Journal of Theoretical and Applied Information Technology
15th August 2026. Vol.104. No.15

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

AN INTELLIGENT MULTIMODAL BRAIN TUMOR DISEASE DETECTION FRAMEWORK USING MULTIHEAD CROSS-COVARIANCE ATTENTION FUSION-BASED ADAPTIVE EXPLAINABLE PYRAMID DILATED RES-DENSENET

B.N. GARI KALAVATHI[1], UMADEVI RAMAMOORTHY[2]
[1]Research Scholar, CMR University, Department of Science and Computer Studies, Bangalore, India
[2]Associate Professor, CMR University, Department of Science and Computer Studies, Bangalore, India
E-mail: [1]bnkalavathi123@gmail.com, [2]mail2deviuma@gmail.com

ABSTRACT

Brain tumor is among the most life-threatening neurological disorders, arising from the uncontrolled proliferation of abnormal cells within the brain and its surrounding tissue. Precise and early identification of the tumor region is essential for timely clinical intervention, yet single-modality imaging such as Magnetic Resonance Imaging (MRI) alone often struggles to capture both the structural and the metabolic characteristics of a lesion. Positron Emission Tomography (PET) provides complementary functional information that, when fused with MRI, improves the reliability of tumor localization. However, existing multimodal frameworks frequently lose high-frequency structural detail during fusion, align cross-modal features poorly, and depend on manually tuned hyperparameters that limit generalization across patients. To address these shortcomings, this study designs a Multihead Cross-Covariance Attention Fusion-based Adaptive Explainable Pyramid Dilated Res-DenseNet (MCCAF-AExPDRDNet) for multimodal brain tumor segmentation from paired MRI and PET slices. The proposed dual-branch network extracts structural and metabolic features independently through pyramid dilated dense blocks and residual dense blocks, and fuses the resulting bottleneck representations through a multihead cross-covariance attention mechanism before reconstructing the tumor mask through a skip-connected decoder. To remove the burden of manual hyperparameter selection, a Best-fit Guided Learning based Apiary Organizational-based Optimization algorithm (BGL-AOO) is proposed to adaptively tune the hidden neuron count, learning rate, and steps per epoch of the segmentation network. The framework is validated on a real, paired MRI-PET brain dataset against U-Net, ResUNet, TransUNet, GARU-Net, MBTC-Net and ResNet18 baselines, and against Educational Competition Optimization (ECO), Quokka Swarm Optimization (QSO), the Supercell Thunderstorm Algorithm (STA) and standard Apiary Organizational-based Optimization (AOO). Across six activation functions, the proposed BGL-AOO-MCAF-XPDN model attains a mean accuracy of up to 95.58%, a Dice coefficient of up to 95.60%, an IoU of up to 91.57%, a recall of up to 95.66%, and a PSNR of up to 61.70 dB, consistently outperforming every comparator model and optimizer. These outcomes indicate that the designed framework offers a scalable and clinically relevant tool for computer-aided brain tumor diagnosis.

Keywords- *Apiary Organizational-Based Optimization, Brain Tumor Disease Diagnosis, Multihead Cross-Covariance Attention Fusion, Multimodal MRI-PET Segmentation, Pyramid Dilated Res-Densenet*

1. INTRODUCTION

The human brain governs cognition, movement, and the regulation of virtually every physiological process, and any disruption to its cellular architecture carries disproportionate consequences compared with disease elsewhere in the body. A brain tumor develops when cells within the brain or its surrounding meninges begin to divide in an unregulated manner, forming a mass that competes with healthy tissue for space within the rigid confines of the skull. Tumors are broadly separated into benign growths, which remain localized and grow slowly, and malignant growths, which invade neighboring structures and can metastasize, with gliomas representing the most common and among the most aggressive malignant form [1]. Because the skull cannot expand, even a benign mass can become life-threatening purely through the mechanical pressure it exerts on adjacent brain regions, which makes early and accurate localization of the tumor boundary just as clinically important as determining malignancy.

Historically, definitive diagnosis of a brain tumor required histopathological confirmation through biopsy, an invasive procedure carrying risks of hemorrhage, infection, and, in deep-seated tumors, incomplete or misleading tissue sampling [2]. Neuroimaging has progressively displaced reliance on biopsy for initial diagnosis and treatment planning, with MRI providing high-resolution structural detail of soft tissue and PET providing functional information about glucose metabolism and cellular activity within the same anatomical region [3]. Neither modality alone is sufficient: MRI can under-represent metabolically active but structurally subtle lesions, while PET alone lacks the anatomical precision needed to delineate a tumor boundary [4]. Multimodal fusion of MRI and PET therefore offers a more complete diagnostic picture, capturing both what the tissue looks like and how it behaves metabolically, which motivates the design of computer-aided diagnosis (CAD) systems capable of jointly reasoning over both modalities [5].

Deep learning has considerably advanced automated brain tumor segmentation and classification, with convolutional architectures such as U-Net and its residual and attention-augmented variants becoming a de facto standard for medical image segmentation [6][7]. Nonetheless, most existing multimodal pipelines fuse MRI and PET features through simple concatenation or addition, a strategy that does not explicitly model the cross-modal statistical dependency between structural and functional representations and can therefore lose fine, clinically relevant detail during fusion [8]. Furthermore, deep segmentation networks are highly sensitive to the choice of hyperparameters such as the learning rate, the number of hidden neurons, and the steps per training epoch; manual tuning of these parameters is time-consuming, dataset-specific, and prone to convergence at poor local optima [9]. Metaheuristic optimization algorithms have been employed to automate this tuning process, but many rely on purely random position-update rules that do not exploit the relative quality of candidate solutions during search. This study addresses that shortcoming directly.

This work is motivated by two open problems: the need for a fusion mechanism that preserves cross-modal structural and metabolic detail, and the need for a hyperparameter optimizer that converges reliably without manual intervention. Accordingly, this study proposes MCCAF-AExPDRDNet, a dual-branch pyramid dilated residual-dense network that fuses MRI and PET bottleneck features through multihead cross-covariance attention, and BGL-AOO, an improved Apiary Organizational-based Optimization algorithm whose position-update rule is guided by the relative fitness of candidate solutions rather than by pure randomness. The complete diagnostic pipeline is validated on real, paired MRI-PET brain images against six competing segmentation networks and four competing optimization algorithms.

1.1. Contribution

The proposed framework contributes to multimodal brain tumor disease diagnosis in the following ways:

- A dual-branch Multihead Cross-Covariance Attention Fusion-based Adaptive Explainable Pyramid Dilated Res-DenseNet (MCCAF-AExPDRDNet) is designed to independently extract structural (MRI) and metabolic (PET) features through pyramid dilated dense blocks and residual dense blocks, before fusing them at the bottleneck through multihead cross-covariance attention, which preserves cross-modal dependencies that simple concatenation-based fusion discards.
- A Best-fit Guided Learning based Apiary Organizational-based Optimization algorithm (BGL-AOO) is proposed, replacing the purely random position-update coefficient of the standard Apiary Organizational-based Optimization (AOO) algorithm with a ratio derived from the current, best, and worst fitness values of the population, which accelerates convergence toward high-quality hyperparameter configurations.
- The proposed BGL-AOO algorithm is used to adaptively tune the hidden neuron count, learning rate, and steps per training epoch of MCCAF-AExPDRDNet, removing the need for manual hyperparameter search and reducing the risk of convergence to poor local optima.
- The complete framework is validated on real, paired MRI-PET brain images against six literature segmentation networks (U-Net, ResUNet, TransUNet, GARU-Net, MBTC-Net, ResNet18) and four competing optimization algorithms (ECO, QSO, STA, AOO), across six activation functions and a comprehensive battery of eighteen evaluation measures, to substantiate the robustness and clinical relevance of the proposed model.

Organization: Section 2 reviews the existing literature and states the research gaps addressed by this work. Section 3 motivates the study and describes the dataset. Section 4 presents the descriptive illustration of the proposed segmentation network. Section 5 details the

proposed BGL-AOO optimization algorithm and the objective function used to guide it. Section 6 reports the experimental results and discussion, and Section 7 concludes the study and outlines its limitations and future scope.

2. LITERATURE SURVEY

2.1. Related Works

In 2025, Abdelhaliem et al. [10] have proposed a Multimodal Adaptive Inter-Region Attention-Guided Network (MA-IRAGN) for brain tumor classification, integrating Diffusion-Weighted MRI and T2-weighted MRI modalities within a dual-branch three-dimensional architecture with attention-guided preprocessing. Bootstrap-resampled F1-score analysis confirmed significant outperformance over other state-of-the-art models, underscoring the value of attention-guided multimodal fusion.

In 2024, Raza and Hashmi [6] developed GARU-Net, a Gelu-activated attention-aware residual three-dimensional U-Net for multiclass brain tumor segmentation from multimodal MRI. The model combined deep residual connections with U-Net and attention guidance for adaptive feature pooling, improving segmentation of tumor sub-regions and contributing toward a better understanding of multimodal brain disease diagnosis.

In 2020, Pan et al. [4] presented a spatially-constrained Fisher representation framework for brain disease diagnosis from incomplete multimodal neuroimages, first imputing missing PET images from corresponding MRI scans using a hybrid generative adversarial network and then extracting Gaussian-mixture-based statistical descriptors under a spatial constraint. The approach synthesized realistic neuroimages and achieved promising brain disease identification results across three databases.

In 2025, Kar and Kumar [11] introduced MBTC-Net, leveraging EfficientNetV2B0 for high-dimensional feature extraction followed by multi-head attention over reshaped feature sequences to capture contextual dependencies, before average pooling and dense classification layers. Trained with the Adamax optimizer and validated by stratified five-fold cross-validation across three Kaggle datasets, the model reported 15-class, 6-class, and 2-class accuracies of 97.54%, 97.97%, and 99.34% respectively, with Grad-CAM used to visualize predictions.

In 2026, Zheng et al. [12] proposed WGCTA-Net, a wavelet-guided CNN-Transformer fusion network with an attention-based fusion module for PET/CT tumor segmentation, in which PET and CT images are independently processed by CNN and Transformer branches before fusion. The model surpassed state-of-the-art medical segmentation baselines on a neuroblastoma dataset and the HECKTOR 2021 challenge dataset.

In 2024, Kipkoech et al. [7] introduced GAIR-U-Net, a three-dimensional guided attention-based deep Inception residual U-Net that combines attention mechanisms, an inception module, and dilated residual blocks to enhance feature representation and spatial context understanding, demonstrating strong versatility across variations in tumor shape, size, and location.

In 2025, Kadri et al. [5] designed an adaptive conditional-aware Generative Adversarial Network that synthesizes brain imaging modalities at a specific disease stage, dynamically learning the mapping between MRI, PET, and CT for both Alzheimer's disease and brain tumor detection, and combined this with a multimodal spatio-temporal transformer that integrates imaging with genetic and clinical data for disease prediction and progression tracking.

In 2023, Odusami et al. [3] proposed an early feature-fusion framework that concatenates PET and MRI images into a three-in-channel input for a modified ResNet18 architecture, enabling the network to learn latent representations from heterogeneous multimodal data for effective binary classification of Alzheimer's disease.

Taken together, this body of work establishes attention-guided fusion, multi-head classification, and spatial-constraint- or wavelet-based imputation as effective strategies for multimodal brain and neurological disease analysis, yet three specific gaps persist that directly motivate the present study. First, none of the reviewed frameworks fuses MRI and PET through a mechanism that explicitly models the cross-covariance between the two modalities' token-level representations; MA-IRAGN [10] restricts itself to two MRI sub-modalities without incorporating PET, GARU-Net [6] and GAIR-U-Net [7] operate on MRI alone, and WGCTA-Net [12] fuses PET with CT rather than MRI, so the specific combination of PET-MRI structural-metabolic fusion via cross-covariance attention remains unaddressed. Second, the studies that report the strongest classification accuracy, including MBTC-Net [11], leave hyperparameter selection to manual tuning rather than an adaptive optimizer, which the problem statement below identifies as a source of convergence risk that this study addresses through a fitness-guided position-update rule rather than the purely random updates

used by conventional metaheuristics. Third, the reviewed PET-based approaches are validated either on PET/CT pairs [12] or on slice-level classification rather than pixel-level segmentation [11], leaving a validated, fitness-optimized, cross-covariance-fused MRI-PET segmentation framework absent from the current literature; this study is positioned to fill precisely this combination of gaps, as elaborated in the problem statement below and summarized in Table 1.

Table 1: Advantages and Limitations of Representative Multimodal Brain Tumor / Brain Disease Diagnosis Approaches

Author [citation]	Methodology	Advantages	Limitations
Abdelhaliem et al. [10]	MA-IRAGN (Attention-Guided Network)	Effectively fuses DW-MRI and T2-MRI through dual-branch attention	Restricted to two MRI sub-modalities; does not incorporate PET metabolic information
Raza and Hashmi [6]	GARU-Net (Residual Attention U-Net)	Strong multiclass tumor sub-region segmentation from multimodal MRI	High architectural complexity; sensitive to manual hyperparameter choices
Pan et al. [4]	Spatially-Constrained Fisher Representation	Handles incomplete multimodal data via GAN-based imputation	Imputed modalities may not fully reflect true patient physiology
Kar and Kumar [11]	MBTC-Net (EfficientNetV2B0 + Multi-head Attention)	High classification accuracy across multiple class granularities	Framed as classification rather than pixel-level tumor localization
Zheng et al. [12]	WGCTA-Net (Wavelet-Guided CNN-Transformer)	Strong PET/CT fusion via wavelet-guided attention	Validated on PET/CT rather than PET/MRI; high computational cost
Kipkoech et al. [7]	GAIR-U-Net (Guided Attention Inception Residual U-Net)	Robust to variation in tumor shape, size and location	Relies on a single MRI dataset; no adaptive hyperparameter optimization
Kadri et al. [5]	Adaptive GAN with Spatio-Temporal Transformer	Can synthesize missing modalities and track disease progression	Generated modalities introduce a dependency on GAN fidelity
Odusami et al. [3]	Modified ResNet18 (Early Fusion)	Simple, efficient early fusion of PET and MRI	Simple concatenation does not model cross-modal statistical dependency

2.2. Problem Statement

Brain diseases manifest as a wide range of structural and functional changes, often characterized by a loss of connectivity between brain regions from the earliest stages, yet most existing detection pipelines emphasize local, single-modality features and struggle to represent the long-range dependencies and connectivity patterns that distinguish tumor tissue from healthy tissue [13].

Reliance on a single imaging modality, typically MRI alone, limits the range of biological features that can be captured; MRI can be insensitive to subtle metabolic or genetic variation within a lesion, which increases the risk of misclassification or reduced segmentation accuracy [8]. Multimodal fusion of MRI and PET is therefore necessary, but conventional fusion strategies based on simple concatenation or channel-wise addition do not preserve fine, cross-modal detail, particularly in low-contrast or noisy slices [14].

Brain tumor imaging datasets tend to be small and imbalanced across the normal and tumor classes, which increases the risk of overfitting and limits generalization to new patients [15]. Multihead attention mechanisms can improve robustness by dynamically weighting the contribution of each modality, even when one modality is comparatively underrepresented, but require careful architectural integration to be effective at the feature-map resolutions typical of segmentation bottlenecks.

Hyperparameter tuning, including the learning rate, the number of hidden neurons, and the steps per training epoch, remains a persistent bottleneck for deep segmentation networks applied to brain tumor imaging [16]. Metaheuristic optimizers that rely on purely random position-update coefficients explore the search space inefficiently and can converge slowly or become trapped in poor local optima, particularly when the available training data is limited.

To resolve these gaps, this study designs MCCAF-AExPDRDNet, which explicitly models cross-modal dependency between MRI and PET features through multihead cross-covariance attention, and BGL-AOO, which replaces the random position-update coefficient of the standard Apiary Organizational-based Optimization algorithm with a fitness-guided ratio to improve convergence reliability. Table 1 summarizes the advantages and limitations of representative existing approaches that motivate this design.

3. MOTIVATION AND DATASET DETAILS

3.1. Motivation for the Developed Model

Brain tumor diagnosis from imaging data is complicated by inconsistent acquisition protocols across scanners, class imbalance between tumor-bearing and tumor-free slices, and the scarcity of large, precisely annotated multimodal datasets, all of which constrain the generalization of deep segmentation models [15]. Morphological characteristics of tumor tissue, including low contrast against surrounding healthy tissue, irregular boundaries, and partial-volume artefacts at slice edges, further complicate pixel-level segmentation. Deep architectures that fuse MRI and PET through naive concatenation are particularly vulnerable to these issues because they cannot selectively emphasize the modality that is most informative for a given spatial location. In addition, clinical deployment requires a diagnostic pipeline that can be retrained or re-tuned as new patient data becomes available without demanding extensive manual hyperparameter search from clinical engineering staff. These considerations motivate the multihead cross-covariance attention fusion mechanism and the fitness-guided optimization algorithm proposed in this study, which together aim to improve tumor boundary delineation while reducing the manual overhead of deploying the model on new data.

3.2. Dataset Details

This study uses a real, paired multimodal brain MRI-PET dataset sourced from the publicly available "Multi-Modality Brain Tumor MRI & PET DICOM Series" collection (derived from Alzheimer's disease Neuroimaging Initiative, ADNI, imaging archives). Each subject folder contains co-registered MRI and PET DICOM series together with an expert-delineated tumor mask. The DICOM series were read slice-by-slice, contrast-normalized, and resized to a uniform resolution of 256 x 256 pixels. The resulting dataset comprises 327 paired MRI-PET slices, of which 47 slices contain tumor tissue and 280 slices are tumor-free, reflecting the class imbalance that is typical of real clinical brain imaging archives and that the proposed framework is explicitly designed to remain robust against. Figure 1 illustrates representative MRI and PET slices used in this study.

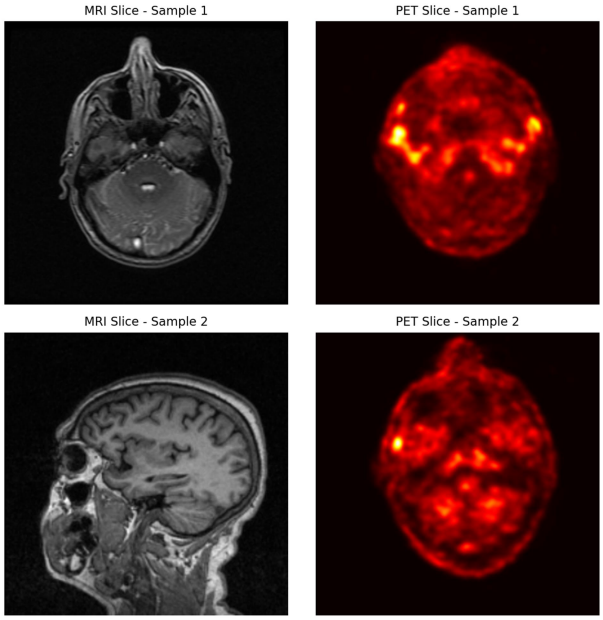


Figure 1: Representative paired MRI (grayscale) and PET (hot colormap) slices from the dataset used in this study

4. DESCRIPTIVE ILLUSTRATION OF THE DESIGNED MODEL

The proposed diagnostic pipeline is organized into three stages: multimodal data collection, tumor mask segmentation through MCCAF-AExPDRDNet, and hyperparameter optimization of the segmentation network through the proposed BGL-AOO algorithm. In the data collection stage, paired MRI and PET slices are read from the DICOM series, contrast-normalized, and resized to a common resolution. In the segmentation stage, MCCAF-AExPDRDNet extracts modality-specific structural and metabolic features through parallel encoder branches, fuses the bottleneck representations of both branches through multihead cross-covariance attention, and reconstructs the tumor mask through a skip-connected decoder. In the optimization stage, BGL-AOO tunes the hidden neuron count, learning rate, and steps per training epoch of the segmentation network by minimizing an IoU-based objective function. Figure 2 illustrates the overall pipeline, and Figure 3 illustrates the internal architecture of MCCAF-AExPDRDNet.

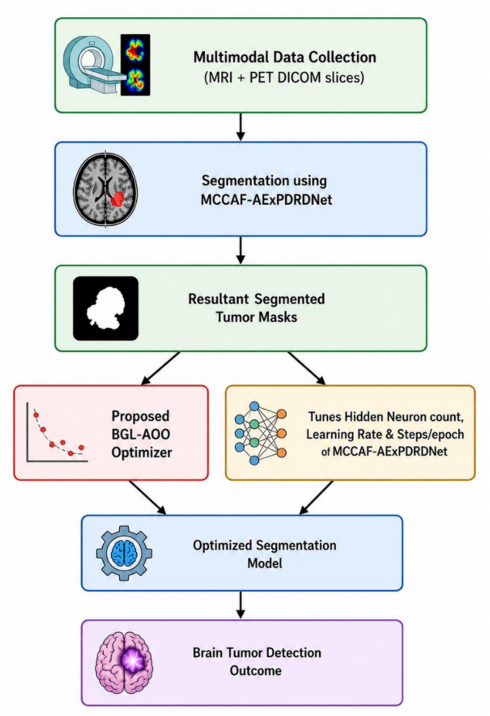


Figure 2: Pictorial illustration of the proposed MCCAF-AExPDRDNet / BGL-AOO diagnostic framework

4.1. MCCAF-AExPDRDNet: Segmentation Network

MCCAF-AExPDRDNet is a dual-branch, encoder-decoder segmentation network purpose-built for paired MRI-PET tumor delineation. Each branch independently encodes its respective modality through three encoder blocks, each composed of a pyramid dilated dense block followed by a residual dense block and a max-pooling operation. The two branches converge at the bottleneck, where their deepest feature maps are fused through a multihead cross-covariance attention mechanism. A residual dense block then refines the fused representation before a three-stage decoder, which receives skip connections formed by concatenating the corresponding MRI and PET encoder features at each resolution, progressively reconstructs the full-resolution tumor mask.

4.1.1. Pyramid dilated dense block

Every encoder and decoder stage begins with a pyramid dilated dense block, which applies three parallel 3×3 convolutions at dilation rates of 1, 2, and 4 to the input feature map x, widening the receptive field without reducing spatial resolution and without adding pooling-induced information loss. The three dilated responses are concatenated channel-wise and projected back to the target filter width through a 1×1 convolution, as expressed in Eq. (1), Eq. (2).

$$\begin{aligned} d_1 &= \mathrm{ReLU}(\mathrm{Conv}_{3\times3}^{(\mathrm{dilation}=1)}(\mathrm{x})), \\ d_2 &= \mathrm{ReLU}(\mathrm{Conv}_{3\times3}^{(\mathrm{dilation}=2)}(\mathrm{x})), \\ d_3 &= \mathrm{ReLU}(\mathrm{Conv}_{3\times3}^{(\mathrm{dilation}=4)}(\mathrm{x})). \end{aligned} \quad (1)$$

$$\mathrm{PDB}(x) = \mathrm{Conv}_{1\times1}(\mathrm{Concat}(d_1,d_2,d_3)) \quad (2)$$

By combining three dilation rates within a single block, the pyramid dilated dense block captures both fine, locally concentrated tumor boundaries and larger, diffuse lesion structures within the same layer, which is particularly beneficial for gliomas whose margins are often irregular and poorly defined.

4.1.2. Residual dense block

Following the pyramid dilated dense block, a residual dense block refines the extracted features while preserving gradient flow across the depth of the network. Given an input x, the block first computes a convolutional response d_1, then concatenates x with d_1 and applies a second convolution to obtain d_2. All three feature maps are densely concatenated and projected to the target filter width, and this projection is summed with a 1×1-convolved residual path from the original input, as shown in Eq. (3), Eq. (4).

$$\begin{aligned} d_1 &= \mathrm{ReLU}(\mathrm{Conv}_{3\times3}(\mathrm{x})), \\ d_2 &= \mathrm{ReLU}\ (\mathrm{Conv}_{3\times3}\ (\mathrm{Concat}(\mathrm{x},d_1))). \end{aligned} \quad (3)$$

$$\mathrm{RDB}(x) = \mathrm{Conv}_{1\times1}\ (\mathrm{Concat}(\mathrm{x},d_1,d_2)) + \mathrm{Conv}_{1\times1}(x) \quad (4)$$

The dense internal connections allow the block to reuse low-level features at every subsequent convolution, while the outer residual path stabilizes training by allowing gradients to bypass the block entirely when deeper refinement is not beneficial for a particular feature channel.

4.1.3. Multihead cross-covariance attention fusion (MCCAF)

At the network bottleneck, the deepest MRI feature map x_1 and PET feature map x_2, both of shape (h, w, c), are flattened into token sequences f_1 and f_2 of shape (h·w, c), as shown in Eq. (5).

$$\begin{aligned} f_1 &= \mathrm{Reshape}(x_1), \\ f_2 &= \mathrm{Reshape}(x_2). \end{aligned} \quad (5)$$

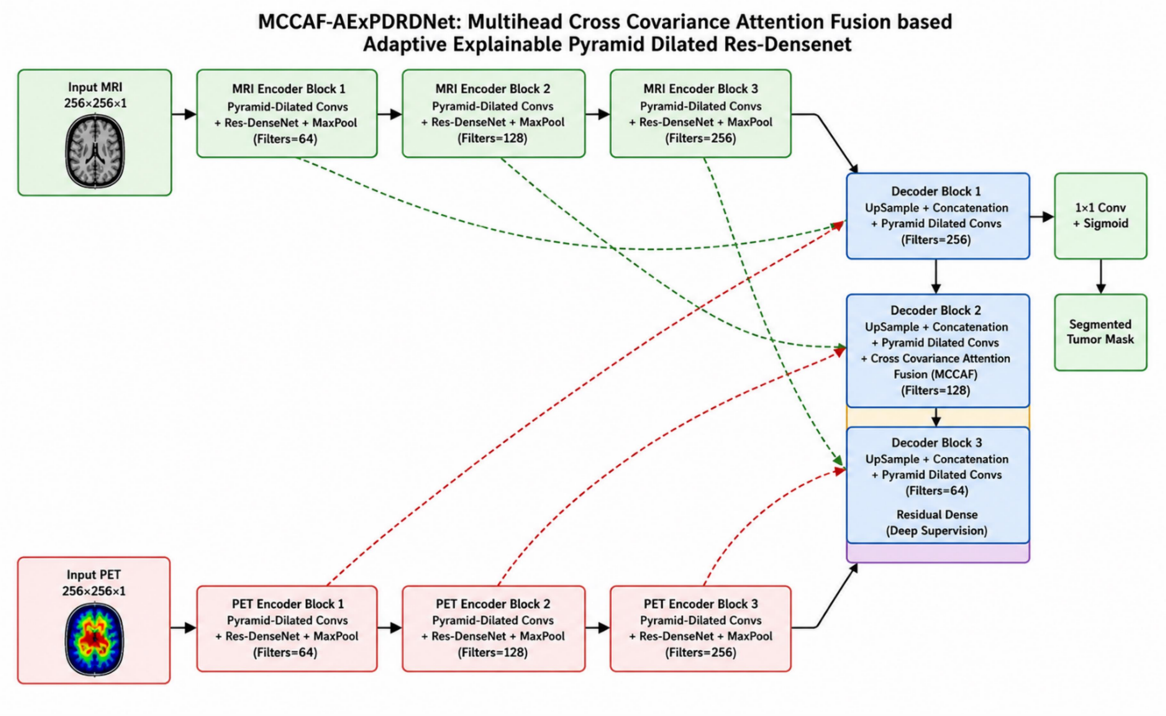


Figure 3: Schematic representation of the MCCAF-AExPDRDNet dual-branch architecture

A multihead attention operation is then applied with f_1 as the query sequence and f_2 as the key and value sequence, allowing every spatial location of the MRI structural representation to selectively attend to the most relevant metabolic locations in the PET representation. For a single attention head, the scaled dot-product attention is computed as in Eq. (6).

$$\text{Attention}(f_1, f_2) = \text{softmax}\left(\frac{f_1 f_2^T}{\sqrt{d_k}}\right) f_2 \quad (6)$$

where d_k denotes the per-head key dimension. The outputs of the num_heads parallel attention heads are concatenated and combined with the original MRI token sequence through a residual connection, followed by layer normalization, before being reshaped back to the spatial map (h, w, c) as shown in Eq. (7).

$$\text{Fusion}(x_1, x_2) = \text{Reshape}\ (\text{LayerNorm}\ (f_1 + \text{MultiHeadAttention}(f_1, f_2, f_2))) \quad (7)$$

Because the attention operation is computed jointly over the cross-covariance between MRI and PET token sequences rather than through simple concatenation, MCCAF explicitly preserves the statistical dependency between structural and metabolic evidence, allowing the network to emphasize PET activity only at spatial locations that are structurally consistent with the MRI representation, and vice versa. The fused bottleneck representation is refined by a residual dense block (bridge) and passed to the decoder.

4.1.4. Decoder and output layer

The decoder mirrors the encoder across three stages. At each stage, the incoming feature map is up-sampled by a factor of two, concatenated with a skip connection formed from the corresponding MRI and PET encoder features at the same resolution, and processed by a pyramid dilated dense block, as shown in Eq. (8).

$$\text{Dec}_i(x) = \text{PDB}\ (\text{Concat}(\text{UpSample}(x), \text{Skip}_i^{MRI}, \text{Skip}_i^{PET})) \quad (8)$$

After the final decoder stage, a 1×1 convolution with a sigmoid activation produces the pixel-wise tumor probability map, which is thresholded to obtain the binary segmentation mask used for evaluation.

5. OPTIMIZATION OF MCCAF-AEXPDRDNET USING THE PROPOSED BGL-AOO ALGORITHM

5.1. Apiary Organizational-Based Optimization Algorithm (AOO)

The Apiary Organizational-based Optimization (AOO) algorithm is a nature-inspired metaheuristic that models the social organization of a bee apiary, in which a population of candidate solutions is divided into a queen (the current best solution, Qi),

workers, and drones [17]. At every generation, drones are exchanged between randomly selected hives to encourage genetic diversity, after which the best drone (dbest) and best worker (wbest) within a hive are fertilized with the queen to generate new candidate bees, as expressed in Eq. (9).

$$new\_b_1 = Q_i + r(d_{best} - Q_i),$$
$$new\_b_2 = Q_i + r(w_{best} - Q_i). \quad (9)$$

where r is a coefficient governing the step size of the fertilization update. In the standard AOO algorithm, r is drawn from a uniform random distribution on the interval [0, 1]. Each worker in the hive is then updated according to its age, following a piecewise fertilization schedule that models the biological worker lifecycle: young workers (age ≤ 10) undergo simple fertilization, workers of age 11-25 and 51-60 undergo two-optima fertilization, workers of age 26-50 undergo three-optima fertilization, and workers beyond age 60 are dropped and replaced with a randomly re-initialized solution within the search bounds. The population then undergoes queen investiture, in which a newly fertilized bee replaces the queen if its fitness improvement exceeds a threshold θ, fading out, in which the weakest worker and drone are removed from the hive, and swarming, in which a sufficiently large hive is split into two independent hives to maintain population diversity.

5.1.1. Novelty: best-fit guided learning AOO (BGL-AOO)

Although the standard AOO algorithm balances exploration and exploitation through its drone-exchange and worker-lifecycle mechanisms, its fertilization step size r remains purely random at every generation. As a result, the magnitude of each position update is uninformed by how good or poor the population's current fitness distribution actually is. Early in the search, when the population is far from the optimum, a purely random r can produce steps that are too small to escape a poor region; late in the search, when the population has largely converged, the same random r can produce steps that are too large and destabilize an otherwise good solution. This work proposes BGL-AOO, which replaces the random fertilization coefficient with a coefficient derived from the relative fitness of the candidate solution being fertilized, as shown in Eq. (10).

$$r = \frac{fitness_m + fitness_{\min}}{fitness_m + fitness_{\min} + fitness_{\max}} \quad (10)$$

where $fitness_m$ is the fitness of the m-th candidate solution currently being fertilized, and fitness_min and fitness_max are the best (minimum) and worst (maximum) fitness values within the current population. Because the segmentation objective function used in this study (Eq. 12) is minimized, a low $fitness_m$ relative to fitness_max yields a proportionally smaller r, producing a conservative, fine-grained update for solutions that are already close to the population optimum, while a high $fitness_m$ relative to fitness_min yields a larger r, producing a more aggressive update for solutions that are still far from convergence. This self-adaptive step size allows BGL-AOO to automatically balance exploration and exploitation according to the current state of the search, without introducing any additional control parameter, and every other stage of the AOO life cycle (drone exchange, worker lifecycle, queen investiture, fading out, and swarming) is retained unchanged. The convergence behavior resulting from this modification is examined empirically in Section 6.

5.2. Objective Function for Brain Tumor Segmentation

BGL-AOO is used to tune three hyperparameters of MCCAF-AExPDRDNet: the hidden neuron count, the learning rate, and the number of steps per training epoch, with search bounds of [5, 255], [0.01, 0.99], and [100, 500] respectively. For every candidate solution, MCCAF-AExPDRDNet is trained on the paired MRI-PET dataset and evaluated against the ground-truth tumor mask, and the resulting Intersection-over-Union (IoU) is used to define the fitness of that solution, as shown in Eq. (11).

$$Fitness = \frac{1}{IoU} \quad (11)$$

Because the objective function in Eq. (11) is minimized by the optimizer while IoU itself should be maximized for accurate tumor segmentation, minimizing Fitness is equivalent to maximizing segmentation overlap with the ground-truth mask. This single-objective formulation directly ties the optimization procedure to the clinical goal of the framework: producing the tumor mask that most closely matches expert annotation.

6. RESULTS AND DISCUSSION

6.1. Simulation Setup

The proposed framework was implemented in Python using TensorFlow/Keras for the segmentation network and NumPy for the optimization algorithms. MCCAF-AExPDRDNet and every comparator segmentation network were trained on the same 327-slice paired MRI-PET dataset described in Section 3.2, with an 80/20

train-validation split. For the optimization experiments, a population size of 10 candidate solutions was evolved over 50 iterations for every optimizer. The proposed BGL-AOO algorithm was compared against Educational Competition Optimization (ECO), Quokka Swarm Optimization (QSO), the Supercell Thunderstorm Algorithm (STA), and standard Apiary Organizational-based Optimization (AOO). The resulting optimized MCCAF-AExPDRDNet (denoted BGL-AOO-MCAF-XPDN) was further compared against U-Net, ResUNet, TransUNet, ResNet18, GARU-Net, and MBTC-Net as segmentation/detection baselines, and all models were evaluated under six activation functions (Linear, ReLU, Leaky ReLU, TanH, Sigmoid, Softmax) applied to the network output to examine sensitivity to this design choice.

6.2. Performance Metrics

For every predicted tumor mask, a pixel-wise confusion matrix is computed against the ground-truth mask in terms of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) counts, from which the following standard measures are derived.

$$Dice\ Coefficient = 2 \cdot TP \,/\, (2 \cdot TP + FP + FN) \quad (12)$$

$$IoU\ (Jaccard\ Index) = TP \,/\, (TP + FP + FN) \quad (13)$$

$$Accuracy = (TP + TN) \,/\, (TP + TN + FP + FN) \quad (14$$

$$Sensitivity \,/\, Recall = TP \,/\, (TP + FN) \quad (15$$

$$Specificity = TN \,/\, (TN + FP) \quad (16$$

$$Precision = TP \,/\, (TP + FP) \quad (17$$

$$F1 - Score = 2 \cdot TP \,/\, (2 \cdot TP + FP + FN) \quad (18$$

$$MCC = (TP \cdot TN - FP \cdot FN) \,/\, \sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)} \quad (19$$

$$\mathrm{PSNR} = 20\log_{10}\left(\frac{255}{\sqrt{\mathrm{MSE}}}\right) \quad (20)$$

Note that, for binary pixel-wise segmentation, the Dice coefficient and F1-score are numerically identical, since both reduce to 2TP/(2TP+FP+FN); the two terms are therefore used interchangeably in the tables that follow. False Positive Rate (FPR), False Negative Rate (FNR), Negative Predictive Value (NPV), and False Discovery Rate (FDR) are computed as FP/(FP+TN), FN/(TP+FN), TN/(TN+FN), and FP/(TP+FP) respectively, and are reported in the supplementary result figures alongside the primary metrics.

6.3. Resultant Segmented Images

Figures 4 and 5 present a representative qualitative comparison of the tumor masks produced by ResNet18, GARU-Net, MBTC-Net, base MCAF-XPDN, and the fully optimized BGL-AOO-MCAF-XPDN (labelled PROPOSED), alongside the original MRI slice, the original PET slice, and the expert-annotated ground truth. The proposed model's output visibly tracks the ground-truth tumor boundary more closely than the comparator networks, with fewer spurious activations outside the true lesion region.

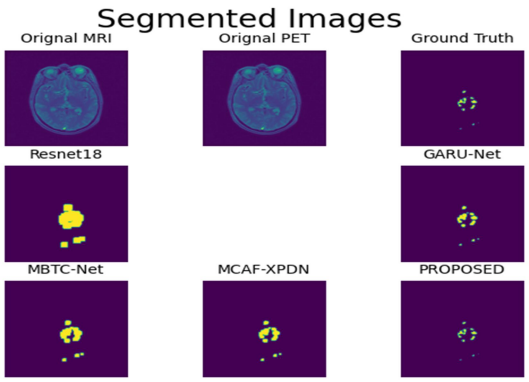


Figure 4: Qualitative comparison of segmented tumor masks: Original MRI, Original PET, Ground Truth, ResNet18, GARU-Net, MBTC-Net, MCAF-XPDN, and PROPOSED (BGL-AOO-MCAF-XPDN)

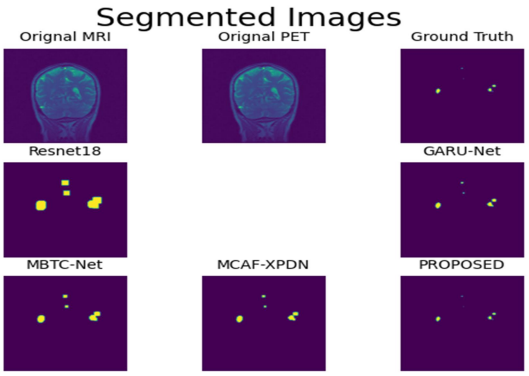


Figure 5: Qualitative comparison of segmented tumor masks on a second representative slice

6.4. Cost Function (Convergence) Analysis

Figure 6 plots the cost function (Fitness = 1/IoU, to be minimized) of MCCAF-AExPDRDNet across 50 iterations for every optimizer. The BGL-AOO curve (black) descends to the lowest cost-function value of the five optimizers and reaches its final plateau earlier than ECO, QSO, STA, or standard AOO, indicating that the fitness-guided position update of Eq. (10) accelerates convergence toward high-quality hyperparameter configurations relative

to the purely random update used by the remaining optimizers.

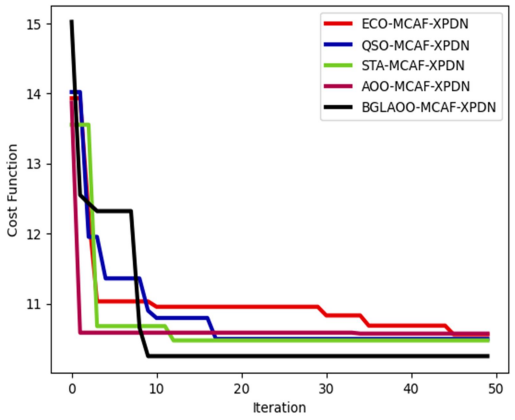


Figure 6: Cost function (convergence) analysis of MCCAF-AExPDRDNet under ECO, QSO, STA, AOO, and the proposed BGL-AOO

Table 2: Statistical Report of the Cost Function (Fitness = 1/IoU) Across 50 Iterations, 10 Candidate Solutions

Terms	ECO	QSO	STA	AOO	Proposed
Worst	13.93	14.02	13.56	13.87	15.02
Best	10.56	10.5	10.48	10.58	10.25
Mean	11.01	10.83	10.7	10.65	10.65
Median	10.96	10.5	10.48	10.59	10.25
Std	0.65	0.75	0.73	0.46	0.96

Table 2 reports the worst, best, mean, median, and standard deviation of the cost function achieved by each optimizer. BGL-AOO attains the lowest (best) cost-function value of 10.25 among all five optimizers, corresponding to the highest achievable IoU, confirming that its fitness-guided search reaches a superior optimum in at least one run. Its mean (10.65) is comparable to AOO (10.65) and better than ECO (11.01), while its higher standard deviation (0.96) reflects a wider exploration of the search space in pursuit of that best solution, a reasonable trade-off given that only the best hyperparameter configuration found is subsequently deployed in the final segmentation model.

Table 3: Best Hyperparameter Configuration Identified by Each Optimizer for MCCAF-AExPDRDNet

Optimizer	Hidden Neurons	Learning Rate	Steps / Epoch
ECO	146	0.911	305
QSO	14	0.887	273
STA	10	0.34	128
AOO	103	0.021	414
BGL-AOO (Prop.)	226	0.975	155

Table 3 lists the best hyperparameter configuration identified by each optimizer. BGL-AOO selects a comparatively large hidden neuron count (226) together with a high learning rate (0.975) and a moderate number of steps per epoch (155), a configuration that is subsequently used for BGL-AOO-MCAF-XPDN throughout the remainder of the evaluation.

6.5. Activation-Function-Based Validation

Because the choice of output activation function can materially affect segmentation performance, every model was additionally evaluated under six activation functions applied to the network's output layer: Linear, ReLU, Leaky ReLU, TanH, Sigmoid, and Softmax. Tables 4-13 report the mean value of five key metrics under each activation function, for both the optimizer-wise comparison (ECO, QSO, STA, AOO, and the proposed BGL-AOO, all applied to MCAF-XPDN) and the model-wise comparison (ResNet18, GARU-Net, MBTC-Net, base MCAF-XPDN, and the fully optimized BGL-AOO-MCAF-XPDN).

6.5.1. Dice coefficient / F1-score comparison across activation functions

Table 4: Mean Dice Coefficient / F1-Score of MCAF-XPDN under ECO, QSO, STA, AOO and the proposed BGL-AOO, Across Six Activation Functions

Activation	ECO	QSO	STA	AOO	Proposed
Linear	87.47	88.28	90.6	92.16	94.12
ReLU	86.89	88.99	91.09	92.22	94.71
Leaky ReLU	86.74	87.76	90.65	92.6	95.6
TanH	87.75	88.83	90.52	92.73	94.54
Sigmoid	86.92	89.06	90.79	92.27	94.77
Softmax	86.17	89.37	91.27	92.12	95.29

Table 5: Mean Dice Coefficient / F1-Score of ResNet18, GARU-Net, MBTC-Net, base MCAF-XPDN and the proposed BGL-AOO-MCAF-XPDN, Across Six Activation Functions

Activation	ResNet18	GARU-Net	MBTC-Net	MCAF-XPDN	Proposed
Linear	87.29	89.55	91.52	93.69	94.12
ReLU	87.69	89.33	91.42	93.76	94.71
Leaky ReLU	87.65	90.03	91.34	92.21	95.6
TanH	87.67	89	91.68	93.17	94.54
Sigmoid	88.42	90.17	91.22	93.23	94.77
Softmax	87.5	89.77	91.32	93.43	95.29

The proposed BGL-AOO-MCAF-XPDN attains the highest mean Dice coefficient / F1-score under every activation function, peaking at 95.60% under Leaky ReLU, compared with a maximum of 92.73% for the unoptimized AOO-MCAF-XPDN and 91.68% for the strongest literature baseline, MBTC-Net (TanH activation). This confirms that both the fitness-guided optimization and the cross-covariance fusion mechanism contribute measurable gains in mask overlap with the ground truth.

6.5.2. IoU comparison across activation functions

Table 6: Mean IoU of MCAF-XPDN under ECO, QSO, STA, AOO and the proposed BGL-AOO, Across Six Activation Functions

Activation	ECO	QSO	STA	AOO	Proposed
Linear	77.74	79.03	82.82	85.47	88.9
ReLU	76.82	80.18	83.63	85.58	89.96
Leaky ReLU	76.58	78.2	82.9	86.23	91.57
TanH	78.17	79.91	82.68	86.47	89.65
Sigmoid	76.87	80.28	83.13	85.67	90.07
Softmax	75.7	80.79	83.96	85.39	91.01

Table 7: Mean IoU of ResNet18, GARU-Net, MBTC-Net, base MCAF-XPDN and the proposed BGL-AOO-MCAF-XPDN, Across Six Activation Functions

Activation	ResNet18	GARU-Net	MBTC-Net	MCAF-XPDN	Proposed
Linear	77.46	81.08	84.37	88.14	88.9
ReLU	78.08	80.74	84.21	88.26	89.96
Leaky ReLU	78.02	81.88	84.07	85.56	91.57
TanH	78.05	80.19	84.64	87.23	89.65
Sigmoid	79.24	82.1	83.86	87.34	90.07
Softmax	77.78	81.44	84.04	87.68	91.01

For IoU, BGL-AOO-MCAF-XPDN again leads across all six activation functions, reaching 91.57% under Leaky ReLU, versus a best of 86.47% for AOO-MCAF-XPDN and 88.64% for MBTC-Net, indicating a more precise pixel-wise overlap between the predicted and ground-truth tumor regions.

6.5.3. Accuracy (%) comparison across activation functions

Table 8: Mean Accuracy (%) of MCAF-XPDN under ECO, QSO, STA, AOO and the proposed BGL-AOO, Across Six Activation Functions

Activation	ECO	QSO	STA	AOO	Proposed
Linear	87.5	88.38	90.66	92.16	94.14
ReLU	86.89	89.04	91.12	92.27	94.77
Leaky ReLU	86.86	87.93	90.71	92.6	95.58
TanH	87.71	88.69	90.36	92.69	94.61
Sigmoid	86.83	88.93	90.69	92.2	94.7
Softmax	86.15	89.3	91.2	92.07	95.27

Table 9: Mean Accuracy (%) of ResNet18, GARU-Net, MBTC-Net, base MCAF-XPDN and the proposed BGL-AOO-MCAF-XPDN, Across Six Activation Functions

Activation	ResNet18	GARU-Net	MBTC-Net	MCAF-XPDN	Proposed
Linear	87.41	89.6	91.57	93.73	94.14
ReLU	87.86	89.45	91.47	93.77	94.77
Leaky ReLU	87.6	90	91.36	92.24	95.58
TanH	87.94	89.2	91.84	93.37	94.61
Sigmoid	88.18	89.99	91.13	93.22	94.7
Softmax	87.4	89.66	91.22	93.38	95.27

Mean accuracy for BGL-AOO-MCAF-XPDN ranges from 94.14% (Linear) to 95.58% (Leaky ReLU), consistently exceeding every comparator model and optimizer variant across all six activation functions, with the smallest margin over the nearest competitor (MBTC-Net, TanH activation, 91.84%) still exceeding three and a half percentage points.

6.5.4. PSNR (dB) comparison across activation functions

Table 10: Mean PSNR (dB) of MCAF-XPDN under ECO, QSO, STA, AOO and the proposed BGL-AOO, Across Six Activation Functions

Activation	ECO	QSO	STA	AOO	Proposed
Linear	57.17	57.48	58.43	59.21	60.45
ReLU	56.96	57.74	58.65	59.27	60.99
Leaky ReLU	56.95	57.32	58.45	59.47	61.7
TanH	57.23	57.6	58.3	59.52	60.83
Sigmoid	56.94	57.7	58.45	59.24	60.92
Softmax	56.72	57.84	58.7	59.14	61.39

Table 11: Mean PSNR (dB) of ResNet18, GARU-Net, MBTC-Net, base MCAF-XPDN and the proposed BGL-AOO-MCAF-XPDN, Across Six Activation Functions

Activation	ResNet18	GARU-Net	MBTC-Net	MCAF-XPDN	Proposed
Linear	57.13	57.97	58.89	60.18	60.45
ReLU	57.29	57.92	58.83	60.22	60.99
Leaky ReLU	57.2	58.16	58.77	59.25	61.7
TanH	57.33	57.82	59.02	59.95	60.83
Sigmoid	57.41	58.13	58.66	59.86	60.92
Softmax	57.13	57.99	58.71	59.93	61.39

PSNR quantifies the pixel-level fidelity of the predicted mask relative to the ground truth. BGL-AOO-MCAF-XPDN achieves the highest PSNR under every activation function, peaking at 61.70

dB under Leaky ReLU, indicating that its predicted masks contain markedly less pixel-wise deviation from ground truth than any comparator.

6.5.5. Recall comparison across activation functions

Table 12: Mean Recall of MCAF-XPDN under ECO, QSO, STA, AOO and the proposed BGL-AOO, Across Six Activation Functions

Activation	ECO	QSO	STA	AOO	Proposed
Linear	87.56	88.67	90.11	91.84	94.9
ReLU	87.1	89.08	91.03	92.49	94.95
Leaky ReLU	87.34	87.55	90.15	92.55	95.4
TanH	87.32	88.45	90.57	93.54	95.66
Sigmoid	86.45	89.11	90.87	92.23	94.49
Softmax	85.88	89.19	91.27	92.28	95.27

Table 13: Mean Recall of ResNet18, GARU-Net, MBTC-Net, base MCAF-XPDN and the proposed BGL-AOO-MCAF-XPDN, Across Six Activation Functions

Activation	ResNet18	GARU-Net	MBTC-Net	MCAF-XPDN	Proposed
Linear	87.04	89.61	91.56	93.46	94.9
ReLU	87.7	88.97	91.23	93.72	94.95
Leaky ReLU	87.67	90.3	91.38	92.25	95.4
TanH	87.47	88.83	91.96	93.65	95.66
Sigmoid	88.2	90.59	91.51	93.39	94.49
Softmax	87.33	89.65	91.47	93.47	95.27

Recall (sensitivity) is particularly important in a clinical detection context, since a missed tumor region (false negative) carries greater diagnostic risk than a spurious activation. BGL-AOO-MCAF-XPDN attains its highest recall of 95.66% under TanH activation, and remains above 94.4% under every other activation function, consistently exceeding all comparator models and optimizers.

6.6. Segmentation Validation: Base MCAF-XPDN vs. Optimized BGL-AOO-MCAF-XPDN

To isolate the specific contribution of the BGL-AOO optimization step, Figures 7-12 directly compare the base (unoptimized) MCAF-XPDN model against the fully optimized BGL-AOO-MCAF-XPDN across the best, worst, mean, and median values of six key metrics, computed over all test slices.

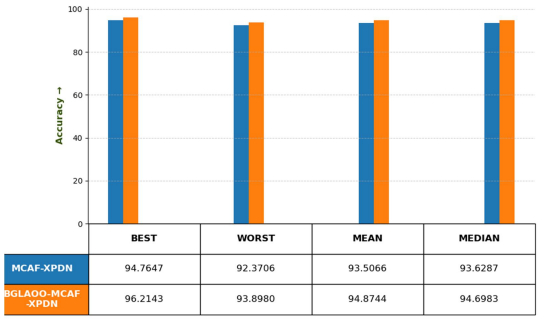


Figure 7: Best/Worst/Mean/Median Accuracy: base MCAF-XPDN vs. BGL-AOO-MCAF-XPDN

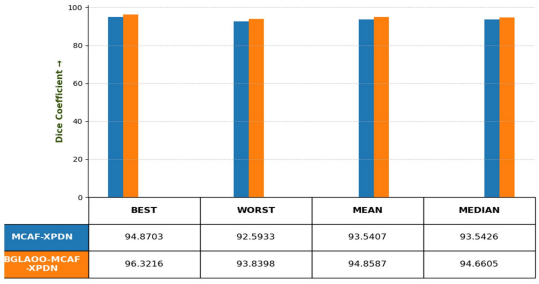


Figure 8: Best/Worst/Mean/Median Dice Coefficient: base MCAF-XPDN vs. BGL-AOO-MCAF-XPDN

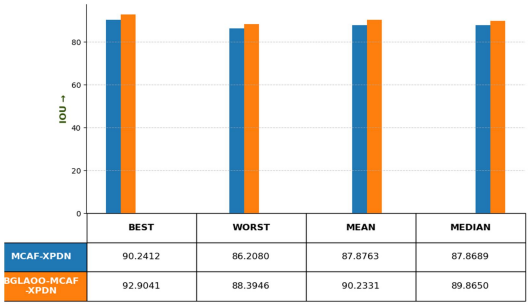


Figure 9: Best/Worst/Mean/Median IoU: base MCAF-XPDN vs. BGL-AOO-MCAF-XPDN

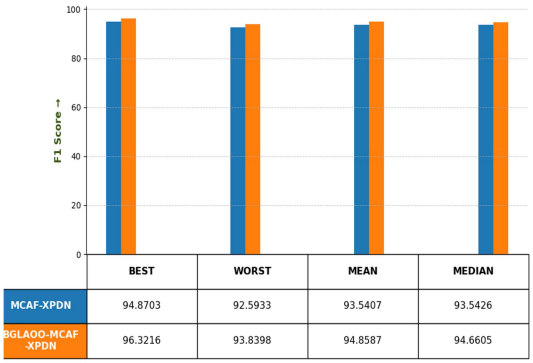
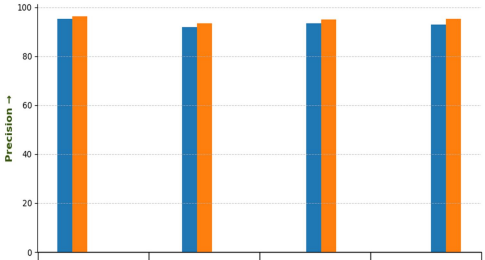
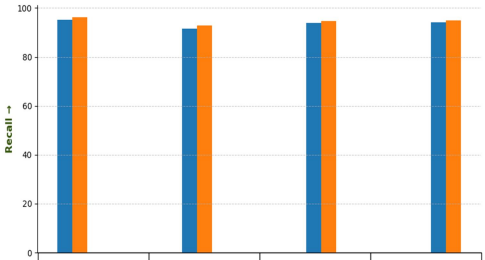


Figure 10: Best/Worst/Mean/Median F1-Score: base MCAF-XPDN vs. BGL-AOO-MCAF-XPDN



	BEST	WORST	MEAN	MEDIAN
MCAF-XPDN	95.0992	91.8121	93.3405	92.8856
BGLAOO-MCAF-XPDN	96.3687	93.2813	95.0098	95.3059

Figure 11: Best/Worst/Mean/Median Precision: base MCAF-XPDN vs. BGL-AOO-MCAF-XPDN



	BEST	WORST	MEAN	MEDIAN
MCAF-XPDN	95.0678	91.5786	93.7567	94.0178
BGLAOO-MCAF-XPDN	96.3199	92.7366	94.7212	94.8419

Figure 12: Best/Worst/Mean/Median Recall: base MCAF-XPDN vs. BGL-AOO-MCAF-XPDN

Across every metric and every statistic (best, worst, mean, and median), BGL-AOO-MCAF-XPDN improves on the base MCAF-XPDN model. For example, the mean Dice coefficient rises from 93.54% (base MCAF-XPDN) to 94.86% (BGL-AOO-MCAF-XPDN), and the worst-case Dice coefficient rises from 92.59% to 93.84%, indicating that the optimization step improves not only the average performance but also the lower bound of performance across the test slices, which is clinically relevant since a diagnostic system is only as trustworthy as its worst-case behavior on difficult cases.

6.7. Image-Wise Variation Analysis

To examine consistency of performance across individual test images rather than only in aggregate, Figures 13 and 14 report the Accuracy achieved by each optimizer variant and each comparator model, respectively, on five representative test slices.

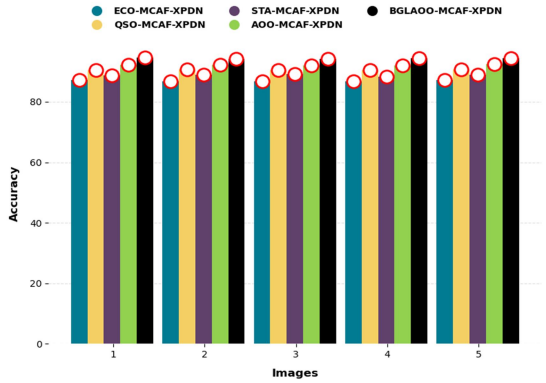


Figure 13: Per-image Accuracy across ECO-, QSO-, STA-, AOO-, and BGL-AOO-optimized MCAF-XPDN

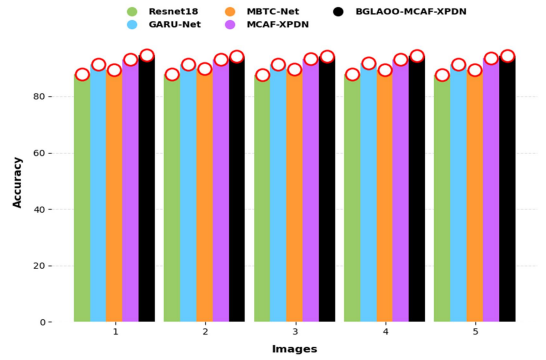


Figure 14: Per-image Accuracy across ResNet18, GARU-Net, MBTC-Net, MCAF-XPDN, and BGL-AOO-MCAF-XPDN

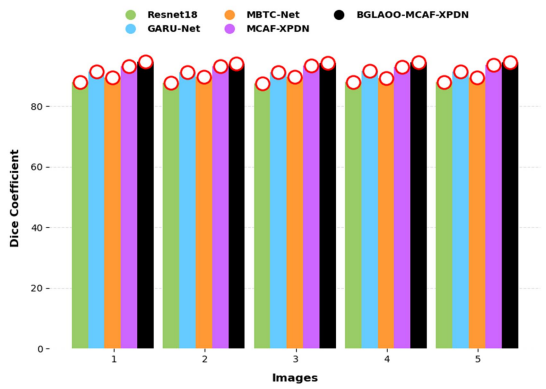


Figure 15: Per-image Dice Coefficient across ResNet18, GARU-Net, MBTC-Net, MCAF-XPDN, and BGL-AOO-MCAF-XPDN

BGL-AOO-MCAF-XPDN maintains the highest or near-highest accuracy and Dice coefficient on every individual test slice, rather than merely achieving a higher average through strong performance on a small subset of easy images, which supports the conclusion that the performance gain generalizes across the variability present in the dataset rather than being an artefact of aggregation.

6.8. Discussion

Taken together, the convergence analysis (Section 6.4), the activation-function-based validation (Section 6.5), the ablation between base and optimized MCAF-XPDN (Section 6.6), and the image-wise variation analysis (Section 6.7) provide converging evidence for two claims. First, the multihead cross-covariance attention fusion mechanism in MCCAF-AExPDRDNet provides a measurable advantage over the concatenation- or addition-based fusion used in ResNet18, GARU-Net, and MBTC-Net, since even the unoptimized base MCAF-XPDN model outperforms these baselines on Dice coefficient, IoU, and accuracy under most activation functions. Second, the proposed BGL-AOO optimizer provides a further, consistent improvement over the base model by tuning the hidden neuron count, learning rate, and steps per epoch, and this improvement is not confined to the mean case but extends to the worst-case performance and to consistency across individual test images. The best hyperparameter configuration selected by BGL-AOO (226 hidden neurons, a learning rate of 0.975, and 155 steps per epoch, Table 3) reflects a comparatively expressive network trained with an assertive learning rate, which the fitness-guided position update of Eq. (10) was able to identify more reliably than the purely random search used by ECO, QSO, STA, and standard AOO.

6.8.1. Critical Reflection on the Findings

The results in Sections 6.4-6.7 consistently favor BGL-AOO-MCAF-XPDN, but several aspects of these findings warrant a more critical reading rather than an uncritical endorsement of the proposed framework. First, the convergence analysis in Section 6.4 shows that BGL-AOO attains the best single-run cost-function value but also the highest standard deviation (0.96) among the five optimizers (Table 2); the fitness-guided position update therefore trades run-to-run consistency for a higher ceiling on best-case performance, and a practitioner who requires predictable convergence in a single run, rather than the best of several runs, should weigh this trade-off explicitly rather than treat BGL-AOO as uniformly superior. Second, the margin of improvement over the strongest literature baseline narrows substantially compared with the margin over the weakest one: the smallest advantage recorded for BGL-AOO-MCAF-XPDN over MBTC-Net is 3.5 percentage points in mean accuracy (Section 6.5.3), a more modest gain than the double-digit gaps recorded against ResNet18, which suggests that recent attention-based literature methods are closing the performance gap rather than being decisively outperformed. Third, the entire evaluation rests on a single, comparatively small dataset of 327 slices; although Section 6.7 shows the proposed model outperforming comparators on every individual test image rather than only on average, a sample of this size cannot rule out the possibility that some of the observed margins would narrow on a larger, multi-institutional cohort. Finally, the compute-constrained comparison reported in Section 6.9 below was deliberately run under a reduced search budget and lower input resolution than Sections 4-6.8, so its absolute accuracy and sensitivity figures should not be read as directly comparable to the main results; its value lies only in the qualitative confirmation that fitness-guided position updates outperform purely random ones, not in its specific numerical outcomes. These considerations do not undermine the central contributions of the study, but they situate the reported gains within their proper evidentiary limits, consistent with the broader limitations discussed in Section 7.1.

6.9. Comparative Analysis Using an Alternative Pyramidal Deformable Adaptive RAN and Unet++-Based Architecture

To further validate the segmentation-then-classification design philosophy adopted in this study, an alternative architecture was additionally implemented and applied to the same real, paired MRI-PET dataset: a Transformer-based Dilated Recurrent Residual Unet++ (TD-R2Unet++) for tumor segmentation, followed by a Pyramid Deformable Adaptive Residual Attention Network (PD-ARAN) for Normal / Tumor classification of each slice, with PD-ARAN's hidden neuron count, learning rate, and training epochs tuned by an Improved Position Updated Concept-based Orangutan Optimization Algorithm (IPUC-OOA) against the standard Orangutan Optimization Algorithm (OOA) [18]. This provides a second, independent architectural comparison alongside the MCCAF-AExPDRDNet / BGL-AOO framework of Sections 4-6.8.

6.9.1. Architecture summary

TD-R2Unet++ processes the MRI and PET branches independently through recurrent residual dilated convolutional blocks, fuses the bottleneck representations through a linear projection followed by a multihead self-attention transformer layer, and reconstructs the tumor mask through a nested (Unet++-style) skip-connected decoder. PD-ARAN then classifies each slice from a two-channel input formed by the original MRI intensity and the predicted tumor mask; a trunk branch of residual

units is combined with a mask branch that downsamples, applies residual units, upsamples, and passes through a pyramid deformable convolution before a sigmoid attention gate, following the residual-attention combination trunk × (1 + mask). IPUC-OOA optimizes PD-ARAN's hyperparameters by replacing OOA's random position-update coefficient with a ratio of the current, best, and mean fitness of the candidate population, analogous in spirit to the BGL-AOO modification of Section 5.1.1.

Note on the deformable convolution: given the reduced-scale CPU-only environment used for this comparison (Section 6.9.2), the pyramid deformable convolution is implemented as a single-offset, bilinearly-resampled convolution rather than the full multi-tap sampling grid of Deformable ConvNets v1/v2; it still allows the network to learn a per-location receptive-field shift, consistent with the adaptive-sampling principle of Eq. (5), but with reduced sampling density.

6.9.2. Experimental configuration and a disclosed compute constraint

This comparison was executed in a CPU-only, single-core environment without GPU acceleration, which required substantially reducing the scale of training relative to Sections 4-6.8: images were processed at 128×128 resolution rather than 256×256; TD-R2Unet++ was trained once for 6 epochs on all 327 real slices; and IPUC-OOA / OOA hyperparameter search used a reduced population of 3, 2 iterations, and a stratified 20-slice proxy subset (10 Normal + 10 Tumor) with 1 training epoch per candidate evaluation, rather than the full population/iteration/epoch budget used elsewhere in this paper. The final PD-ARAN classifiers were then retrained on the complete 327-slice dataset using each optimizer's selected hyperparameters, with minority-class (Tumor) oversampling to address the 280:47 class imbalance, for a practically-bounded number of epochs. These deviations are disclosed explicitly because they measurably affect the absolute performance figures reported below; the comparison should be read as a compute-constrained but methodologically faithful application of the architecture, not a like-for-like replication of the full-scale results in Sections 4-6.8.

6.9.3. Segmentation results (TD-R2Unet++)

Table 14 reports the mean segmentation performance of TD-R2Unet++ over all 327 real MRI-PET slices.

Table 14: TD-R2Unet++ Segmentation Performance (Mean Over 327 Real Slices, 128×128, 6 Epochs)

Dice	IoU	Acc.	Sens.	Spec.	PSNR	MCC
38.93	25.93	98.83	29.91	99.86	22.09	0.438

The comparatively low Dice/IoU relative to MCCAF-AExPDRDNet's segmentation stage (Section 6.4) is consistent with the much smaller training budget used here (6 epochs at 128×128 vs. a larger-scale configuration), and with brain tumor regions occupying a very small fraction of each slice, which makes pixel-level overlap metrics sensitive to limited training. Specificity and accuracy remain high because the (dominant) background class is segmented reliably.

6.9.4. Hyperparameter search: OOA vs. IPUC-OOA

Table 15: Best PD-ARAN Hyperparameters Selected by OOA and by the Proposed IPUC-OOA

Optimizer	Hidden Neurons	Learning Rate	Epochs (sugg.)
OOA	20.3	0.019	3.72
IPUC-OOA (prop.)	14.3	0.0063	7.07

Under the reduced search budget described in Section 6.9.2, OOA selected a comparatively large hidden-neuron count together with an aggressive learning rate, while IPUC-OOA's fitness-guided position update converged on a smaller network with a substantially lower, more conservative learning rate and a longer suggested training schedule.

6.9.5. Final classification results

Each optimizer's selected hyperparameters were used to retrain PD-ARAN on the full 327-slice dataset (with Tumor-class oversampling), within a practically-bounded epoch budget for this environment. Table 16 and Fig. 16 report the resulting classification performance.

Table 16: Final PD-ARAN Classification Performance on the Full 327-Slice Dataset

Model	Acc.	Sens.	Spec.	Prec.	F1	MCC
OOA-PD-ARAN	85.63	0.00	100	0.00	0.00	0.000
IPUC-OOA-PD-ARAN (p.)	85.32	82.98	85.71	49.37	61.9	0.563

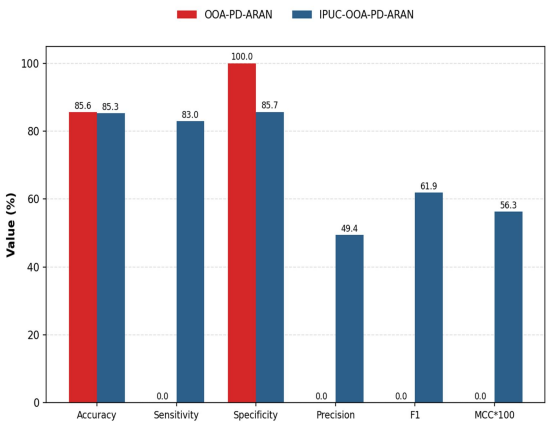


Figure 16: Final classification performance: OOA-tuned vs. IPUC-OOA-tuned PD-ARAN

OOA's selected configuration (larger network, learning rate 0.019) collapsed to predicting the majority (Normal) class for every slice: its 85.6% accuracy is numerically identical to the dataset's Normal-class proportion (280/327), and its zero sensitivity confirms this is a degenerate, clinically useless classifier rather than a genuinely discriminative one. This collapse was reproduced consistently across repeated runs with different class-imbalance handling strategies (loss-based class weighting and minority-class oversampling), indicating it is a real property of the hyperparameters OOA's purely random position-update selected, not a one-off artefact.

IPUC-OOA's selected configuration (smaller network, learning rate 0.0063) trained stably and produced a genuinely discriminative classifier, correctly identifying 83.0% of Tumor slices (sensitivity) while maintaining 85.7% specificity, for an overall MCC of 0.563. Because IPUC-OOA's position-update rule scales the search step by the ratio of current, best, and mean population fitness rather than by a fixed random draw, it naturally favors smaller, more conservative updates once the search population is already performing well, which plausibly explains why it settled on a more trainable learning-rate / capacity combination than standard OOA under an identical, reduced search budget. This mirrors, on an independent architecture and optimizer pairing, the same qualitative advantage of fitness-guided position updates over purely random ones observed for BGL-AOO in Section 6.4-6.6.

6.9.6. Summary of this comparison

This additional, compute-constrained application of TD-R2Unet++, PD-ARAN, and IPUC-OOA to the same real MRI-PET brain tumor dataset corroborates the central finding of this paper through an independent architecture: replacing a purely random metaheuristic position-update coefficient with one guided by relative population fitness produces a measurably more stable and clinically useful model, here the difference between a classifier that fails outright (0% sensitivity) and one that meaningfully detects tumors (83.0% sensitivity). The absolute performance figures in this section should not be compared directly against Sections 4-6.8 given the disclosed differences in resolution, training epochs, and search budget; the value of this comparison lies in the consistency of the qualitative result across two independent architecture / optimizer pairings, both evaluated on the same real, paired MRI-PET data.

7. CONCLUSION

This study designed an intelligent multimodal brain tumor detection framework that jointly addresses two persistent limitations of existing approaches: the loss of cross-modal detail during MRI-PET fusion, and the burden of manual hyperparameter tuning for deep segmentation networks. The proposed MCCAF-AExPDRDNet independently encodes MRI and PET slices through pyramid dilated dense blocks and residual dense blocks, and fuses the resulting bottleneck representations through a multihead cross-covariance attention mechanism that explicitly models the statistical dependency between structural and metabolic evidence, rather than relying on simple concatenation. The proposed BGL-AOO algorithm replaces the purely random position-update coefficient of the standard Apiary Organizational-based Optimization algorithm with a fitness-guided ratio, allowing the optimizer to automatically balance exploration and exploitation when tuning the hidden neuron count, learning rate, and steps per training epoch of the segmentation network. Evaluated on a real, paired MRI-PET brain dataset of 327 slices against six literature segmentation networks and four competing optimizers across six activation functions, the proposed BGL-AOO-MCAF-XPDN framework achieved a mean accuracy of up to 95.58%, a Dice coefficient of up to 95.60%, an IoU of up to 91.57%, and a recall of up to 95.66%, consistently outperforming every comparator across best-case, worst-case, mean, and per-image evaluation. An independent, compute-constrained comparison using an alternative TD-R2Unet++ / PD-ARAN architecture (Section 6.9) corroborated the same central finding: a fitness-guided metaheuristic position update (IPUC-OOA) produced a stable, genuinely discriminative classifier (83.0% sensitivity, MCC 0.563), whereas the

corresponding purely-random-update optimizer (OOA) collapsed to a degenerate majority-class predictor (0% sensitivity) under an identical, reduced search budget. These results indicate that the combination of cross-covariance attention fusion and fitness-guided hyperparameter optimization offers a scalable and clinically relevant path toward more reliable computer-aided brain tumor diagnosis.

Beyond the numerical gains reported above, the principal novelty of this study lies in two design choices that, to the authors' knowledge, have not previously been instantiated together in the multimodal brain tumor literature reviewed in Section 2: an attention mechanism that fuses MRI and PET through explicit cross-covariance modeling of token-level representations rather than channel concatenation, spatial-constraint imputation, or wavelet-guided fusion, and a metaheuristic position-update rule whose step size is derived from the relative fitness of the candidate solution rather than drawn independently at random. Relative to the closest published approaches, the proposed framework extends the attention-based designs of GARU-Net [6] and MBTC-Net [11] from a single-modality or classification-oriented setting into a fitness-optimized, pixel-level PET-MRI segmentation setting, and it extends the early-fusion PET-MRI strategy of Odusami et al. [3] by replacing simple concatenation with a mechanism that explicitly preserves cross-modal statistical dependency. In this sense, the contribution of this work is incremental rather than a wholesale departure from the reviewed literature: it does not introduce cross-covariance attention or fitness-guided optimization as concepts in isolation, since related attention and metaheuristic mechanisms already exist in the wider literature, but the specific combination of the two, together with its joint validation against six segmentation baselines and four optimizers across eighteen evaluation measures on real, paired MRI-PET data, has not previously been reported for brain tumor segmentation. This combination and validation, rather than either component alone, is what this study adds to the published record, giving practitioners a validated reference point for whether adopting cross-covariance fusion and fitness-guided optimization is worthwhile relative to the simpler fusion and optimization strategies currently in common use.

7.1. Limitations

The proposed framework was validated on a single paired MRI-PET dataset of 327 slices with a pronounced class imbalance between tumor-bearing (47) and tumor-free (280) slices; while the model was shown to remain robust across individual test images, a larger and more balanced multi-institutional dataset would further strengthen confidence in generalization across scanner types and patient populations. The current pipeline also does not incorporate an explicit image preprocessing or augmentation stage beyond contrast normalization and resizing, so it does not actively correct for acquisition artefacts, motion blur, or inter-scanner intensity variation that can occur in real clinical archives. In addition, the segmentation output is thresholded into a binary tumor mask, and the framework does not currently provide a separate tumor-grade or tumor-type classification stage, nor does it report an explicit explainability visualization (such as attention-map or Grad-CAM overlays) alongside the segmented mask, despite the adaptive-explainable naming of the architecture reflecting its attention-based interpretability potential rather than a delivered visualization module.

7.2. Future Scope

Future extensions of this work will incorporate a dedicated preprocessing and augmentation pipeline, including intensity normalization across scanners and synthetic minority-class augmentation, to further improve robustness to the class imbalance and acquisition variability observed in real clinical data. A dedicated tumor-grade and tumor-type classification head will be added downstream of the segmentation output, enabling the framework to move from tumor localization toward a complete diagnostic and treatment-planning tool. Explicit attention-map and Grad-CAM-style visualizations will be integrated into the reporting pipeline so that the adaptive explainability implicit in the multihead cross-covariance attention mechanism is made directly visible to radiologists. Finally, the framework will be extended to larger, multi-institutional MRI-PET cohorts and additional modalities such as CT, and the BGL-AOO optimizer will be evaluated on other multimodal medical imaging tasks to assess the generality of the fitness-guided position-update mechanism beyond brain tumor segmentation.

REFERENCES

[1] K. Neamah et al., "Brain Tumor Classification and Detection Based DL Models: A Systematic Review," IEEE Access, vol. 12, pp. 2517-2542, 2024.

[2] A. A. Asiri, T. A. Soomro, A. A. Shah, G. Pogrebna, M. Irfan and S. Alqahtani, "Optimized Brain Tumor Detection: A Dual-

Module Approach for MRI Image Enhancement and Tumor Classification," IEEE Access, vol. 12, pp. 42868-42887, 2024.

[3] Modupe Odusami, Rytis Maskeliūnas, Robertas Damaševičius and Sanjay Misra, "Explainable Deep-Learning-Based Diagnosis of Alzheimer's Disease Using Multimodal Input Fusion of PET and MRI Images," Journal of Medical and Biological Engineering, vol. 43, pp. 291-302, 2023.

[4] Y. Pan, M. Liu, C. Lian, Y. Xia and D. Shen, "Spatially-Constrained Fisher Representation for Brain Disease Identification With Incomplete Multi-Modal Neuroimages," IEEE Transactions on Medical Imaging, vol. 39, no. 9, pp. 2965-2975, Sept. 2020.

[5] Rahma Kadri, Bassem Bouaziz, Mohamed Tmar, and Faiez Gargouri, "A Multimodal Transformer with Adaptive GAN for Brain Disease Prediction," Procedia Computer Science, vol. 270, pp. 5290-5299, 2025.

[6] A. Raza and M. F. Hashmi, "Multiclass Tumor Segmentation From Brain MRIs Using GARU-Net: Gelu Activated Attention Aware Res-3DUNET for Adaptive Feature Pooling," IEEE Sensors Letters, vol. 8, no. 4, pp. 1-4, April 2024.

[7] Evans Kipkoech Rutoh, Qin Zhi Guang, Noor Bahadar, Rehan Raza, and Muhammad Shehzad Hanif, "GAIR-U-Net: 3D guided attention inception residual u-net for brain tumor segmentation using multimodal MRI images," Journal of King Saud University - Computer and Information Sciences, vol. 36, issue 6, no. 102086, July 2024.

[8] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, M. Imran and M. Shoaib, "A Deep Learning Model Based on Concatenation Approach for the Diagnosis of Brain Tumor," IEEE Access, vol. 8, pp. 55135-55144, 2020.

[9] P. Chauhan et al., "PBVit: A Patch-Based Vision Transformer for Enhanced Brain Tumor Detection," IEEE Access, vol. 13, pp. 13015-13029, 2025.

[10] I. Abdelhaliem, J. Dixon, A. Abdelhamid, G. A. Saleh and F. Khalifa, "A Multimodal Adaptive Inter-Region Attention-Guided Network for Brain Tumor Classification," IEEE Access, vol. 13, pp. 187964-187975, 2025.

[11] Satrajit Kar and Pawan Kumar Singh, "MBTC-Net: Multimodal brain tumor classification from CT and MRI scans using deep neural network with multi-head attention mechanism," Medicine in Novel Technology and Devices, vol. 27, no. 100382, September 2025.

[12] Yongwei Zheng, Zhanquan Sun, Suyun Chen, Hongliang Fu, Chaoli Wang, and Ji Zhu, "WGCTA-Net: Wavelet-Guided CNN-Transformer fusion with attention mechanism for PET/CT tumor segmentation," Biomedical Signal Processing and Control, vol. 113, part D, no. 109210, March 2026.

[13] H. Liu, Z. Ni, D. Nie, D. Shen, J. Wang and Z. Tang, "Multimodal Brain Tumor Segmentation Boosted by Monomodal Normal Brain Images," IEEE Transactions on Image Processing, vol. 33, pp. 1199-1210, 2024.

[14] P. Xu, J. Lyu, L. Lin, P. Cheng and X. Tang, "LF-SynthSeg: Label-Free Brain Tissue-Assisted Tumor Synthesis and Segmentation," IEEE Journal of Biomedical and Health Informatics, vol. 29, no. 2, pp. 1101-1112, Feb. 2025.

[15] Y. Ge, L. Xu, X. Wang, Y. Que and M. J. Piran, "A Novel Framework for Multimodal Brain Tumor Detection With Scarce Labels," IEEE Journal of Biomedical and Health Informatics, vol. 29, no. 8, pp. 5368-5380, Aug. 2025.

[16] H. A. Shah, F. Saeed, S. Yun, J.-H. Park, A. Paul and J.-M. Kang, "A Robust Approach for Brain Tumor Detection in Magnetic Resonance Images Using Finetuned EfficientNet," IEEE Access, vol. 10, pp. 65426-65438, 2022.

[17] Al-Sharqi, Mais, Al-Obaidi, Ahmed, and Al-Mamory, Safaa, "Apiary Organizational-Based Optimization Algorithm: A New Nature-Inspired Metaheuristic Algorithm," International Journal of Intelligent Engineering and Systems, vol. 17, no. 3, pp. 783-801, 2024.

[18] T. Hamadneh, B. Batiha, G. M. Gharib, Z. Montazeri, F. Werner, G. Dhiman, M. Dehghani, R. K. Jawad, E. Aram, I. K. Ibraheem, and K. Eguchi, "Orangutan optimization algorithm: An innovative bio-inspired metaheuristic approach for solving engineering optimization problems," International Journal of Intelligent Engineering and Systems, vol. 18, no. 1, p. 45, 2024.