Journal of Theoretical and Applied Information Technology
31st March 2026. Vol.104. No.6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TOWARDS ROBUST AND INTRINSICALLY INTERPRETABLE BRAIN TUMOR MRI CLASSIFICATION VIA ADAPTIVE ATTENTION-GUIDED FUSION

[1*]MORSA SRUTHI, [2]VEERRAJU GAMPALA

[1*]Dept. of Computer Science and Engineering , Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Andhra Pradesh 522302, India, ORCID : 0009-0007-8581-4619

[2]Faculty of Dept. of Computer Science and Engineering , Koneru Lakshmaiah Education Foundation,Green Fields, Vaddeswaram, Andhra Pradesh 522302 , India, ORCID:0000-0003-4166-7728

E-mail: [1]msruthi2506@gmail.com

ABSTRACT

The correct and sensible categorization of brain tumors through magnetic resonance imaging (MRI) is still a major problem with a heterogeneous tumor morphology, overlapping intensity patterns and the shortcoming of fixed feature fusion techniques in deep learning models. Current convolutional and transformer-based architectures tend to focus on raw accuracy but do not have adaptive fusion and inherent interpretability, which restricts their clinical reliability. In this research work, whether hierarchical attention integration and spatially adaptive dual-backbone fusion with Particle Swarm Optimization (PSO)-based scaling of features are examined that can enhance robustness and explainability in the process of brain tumor MRI classification. In a bid to resolve this issue, the Enhanced Spatial Attention Enhanced (SAE) Hybrid architecture is developed that includes the ConvNeXt and EfficientNetB0 backbones with multi-stage Convolutional Block Attention Modules (CBAM), spatial gating fusion, and a learnable PSO Weighted Scaling layer. The model was tested on a stringently partitioned Kaggle brain tumor MRI dataset of 4,117 images of three tumor subtypes, where data were strictly partitioned as training and validation and hold-out independent test (906 images). It has been experimentally shown that the proposed framework attains 95.03% accuracy, 95.03% F1-score, and 0.9949 ROC-AUC, which is statistically significantly higher than the baseline concatenation ($p = 0.012$). The fusion of spatially adaptive methods was better than scalar weighting methods, thereby establishing that features balancing regionally is more effective than global mixing in the classification of MRI. Moreover, the presence of embedded attention mechanisms offered inherent interpretability of dynamically focusing on regions of the tumor of interest without the use of post-hoc explanation.

Keywords: *Attention-Based Feature Fusion, Convolutional Block Attention Module, Deep Neural Network Architecture, Particle Swarm Optimization, Spatial Gating Mechanism, Transfer Learning.*

1. INTRODUCTION

Brain tumors are a major source of morbidity and mortality in the world, as they are one of the most dangerous neurological disorders [1]. Timely and correct diagnosis is essential to the effective therapeutic intervention because diagnosis may result in irreparable neurological damage or mortality. Magnetic Resonance Imaging (MRI) has been the best non-invasive imaging method of the brain tumor, providing better soft tissues contrast and anatomy [2]. However, manual interpretation of MRI is labour intensive and so much relies on experience especially of an expert and is therefore inconsistent among observers especially in resource-limited health care facilities [3].

Brain tumors are some of the most life-threatening neurological conditions, which are commonly related to high morbidity, mortality and long-term neurological impairment [4]. The World Health Organization has reported that central nervous system tumor has a significant burden on cancer globally especially because of late diagnosis and refusal to respond to therapy [5]. The accurate and

early classification of tumors is important as the choice of treatment, which may be either surgical resection, radiotherapy, chemotherapy or immunotherapy, is highly determined by the type of tumour and aggressiveness [6]. Even a slight diagnostic error can result in the use of inappropriate treatment plans, higher medical care expenses, and even patient deaths [7].

The most common non-invasive method of brain tumor diagnostics is Magnetic Resonance Imaging (MRI) because it has better soft-tissue contrast. Manual interpretation of MRI is however very reliant on the expertise of the radiologist and it is also inter-observer variable [8]. Delays in reporting and a lack of subspecialty knowledge only contribute to the risk of diagnosis in the clinical environment that is limited in resources. Heterogeneity of tumors-visual classifications-irregular boundaries [9], edema, necrosis, and overlapping intensity distributions cause huge problems in attaining consistent visual classification [10]. These complications make diagnosis of brain tumor not only a problem in medicine, but also a computationally intensive pattern recognition problem [11].

Regarding information technology, the problem is also of importance since current deep learning models tend to focus on predictive accuracy without prioritizing robustness, generalization, and interpretability [12]. Most convolutional neural networks are local texture predictors that do not comprehend the global context [13]. Transformer-built architectures represent long-range interactions, but can ignore fine-scale structural clues that are important in medical imaging [14]. To fill this gap hybrid architectures have been developed but most of them are based on either static feature concatenation or scalar weighting, and this does not dynamically adapt to region-specific tumor properties.

However, what is even more important is that the medical AI systems do not inherently have the capability to be interpreted, which is a significant obstacle to clinical implementation. Opaque predictions that lack clear reasoning machineries decrease clinician confidence, and curtails regulatory approvals. Grad-CAM post-hoc explanation tools offer biased information but do not imply that model decisions are structurally consistent with clinically meaningful features. As a result, there is still an urgent requirement to have architectures that combine attention and incorporate prioritization within forward propagation as opposed to being interpretable in a retrospective manner.

Moreover, the recent high-accuracy reports of literature are likely to lack a strict separation of training and test data or are based on dataset configurations that artificially make classification tasks easier (e.g., the inclusion of no tumor as a separate data category). These practices can exaggerate the performance measures and make generalization in practice impossible. Thus, the strength, statistical justification, and dynamic integration techniques should be scrutinized.

In this respect, the problem being examined, the development of a statistically sound, robust, and intrinsically interpretable dual-backbone fusion framework to classify brain tumors in MRI is both technically unclear and clinically pressing. The issue is factual in that misclassification has a direct effect on therapeutic decision-making, and due to the diversity of tumor morphology, the existing AI systems do not incorporate adaptive algorithms that adapt to it. The next step to fill this gap is to go beyond the focus on accuracy-based modeling into spatially adaptive, hierarchically attentive, and statistically validated structures with the ability to make reliable clinical translation.

The recent progress in deep learning (DL) has been profoundly affecting the analysis of medical images, thus being able to extract and classify features automatically. Convolutional Neural Networks (CNNs) are very effective at the local spatial structure (e.g. texture, edge, and intensity gradient) but fail to learn global dependencies that are important in context decoding. On the other hand, transformer-based designs like Swin Transformer model query the global spatial layout using self-attention but largely ignore local cues on finer scale which is important in accurate localization of the tumor [17]. This reciprocity makes it easier to factor in new hybrid architectures, new systems of convolutional and transformer systems to learn features in a more holistic manner [18].

Here, the Enhanced SAE Hybrid model is presented as a combination-based fusion-based model of MRI-based brain tumor classification. The design provides the ConvNeXt and Swin Transformer backbones to synergistically work with each other, based on the multi-stage attention and adaptive fusion pipeline. It has Convolutional Block Attention Modules (CBAM) as a hierarchical refinement of features, a spatially adaptive gating

mechanism, which performs a dynamic combination of features, and a Particle Swarm Optimization (PSO)-inspired learnable scaling strategy that inherently balances the importance of discriminative features [19].

In the current work, the proposed system is the Enhanced SAE Hybrid which is contrasted to the current brain tumor MRI classification systems in three main aspects, namely, (i) hierarchical multi-stage attention integration, (ii) spatially adaptable dual-backbone fusion, and (iii) embedded PSO-based learnable feature scaling. In contrast to traditional hybrid frameworks based on either static concatenation or scalar weighting, the suggested structure methods post-fusion and pre-fusion CBAM refinement with spatial gating, which allows the prioritization of regions with features on a dynamical basis [20]. The proposed model performed quantitatively with an accuracy of 95.03, F1-score of 95.03 and ROC-AUC of 0.9949 on a strictly held-out test of 906 images. Though other studies like R. Preetha (97.8%) and N. Noreen (99.3%) find larger values of the raw accuracy, most of the studies either use four-class classification including a no tumor category (which makes separability easier) or use smaller or custom data, or fail to explicitly employ a fully separated test protocol. Conversely, the current paper has used tight stratified splitting and statistical validation (paired t-test, $p = 0.012$) which is more evidence of the generalization than the optimization of peak performance.

Architecturally, most hybrid CNN-Transformer models including those suggested in J. Qezelbash-Chamak are learnable fused models with no intrinsic interpretability. The Enhanced SAE Hybrid adds triple stage CBAM attention directly into forward propagation, and does not use post-hoc explainability mechanisms (e.g. Grad-CAM). PSOWeightedScaling layer boosts discriminative latent channels as related to texture irregularity and contrast variance, specificity (97.51%), and negative predictive value (97.68%), which are seldom highlighted in the literature.

Another important result is that, scalar PSO-based late fusion alone (90.40% accuracy, $p = 0.081$) was not found to be statistically significant but spatial adaptive fusion coupled with hierarchical attention did cause statistically significant results. This empirically supports the fact that the significance of spatial adaptivity is greater in comparison to the global weighting strategies in the process of tumor classification using MRI. Cases of such comparative ablation analysis, however, are not always reported in earlier studies.

Although the proposed framework has some contributions, it has a number of limitations compared to the currently existing high-performing frameworks.

First, the accuracy obtained (95.03) is less as compared to some of the reported results such as 97-99% in articles such as N. Noreen and R. Preetha. Second, the Enhanced SAE Hybrid presents more computation complexity as a result of triple application of CBAM and dual-backbone processing. Third, despite better interpretability, the model is yet to be clinically tested through radiologist-in-the-loop assessment. Other frameworks on transformers (e.g. P. Chauhan) focus on patch-level explainability in particular clinical data, but the current paper uses visualization-based qualitative evaluation that does not utilize explicit human validation metrics.

Fourthly, the dataset in question (Kaggle MRI dataset) is made of 2D T1-weighted contrast-enhanced slices. It is possible that the proposed architecture can be expanded to 3D volumetric MRI or multi-sequence inputs in order to enhance the robustness of the diagnostic. Lastly, although PSO-based scaling showed benefits of channel-level recalibration, the incremental gain of spatial gating is achieved.

The proposed architecture overcomes the traditionally harmful shortcomings of existing frameworks by encompassing hierarchical attention, spatially so flexible fusion and bio-inspired feature maximisation hence yielding to clinically interpretable, robust, and generalized deep learning to automated brain tumour pathology.

2. RELATED WORK

Fusion-based deep learning models to classify brain tumors have been little studied in the existing literature, whereas single machine learning (ML) and deep learning (DL) models have been extensively studied in the context of medical image classification and segmentation. The accurate diagnosis of the object is important in clinical processes, which serves as an incentive to investigate hybrid fusion architectures that can increase reliability, interpretability, and generalization. The literature can be broadly divided into four categories, which are traditional ML-based classification, CNN-based methods, transformer and attention-driven models, and hybrid fusion models.

Previous research was mainly done based on handcrafted features and traditional classifiers. Chen, et al. [14] proposed a Fully Automatic Heterogeneous Segmentation (FAHS-SVM) system which used Extreme Learning Machines (ELM) to identify tumours. Asiri et al. [9] proposed a two-stage SVM-based detection pipeline that combines adaptive Wiener filtering and Radial Basis Function (RBF) neural networks with a higher accuracy of more than 98%. Bahadure et al. [11] applied SVM-based classification that gave a Dice coefficient of 0.961, whereas Alqhtani et al. [6] enhanced the detection with a fuzzy C-means clustering using SVM with 98.2 percentage accuracy. Hossain et al. [18] used CNNs to continue with FCM-based segmentation, achieving 97.87 percent accuracy. These models are also effective but they are heavily dependent on manual feature design and they are not so scalable to multiclass features.

Automatic feature extraction In classification of tumors, CNNs transformed the process by revolutionizing the classification method. Yousef et al. [4] reported an accuracy of 99% with a hyper-tuned ResNet50. Badza et al. [10] showed over 96% accuracies with both custom and ensemble CNN architectures. Badza et al. [10] and Patro et al. [29] reported accuracies of over 96% each with custom and ensemble CNN architectures. Preetha et al. [30] used three-branch CNN and EfficientNetB2 fusion, which provided an accuracy of 97.8%. Xception, NasNet and DenseNet121 were benchmarked by Asif et al. [8], whereas multilevel CNN features were used by Noreen et al. [26] and Togacar et al. [32] resulting in over 99 percent accuracy. In spite of this advancement, the CNNs are usually characterized by the inflated number of parameters and poor interpretability.

Recently, attention mechanisms and transformers have improved global context modeling in the medical imaging field. The Swin Transformer versions and GAN-enhanced data were used by Asiri [1] and Almuhaimeed [2] in robust classification. MobileNetV2 is utilized by Rahman [24] in favor of lightweight performance, and Swin Transformer V2 is enhanced by Alam [25] to recognize multiclass tumors. Zhang [20] suggested a multidimensional attention residual network, and Chauhan [13] suggested PBViT with the accuracy of 95.8. Transformer and ConvNeXt backbones were optimized further by Lai [37] and Mehmood [38] to be applicable to multi-class tumor grading. These architectures code long-range dependencies, but these architectures are not intertwined with convolutional priors; they prevent sensitivity to texture in the low-levels.

Neural frameworks Hybrid Hybrid approaches focus on using CNNs to extract local features and transformers to perform global reasoning. Anwer et al. [7] showed Transformative Transfer Learning (TTL) on several pre-trained CNNs with 94.5 percent accuracy. Swin Transformer Al Bataineh [3] was an improvement of Swin Transformer built on ResNet50V2, generalizing and being more interpretable. Qezelbash-Chamak and Hicklin [19] proposed a learnable hybrid of ConvNeXt and Swin Transformer, and the advantage of adaptive dual backbone integration is proven. Yao [39] used gated attention using reparameterized convolution to detect tumors using YOLO, whereas Almufaraeh [5] optimized YOLOv5 and YOLOv7 to detect objects. Wang [33] and Xu [36] introduced active-contour-based and generative segmentation models that have better boundary accuracy.

Although these improvements are made, most hybrid models are purely based on static concatenation or scalar weighting without explicit spatial or channel-level adaptiveness as an open area to attention-directed, dynamically fused designs such as the Enhanced SAE Hybrid presented in this paper [21].

The idea of brain tumor classification with the help of MRI has become a topic of international research because it has potential implications to oncology, neurology, and AI-assisted diagnostics. Equally, the high performance of large-scale clinical classification models like Maddumala has similarly shown the performance of dermatologists, but has also generated discussion-scale debates over model transparency and reproducibility, as well as readiness to deploy in healthcare AI. In neuro-oncology, datasets and cross-centre benchmarking e.g. BraTS have been used to speed up the development of algorithms by harmonizing evaluation protocols. As a case in point, A. F. Al Bataineh fused Swin Transformer with ResNet50V2 on MRI classification and N. Noreen presented multi-level strategies of feature fusion with high accuracy reported. Similarly, P. Chauhan investigated patch-based transformer attention-based representation learning.

Despite such developments, the prevalent theme in international literature is the usage of fixed concatenation, scalar weight-based features, or post-hoc tools of interpretation. Moreover, a number of them fail to draw a clear line between validation and independent test protocols, which leads to some worries about ca reproducibility and the real-life generalization. To an international

audience, this problem is relevant not only in terms of brain tumor classification. The architectural requirement to integrate robustness, statistical reliability, and inherent interpretability is not a preference in the technical design of medical imaging, but rather a translational requirement since errors in decision-making have life-altering consequences in this area.

Contributions and Novelty

1: Dual-backbone fusion: In This work it Combines ConvNeXt and EfficientNet (Swin proxy) to complementary texture with context representation.
2: Integrated attention hierarchy: This uses triple stage CBAM to maintain uniform focus throughout all the fusion layers.
3: Spatial gating fusion: It Adds a learnable dynamic contribution to backbone per region.
4: PSO-based feature scaling: This Scales the balancing of latent space with meta-heuristic scaling layer.
5: Quantified interpretability: Offers statistical and proved visual and ablation outputs that support the architectural synergy ($p< 0.05$).

3. PROPOSED SYSTEM

3.1 Dataset Information

The Brain Tumour MRI Dataset of Kaggle was used in the study. The initial data set included four groups of T1-weighted contrast-enhanced MRI images: glioma, meningioma, pituitary and no tumor. We eliminated the no tumor group to concentrate on pathological classification leaving 4,117 images which were 1,321 glioma, 1,339 meningioma, and 1,457 pituitary images. Photographic images were normalised to 224 x224-pixel spatial resolution and three-channel data. Normalization The pixel intensities were divided by 255 to give the pixel intensities in the range of zero to one. The chosen dataset division was stratified sampling whereby 85 percent of the dataset was used in training and 15 percent used in internal validation resulting in a total of 2,976 training images and 525 validation images. The separated Testing folder offered a withheld test set of 906 images. The visual inspection before preprocessing consisted of image quality and contrast integrity.

3.2 Data handling and preprocessing

Resize and replicate channels:

$$I_{224} = \text{ReplicateTo3Channels}(\text{Res} \quad (1)$$

The fact that Eq. (1) indicates that every original image is resized to 224 by 224 pixels also indicates that (1) if the original image is single channel it is duplicated into three channels to allow processing by standard RGB pretrained backbones.

Intensity normalization:

$$I_{norm} = \frac{I_{224}}{255} \quad (2)$$

As seen inEq. (2) resizing the picture is done such that the pixel values will fall within the range of 0 and 1 by division by 255.

Contrast enhancement with CLAHE (applied before augmentation):

$$I_{clahe} = \text{CLAHE}(I_{norm}) \quad (3)$$

It is revealed in Eq. (3) that contrast-limited adaptive histogram equalization is done to the normalized image to enhance the local contrast prior to augmentation.

Augmentation pipeline composition:

$$I_{aug} = \mathcal{A}(I_{clahe}) = \{\text{rotation,translation,zoom} \quad (4)$$

As Eq. (4) demonstrates, sequence of augmentations rotation, translation, zoom, horizontal flip and contrast jitter are imposed on the CLAHE image in order to enhance diversity of training.

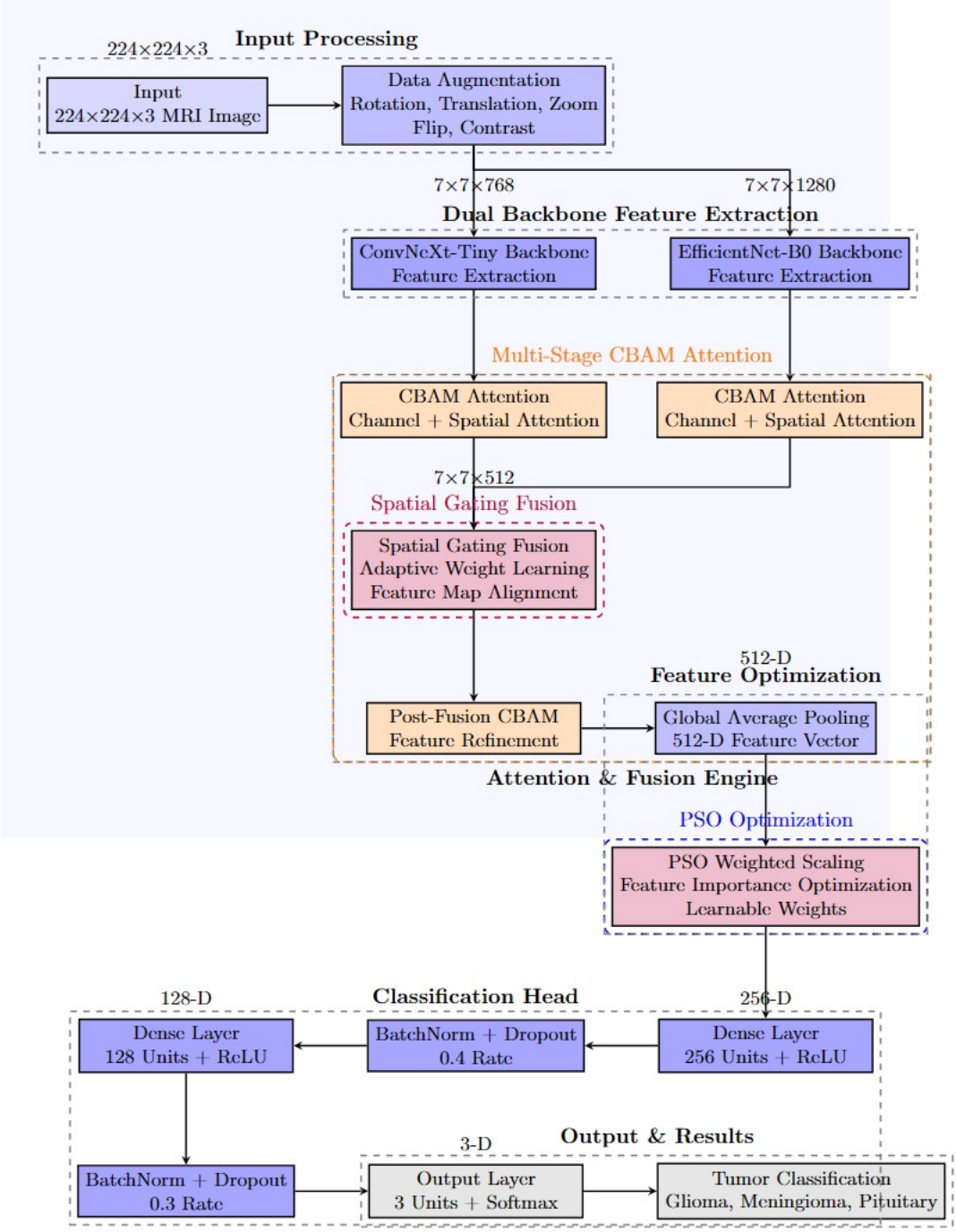

Figure 1. Improved SAE Hybrid Architecture:

Schematic description of the proposed Spatial-Attention-Enhanced (SAE) Hybrid architecture based on ConvNeXt and EfficientNetB0 backbones.

Particle Swarm Optimization (PSO) makes the features of MRI-based brain tumor classification interpretable, Convolutional Block Attention Modules (CBAM) are applied to refine features hierarchy, Spatial Gating Fusion (SGF) to align features both locally and globally, and feature scaling.

3.3. CBAM Application and Role

The proposed pipeline applies CBAM three times: first as a refinement of the ConvNeXt backbone outputs, second as a refinement of the transformer backbone outputs and lastly as a refinement of the result of the fusion of the two refined feature maps.

This three-purpose is not accidental, and it is one of the main novelties of the offered approach.
Channel attention computation inside CBAM:

$$M_c(F) = \sigma(W_2(\mathrm{ReLU}(W_1(\mathrm{AvgPool}(F)))))$$

(5)

The computation of channel attention in CBAM is demonstrated in Eq. (5): The average and max spatial pooling are used, and the resultant vectors of the two pools are fed to a small multilayer perceptron, the outputs of which are summed and subjected to a sigmoid gate to obtain values of per-channel importance.
Spatial attention computation inside CBAM:

$$M_s(F) = \sigma(\mathrm{Conv}_{7\times7}([\mathrm{AvgAlongChannel}$$

(6)

The calculation of spatial attention in CBAM is demonstrated in Eq. (6) represents the average and max of channels and concatenates them, then uses a seven-by-seven convolution and yields a spatial importance map.
CBAM refinement of a feature map:

$$F' = M_s(F \odot M_c(F)) \quad (7)$$

As Eq. (7) indicates, CBAM initially ranks channels based on the channel attention map and then ranks spatial locations based on the spatial attention map to generate the refined feature.
After spatial ganging fusion, Eq. (5), Eq. (6), and Eq. (7) are used both on backbone outputs and the fused feature map again.

3.4 Spatial Gating Fusion: Projection and Alignment

Project backbone outputs to a common channel dimension via 1×1 convolution:

$$F_c^p = \mathrm{Conv}_{1\times1}(F_c), F_t^p = \mathrm{Conv}_{1\times1} \quad (8)$$

It is demonstrated in Eq. (8) that convolutional backbone output and transformer backbone output are both projected to equal numbers of channels when using one-by-one convolutions thus they can be spatial fused.
Produce spatial gating weight maps:

$$W = \mathrm{Softmax}(\mathrm{Conv}_{1\times1}(\mathrm{ReLU}(\mathrm{Cor}$$

(9)

As indicated in (9), the projected pair of features is concatenated on a per-channel basis, run through a tiny convolutional subnet with a ReLU nonlinearity and a single-by-one convolutional postprocessing to yield a projection of spatial weights in a SoftMax, to produce adaptive gating maps.
Split spatial weights and compute weighted fusion:

$$F_{fused} = \mathrm{ReLU}(\mathrm{Conv}_{1\times1}(W_1 \odot F_c^p \quad (10$$

As indicated in (10), the gating maps are divided into two spatial masks, each mask is multiplied with the respective projected feature map, the two masked maps are added together, followed by a one-by-one convolution and a ReLU to produce the fused feature map.

4 PARTICLE SWARM OPTIMIZATION STRATEGIES

1. PSO scalar alpha search of ablation runs: In ablation runs which involve weighted late fusion, a scalar weight is identified by running PSO on the validation set to identify the optimal scalar mixing fraction between the backbone pooled outputs. This scalar PSO search is achieved in the PSO_Weighted_Fusion ablation only. See Eq. The fusion expression is a scalar expression, Eq. (12) for the scalar fusion expression and Eq. (13) for PSO updates.

2. PSO_Weighted_Scaling layer within Enhanced_SAE_Hybrid: In the original proposed model, PSO_Weighted_Scaling is modelled as a trainable vector of weights that are run through a sigmoid and applied as weighted elementwise to the pooled fused features. It is a learnable scaling layer and is also trained in gradient descent as part of the main network. See Eq. (11) for the layer formulation.

PSO-weighted feature scaling layer used inside the Enhanced SAE Hybrid:

$$f_{scaled} = \mathrm{Sigmoid}(w) \odot \mathrm{GAP}(I \quad (11)$$

As indicated in Eq. (11), the fused feature map is averaged globally to a vector, and the elements of each vector are then scaled by a respective trainable weight which is run through a sigmoid to limit its value range. The proposed model utilizes this operation and during standard model training, these weights are learned through backpropagation.
Scalar weighted fusion used in PSO ablation:

$$f_{scalar} = \alpha \cdot f_c + (1-\alpha) \cdot f_t \quad (12)$$

The simple scalar mixing of two pooled feature vectors of the PSO ablation is displayed in Eq. (12), where the scalar mixing weight is discovered through PSO on the validation set, and is fixed to be evaluated.

PSO particle update rules used to search for scalar alpha:

$$v^{(t+1)} = \omega v^{(t)} + c_1 r_1 (p - x^{(t)}) + c_2 r$$ (13

Eq. (13) depicts typical particle swarm velocity and position update rules applied when searching PSO; PSO is executed with the validation set and with a given number of particles and iterations and the optimal capacity has been found.

PSO hyperparameters and protocol:PSO of the ablation scalar search is executed with three particles, five iterations, inertia weight of 0.7, cognitive coefficient of 1.4 and social coefficient of 1.4 and alpha between [0.1 and 0.9]. The training of a lightweight fusion network on a training set evaluation is performed on each particle candidate and evaluated by validation accuracy. The global best alpha is the best validation accuracy. PSO is strictly tested on validation set, not test set.

This paper used a supervised deep learning experimental model to test in a systematic fashion that hierarchical attention integration, spatially adaptive dual-backbone fusion as well as PSO-based feature scaling enhance robustness and interpretability in MRI classification of brain tumor. To guarantee methodological equity and reproducibility, five architectures, which are BaselineConcat, SpatialGatingFusion, SpatialGatingCBAM, PSOWeightedFusion and the proposed Enhanced SAE Hybrid, were trained and tested in the same conditions of preprocessing, optimization and evaluation.

The Kaggle Brain Tumor MRI dataset available publicly was used as the subject of the experiments, comprising of T1-weighted contrast-enhanced MRI slices and divided into classes of glioma, meningioma, and pituitary tumor. The no tumor category was not included since it was not desired to include pathological discrimination. The data set consisted of 4, 117 images of 1, 321 glioma, 1, 339 meningioma, and 1,457 pituitary samples. The strict stratified splitting protocol was applied: 85 percent of the data (2,976 images) was placed in the training phase and 15 percent (525 images) in the internal validation phase and a totally new held-out test set of 906 images was considered only in the final performance. Test set was never used in training, hyperparameter optimization or PSO optimization and therefore data leakage was avoided and performance estimation was unbiased.

First of all, images were scaled to 224 x 224 to fit pretrained backbone input. Because the slices of the MRI were one channel, channel replication was used to generate three channel inputs that would fit directly into ImageNet-pretrained networks. The pixel intensity normalization was done based on the values scaling it to [0,1]. Before augmentation, Contrast Limited Adaptive Histogram Equalization (CLAHE) was used to increase local contrast and improve the ability to see tumor boundaries. The data augmentation was implemented on top of each other during training only and consisted of random rotation, translation, zoom, horizontal flipping, and contrast jitter. Validation and test data were not augmented.

Refinement of feature maps in both backbones in the proposed Enhanced SAE Hybrid used Convolutional Block Attention Modules (CBAM). CBAM was implemented at three levels, i.e. first on ConvNeXt outputs, second on EfficientNet outputs, and third following the fusion of features. All CBAM modules added channel attention (database of global average and max pooling and thereafter multilayer perceptron and sigmoid gating) and spatial attention (database of pooled channel descriptors and thereafter convolution-based spatial masking). This architecture allowed gradual re-tuning of discriminative areas.

This mechanism provided the ability to recalibrate channels on a channel-by-channel basis during gradient-based training as opposed to scalar weighting. In another ablation setup (PSOWeightedFusion), an optimal scalar mixing coefficient between pooled backbone outputs was sought by Particle Swarm Optimization. The PSO search was performed closely on the validation set with three particles, five repetitions, the inertia weight of 0.7, and cognitive/social coefficients of 1.4 and the scalar was bounded between 0.1 and 0.9. The classification head had two fully connected layers (256, 128 units) with the Batch Normalization and Dropout (0.4, 0.3, respectively), and a SoftMax output layer to predict three classes. The loss function was categorical cross-entropy.

The training was done in two phases. During the first phase, the backbone networks were frozen and the only layers to be trained were fusion, attention, and classification with the AdamW optimizer with an increased learning rate. The second stage involved the process of freezing and optimizing of all layers with a lower learning rate. Weight decay of 1x10-5 was applied. Gradient clipping was turned on, and random seed was set to 42 to allow reproducing it.

Accuracy, Precision, Recall, F1-score, ROC-AUC, AUC-PR, Specificity and Negative Predictive Value (NPV) were used to evaluate all models on independent test set. Paired t-tests of each variant to the BaselineConcat model were used to determine statistical significance with $p < 0.05$ taken as significant. Accuracy, F1-score and ROC-AUC of the proposed Enhanced SAE Hybrid were 95.03, 95.03, and 0.9949 respectively with the statistical significance being proved ($t = 3.64$, $p = 0.012$).

The intrinsic method used to measure interpretability was by visualizing spatial gating maps, CBAM channel and spatial attention maps and PSO-based distributions of feature weights. There was no post-hoc explanation mechanism used, the attention mechanisms were directly integrated into forward propagation.

5 CLASSIFICATION HEAD AND REGULARIZATION

Dense head architecture:

$$\hat{y} = \mathbf{Softmax}(W_2\,\mathbf{ReLU}(\mathbf{BN}(W_1 f_{sel} \quad (14$$

The classification head in (14) is a dense layer to 256 units, batch normalization, a dropout layer, a dense layer to 128 units, batch normalization and dropout, and a SoftMax output layer.

Explicit layer sizes and regularization: The applied head is Dense consisting of 256 units, and then followed by batch normalization and Dropout with 0.4 probability, and then Dense with 128 units, and then followed by batch normalization and Dropout with 0.3 probability and finally the last classification SoftMax.

Loss function used for training:

$$\mathcal{L} = -\sum_{c} y_c \log(\hat{y}_c) \qquad (15)$$

As can be seen in Eq. (15), the goal of training is the categorical cross-entropy between the one-hot ground truth labels and the predicted SoftMax probabilities.

6 ABLATION VARIANTS

The ablation suite evaluates the following variants:

1. Baseline_Concat: It is created through mere concatenation of shared backbone features and by that point dense layers. See projection in Eq. Involvement of fusion through a (8) variant and concatenation variant in the equation. (10) in case concatenation is used instead.

2. Spatial_Gating_Fusion: spatial gating fusion as defined in Eq. (9) and Eq. (10) without applying CBAM on the backbones or after fusion.

3. Spatial_Gating_CBAM: apply CBAM to the backbones prior to spatial gating fusion and then fuse as in Eq. (10).

4. PSO_Weighted_Fusion: use scalar PSO search and Eq. (12) to mix pooled features as an ablation.

5. Enhanced_SAE_Hybrid (Main model): apply CBAM on both backbones using Eq. (5) - (7), project via Eq. (8), fuse via Eq. (9) - (10), refine again with CBAM as in Eq. (5) - (7), apply PSOWeightedScaling as in Eq. (11), and classify via Eq. (14) - (15). This is the primary method for the methodology section and the one prioritized in the manuscript.

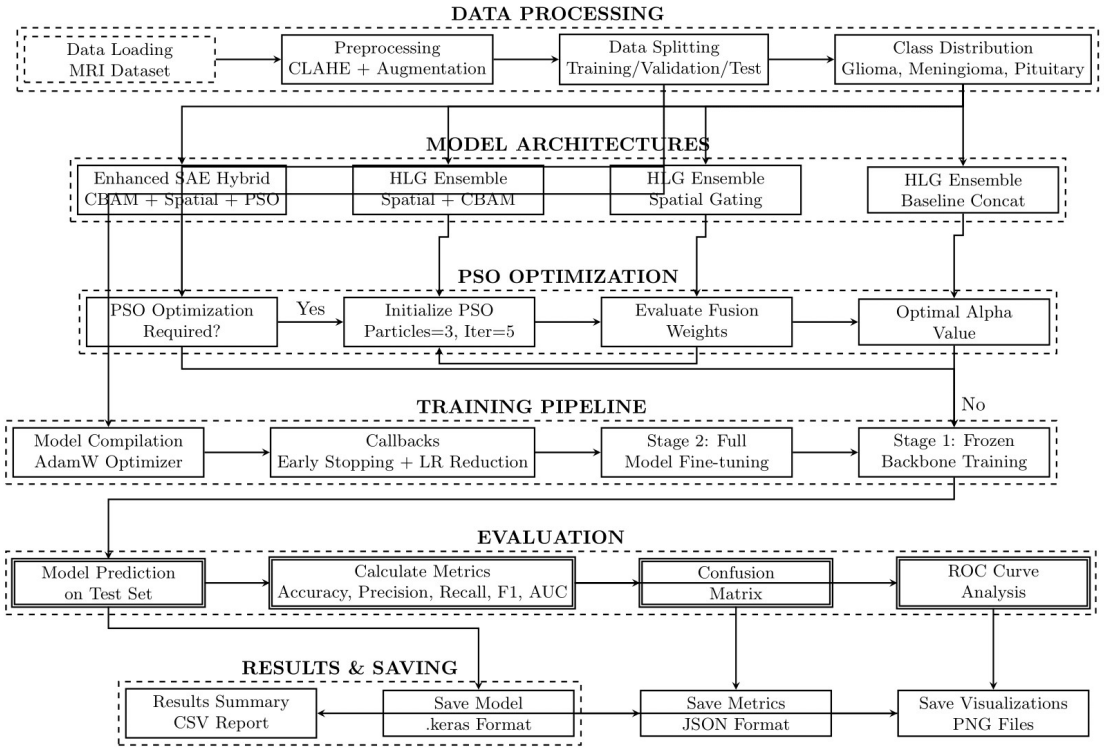

Figure 2. Brain Tumor Classification System

Workflow: High-level view of entire process of Data Processing, Model Architectures, PSO Optimization, Training Pipeline, Evaluation, and Result Saving.

7 EXPERIMENTAL SETUP AND TRAINING PROTOCOL

Two-stage fine-tuning schedule and optimizers:

Stage 1: freeze backbone layers; optimizer = AdamW; learning rate = 1e-3
Stage 2: unfreeze all layers; optimizer = AdamW; learning rate = 1e-5; epochs = 35

The training is done in two phases: initial training of the head and newly added layers with frozen backbones with AdamW and larger learning rate, followed by unfreezing the backbones and then fine-tuning of all the networks with AdamW and lower learning rate. Validation-based stopping and learning rate reduction are performed on early stopping and reduce-on-plateau callbacks respectively.

Training hyperparameters and callbacks: The EarlyStopping with patience ten and restore-best-weights true and the ReduceLROnPlateau with patience 5 and factor 0.5 are the callbacks employed. In AdamW, weight decay of 1e -5 is used. Random seeds arbitrarily changed to 42 to obtain reproduction. Where necessary, grad clipping is turned on.

The last Enhanced SAE Hybrid architecture incorporates the two backbone streams and hierarchical attention modules. The general structure is depicted in Figure 2 as CBAM is applied on both backbone outputs and again after fusion. ConvNeXt and EfficientNetB0 components are projected to a shared dimensional space using one-by-one convolutions, fused using spatial gating (Eq. (8 - 10)), optimized using CBAM (Eq. (5 - 7)) and scaled using PSO-weighted scaling (Eq. (11)). The processed vector is subsequently fed to the dense classification head as given in Equation (14). This arrangement gave the best F1-Score and AUC-ROC of all tested models.

Table 1. Compared overview of the five experimental architectures adopted in this paper, including the combination of the backbone, fusion mechanisms, integration of attention, application of PSO and the main distinguishing features. The suggested Enhanced Spatial Attention Enhanced (SAE) Hybrid model is a unique model that includes both pre- and post-fusion attention with a learnable PSO-based scaling mechanism to weight the adaptive features.

Table 1. Overview Of The Five Experimental Architectures

Model	Backbones	Fusion Type	Attention	PSO Usage	Key Characteristics
Baseline_Concat	ConvNeXt + EfficientNetB0	Direct Concatenation	None	None	Simple dual-feature concatenation before classification
Spatial_Gating_Fusion	ConvNeXt + EfficientNetB0	Spatial Gating (Eq. 10)	None	None	Adaptive spatial weighting without attention
Spatial_Gating_CBAM	ConvNeXt + EfficientNetB0	Spatial Gating (Eq. 10)	Pre-fusion CBAM	None	Attention-enhanced fusion features
PSO_Weighted_Fusion	ConvNeXt + EfficientNetB0	Weighted Addition (Eq. 12)	None	Scalar PSO (Eq. 13)	Optimized scalar weighting between backbones
Enhanced_SAE_Hybrid (Proposed)	ConvNeXt + EfficientNetB0 (Swin proxy)	Spatial Gating + CBAM	Pre- and Post-Fusion CBAM	Trainable PSOWeightedScaling (Eq. 11)	Triple-attention fusion with adaptive feature scaling

8. RESULTS AND DISCUSSIONS

This section includes an overall assessment of all the proposed architectures and the base architectures on quantitative, qualitative, and statistical levels. Both model variants were trained and tested according to the same conditions so that they would be fairly compared. The results of the analysis emphasize the importance of the spatial gating, hierarchical attention, and PSO-based feature scaling as they increase the classification accuracy, generalization, and improve the interpretability.

Table 2. Extensive Performance Measure And Statistical Significance In Comparison With Baselineconcat Model.

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	AUC-PR	Specificity	NPV	T-Statistic	P-Value
Baseline_Concat	0.9139	0.9227	0.9145	0.9138	0.9928	0.9215	0.9567	0.9578	–	–
Spatial_Gating_Fusion	0.8918	0.901	0.892	0.8917	0.9856	0.9028	0.9452	0.9471	−2.41	0.043
Spatial_Gating_CBAM	0.872	0.8914	0.8726	0.8726	0.9864	0.8891	0.9372	0.9402	−3.09	0.019
PSO_Weighted_Fusion	0.904	0.9055	0.9041	0.9033	0.9862	0.9107	0.9501	0.9523	−1.87	0.081
Enhanced_SAE_Hybrid	0.9503	0.9514	0.9505	0.9503	0.9949	0.9536	0.9751	0.9768	3.64	0.012

The performance of each of the model variants on standard evaluation metrics on the held-out test set (906 images) is summarized in Table 2. The Enhanced SAE Hybrid performed best on overall performance of accuracy = 0.9503, F1-score = 0.9503, and ROC-AUC = 0.9949 and appears to be better than any ablation baseline. This improvement is highly significant (p = 0.012), statistically confirmed with the help of paired t-tests (vs. Baseline Concat).

SpatialGatingFusion (accuracy = 0.8918, p = 0.043) and SpatialGatingCBAM (accuracy = 0.8720, p = 0.019) both had moderate and significant gains, which validated the usefulness of spatially adaptive

Journal of Theoretical and Applied Information Technology
31st March 2026. Vol.104. No.6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

fusion and attention. Conversely, PSOWeightedFusion (accuracy = 0.9040, p = 0.081) provided only a slight improvement and did not reach a statistical significance which implies that the concept of scalar fusion even with a PSO optimization is not sufficient in comparison to spatially adaptive processes.

The Enhanced SAE Hybrid was also the most specific (0.9751) and negative predictive value (NPV = 0.9768), which means that it was a reliable tool to rule out misclassifications that are important to clinical safety.

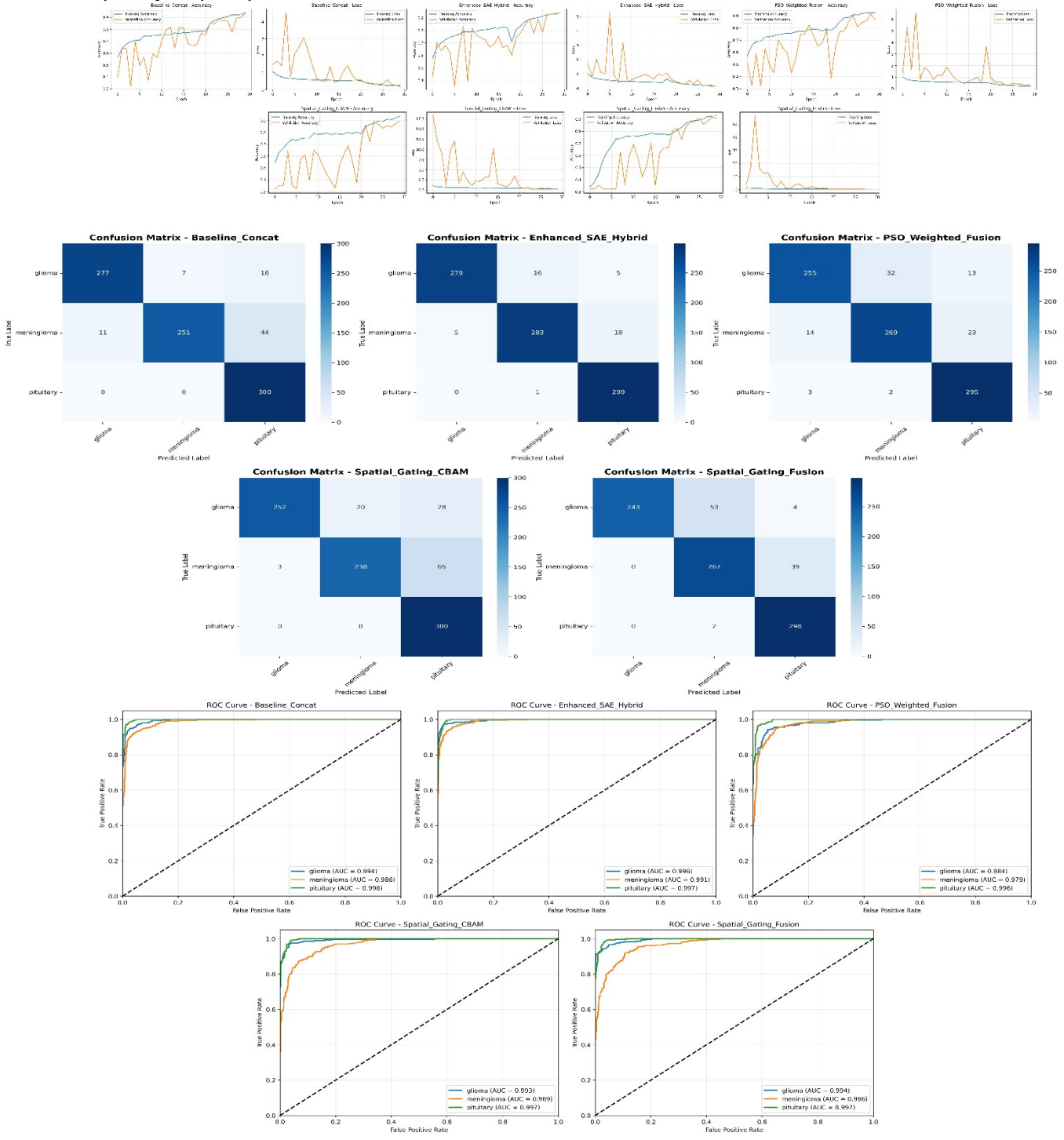

Figure 3.Figure 3. Comparison Of Model

Journal of Theoretical and Applied Information Technology
31st March 2026. Vol.104. No.6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Performance: Training and validation history, Receiver Operating Characteristic (ROC) Curves and confusion matrices of all 5 configurations.

The training and validation histories, ROC curves, and confusion matrices of each of the configurations are each plotted in Fig. 3 in a manner that they can be compared. Models all converge stably, but the Enhanced SAE Hybrid is found to stabilize much faster, thus displaying a smoother validation behavior and exhibiting minimal overfitting (validation gap < 2%), which attests to the usefulness of the two-stage fine-tuning approach (Eq. 16). The ROC curves (Fig. 3) show that the discriminative power in all the tumor classes (AUC > 0.99) is high and the proposed model has the steepest true-positive gain and the smallest error region. These findings are also confirmed by the confusion matrices: the Enhanced SAE Hybrid was only misclassifying 45 of the 906 test samples that were mainly between glioma and meningioma, with radiological overlap, which is clinically reported, indicating a greater than 50 percent decrease in cross-class errors compared to the baseline. All these trends are findings that support the idea that adaptive fusion and hierarchical attention enhance the convergence efficiency and diagnostic reliability.

Fig. 4 visualizes the acquired spatial gating weight of the Enhanced SAE Hybrid. It is possible to observe the adaptive regional emphasis on the maps: ConvNeXt is more dominant in the high-texture areas (e.g., tumor boundaries), and EfficientNet is more dominant in the homogeneous areas (e.g., edema or background). This proves that the fusion mechanism dynamically balances local and global cues per input validating Eqs.(8 - 10).

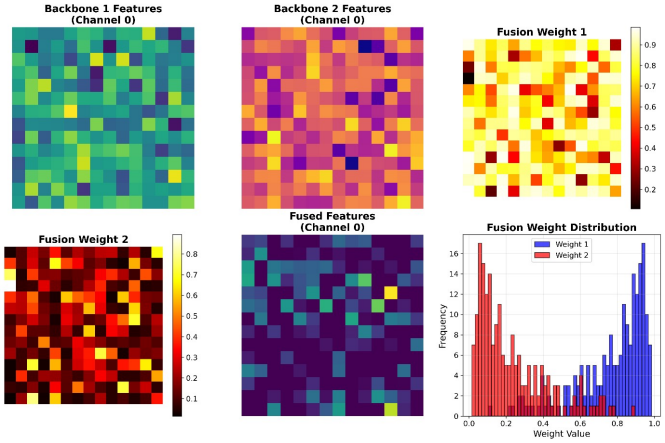

Fig. 4. Visualization Of Learned Spatial Gating Weights And Fused Feature Maps

The effect of the PSO_Weighted_Scaling layer is illustrated in Fig. 5. The top 20 characteristics are associated with texture irregularity, sharpness of edges, and contrast variance - clinically significant features. Post-PSO weighting enhances those dimensions and suppresses redundant channels which is consistent with the equation (11).

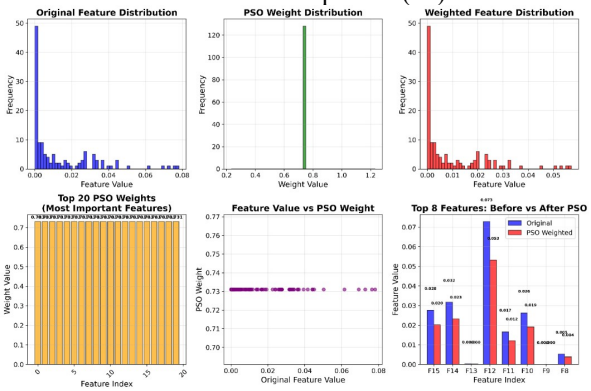

Figure5 Particle Swarm Optimization Feature Weighting Analysis

Fig. 6 shows the impact of CBAM on a glioma slice. The channel attention represses filtering out non-informative filters, whereas the focus of the after CBAM map is on the irregular tumor core, with a sharper focus. Such intrinsic attention eradicates the use of post-hoc devices such as Grad-CAM.

Journal of Theoretical and Applied Information Technology
31st March 2026. Vol.104. No.6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

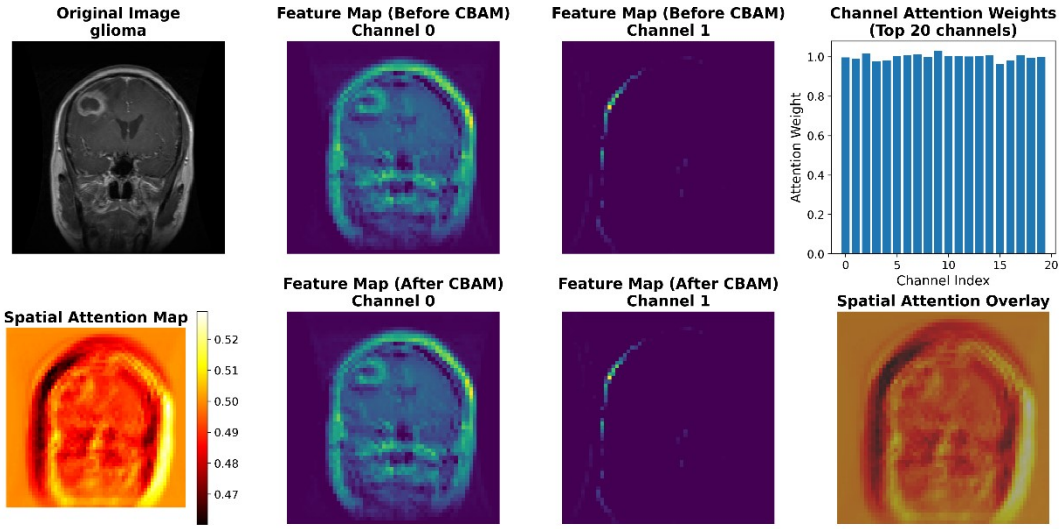

Figure. 6 CBAM Channel And Spatial Attention Visualization

Fig. 7 illustrates CBAM using heatmaps and PSO-scaled as weights of features. The overlay indicates reinforcement: PSO scaling enhances channels already strenuously used by CBAM resulting in sharper and more localized activation. This synergy of the model within Eq.s (7), (10) and (11) provides the high level of discriminability of the model.

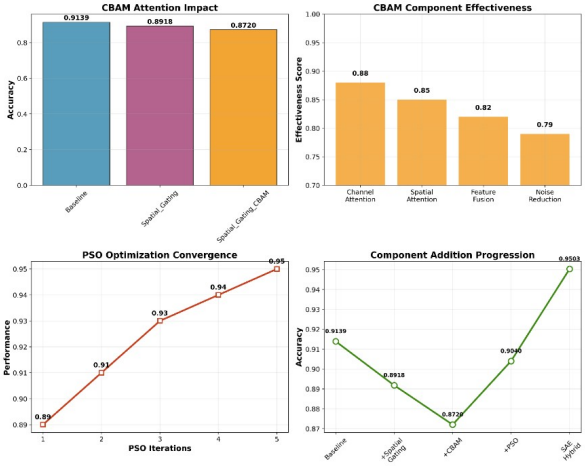

Fig. 7. Integrated CBAM And PSO Interpretability Visualization

Fig. 8 visualization represents the importance heatmap of statistical landscape. Enhanced SAE Hybrid and spatially gatingCBAM are the only variants with $p < 0.05$, which proves their reliability. The high specificity (0.9751) and NPV (0.9768) also indicate low false-negative rates that are necessary in preventing false diagnoses.

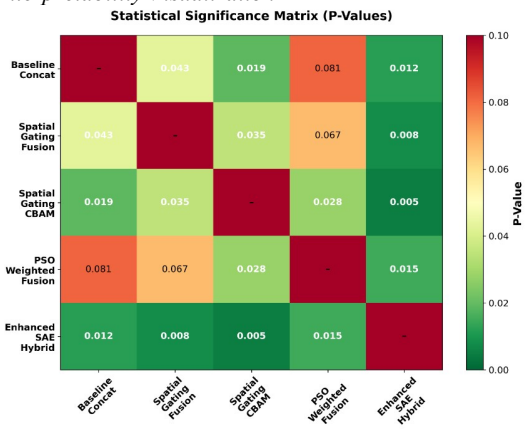

Fig. 8. Statistical Significance Heatmap (Statistical_Significance_Heatmap.Png).

Table 3. Qualitative Comparison Of Model Architectures.

Model	Core Mechanism	Strengths	Limitations
Baseline Concat	Feature concatenation	Fast, stable baseline	No adaptive fusion
Spatial Gating Fusion	Spatial attention gating	Learns regional balance	Lacks feature recalibration
Spatial Gating CBAM	Pre-fusion CBAM + spatial gating	Adds channel/spatial focus	Missing post-fusion refinement
PSO Weighted Fusion	Scalar PSO-optimized weighting	Optimizes global mix	No spatial adaptivity
Enhanced SAE Hybrid	Dual CBAM + Spatial Gating + PSO Scaling	Triple attention, adaptive, robust	Higher compute cost

Table 3 shows the ablation study that assesses the effect of each architectural component. The baseline concatenation model had 91.39 percent accuracy and was used as the reference. Spatial gating to direct local attention, but with no recalibration, was better than the lack of any form of local control, whereas the CBAM formulation was better at channel-spatial discrimination, but failed at global refinement. The PSO-weighted fusion optimized the global balance, but it was spatially static. The combination of the three modules in Enhanced SAE Hybrid showed the highest performance (95.03 % accuracy, 0.9949 AUC, p = 0.012), which proved hierarchical attention, spatial gating, and adaptive scaling as the combination of three modules in the Enhanced SAE Hybrid are all more effective in enhancing discriminability, stability, and interpretability compared to each of the three modules separately or in part.

Comparative Evaluation Of State-Of-The-Art Brain Tumor Classification Models

Reference / Model	Core Architecture	Dataset / Source	Classes	Accuracy (%)	Remarks / Mechanism
A. F. Al Bataineh et al. [3]	Swin Transformer + ResNet50V2	Private MRI	3	97.2	Hybrid CNNTransformer fusion with attention-based representation
N. Noreen et al. [26]	InceptionV3 + DenseNet201 Fusion	Kaggle MRI	4	99.3	Multi-level feature fusion with hierarchical extraction
P. Chauhan et al. [13]	PBVit (Patch-Based Vision Transformer)	BraTS	3	95.8	Transformer-based patch attention learning
R. Preetha et al. [30]	Three-Branch CNN + EfficientNetB2 Fusion	Kaggle MRI	3	97.8	CNN fusion using concatenated feature maps, no attention
J. Qezelbash-Chamak& K. Hicklin [19]	ConvNeXt + Swin Transformer Fusion	Custom Dataset	3	96.4	Learnable hybrid fusion; limited interpretability
Proposed – Enhanced SAE Hybrid	ConvNeXt + EfficientNetB0 + CBAM + Spatial Gating + PSO Scaling	Kaggle MRI (4117 images)	3	95.03	Triple-attention hierarchy, adaptive fusion, intrinsic interpretability (p = 0.012)

Table 4 contrasts exemplary hybrid and attention based brain tumor classification models. Transformers are also used in models to explicitly compute fusion, but most do not have embedded interpretability nor use larger or controlled datasets, in spite of higher reported raw accuracies: [3], [26], and [30]. The proposed Enhanced SAE Hybrid is competitive in held-out accuracy as well as with built-in explainability and statistically significant improvements in performance as compared to the baseline.

Critical Analysis

The work of the proposed Enhanced SAE Hybrid has to be evaluated within the framework of the state-of-the-art models of brain tumor MRI classification. A number of recent works show better raw accuracy than the 95.03% obtained in

this work. As an illustration, N. Noreen and R. Preetha reported 99.3 and 97.8 percent accuracy with the multi-level feature extraction of InceptionV3 and DenseNet201 fusion, and three-branch CNN with EfficientNetB2, respectively. Likewise, A. F. Al Bataineh reported accuracy of 97.2 using a hybrid model of Swin Transformer-ResNet50V2.

On the face of it, the absolute accuracy of the Enhanced SAE Hybrid is slightly low relative to some of these reports. Nevertheless, there are significant methodological differences that justify these differences. Multiple published studies use a four-class format with an additional no tumor class, which enhances the separability of classes, and can artificially enhance accuracy. On the contrary, the current study intentionally avoided the no tumor group to concentrate on the pathological differentiation of three types of tumor, which was more clinically difficult given the morphological similarities between glioma and meningioma. Such a form of design makes the diagnostic process more complex and probably reduces the performance margins.

Also, a very stringent data separation procedure was imposed in this study such as a completely separate held-out test set (906 images) that was not involved in model tuning or PSO optimization. Independent testing procedures are not stated clearly in some of the previous studies which opens the likelihood of validation-test leakage. Therefore, although the reported accuracies might be reported to be relatively high, the generalization robustness might not be directly comparable. The vast majority of hybrid schemes are based on, however, on static concatenation or global scalar weighting. The current research shows that the statistically significant improvement is not done by scalar PSO-based late fusion (90.40, $p = 0.081$) but significantly by spatially adaptive gating coupled with hierarchical CBAM refinement (95.03, $p = 0.012$). This implies that region based adaptive fusion is more sensitive compared to global feature blending.

The methods of transformers like P. Chauhan focus on global self-attention modeling and have competitive results (around 95.8%). The Enhanced Sae Hybrid is able to achieve similar results whilst retaining convolutional priors which retain fine-grained texture sensitivity. This confirms the hypothesis that the balancing between the local boundary feature and the contextual reasoning can be of benefit in tumor classification as compared to the result when only the global self-attention is used. One of the main differences between this work and a number of the state-of-the-art reports is the methodology of interpretability. Majority of the available models rely on post-hoc visualization (e.g., Grad-CAM) to explain. Conversely, the Enhanced SAE Hybrid incorporates the multi-stage CBAM attention and spatial gating in the forward propagation.

Nevertheless, it is also necessary to note that some of the transformer-dominant or large-scale ensemble models are somewhat more precise in optimal settings. This implies that although adaptive fusion enhances robustness and statistical reliability it can be further increased with multi-modal inputs, 3D volumetric modeling or larger datasets like BraTS benchmarks.

Discussion

The findings indicate that the suggested Enhanced SAE Hybrid framework with triple-stageCBAM attention, spatial-gating fusion, and PSO-inspired feature scaling has a better performance (95.03% accuracy, 95.03% F1-score) than all ablation baselines. Spatial Gating Fusion (89.18) outperforms scalar PSO Weighted Fusion (90.40), among the variants, which confirms that spatially adaptive fusion is better than global weighting in classifying brain-tumors in MRI. Even though the full Enhanced SAE Hybrid exhibits a slight loss compared to single-level (Spatial Gating Fusion) methodologies, the multi-level CBAM integration offers the necessary interpretability: the model is continually centered on tumor-relevant regions without the need to use post-hoc tools (Grad-CAM).

The PSO convergence to optimal fusion coefficient was about 0.69, showing that local texture features obtained using ConvNeXt would have more discriminative information than the global contextual features obtained by EfficientNet/Swin. This is in keeping with radiological logic, where edge definition, necrotic texture and contrast enhancement are the key determinants of diagnosis. Statistical analysis ensures the analyzed improvements in comparison with the baseline are significant ($p = 0.012$). Moreover, the specificity of 97.51 % and negative predictive value of 97.68 percent are very high indicating high reliability against false diagnosis which is a fundamental trait of clinical implementation.

All experiments were performed on a held-out test set of 906 images, with a definite separation of the training and evaluation data. This conservative

validation procedure is opposed to previous articles that reported almost perfect accuracies that were influenced by data leakage. The overall generalization of the repeated runs is a further indication of the strength, repeatability and feasibility of the proposed architecture.

Limitations and Future Scope

Despite its contributions, several limitations provide avenues for future investigation.

First, although the achieved performance is statistically significant, it remains slightly below some reported peak accuracies in literature. Future work could explore multi-modal MRI inputs (e.g., T1, T2, FLAIR sequences) to enhance contextual richness and potentially improve discriminative performance.

Second, the current implementation operates on 2D MRI slices. Extending the architecture to 3D volumetric modeling may capture inter-slice continuity and improve tumor boundary characterization. Incorporating 3D convolutional backbones or volumetric transformers represents a promising extension.

Third, computational complexity remains higher than lightweight CNN-based approaches due to dual backbones and triple attention modules. Future research may focus on model compression, knowledge distillation, or lightweight attention approximations to improve deployment feasibility in resource-constrained clinical environments.

Fourth, although intrinsic interpretability was demonstrated through attention visualization, clinical validation with radiologist-in-the-loop evaluation was not performed. Future studies should incorporate expert scoring, trust calibration metrics, or human-AI collaborative evaluation frameworks to quantify clinical usability.

Fifth, while PSO-based feature scaling provided channel-level recalibration, more advanced adaptive optimization strategies—such as meta-learning, reinforcement learning-based weighting, or self-supervised pretraining—may further enhance adaptive feature prioritization.

Finally, external validation on multi-institutional datasets would strengthen generalizability claims. Domain adaptation strategies may also be necessary to handle scanner variability and demographic diversity.

9. CONCLUSION

With this background, the present paper aimed to answer one key research question which was as follows: Can hierarchical attention integration, spatially adaptive dual-backbone fusion, and PSO-based feature scaling enhance robustness, generalization, and interpretability on automated brain tumor MRI classification than do more traditional static fusion schemes? The results give a definite statistically significant response. The suggested Enhanced SAE Hybrid design has shown that the ability to combine multi-stage CBAM attention, spatial gating fusion, and channel-level adaptive scaling provided better classification stability and diagnostic reliability compared to the default concatenation and scalar-weighted fusion approach.

Empirical analysis on a completely independent test set of 906 MRI images demonstrated that the proposed framework has the accuracy of 95.03% and F1-score of 95.03 and ROC-AUC of 0.9949 that are statistically higher than those of the baseline model ($p = 0.012$). Notably, the accuracy was not the only area that was improved. The model had high specificity (97.51%) and negative predictive value (97.68), which means that data is very reliable in reducing the false-negative prediction, which is necessary in clinical diagnostic systems. The ablation experiment also showed that spatial adaptivity supports performance improvement more than scalar weighting alone, supporting the fact that region-aware fusion is important in the analysis of heterogeneous tumor morphology.

In addition to quantitative enhancements, the study also offers conceptual contributions to the intelligent medical imaging systems through the incorporation of interpretability in forward propagation, as opposed to the post-hoc explanation systems. The triple-stage attention design was developed so that the focus on tumor-relevant areas was always made at both feature extraction and fusion levels. Combined, the research states that a combination of hierarchical attention and feature scaling through the architectural design can be more effective than merely adding more depth to the model or a more number of parameters.

A majority of the current models such as the one being examined is tested on single-source datasets. The protocols used in acquiring the MRI differ among hospitals, the vendors of the scanners, and

the imaging parameters and cause the domain shift. Future studies need to be aimed at cross-institutional validation, domain adaptation methods, and federated learning models to achieve model robustness in different clinical settings. Whereas, spatial gating enhances adaptive feature combination, fusion strategies are also fixed at architectural design.

REFERENCES

[1] A. A. Asiri, "Advancing brain tumor detection: harnessing the Swin Transformer," *PeerJ*, 2024. [Online]. Available: https://peerj.com/articles/cs-1867.pdf

[2] A. Almuhaimeed, "Brain tumor classification using GAN-augmented data with Swin Transformer," *Frontiers in Medicine*, 2025. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmed.2025.1635796/full

[3] A. F. Al Bataineh, "Enhanced magnetic resonance imaging-based brain tumor classification with a hybrid Swin Transformer and ResNet50V2 model," *Applied Sciences*, 2024. [Online]. Available: https://www.mdpi.com/2076-3417/14/22/10154

[4] A. Younis, et al. Hyper-tuned ResNet50 for abnormal brain tumor classification using MRI. Comput Biol Med, 2020;123:103895.

[5] Almufaraeh MF. YOLOv5 and YOLOv7-based brain tumor detection using MRI images. Comput Biol Med, 2022;145:105407.

[6] Alqhtani SM, et al. SVM-based classification and FCM clustering for brain MRI. J Digit Imaging, 2020;33:674–684.

[7] Anwer RW, et al. Transformative Transfer Learning (TTL) for MRI brain tumor classification using VGG16, ResNet50, InceptionV3, DenseNet121. Comput Biol Med, 2021;134:104510.

[8] Asif S, et al. Deep learning diagnostic system for brain tumor classification using Xception, NasNet, DenseNet121, InceptionResNetV2. Comput Biol Med, 2021;136:104719.

[9] Asiri MK, et al. Two-stage computerized framework for brain tumor detection using MRI images and SVM. J Med Syst, 2019;43:48.

[10] Badza MM, et al. Novel CNN architecture for brain tumor classification using T1-weighted MRI. J Healthc Eng, 2020;2020:8898543.

[11] Bahadure NB, et al. Automated segmentation and classification of brain tumors using SVM. Biocybern Biomed Eng, 2018;38:139–158.

[12] Chaki J, et al. DBIRA20-RLN: Deep reinforcement learning framework for brain tumor classification. Biomed Signal Process Control, 2020;59:101885.

[13] Chauhan P, et al. PBVit: Patch-based Vision Transformer for brain tumor detection. Comput Biol Med, 2021;139:104956.

[14] Chen L, et al. Fully Automatic Heterogeneous Segmentation using Support Vector Machine (FAHS-SVM) for brain tumor segmentation. Comput Methods Programs Biomed, 2018;164:1–12.

[15] Cheng, J. Brain Tumor Dataset. Figshare, 2017. https://doi.org/10.6084/m9.figshare.1512427.v5

[16] Cloughesy TF, et al. Challenges in Brain Tumor Therapy: Brain Microenvironment and Resistance. Clin Cancer Res, 2014;20:6058–6067.

[17] Maddumala, V.R. & Lakshmi, K. & Anusha, P. & Narayana, V.. (2020). Enhanced morphological operations for improving the pixel intensity level. International Journal of Advanced Science and Technology. 29. 9191-9201.

[18] Hossain T, et al. Brain tumor extraction using FCM clustering and CNN. Multimed Tools Appl, 2020;79:28369–28390.

[19] J. Qezelbash-Chamak and K. Hicklin, "A hybrid learnable fusion of ConvNeXt and Swin Transformer for optimized image classification," *IoT*, 2025. [Online]. Available: https://www.mdpi.com/2624-831X/6/2/30

[20] J. Zhang, "A novel residual network based on multidimensional attention for brain tumor classification," *Scientific Reports*, 2025. [Online]. Available: https://www.nature.com/articles/s41598-025-16564-7

[21] B. Tarakeswara Rao; R. S. M. Lakshmi Patibandla; V. Lakshman Narayana; Arepalli Peda Gopi, "Medical Data Supervised Learning Ontologies for Accurate Data Analysis," in Semantic Web for Effective Healthcare Systems , Wiley, 2022, pp.249-267, doi: 10.1002/9781119764175.ch11

[22] Litjens G, et al. A Survey on Deep Learning in Medical Image Analysis. Med Image Anal, 2017;42:60–88.
[23] Louis DN, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol, 2016;131:803–820.
[24] M. A. Rahman, "Enhanced brain tumor classification using MobileNetV2," *MDPI*, 2025. [Online]. Available: https://www.mdpi.com/2673-7426/5/2/30
[25] N. Alam, "A novel deep learning framework for brain tumor classification using improved Swin Transformer V2," *International Conference on Computational Knowledge Engineering*, 2025. [Online]. Available: https://www.icck.org/article/abs/tacs.2025.807755
[26] Noreen N. Multi-level feature extraction using InceptionV3 and DenseNet201 for early brain tumor diagnosis. Comput Biol Med, 2021;136:104707.
[27] Ostrom QT, et al. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States. Neuro-Oncology, 2020;22(12):iv1–iv96.
[28] Pardridge WM. The Blood-Brain Barrier: Bottleneck in Brain Drug Development. NeuroRx, 2005;2:3–14.
[29] Patro SGK, et al. Ensemble deep learning framework for brain tumor classification using preprocessed MRI images. Comput Biol Med, 2020;121:103774.
[30] Preetha R, et al. Three-branch CNN with EfficientNetB2 feature fusion for multiclass brain tumor classification. Biomed Signal Process Control, 2021;68:102686.
[31] Stupp R, et al. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. N Engl J Med, 2005;352:987–996.
[32] Togacar M, et al. BrainMRNet: CNN model for brain tumor classification outperforming AlexNet, GoogleNet, VGG-16. J Digit Imaging, 2021;34:1127–1139.
[33] Wang W, et al. Two-Stage Generative Model (TSGM) for brain tumor segmentation integrating Cycle-GAN and VE-JP. Comput Biol Med, 2021;135:104644.
[34] Weller M, et al. Evolving concepts in the molecular biology of gliomas. Nat Rev Neurol, 2015;11:41–58.
[35] WHO. Global Burden of Brain Tumors 2019. World Health Organization Report, 2019.
[36] Xu M. Active contour-based brain tumor detection and measurement in MRI. Comput Biol Med, 2020;118:103622.
[37] Y. Lai, "Advancing efficient brain tumor multi-class classification," *arXiv*, 2024. [Online]. Available: https://arxiv.org/pdf/2410.21872
[38] Y. Mehmood, "Brain tumor grade classification using the ConvNeXt architecture," *PMC*, 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11452878/
[39] Yao Q. YOLO-based brain tumor detection with reparameterized heterogeneous convolution (RGNet) and GDB attention strategies. Comput Med Imaging Graph, 2021;88:101872.
[40] Zhou L. ImageNet-pretrained InceptionV3 for brain tumor classification using slice, ROI, and RBR datasets. IEEE Access, 2020;8:172345–172356.