a larger gap between training and test results, suggesting some degree of overfitting. These observations are similar to those reported in other works where only a few annotations are available for medical images [3, 13, 19], and pretrained CNN representations and transfer learning have been found to be more useful. The consistent validation performance across different fine-tuning stages also corroborates previous work showing that subtle adaptation of pretrained representations can enhance generalization and reduce overfitting [19, 21]. Moreover, the results of DenseNet121, ResNet50, MobileNetV2, and EfficientNetB0 show that architectural complexity is not the key to clinical classification performance; preserving and adapting discriminative features are also essential. However, the current findings are based on a single dataset (1,137 two-dimensional MRI images), and independent validation in larger, multi-center datasets and across different MRI acquisition protocols is needed to

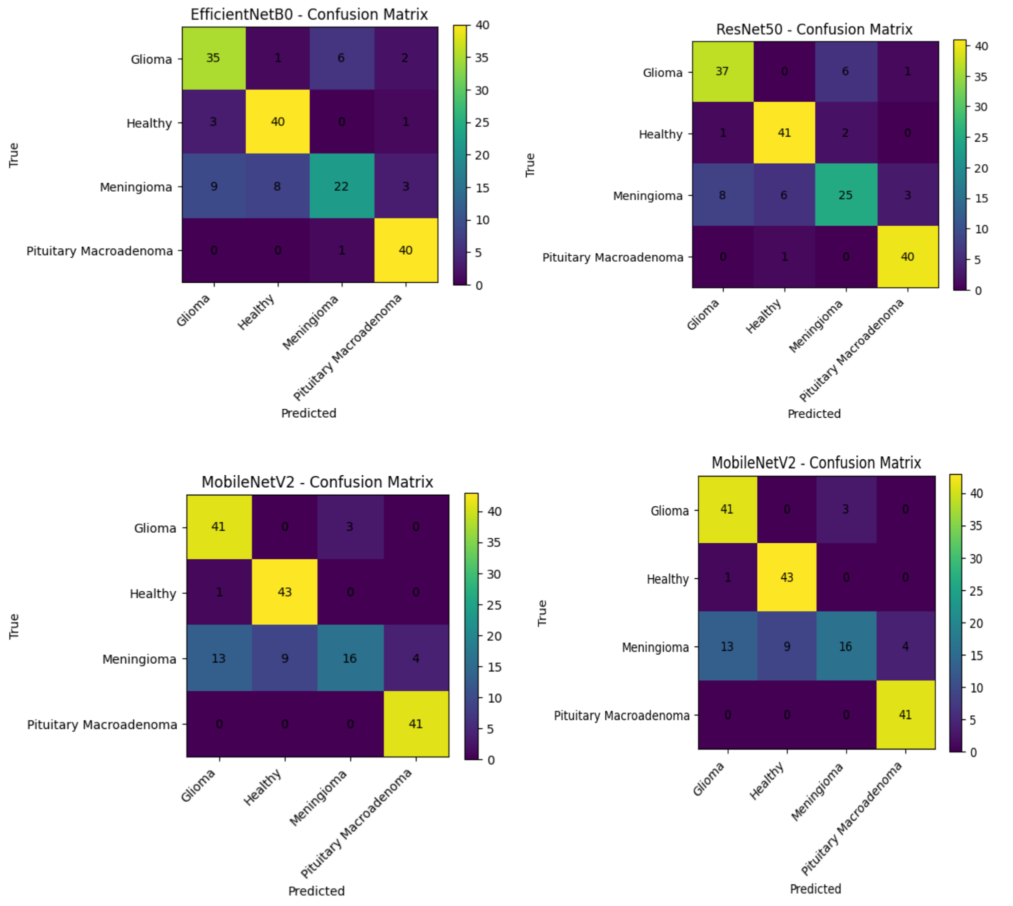


Figure 7. Confusion matrices for EfficientNetB0, ResNet50, MobileNetV2, and DenseNet121.

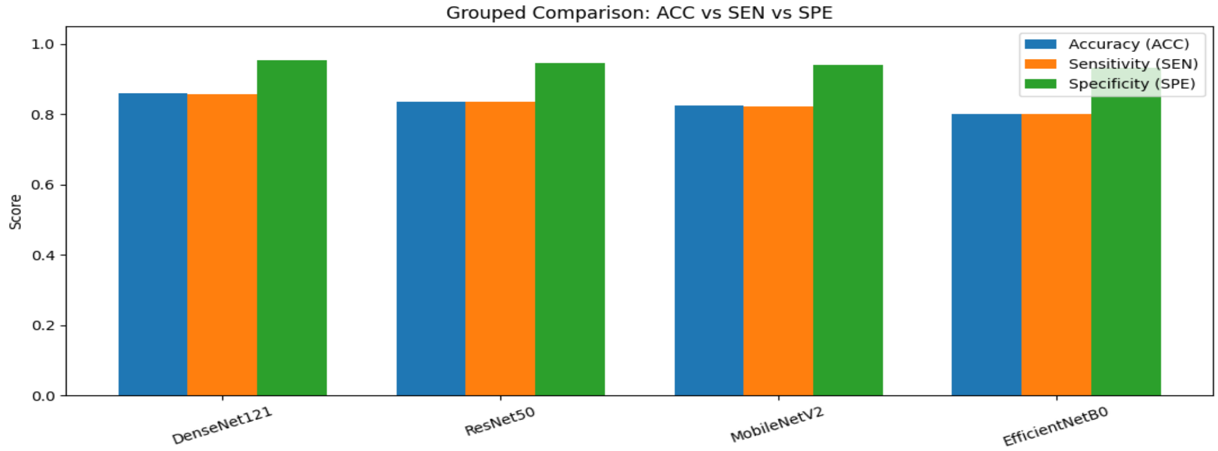


Figure 8. Grouped comparison of accuracy, sensitivity, and specificity for EfficientNetB0, ResNet50, MobileNetV2, and DenseNet121.

confirm the findings and further advance the imaging technique into clinical use.

Conclusion

The paper demonstrated a Hybrid Transfer Learning Attention Network (HTLA-Net) for automatically classifying brain tumors across multiple

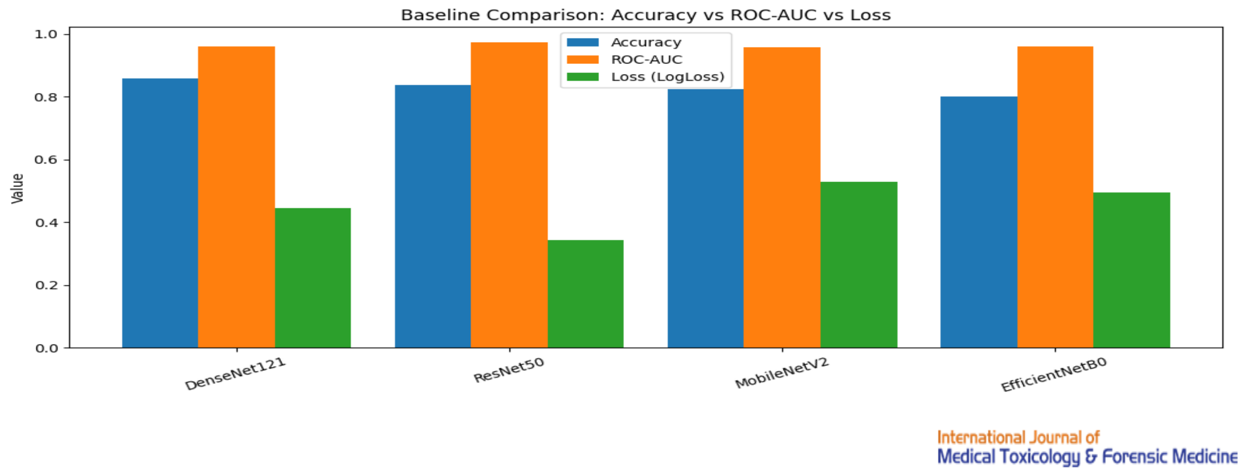


Figure 9. Baseline comparison of accuracy, ROC-AUC, and log loss for EfficientNetB0, ResNet50, MobileNetV2, and DenseNet121.

classes using MRI images. The suggested structure combines both pretrained convolutional backbones and staged fine-tuning and adaptive optimization to improve the search for discriminative features and the stability of generalization. Four state-of-the-art modern architectures, EfficientNetB0, ResNet50, MobileNetV2, and DenseNet121, were thoroughly examined using a single training protocol. It was experimentally shown that DenseNet121 has the most balanced performance, with the highest test accuracy (85.96%), macro F1-score (0.8557), specificity, and ROC-AUC. ResNet50 was much better at class separability in probabilistic space, whereas MobileNetV2 is a more desirable balance between performance and computational efficiency. The minimal difference in generalization between the two DenseNet121 models vindicates the success of the two-stage transfer learning procedure in preventing overfitting when limited medical information is available. Several main advantages of the suggested approach include: (i) good use of pretrained representations with smaller MRI datasets, (ii) good stability in convergence due to staged optimization, (iii) good improvement in various clinically relevant measures such as sensitivity and specificity, and (iv) flexibility in backbone settings. Altogether, the results confirm the soundness and applicability of hybrid transfer learning for effective computer-aided brain tumor diagnosis. Multi-modal learning can also be examined in future research by incorporating other imaging modalities, such as FLAIR and T2-weighted MRI, or clinical data to enhance diagnostic robustness. Generative adversarial network (GAN)-based data augmentation techniques may be another way to address data scarcity and improve generalization performance.

Ethical Approval

Not applicable.

Acknowledgment

None.

Funding

No external funding was received for this research.

Conflicts of Interest

The authors report that there are no competing interests to declare.

References

[1] Çevik N, Çevik T, Osman O, Alsubai S, Rasheed J. Advancing patient care with AI: a unified framework for medical image segmentation using transfer learning and hybrid feature extraction. Frontiers in Medicine. 2025;12. [DOI: 10.3389/fmed.2025.1589587]

[2] Messaoudi H, Belaid A, Salem D, Conze P. Cross-dimensional transfer learning in medical image segmentation with deep learning. Medical Image Analysis. 2023;88:102868. [DOI: 10.1016/j.media.2023.102868]

[3] Kora P, Ooi C, Faust O, Raghavendra U, Gudigar A, Chan W, et al. Transfer learning techniques for medical image analysis: A review. Biocybernetics and Biomedical Engineering. 2021. [DOI: 10.1016/j.bbe.2021.11.004]

[4] Lei T, Wang R, Wan Y, Du X, Meng H, Nandi A. Medical Image Segmentation Using Deep Learning: A Survey. ArXiv. 2020;abs/2009.13120. [DOI: 10.1049/ipr2.12419]

[5] Zhang L, Wang X, Yang D, Sanford T, Harmon S, Turkbey B, et al. Generalizing Deep Learning for Medical Image Segmentation to Unseen Domains via Deep Stacked Transformation. IEEE Transactions on Medical Imaging. 2020;39:2531-40. [DOI: 10.1109/tmi.2020.2973595]

[6] Lu L, Liang M. Deep learning-driven medical image analysis for computational material science applications. Frontiers in Materials. 2025. [DOI: 10.3389/fmats.2025.1583615]

[7] Haque I, Neubert J. Deep learning approaches to biomedical image segmentation. Informatics in Medicine Unlocked. 2020;18:100297. [DOI: 10.1016/j.imu.2020.100297]

[8] Verma A, Yadav A. Brain tumor segmentation with deep learning: Current approaches and future perspectives. Journal of Neuroscience Methods. 2025;418. [DOI: 10.1016/j.jneumeth.2025.110424]

[9] Kourounis G, Elmahmudi A, Thomson B, Nandi R, Tingle S, Glover E, et al. Deep learning for automated boundary detection and segmentation in organ donation photography. Innovative Surgical Sciences. 2024;10:131-41. [DOI: 10.1515/iss-2024-0022]

[10] Xia Q, Zheng H, Zou H, Luo D, Tang H, Li L, et al. A comprehensive review of deep learning for medical image segmentation. Neurocomputing. 2024;613:128740. [DOI: 10.1016/j.neucom.2024.128740]

[11] Isensee F, Jaeger P, Kohl S, Petersen J, Maier-Hein K. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods. 2020;18:203-11. [DOI: 10.1038/s41592-020-01008-z]

[12] Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D. Image Segmentation Using Deep Learning: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2020;44:3523-42. [DOI: 10.1109/tpami.2021.3059968]

[13] Wang J, Zhu H, Wang S, Zhang Y. A Review of Deep Learning on Medical Image Analysis. Mobile Networks and Applications. 2020;26:351-380. [DOI: 10.1007/s11036-020-01672-7]

[14] Jiang X, Hu Z, Wang S, Zhang Y. Deep Learning for Medical Image-Based Cancer Diagnosis. Cancers. 2023;15. [DOI: 10.3390/cancers15143608]

[15] Liu X, Song L, Liu S, Zhang Y. A Review of Deep-Learning-Based Medical Image Segmentation Methods. Sustainability. 2021;13(3):1224. [DOI: 10.3390/su13031224]

[16] Wang S, Li C, Wang R, Liu Z, Wang M, Tan H, et al. Annotation-efficient deep learning for automatic medical image segmentation. Nature Communications. 2020;12. [DOI: 10.1038/s41467-021-26216-9]

[17] Zhou W. Medical Image Information Segmentation Algorithm based on Deep Learning. 2025 International Conference on Intelligent Computing and Knowledge Extraction (ICICKE). 2025:1-6. [DOI: 10.1109/icicke65317.2025.11136622]

[18] Jamil N, Khan AA. Multi-Class Brain Tumor MRI Dataset: Glioma, Healthy Brain, Meningioma, and Pituitary Macroadenoma. Mendeley Data. 2025;V2. [DOI: 10.17632/82mtzd8x72.2]

[19] Osmani N, Rezayi S, Esmaeeli E, Karimi A. Transfer Learning from Non-Medical Images to Medical Images Using Deep Learning Algorithms. Frontiers in Health Informatics. 2024. [DOI: 10.30699/fhi.v13i0.549]

[20] Fu Y, Lei Y, Wang T, Curran W, Liu T, Yang X. A review of deep learning based methods for medical image multi-organ segmentation. Physica Medica. 2021;85:107-22. [DOI: 10.1016/j.ejmp.2021.05.003]

[21] Renard F, Guedria S, Palma N, Vuillerme N. Variability and reproducibility in deep learning for medical image segmentation. Scientific Reports. 2020;10. [DOI: 10.1038/s41598-020-69920-0]

[22] Singh M, Kr V, Gupta S, HoD D. Deep Learning-Driven Knowledge Transfer for Automated Medical Image Segmentation. 2025 3rd International Conference on Disruptive Technologies (ICDT). 2025:76-82. [DOI: 10.1109/icdt63985.2025.10986641]