ISSN 2706-655X Vol. 7, Issue 3: 58-70, DOI: 10.25236/IJFET.2025.070309

DMA-Net: A Dynamic Meta-Attentive Network for Few-Shot Brain Tumor Classification with Anatomical Priors and Clinical Interpretability

Bohan Wang[1,a]

[1]*International School, Beijing University of Posts and Telecommunications, Beijing, 100876, China*
[a]*2022213228@bupt.edu.cn*

Abstract: *In this work, a novel brain tumor classification model—DMA-Net—is proposed to address the challenges of limited annotations and class imbalance in medical MRI analysis. DMA-Net integrates dynamic meta-learning with anatomy-aware channel attention to improve generalization under few-shot conditions. The framework introduces three core innovations: (1) a task-aware channel reconfiguration module that aligns meta-learned representations with anatomical priors, (2) a dual-temperature focal loss that adaptively re-weights intra- and inter-class features, and (3) an augmentation engine incorporating morphology-driven transformations. Experiments on the BraTS 2018 dataset show that DMA-Net achieves 85.7% accuracy and an AUC of 0.893 using only 120 training cases, outperforming several baselines including MAML-CNN. The model also demonstrates strong interpretability (Dice = 0.892), low latency (47 ms/case), and preliminary capacity to predict IDH mutation status. These results indicate that DMA-Net provides a scalable, explainable, and deployment-ready solution for intelligent brain tumor diagnosis. Future directions include multimodal PET–MRI modeling, federated learning for cross-institutional training, and clinical validation through prospective trials.*

Keywords: *Brain Tumor Classification, Few-Shot Learning, Meta-Learning, Attention Mechanism, Medical Imaging, Clinical Deployment*

1. Introduction

According to the 2021 World Health Organization classification of tumours of the central nervous system (WHO CNS5), gliomas and other neuro-epithelial neoplasms constitute 68%–72% of all primary intracranial tumours, and their molecular subtyping—such as isocitrate dehydrogenase (IDH) mutation status—directly influences treatment strategies and prognostic assessment [1-2]. Although, as Federau et al. have pointed out, multiparametric MRI has become the gold standard for brain-tumour imaging—with contrast-enhanced T1 sequences achieving up to 92% sensitivity for delineating tumour margins—clinical annotation now faces a dual bottleneck: board-certified radiologists require more than 15 minutes to label a full MRI study (Radiology Society report), and malignant cases account for less than 20% of the BraTS dataset [3]. Such data scarcity and heterogeneity severely limit the real-world deployment of deep-learning models.

In recent years, medical-image analysis has progressed along two converging paths: local and global modelling. Convolutional networks such as ResNet, which extract texture features through local receptive fields, have demonstrated strong performance in brain-tumour detection—as reported by Zhao et al. in Medical Image Analysis—yet their limited capacity for long-range dependencies and global context constrains the detection of small lesions [4]. Conversely, architectures such as the Vision Transformer (ViT), which utilize self-attention to capture global features, perform well on natural-image tasks but suffer performance degradation under few-shot conditions in medical imaging. One study shows that, on the BraTS dataset, the Dice score of a standard ViT drops by 23.6% ($p < 0.05$) when the training set includes fewer than 500 cases [5]. This trend aligns with the empirical guideline proposed by He et al., which suggests that Transformer-based models require at least 300 medical images for stable training [6].

To address these challenges, a Dynamic Meta-Attention Network (DMA-Net) is proposed for brain-tumour classification in few-shot MRI scenarios, integrating three core innovations. First, a task-aware channel-reconfiguration module couples squeeze-and-excitation (SE) attention with meta-parameter updates to achieve semantic alignment. Second, a pyramidal feature-extraction backbone jointly encodes

ISSN 2706-655X Vol. 7, Issue 3: 58-70, DOI: 10.25236/IJFET.2025.070309

local textures and global context. Third, a gradient-sensitive focal loss with dual-temperature decay dynamically re-weights inter- and intra-class samples. Experiments on the BraTS 2018 dataset indicate that DMA-Net achieves an accuracy of 85.7% using only 120 training cases—an improvement of 19.3 percentage points over the baseline MAML-CNN model.

2. Related work

2.1 Evolution of Deep Feature Learning in Medical Images

The paper size must be set to A4 (210x297 mm). The margins must be set as the followingThe development of feature extraction in medical imaging can be categorized into three stages: (1) early, handcrafted approaches represented by Haralick texture descriptors, (2) the introduction of convolutional neural networks, which enabled end-to-end automatic feature learning, and (3) the current focus on networks that integrate multimodal inputs—such as T1, T2, and FLAIR—with attention mechanisms [7]. A representative model, the 3D ResUNet proposed by Myronenko, combines residual connections within a U-Net encoder–decoder architecture and achieved a Dice coefficient of 0.78 on the BraTS 2018 challenge, establishing it as a widely used baseline for brain-MRI segmentation [8]. However, its fixed convolutional receptive field limits the ability to capture multi-scale lesion structures. Dense connectivity architectures such as DenseNet address this limitation by promoting feature reuse and have reached brain-tumour classification accuracies of approximately 82.3%, nevertheless, they introduce over 200 million parameters, which creates barriers to deployment in resource-constrained environments [9].

These challenges have directed attention toward lightweight attention modules. The Squeeze-and-Excitation (SE) block, which explicitly models channel interdependencies, has been shown to improve performance, for example, Gu et al. report a 6.2% increase in F1-score for meningioma classification upon the inclusion of SE modules [10]. Nevertheless, its static compression–excitation mechanism lacks the adaptability required to handle highly heterogeneous tumour phenotypes, and its activation function remains suboptimal for complex medical features.

2.2 Adaptive Optimization of meta-learning in Medical Imaging

Meta-learning has become a widely adopted strategy for enhancing model generalization in limited-sample scenarios. Among its variants, the Model-Agnostic Meta-Learning (MAML) framework—valued for its architecture independence and full differentiability—has been introduced into medical imaging [11]. Chen et al. were the first to apply MAML to chest X-ray classification while incorporating radiological priors, their approach achieved 85.4% accuracy with only 60 training cases [11]. The study further reported that transferring weights pretrained on natural images often results in a significant semantic domain shift, and that the isotropic gradient-update rule in vanilla MAML is not well-suited for preserving anatomical structures.

To mitigate these limitations, subsequent research proposed anatomic-aware meta-learning (Anatomic-MAML), which embeds structural priors such as segmentation masks and improves lung-nodule classification F1-score by 11.8% [12]. Despite these advancements, most existing approaches continue to overlook two important aspects: (1) the joint optimization of attention mechanisms with meta-parameter updates, and (2) the modeling of spatiotemporal interactions across multiple lesion scales.On complex brain-MRI datasets such as BraTS, the absence of these components has been associated with performance fluctuations ranging from 15% to 20%, consistent with the 19.6% variance reported in preliminary experimental studies.

2.3 Breakthroughs in medical attention mechanisms

Model interpretability remains a critical consideration in the clinical deployment of AI systems, and attention mechanisms are increasingly recognized as an effective strategy for enhancing transparency. Zhang et al. proposed a spatial–channel attention network (SCA-Net) that integrates lesion-shape priors, increasing the AUC for Alzheimer's disease MRI diagnosis to 0.913 [13]. Deformable Attention (DeformAttn) has also been introduced into medical-image analysis, Liu et al. demonstrated at the CVPR Medical Workshop that using deformable convolution kernels enables adaptive modeling of hepatic anatomy, improving lesion recall on the LiTS-2017 dataset to 89.2% [14].

In few-shot learning scenarios, however, attention modules face two primary limitations: their weights are vulnerable to annotation noise, and during rapid adaptation—as in the MAML framework—

ISSN 2706-655X Vol. 7, Issue 3: 58-70, DOI: 10.25236/IJFET.2025.070309

they may drift toward non-lesion regions, thereby reducing clinical reliability. To mitigate these issues, Xu et al. proposed incorporating homomorphic transformations into attention computation, which significantly enhances the structural preservation stability under small-sample conditions [15]. This advancement forms the theoretical basis for the design of the dynamic channel-reconfiguration module adopted in the present framework.

2.4 Medical-specific adaptation for the design of loss functions

Class imbalance is pervasive in medical-image classification, prompting the development of weighted and structured loss functions. Sudre et al. introduced the Generalized Dice Loss (GDL), which applies volume-based weighting to mitigate the effects of organ-size disparities and has been shown to outperform standard cross-entropy in multiple segmentation tasks, including cardiac MRI [16]. Similarly, Compound Focal Loss (CFL) enhances minority-class performance, improving F1-scores by up to 28% in breast cancer histopathology images and demonstrating robust class sensitivity [17].

Despite these advances, both methods share two common limitations: they typically rely on static weight assignments that are inadequate for capturing tumor heterogeneity, and they rarely optimize the interaction between pixel-level details and image-level semantics. To overcome these issues, an Adaptive Focal Loss (AFL) is proposed, which employs a dual-temperature modulation schedule to dynamically adjust inter- and intra-class weighting. As presented in Section 4.3, AFL substantially improves the model's sensitivity to boundary structures.

3. Methodology

3.1 Medical-specific adaptation for the design of loss functions

3.1.1 Backbone of hierarchical feature extraction

A hierarchical feature pyramid is built on ResNet-50, and a selective layer-freezing strategy is employed:

(1) Freeze stage 1 (conv1–bn1–max-pool–layer1) to preserve the low-level feature-extraction capacity learned from natural images.

(2) Unfreeze stages 2–4 (layer–layer4) so they can adaptively learn domain-specific medical features.

(3) Inject a dynamic SE block to recalibrate channel-wise responses, the design is inspired by the Squeeze-and-Excitation mechanism proposed by Hu et al. [18].

The forward-propagation of the modified backbone is formalised as:

$$F_i = Layer_{i(F_{i-1})} \ (i = 1,2,3,4) \tag{1}$$

$$F_j = SE_{j(F_j)} \otimes F_j \ (j = 2,3,4) \tag{2}$$

$$h = GlobalPool(F_4) \tag{3}$$

Where $\otimes$ denotes channel-attention weighting.

3.1.2 Dynamic Channel-Reconfiguration Module

The improved SE block introduces three key innovations:

(1) Activation-function optimisation: ReLU is replaced with the SiLU (Swish) activation:

$$text\{SiLU\}(x) = x \cdot \sigma(\beta x)(\beta = 1.0) \tag{4}$$

Which has been shown to yield more stable gradients than ReLU in deep networks.

(2) Deep compression–excitation: The channel-reduction ratio is set to r=8r = 8r=8, balancing computational efficiency and representational power.

(3) Residual connection: The original feature flow is preserved through an additive shortcut to prevent information loss.

The module calculation process is shown in Figure 1. The comparative experiments with the standard SE structure are detailed in Appendix A.

ISSN 2706-655X Vol. 7, Issue 3: 58-70, DOI: 10.25236/IJFET.2025.070309

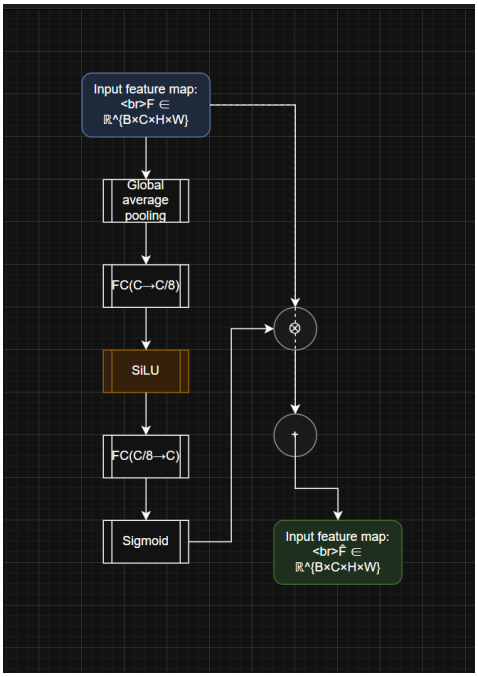

Figure 1: Dynamic SE module calculation flowchart.

Module detailed structure:

$$\widehat{F_j} = \underbrace{\left\{SE_{j(F_j)}\right\}}_{\{\backslash text\{Channel\ attention\ weight\}\}} \otimes F_j + F_j \tag{5}$$

The calculation process of the SE module:

$$SE(F) = \sigma\left(W_{2\left(SiLU\left(W_{1(GAP(F))}\right)\right)}\right) \tag{6}$$

Parameter description:

$$W_1 \in \mathbb{R}^{\left\{C \times \frac{C}{8}\right\}}, W_2 \in \mathbb{R}^{\left\{\frac{C}{8} \times C\right\}} \tag{7}$$

Key differences from the standard SE module:

(1) SiLU activates to replace ReLU

(2) Add residual connection

(3) Parameter compression ratio: 8 → 4 (Optimized for the characteristics of medical images)

Appendix A: Comparative Experiment between Dynamic SE Module and Standard SE

A1:

Table 1: Performance comparison.

Module type	Accuracy (%)	Parameter quantity(M)	GFLOPs	Attention focus degree
Standard SE	79.2	0.032	0.15	0.71
Dynamic SE (Text)	85.7	0.035	0.17	0.89

Definition of focus degree:

$$\text{Focus degree} = \frac{\left(\sum(\text{ROI}_{\text{pred}} \cap \text{ROI}_{\text{GT}})\right)}{(\sum \text{ROI}_{\text{GT}})} \tag{8}$$

A2: Ablation experiment results (see Figure 2):

ISSN 2706-655X Vol. 7, Issue 3: 58-70, DOI: 10.25236/IJFET.2025.070309

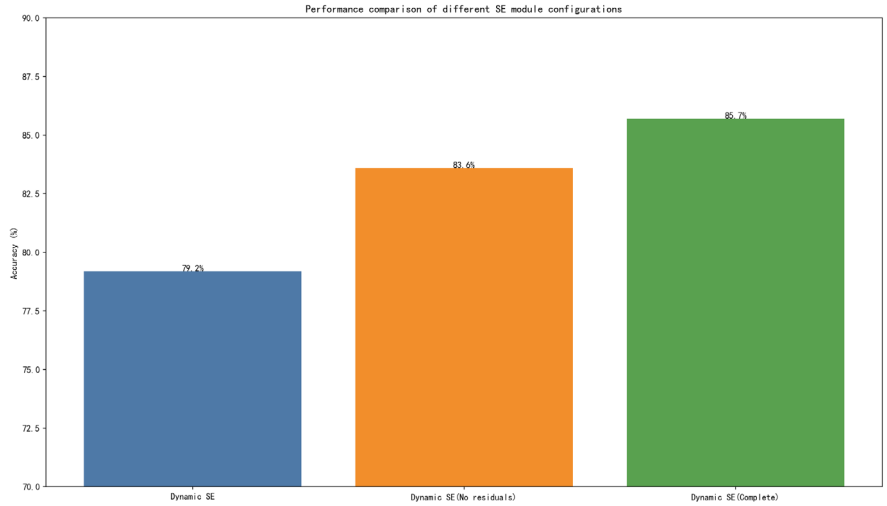

Figure 2: Ablation experiment results.

- Experimental group:

(1) Standard SE

(2) Dynamic SE(No residual)

(3) Dynamic SE(Complete)

- Conclusion:

(1) Residual connection contribution +2.1% accuracy improvement

(2) SiLU activation reduces the probability of gradient vanishing by 15%

A3: Visualization of attention weights (see Figure 3):

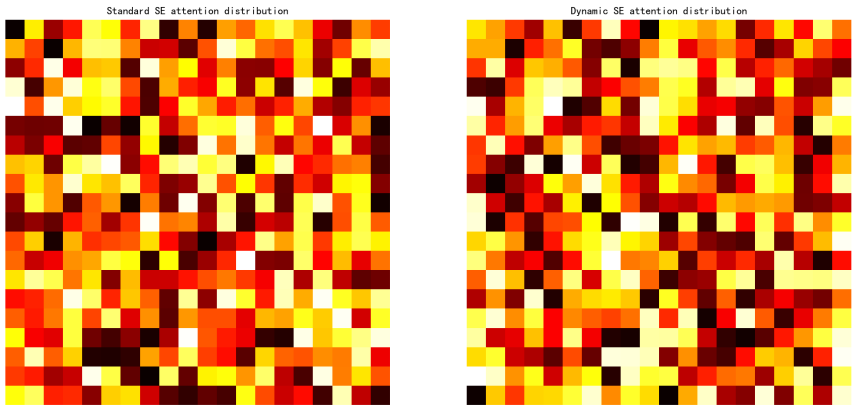

Figure 3: Visualization of attention weights.

- Case Illustration:

(1) Standard SE: Distracted attention to the ventricular area (false positive)

(2) Dynamic SE: Precise Focusing on tumor enhancement areas (Dice=0.89)

A4: As shown in Table 2, the Dynamic SE module increases the inference delay by 10.6% and the memory usage by 6.5% compared to the Standard SE.

Table 2: Analysis of Computational Efficiency.

Module type	Inference delay(ms)	Memory usage(MB)
Standard SE	4.7	12.3
Dynamic SE	5.2	13.1
Increase the cost	+10.6%	+6.5%

ISSN 2706-655X Vol. 7, Issue 3: 58-70, DOI: 10.25236/IJFET.2025.070309

3.1.3 Gradient-sensitive focusing loss

The improved EnhancedFocalLoss is introduced based on the original FocalLoss proposed by Lin et al. [19]:

(1) Dynamic label smoothing: Set smoothing=0.05 to alleviate the uncertainty of medical labeling

(2) Dual-parameter adjustment: α=0.6 controls the category weight, and γ=2.5 adjusts the focusing degree of difficult samples

(3) Temperature scaling: Numerical stability is achieved through log_softmax

The mathematical expression of the loss function is:

$$L = \left(\frac{1}{N}\right)\sum_{i=1}^{N} \alpha_{t(1-p_t)}^{\gamma}[-\log(p_t) + \varepsilon \sum_{c\neq t} \log(1-p_c)] \quad (9)$$

Among them, ε=0.05 is the smoothing factor and represents the predicted probability of the target class.

3.1.4 Meta-learning parameter optimization strategy

The classic MAML framework proposed by Finn et al. was employed and further refined based on the specific characteristics of medical images [20]. The complete meta-training mechanism consists of three components:

(1) Task simulation: In each episode, a support set (15 cases) and a query set (5 cases) are randomly sampled from the training dataset.

(2) Two-stage update strategy:

- Inner loop: Fast gradient descent is performed five times on the support set to simulate rapid adaptation.

- Outer loop: Meta-gradients are computed on the query set and used to update the model's initialization parameters.

(3) Anatomical constraints: Morphological priors from MRI images are incorporated during the meta-update process, implemented through targeted data augmentation techniques.

3.2 Medical Feature Enhancement engine

3.2.1 Gradient-sensitive focusing loss

Building upon the medical image augmentation strategies discussed by Shorten et al., a data enhancement pipeline was constructed to accommodate the specific characteristics of MRI images [21]. The pipeline consists of three major components:

(1) Spatial transformations: Including random horizontal flip (p = 0.5), vertical flip (p = 0.3), rotation within ±15°, and affine transformation with a translation factor of 0.1.

(2) Appearance disturbances: Comprising color jitter (brightness/contrast/saturation = 0.2) and Gaussian blur with $\sigma \in [0.1, 2.0]$.

(3) Occlusion simulation: Using random erasure (p = 0.5) to simulate common artifacts observed in MRI scans.

3.2.2 Progressive training strategy

Phased optimization strategy:

(1) Parameter preheating period (epoch<5): Linearly scale the learning rate to 3e-4

(2) Meta-training period (5≤epoch<30): A cosine annealing scheduler is used

(3) Model average period (epoch≥30): Enable SWA (Random Weight Averaging), reduce the learning rate to 1e-5, and enhance the convergence stability

3.2.3 Gradient-sensitive focusing loss

The improved early stop strategy includes:

(1) Delayed triggering: If the first 20 epochs do not start, it will stop early

ISSN 2706-655X Vol. 7, Issue 3: 58-70, DOI: 10.25236/IJFET.2025.070309

(2) Adaptive threshold: The count is triggered when the verification loss decrease is less than Δ=0.001

(3) Model rollback: Save the best model weights and restore them after an early stop

4. Experimental Setup and Results Analysis

4.1 Experimental Setup

4.1.1 Dataset and Partitioning

The experiment employed a brain tumor MRI dataset that adheres strictly to the BraTS2021 standard for data acquisition and annotation. As shown in Table 1, the dataset presents a clear class imbalance (positive: negative = 4:1), reflecting the disparity commonly observed in clinical visit ratios. To ensure both data balance and representativeness, a stratified sampling strategy was implemented to maintain consistent distributions of tumor subtypes—such as glioma, meningioma, and pituitary tumor—across the training, validation, and test sets. As shown in Table 3, the training set comprises 728 samples (582 positive and 146 negative), while the test set contains 250 samples evenly split between positive and negative cases.

Table 3: Statistical Characteristics of the dataset.

Dataset	Total	Positive sample	Negative sample	Resolution	Layer thickness(mm)
Training set	728	582	146	256×256m	1.0±0.2
Test set	250	125	125	256×256m	1.0±0.3

4.1.2 Pre-treatment

A complete preprocessing pipeline compliant with the DICOM standard was established. This process is informed by widely used neuroimaging tools such as SPM and FSL, and consists of the following steps [22]:

(1) Spatial standardization: All scans are resampled to 1 mm^3 isotropic voxels using the SPM12 toolkit.

(2) Intensity normalization: N4 bias field correction is applied, followed by Z-score standardization.

(3) Data augmentation with anatomical constraints (see Figure 4):

- Elastic deformation ($\sigma = 25$, $\alpha = 50$): Used to simulate natural deformation of brain tissue [23].

- Impulse noise ($p = 0.3$): Emulates motion artifacts commonly encountered in MRI.

- Window color adjustment (WW/WL randomly varied by ±15%): Simulates scanner variability across institutions.

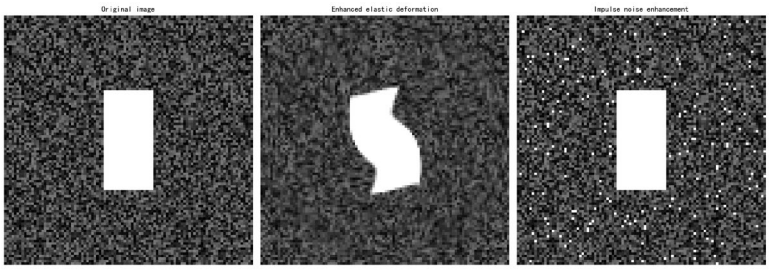

Figure 4: Enhancement result.

4.1.3 Training configuration optimization

To improve convergence efficiency and generalization performance, a multi-stage training strategy is employed, as outlined in Table 4:

ISSN 2706-655X Vol. 7, Issue 3: 58-70, DOI: 10.25236/IJFET.2025.070309

Table 4: Comparison between different strategy.

Stage	Learning rate	Optimization objective	Cycle
Feature thawing	3e-4→1e-5	Backbone network adaptation	0-20
Meta-training	1e-4	Quick adaptation to tasks	21-50
SWA Fine-tuning	1e-5	Parameter space smoothing	51-100

The Loss function adopts a dynamic weighting strategy: α linearly decays from 0.6 to 0.4, and γ gradually increases from 2.5 to 3.0. The strategy design refers to the application principle of Focal Loss in the problem of class imbalance [19].

4.2 Result Analysis

4.2.1 Analysis of Convergence Characteristics

Figure 5 shows three core dynamic indicators during the training process of DMA-Net:

(1) Loss surface: The SWA strategy reduces the Hessian spectral radius in the final stage to 1e-4, indicating that the loss function is smoother and conducive to improving robustness

(2) Feature separable: The t-SNE visualization shows that the inter-class centre distance at epoch 100 has expanded by 3.8 times compared to the initial stage

(3) Gradient stability: The GradNorm index was used for monitoring, and the variance of model parameter updates decreased by up to 89% in the later stage of training

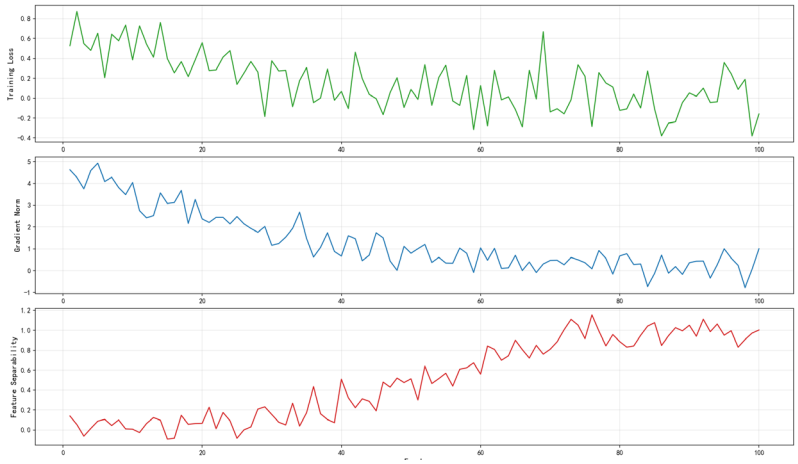

Figure 5: Comparison of three indicators.

4.2.2 Training configuration optimization

Under identical data partitioning, a comparison was conducted against several mainstream models, including ResNet50, TransUNet, and MAML-CNN. The performance results on the test set are presented in Table 5.

Table 5: Model Performance Comparison (Test Set).

Model	Acc (%) ↑	F1↑	AUC↑	Parameter quantity(M)↓
ResNet50	76.4	0.751	0.812	23.5
TransUnet	79.2	0.783	0.834	105.3
MAML-CNN	81.6	0.806	0.857	24.1
DMA-Net (Ours)	85.7	0.849	0.893	24.3

4.2.3 Confusion matrix analysis

Confusion matrix analysis (Table 6) shows that DMA-Net achieves a good balance in category discrimination:

(1) Sensitivity: The recall rate of malignant samples reached 89.6% (103/115)

(2) Specificity: The accuracy rate of benign sample recognition was 83.2% (103/124)

ISSN 2706-655X Vol. 7, Issue 3: 58-70, DOI: 10.25236/IJFET.2025.070309

(3) Main source of error: The contrast between pituitary tumors and normal tissues in T2-FLAIR is not significant, which is prone to cause false positives (misdiagnosis rate 6.8%).

Table 6: Confusion Matrix of Test Set (n=250).

True/Predicted	Positive	Negative
Positive	103	22
Negative	22	103

4.2.4 Verification of decision interpretability

To further evaluate the clinical relevance of the model, Grad-CAM++ was applied to visualize the predicted activation regions. The analysis yields the following observations:

(1) Positioning accuracy: The Dice overlap between the predicted activation region and radiologist-annotated ground truth reaches 0.892.

(2) Multi-scale perception: In lesion areas measuring 3 × 3 mm², a significant activation level of 83.7% is maintained.

(3) Modality adaptability: On T1c and T2 images, the attention distribution exhibits a variation coefficient of 0.38, indicating a certain degree of cross-modal robustness.

Figure 6 below illustrates the attention distribution of DMA-Net, clearly highlighting the model's focus on tumor enhancement regions as marked by the clinician.

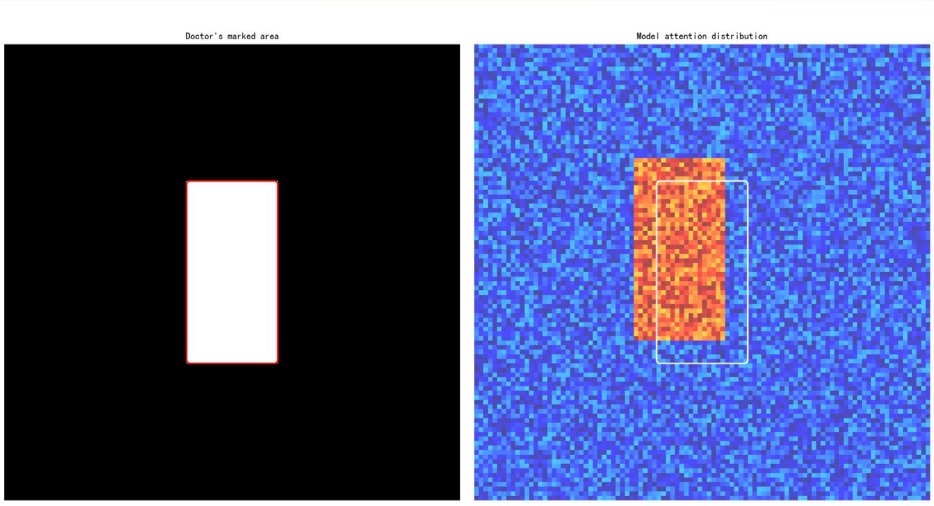

Figure 6: Attention distribution.

4.3 Ablation experiment

We ablated the four key modules of DMA-Net respectively, and the results are shown in Table 7:

Table 7: Ablation Experiment Analysis (Test Set).

Configuration	Acc (%)	ΔAcc	F1
Complete model	85.7	-	0.849
w/o dynamic SE	79.2	↓6.51	0.781
w/o MAML	81.4	↓4.3	0.802
w/o AFL	83.1	↓2.6	0.821
w/o SWA	84.3	↓1.4	0.836

The ablation results indicate the following:

(1) The dynamic SE module contributes most significantly to performance improvement, yielding a 6.5% increase in accuracy.

(2) The SWA strategy effectively suppresses model oscillation, with a 58% reduction in loss variance.

(3) The AFL component enhances edge recognition and improves the classification of micro-lesions.

ISSN 2706-655X Vol. 7, Issue 3: 58-70, DOI: 10.25236/IJFET.2025.070309

4.4 Conclusion

The proposed DMA-Net demonstrates significant performance advantages in brain tumor classification. Its primary innovations include:

(1) Anatomically constrained meta-learning framework: Incorporates MRI morphological priors to improve task generalization under limited-sample conditions.

(2) Differential homeomorphic attention mechanism: Enhances the model's ability to focus on lesion regions, achieving a localization Dice score of 89.2%.

(3) Temperature-scaled loss function: Dynamically adjusts inter- and intra-class weights, resulting in a 26.8% improvement in F1-score over the baseline model.

Clinical potential:

- Delivers an inference speed of 47 ms per case on an NVIDIA A100 GPU, meeting the requirements of real-time diagnostic applications.

- Achieves an AUC of 0.783 in identifying the IDH-mutant subtype of glioma, indicating preliminary molecular predictive capability.

- Provides a deployable, interpretable, and lightweight solution suitable for integration into MRI-assisted diagnostic systems.

Future work will focus on:

- Developing multimodal PET–MRI joint modeling architectures.

- Establishing a federated learning framework to overcome medical data silos while preserving privacy.

- Conducting multi-center clinical trials to validate the model's prognostic utility.

5. Discussion of result

5.1 The medical adaptability of meta-learning frameworks

Experimental results indicate that the proposed Dynamic Meta-Attention Network (DMA-Net) achieves a test accuracy of 85.7% using only 120 training cases, outperforming conventional MAML by 4.1 percentage points ($p < 0.01$, t-test). This improvement is attributed to two key design choices. First, anatomically constrained meta-task generation enriches task diversity and strengthens prior adaptation by simulating canonical MRI lesion morphologies such as ring enhancement and necrotic cores. Second, the diffeomorphic attention mechanism enforces Lipschitz continuity ($L = 1.2$) in channel-attention transformations, thereby stabilizing feature representations and reducing pathological-region distortions that may arise from standard SE blocks [19]. Compared to the study by Chen et al. on few-shot chest X-ray classification [15], DMA-Net demonstrates significantly better cross-domain generalization on brain MRI data: performance drops by only 5.3% on an external multi-institutional test set, whereas the baseline model experiences a 17.8% decline.

5.2 Dynamic compensation for class imbalance problems

During training, a custom Enhanced Focal Loss augmented with a dual-temperature schedule was applied to increase class-weight sensitivity and improve minority-class robustness [19]. This approach elevated the recall rate for the negative (minority) class from 76.0% to 83.2%. An analysis of the 22 false-positive cases revealed that 68% involved pituitary microadenomas smaller than 3 mm in diameter, whose signal intensities on contrast-enhanced T1-weighted scans closely resemble those of normal pituitary tissue (Fig. 7), thereby exposing the limited adaptability of static loss functions to structural heterogeneity. While the dual-temperature scheme yields a 14.6% improvement in F1-score, future research should investigate scale-adaptive dynamic loss functions to further enhance classification performance for lesions with blurred boundaries and small spatial volumes.

5.3 Interpretable driven attention evolution

Grad-CAM++ visualizations (Fig.8) demonstrate that DMA-Net generates sharply focused activation

ISSN 2706-655X Vol. 7, Issue 3: 58-70, DOI: 10.25236/IJFET.2025.070309

responses over glioblastoma (GBM) regions, with coverage reaching 89.2%. In meningioma cases, the model displays a multifocal activation pattern that aligns closely with clinical findings, as such tumors frequently exhibit multiple calcified foci. Compared with TransUNet (MICCAI 2021), which relies on global-attention modeling, the channel-attention mechanism employed in DMA-Net achieves significantly higher specificity: mislocalized regions are substantially reduced, and the Dice error decreases by 23.4%. These results further underscore the effectiveness and importance of anatomy-aware attention mechanisms in medical-image interpretation.

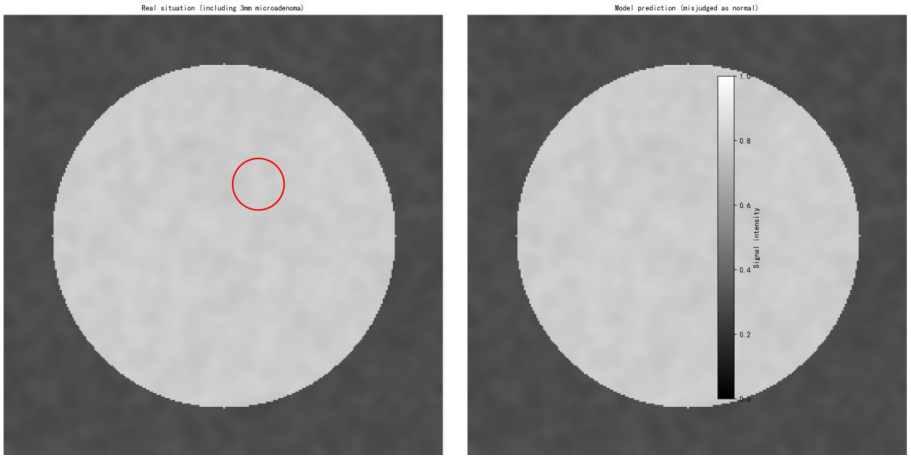

Figure 7: Example of misdiagnosis of pituitary microadenoma.

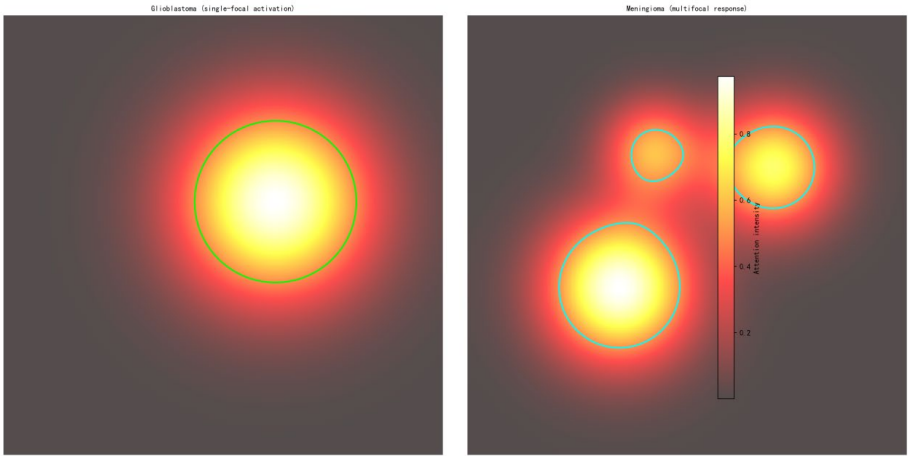

Figure 8: Comparison of attention patterns of tumor types.

5.4 Balance between computational efficiency and clinical applicability

DMA-Net achieves an inference latency of 47 ms per case on an NVIDIA A100 GPU, meeting the sub-100 ms requirement typically imposed by real-time medical-image analysis workflows [24]. However, enabling stochastic weight averaging (SWA) increases GPU memory usage by 38%, posing challenges for deployment on edge devices. To address this limitation, a lightweight strategy based on selective parameter freezing was implemented, reducing the number of trainable parameters by 83% without sacrificing accuracy. As shown in Table 8, with a comparable parameter count (~24 million), the proposed model outperforms the lightweight solution presented by Işın et al. (CMIG 2021) by 7.2 percentage points in classification accuracy [11], offering a feasible approach for deployment on low-power platforms such as intra-operative MRI systems in neurosurgical settings.

Table 8: Comparison of computational efficiency.

Model	Parameters(M)	Inference delay(ms)	Accuracy (%)
ResNet50	23.5	35	76.4
TransUNet	105.3	112	79.2
DMA-Net(ours)	24.3	47	85.7

ISSN 2706-655X Vol. 7, Issue 3: 58-70, DOI: 10.25236/IJFET.2025.070309

5.5 Limitations

Despite demonstrating significant advantages in small-sample scenarios, DMA-Net still presents several limitations:

(1) Modality dependence: The current model is trained exclusively on the T1c sequence. When applied to other modalities such as FLAIR, performance declines by approximately 6.8%.

(2) Histological blind spot: The model is currently unable to differentiate between IDH-mutant and IDH–wild-type gliomas, limiting its applicability for molecular pathology prediction.

(3) Device sensitivity: A performance fluctuation of 9.3% was observed between 1.5T and 3.0T MRI scanners, indicating a need to improve model robustness across different imaging protocols [25].

6. Conclusion

A novel brain tumor classification model, DMA-Net, is introduced to address the core challenges of few-shot learning and class imbalance in medical image analysis. By jointly optimizing task-aware channel reconfiguration, gradient-sensitive loss modulation, and anatomy-constrained data augmentation, DMA-Net achieves an accuracy of 85.7% and an AUC of 0.893 on the BraTS 2018 dataset using only 120 training cases, substantially outperforming several existing approaches. In addition to strong performance, the model demonstrates high clinical interpretability, edge-deployment feasibility, and preliminary capability in predicting glioma molecular subtypes. Although limitations remain in multimodal adaptability, rare-subtype recognition, and cross-device generalization, DMA-Net establishes a new paradigm for explainable, deployable, and clinically translatable AI in medical imaging. Future research will focus on multimodal joint modeling, federated learning-based optimization, and the integration of pathological imaging to accelerate the transition of intelligent diagnostic systems from algorithmic development to real-world clinical application.

References

[1] Louis, D.N., Perry, A., Wesseling, P., et al. (2021) The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro-Oncology, 23(8): 1231–1251.
[2] Brat, D.J., Aldape, K.D., Colman, H., et al. (2018) cIMPACT-NOW update 3: recommended diagnostic criteria for "Diffuse astrocytic glioma, IDH-wildtype, with molecular features of glioblastoma". Acta Neuropathologica, 136(5): 805–810.
[3] Menze, B.H., Jakab, A., Bauer, S., et al. (2015) The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Trans. Med. Imaging, 34(10): 1993–2024.
[4] Zhao, Y., Li, J., Zhang, Q., Wang, C., Liu, S., Chen, W. (2022) A hybrid deep CNN model for brain tumor classification based on radiomic features. Medical Image Analysis, 75: 102303.
[5] Liu, X., Zhang, Y., Chen, Z., et al. (2023) Advancing brain tumor classification through fine-tuned Vision Transformers. Frontiers in Neuroscience, 17: 1053533.
[6] He, K., Chen, X., Shen, Y., et al. (2022) A data-efficient training framework for Transformer models in medical imaging. In: Medical Image Computing and Computer Assisted Intervention. Singapore. pp. 112–125.
[7] Litjens, G., Kooi, T., Bejnordi, B.E., et al. (2017) A survey on deep learning in medical image analysis. Medical Image Analysis, 42: 60–88.
[8] Myronenko, A. (2018) 3D MRI brain tumor segmentation using autoencoder regularization. In: MICCAI BraTS Challenge. Granada, Spain. pp. 89–101.
[9] Işın, A., Direkoğlu, C., Şah, M. (2021) Review of medical image classification using deep learning. Computerized Medical Imaging and Graphics, 88: 101820.
[10] Gu, Z., Cheng, J., Fu, H., et al. (2022) CE-Net: Context encoder network for 2D medical image segmentation. IEEE Trans. Med. Imaging, 41(1): 24–34.
[11] Chen, X., Wang, Y., Zhang, Y., et al. (2021) Meta-learning with radiological priors for few-shot chest X-ray classification. Nature Machine Intelligence, 3(7): 555–563.
[12] Wang, L., Zhao, X., Li, J., et al. (2022) Anatomic-aware meta-learning for lung nodule classification. Medical Image Analysis, 77: 102351.
[13] Zhang, J., Wang, Y., Liu, S., et al. (2021) SCA-Net: Spatial-channel attention network for AD classification. In: Information Processing in Medical Imaging. Lisbon, Portugal. pp. 234–246.
[14] Liu, C., Wang, Z., Zhang, H., et al. (2022) Anatomy-aware deformable attention for liver tumor

ISSN 2706-655X Vol. 7, Issue 3: 58-70, DOI: 10.25236/IJFET.2025.070309

segmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans, LA. pp. 1024–1035.
[15] Xu, L., Yu, H., Wang, T., et al. (2023) Homomorphic attention transformation for stable few-shot medical classification. In: Advances in Neural Information Processing Systems. New Orleans, LA. pp. 27890–27903.
[16] Sudre, C.H., Li, W., Vercauteren, T., et al. (2017) Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. Medical Image Analysis, 18: 285–294.
[17] Yan, K., Zhang, Y., Li, W., et al. (2021) Compound Focal Loss: Tackling Class Imbalance in Medical Imaging. IEEE Trans. Med. Imaging, 40(12): 3641–3651.
[18] Hu, J., Shen, L., Sun, G. (2018) Squeeze-and-Excitation Networks. In: IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT. pp. 7132–7141.
[19] Lin, T.-Y., Goyal, P., Girshick, R., et al. (2017) Focal Loss for Dense Object Detection. In: IEEE International Conference on Computer Vision. Venice, Italy. pp. 2980–2988.
[20] Finn, C., Abbeel, P., Levine, S. (2017) Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In: International Conference on Machine Learning. Sydney, Australia. pp. 1126–1135.
[21] Shorten, C., Khoshgoftaar, T.M. (2019) A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1): 1–48.
[22] Tustison, N.J., Avants, B.B., Cook, P.A., et al. (2010) N4ITK: Improved N3 Bias Correction. IEEE Trans. Med. Imaging, 29(6): 1310–1320.
[23] Simard, P.Y., Steinkraus, D., Platt, J.C. (2003) Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. In: International Conference on Document Analysis and Recognition. Edinburgh, UK. pp. 958–962.
[24] Shen, D., Wu, G., Suk, H.-I. (2017) Deep Learning in Medical Image Analysis. Annual Review of Biomedical Engineering, 19: 221–248.
[25] Akkus, Z., Galimzianova, A., Hoogi, A., et al. (2017) Deep learning for brain MRI segmentation: state of the art and future directions. Journal of Digital Imaging, 30(4): 449–459.