

Article

ETDACVO: Structural-Fidelity-Aware Evolutionary Co-Optimization for Robust and Explainable Brain Tumor MRI Classification

Indrakumar Krishnamurthy ¹, Ravikumar Manjunath ¹, Mohammed A. S. Al-Mohamadi ¹, Lubna A. Gabralla ², Sami F. Karali ³, Mohammed I. Thanoon ³, Abed Saif Ahmed Alghawli ⁴ and Abdulbasit A. Darem ^{5,*}

¹ Department of MCA and Computer Science, Jnanasahyadri, Kuvempu University, Shivamogga 577 451, India; indk214@gmail.com (I.K.); ravi2142@yahoo.co.in (R.M.); almohmdy30@gmail.com (M.A.S.A.-M.)

² Department of Computer Sciences, Applied College, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia; lagabralla@pnu.edu.sa

³ Computer Science Department, College of Engineering and Computers, Umm Al-Qura University, Al-Lith 21961, Saudi Arabia; sfkarali@uqu.edu.sa (S.F.K.); mithanoon@uqu.edu.sa (M.I.T.)

⁴ Computer Science Department, College of Sciences and Humanities, Prince Sattam Bin Abdulaziz University, Aflaj 16278, Saudi Arabia; a.alghawli@psau.edu.sa

⁵ Center for Scientific Research and Entrepreneurship, Northern Border University, Arar 73213, Saudi Arabia

* Correspondence: basit.darem@nbu.edu.sa

Abstract

Background/Objectives: Heterogeneous imaging protocols, a lack of labeled data, and domain shifts continue to make training deep learning models to analyze medical images a challenge. This study presents ETDACVO (Enhanced Tasmanian Devil Anti-Conservative Variable Optimization), a hybrid evolutionary optimization system designed to improve convergence stability and cross-domain robustness in brain tumor MRI classification. **Methods:** ETDACVO combines Tasmanian Devil Optimization (TDO), Anti-Conservative Variable Optimization (ACVO), and Exponentially Weighted Moving Average (EWMA) smoothing to stabilize evolutionary parameter updates. Unlike existing approaches that optimize augmentation policies or optimizer dynamics separately, ETDACVO simultaneously evolves both components within a single evolutionary loop. The framework was evaluated on four MRI datasets (Nickparvar, Mendeley, BRISC, and Figshare), comprising 28,151 images. In addition, a convergence-aware explainability mechanism, CA-EA-GradCAM, was developed by integrating gradient saliency, transformer attention, and evolutionary convergence confidence to generate confidence-sensitive tumor localization maps. **Results:** Experimental results demonstrated that ETDACVO achieved a 2.3–2.5% improvement in classification accuracy and converged 19–22 epochs faster than baseline optimizers. The statistical significance of these improvements was confirmed using paired statistical tests ($p < 1 \times 10^{-5}$). Cross-dataset transfer experiments further showed strong domain-shift resilience, with performance retention reaching 92.8%. The proposed CA-EA-GradCAM mechanism provided interpretable and confidence-aware tumor localization maps. **Conclusions:** ETDACVO provides a robust and computationally efficient optimization framework for deep-learning-based medical image analysis. By jointly optimizing augmentation strategies and optimizer dynamics, the framework enhances convergence stability, cross-domain robustness, and interpretability, making it a promising approach for reliable brain tumor MRI classification under heterogeneous imaging conditions.

Academic Editor: Akira Hara

Received: 24 May 2026

Revised: 23 June 2026

Accepted: 26 June 2026

Published: 29 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: adaptive optimization; evolutionary algorithms; medical image learning; brain tumor MRI; domain generalization; explainable AI; hybrid CNN–transformer models; ETDACVO

1. Introduction

Medical image learning system optimization strategies are essential in enhancing the performance and reliability of the learning systems, especially in the case of heterogeneous and domain-shifted datasets. Magnetic resonance imaging (MRI) data usually vary substantially among scanners, acquisition protocols, and patients. Moreover, sparse-labeled data and class imbalance often decrease the strength of deep neural networks. This complicates the process of achieving stable optimization and robust generalization of clinical reliable diagnostic models. The majority of current medical imaging pipelines use fixed or manually chosen data-augmentation policies and manually constructed learning-rate schedules, usually informed by natural image benchmarks but not by medical constraints [1–4]. Nevertheless, heuristic training methods are unable to change to non-stationary medical data distributions and tend to converge more slowly, overfit, and lack cross-domain robustness when deployed in different clinical settings. Adaptive optimization has thus been given growing interest as a method to dynamically trade off exploration and exploitation during model training. Evolutionary-inspired algorithms offer a dynamic way of changing hyperparameters, learning rates, and augmentation policies based on model performance. Most methods, however, maximize either augmentation policies or optimizer dynamics alone. Despite a few studies having been performed on adaptive augmentation or meta-optimization individually, coherent frameworks that jointly govern the dynamics of learning-rate and augmentation intensity within the same evolutionary cycle are still scarce in medical imaging settings [1–7].

In order to overcome this shortcoming, we propose the Enhanced Tasmanian Devil Anti-Conservative Variable Optimization (ETDACVO) algorithm, which is a hybrid evolutionary optimization framework that simultaneously controls the augmentation parameters and learning-rate schedules. ETDACVO combines three complementary mechanisms, i.e.,: Tasmanian Devil Optimization (TDO) to explore the world by using stochastic predator–prey search dynamics; Anti-Conservative Variable Optimization (ACVO) to ensure population diversity and avoid premature convergence; and Exponentially Weighted Moving Average (EWMA) smoothing to stabilize parameter trajectories in non-stationary training. ETDACVO adapts dynamically to augmentation parameters, optimizer variables (learning rate, momentum, and weight decay), and maintains structural fidelity in medical images. In addition, the proposed framework uses a reliability-conscious explainability that is dubbed as Convergence-Aware Evolutionary Attention Grad-CAM (CA-EA-GradCAM). CA-EA-GradCAM is based on gradient signals in contrast to other conventional post hoc explainability frameworks that do not incorporate optimization stability information based on the evolutionary trajectory, resulting in trust-calibrated tumor localization heatmaps. To measure the efficacy of the proposed method, ETDACVO was measured on four heterogeneous MRI sets with different acquisition protocols and anatomy. These results indicate that ETDACVO offers a sound optimization model to adaptive medical image learning systems. In contrast to the current state of development of evolutionary optimization methods, which can be either tuned to hyperparameters or search augmentation policy, the proposed ETDACVO framework optimizes data augmentation parameters and optimizer dynamics simultaneously and imposes anatomical fidelity using SSIM-based constraints. This co-optimization approach enables the model to vary the

training data distribution and the learning path together, unlike current augmentation search or optimizer tuning strategies. This research has the following key contributions:

1. **Unified Evolutionary Optimization Framework:** We present ETDACVO, a hybrid evolutionary algorithm that can optimize both data augmentation policies and learning-rate dynamics within a single optimization loop to learn to classify brain tumor MRI images.
2. **Structural-Fidelity-Aware Fitness Function:** A new fitness assessment approach that uses structural similarity (SSIM) to ensure that automatically generated augmentations do not disrupt clinically significant anatomical structures.
3. **Hybrid Evolutionary Update Mechanism:** A mathematically consistent update rule that combines TDO exploration, ACVO diversity preservation, and EWMA temporal smoothing to enhance optimization stability.
4. **Better Training Stability and Generalization:** Large-scale experiments on four heterogeneous MRI datasets show better classification accuracy (2.32.5%), quicker convergence (1922 fewer epochs), and less performance variability than with traditional optimization strategies.
5. **Reliability-Aware Explainability:** We present CA-EA-GradCAM, a convergence-conscious explainability model that combines CNN gradient localization, transformer attention context, and evolutionary optimization stability to generate trust-calibrated tumor localization maps.

ETDACVO offers a cohesive evolutionary model compared with current optimization models that co-regulates training data distribution and optimization dynamics, retaining anatomical fidelity. These contributions, combined, form a single evolutionary optimization framework that enhances the stability of training, the generalization ability, and the interpretability of deep-learning-based brain tumor MRI analysis. The rest of this paper is structured in the following way. Section 1.1 provides a review of related literature in adaptive optimization and explainable medical imaging models. Section 2 introduces the suggested ETDACVO framework and CA-EA-GradCAM explainability module. Section 3 contains the description of the experimental setup and evaluation results. Section 4 addresses implications and limitations of the proposed approach, and Section 5 ends the research and suggests future research directions.

1.1. Related Work

1.1.1. Deep Learning for Brain Tumor MRI Classification

Automated brain tumor detection and classification based on magnetic resonance imaging (MRI) has been greatly enhanced with the help of deep learning methods. Initial methods were mainly based on convolutional neural networks (CNNs) to obtain hierarchical information in MRI images. Indicatively, Ismael et al. [1] used a residual network architecture, which enhanced the performance of tumor classification by extracting deep features. In the same vein, Mallick et al. [2] proposed a deep wavelet autoencoder-based neural network to improve the representation of features to classify brain tumors. Advanced architectures to enhance diagnostic performance have been studied more recently. Ahmed et al. [3] suggested a hybrid Vision Transformer (ViT) and gated recurrent unit (GRU) model that combined the explainable AI methods with MRI-based tumor classification. Transformer-based architectures have also been explored in medical image analysis, such as TransMed [8] and MedViT [9], that use self-attention mechanisms to learn long-range spatial dependencies in medical images. Hybrid CNN–transformer networks have also been shown to be very effective in MRI tumor classification, where local convolutional information is used with global contextual information [10]. Though these have been achieved, most of the current models use fixed training settings and optimization

schemes that are manually designed, and this can be too restrictive when used with heterogeneous MRI datasets.

Recent studies have highlighted the importance of structure-aware representation learning and feature fusion in medical image analysis. Liu et al. [11] proposed a structure-aware, context-enhanced, and dual-path synergistic decoding network for atrophic gastritis segmentation, demonstrating that contextual modeling and structural information can substantially improve segmentation accuracy. Similarly, Li et al. [12] introduced a deep feature fusion framework for acute ischemic stroke lesion segmentation that effectively integrates complementary feature representations to improve lesion delineation performance. These studies emphasize the importance of preserving anatomical structures and exploiting multi-scale feature information in medical imaging. While these approaches primarily focus on segmentation network design and feature representation, the proposed ETDACVO framework addresses a complementary problem by optimizing the training process itself through joint evolutionary adaptation of augmentation policies and optimizer dynamics while enforcing structural-fidelity constraints.

1.1.2. Medical Image Data Augmentation

The generalization ability of deep learning models trained on small medical datasets is commonly enhanced by data augmentation. Sajjad et al. [13] have shown that the extensive augmentation strategies are effective to enhance the brain tumor classification using CNNs as a result of augmenting the training data diversity. To maintain structural consistency across MRI images but add meaningful variability, anatomically informed augmentation techniques have also been suggested [6]. Other more recent methods have explored automated augmentation methods based on generative models and optimization-based methods. Tronchin et al. [4] suggested the creation of augmented samples, LatentAugment, which employs the manipulation of the latent space of generative adversarial networks in a controlled manner. In the same manner, Park et al. [14] used GAN-based augmentation to increase training diversity used in medical image classification tasks. Despite enhancing the diversity of datasets, the majority of augmentation approaches are independent of the optimization procedure of the model and do not directly introduce the structural constraints needed to ensure anatomical consistency in medical images.

1.1.3. Optimization Strategies for Medical Image Learning

The learning of medical images can be optimized in the following ways. In the training of deep learning models to analyze medical images, optimization algorithms are important. Conventional pipelines of training are usually based on gradient-based optimizers as well as manually crafted learning-rate schedules. A number of studies have examined the adaptive optimization methods to enhance the stability of training and convergence rate. Aytaç et al. [15] suggested an adaptive momentum-based optimizer of CNN-based medical image classification, and Sun et al. [16] suggested a better Adam-based optimization strategy to increase the convergence of the model. In addition to gradient-based methods, evolutionary and meta-optimization techniques have also been considered in the analysis of medical images. Darwish et al. [17] suggested a quantum-inspired evolutionary system of automated detection of brain tumors in MRI images. Equally, Bayesian optimization has been utilized to optimize CNN hyperparameters to tumor classification problems [18]. Although the current methods are effective at enhancing specific factors of model training, the majority of studies optimize hyperparameters or augmentation policies separately. AutoAugment methods are automated methods of augmentation that are trained using reinforcement learning, but they are computationally expensive, and

RandAugment methods are easier to search but use fixed augmentation schedules. Population-Based Training (PBT) optimizes hyperparameters by applying population resampling to the optimizer but addresses the optimizer parameters (instead of augmentation fidelity) as its main objective. Conversely, the suggested ETDACVO framework co-optimizes the augmentation parameters and the optimizer dynamics in a single evolutionary process and implements structural fidelity by imposing SSIM-based constraints. In contrast to PBT, which only changes optimizer parameters, ETDACVO balances the training path and the input data distribution. This collective adaptation enhances the stability and robustness of optimization to heterogeneous MRI datasets.

Recent advances in evolutionary optimization have also focused on high-dimensional feature selection problems. For example, a modified Gray Wolf Optimization (GWO) algorithm was proposed to improve feature selection performance in high-dimensional search spaces by enhancing exploration and exploitation balance [19]. The study demonstrated that population-based optimization strategies can effectively handle complex optimization landscapes and improve model performance. However, such methods primarily focus on feature subset selection, whereas the proposed ETDACVO framework addresses a different problem by jointly optimizing augmentation policies and optimizer dynamics during deep learning training while incorporating structural-fidelity constraints for medical image analysis.

1.1.4. Explainable Artificial Intelligence in Medical Imaging.

Explainability has emerged as a key condition of clinical adoption of deep learning models in medical imaging. Gradient-based visualization methods like Grad-CAM [5] are popularly employed to indicate areas of interest that affect model predictions to enhance model transparency and model interpretability. Later studies have applied the explainability approach to tumor study using MRI. Zeineldin et al. [20] explored explainable deep learning models to analyze brain tumor MRI and showed the significance of visual explanations in supporting diagnostic forecasts. Most explainability methods, however, are post hoc visualization methods that are based on model parameter settings that are fixed and do not include information about the optimization dynamics that produced those parameter settings.

1.1.5. Research Gap

Despite the fact that much has been achieved in deep-learning-based brain tumor MRI analysis, there are still a number of limitations. Current methods tend to assume that the data augmentation, model optimization, and explainability are separate stages of the training pipeline. Stable training, diversity of datasets, and interpretability are key research problems, and the creation of integrated frameworks capable of enhancing all three aspects of medical imaging is an essential issue, as has been emphasized in recent surveys of deep learning in medical imaging [21,22]. We propose the Enhanced Tasmanian Devil Anti-Conservative Variable Optimization (ETDACVO) to close this gap. The suggested method brings about a combined evolutionary optimization process that collectively modifies data augmentation policies and optimizer dynamics throughout the training. Besides, the framework also involves a Convergence-Aware Evolutionary Attention Grad-CAM (CA-EA-GradCAM) scheme that combines optimization stability with gradient-based explanations to produce credible tumor localization maps. This single design is intended to enhance the stability of convergence, cross-dataset generalization, and interpretability of deep-learning-based brain tumor MRI analysis.

2. Materials and Methods

The pipeline shown in Figure 1 begins with MRI preprocessing and standardization, followed by the ETDACVO evolutionary optimization framework that integrates Tasmanian Devil Optimization (TDO), Anti-Conservative Variable Optimization (ACVO), and EWMA smoothing to adapt augmentation parameters and learning-rate dynamics. The optimized parameters guide hybrid CNN–Transformer training. The framework further incorporates CA-EA-GradCAM to generate reliability-aware tumor localization heatmaps, while performance is evaluated using classification accuracy, optimization stability, and cross-dataset generalization.

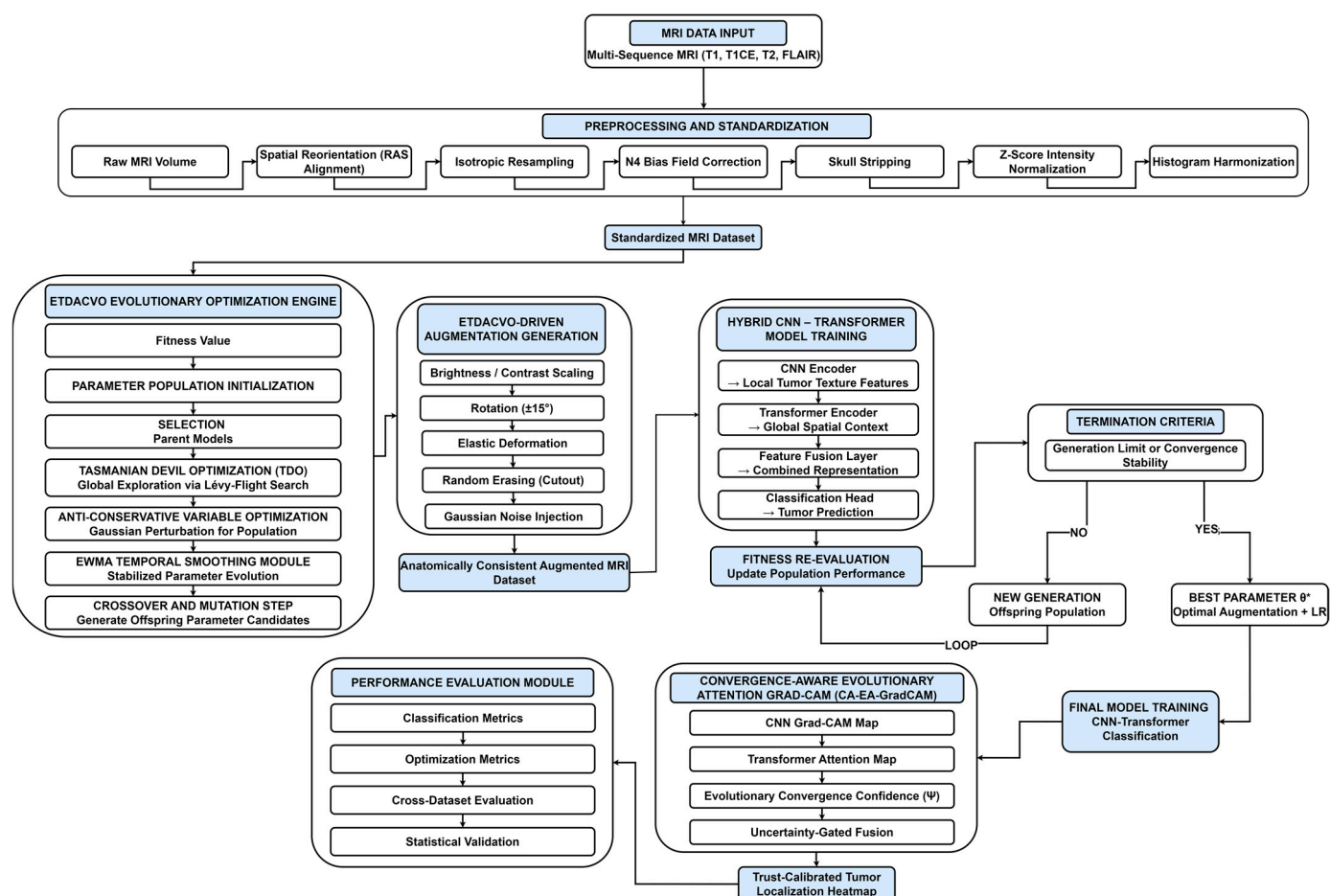


Figure 1. Overall architecture of the proposed ETDACVO framework.

2.1. Problem Formulation

Brain tumor MRI classification can be formulated as a supervised learning problem defined over a dataset:

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

where x_i represents an MRI image and $y_i \in \{1, \dots, C\}$ denotes the corresponding tumor class label. The objective is to learn a classifier $f_\theta(x)$ parameterized by θ that minimizes the expected classification loss across the dataset.

Conventional training procedures optimize network parameters using gradient-based optimizers with fixed augmentation policies. However, in heterogeneous medical datasets, both the training data distribution and optimization dynamics significantly influence model convergence and generalization. Therefore, this study formulates training

as a joint optimization problem where augmentation parameters and optimizer hyperparameters are simultaneously adapted during training.

Let $\theta = [\theta_{\text{aug}}, \theta_{\text{opt}}]$ denote the combined parameter vector controlling augmentation policies and optimizer dynamics. The goal is to find the optimal parameter vector θ^* that minimizes the composite training objective:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} F(\theta) \quad (2)$$

where $F(\theta)$ represents the structural-fidelity-aware loss defined in Section 2.4. The proposed ETDACVO framework performs this joint optimization using a hybrid evolutionary search strategy.

2.2. Preprocessing and Input Standardization

MRI volumes were standardized using a preprocessing pipeline consisting of spatial reorientation to the RAS coordinate system, isotropic resampling, N4 bias-field correction, skull stripping, z-score intensity normalization, and histogram harmonization. These steps reduce scanner-induced intensity variations across datasets. The resulting standardized volumes were used as input for the ETDACVO optimization framework.

2.3. Hybrid CNN–Transformer Backbone

The classification backbone used in this study consists of a hybrid CNN–Transformer architecture designed to capture both local spatial features and global contextual dependencies in MRI data. The CNN encoder extracts hierarchical feature maps from the input MRI slices, capturing local anatomical patterns and tumor texture information. These features are then projected into token representations and processed by a lightweight transformer encoder that models long-range spatial relationships across brain regions using self-attention mechanisms. The final classification head aggregates the fused CNN–Transformer features to predict tumor class probabilities.

For reproducibility, the hybrid CNN–Transformer backbone follows a lightweight architecture designed for medical image classification. The convolutional feature extractor is based on a ResNet-34 backbone that processes MRI slices resized to 224×224 pixels. The CNN encoder produces a $7 \times 7 \times 512$ feature map, which is flattened and projected into a sequence of token embeddings using a linear projection layer with embedding dimension 256. These tokens are processed by a transformer encoder consisting of four self-attention layers with eight attention heads per layer. Each transformer block includes multi-head self-attention followed by a feed-forward network with hidden dimension 1024, residual connections, layer normalization, and dropout regularization ($p = 0.1$). The final token representations are aggregated using global average pooling and passed to a fully connected softmax classifier that predicts the tumor category. This architecture enables the CNN component to capture local anatomical structures while the transformer encoder models long-range spatial dependencies across brain regions.

2.4. ETDACVO: Evolutionary Co-Optimization Framework

The ETDACVO framework is designed to simultaneously optimize augmentation parameters and learning-rate dynamics through a hybrid evolutionary optimization strategy. Let $\theta_t \in \mathbb{R}^d$ denote the parameter vector at iteration t , where d represents the dimensionality of the parameter space. Let P denote the evolutionary population size, representing the number of candidate parameter vectors simultaneously evolved during the optimization process. The vector θ_t contains both augmentation-control parameters and optimizer-control parameters:

$$d = d_{\text{aug}} + d_{\text{opt}} \quad (3)$$

The ETDACVO update rule combines three complementary mechanisms: Tasmanian Devil Optimization (TDO), Anti-Conservative Variable Optimization (ACVO), and Exponentially Weighted Moving Average (EWMA) smoothing.

The unified parameter update rule is defined as:

$$\theta_{t+1} = \theta_t + \Lambda_1 \alpha_1 T_t + \Lambda_2 \alpha_2 A_t + \Lambda_3 \beta S_t \quad (4)$$

where T_t represents the exploration step generated by TDO, A_t represents the diversity perturbation from ACVO, S_t denotes the EWMA-smoothed parameter vector, and $\alpha_1, \alpha_2, \beta$ control the contribution of each mechanism.

Where Λ_1, Λ_2 and Λ_3 denote scalar coefficients controlling the relative influence of the TDO exploration step, ACVO diversity perturbation, and EWMA temporal smoothing components, respectively. Specifically, the coefficients were set to $\Lambda_1 = 1.0$, $\Lambda_2 = 1.0$, and $\Lambda_3 = 1.0$, ensuring that the contribution of each mechanism is governed primarily by the corresponding control parameters α_1 , α_2 , and β .

It should be noted that the parameter vector θ optimized by ETDACVO does not correspond to the trainable weights of the CNN–Transformer classifier. Instead, θ represents a set of augmentation-control and optimizer-control parameters, including learning rate, momentum, weight decay, augmentation intensity, and transformation settings. The objective of ETDACVO is to identify an optimal parameter configuration θ^* that maximizes classification performance while preserving structural fidelity. The resulting θ^* is subsequently used to guide the final training of the CNN–Transformer model through conventional gradient-based optimization.

2.4.1. Tasmanian Devil Optimization (TDO)

The TDO mechanism performs global exploration using Lévy-flight dynamics:

$$u_t = \alpha \cdot \text{Levy}(\lambda) \quad (5)$$

where u_t denotes the exploration step, α controls step magnitude, $\lambda \in (1, 3]$ represents the Lévy exponent, and $\text{Levy}(\lambda)$ denotes a Lévy probability distribution characterized by heavy-tailed step lengths, enabling occasional long jumps in the search space and improving the exploration capability of the optimizer.

The Lévy-flight step length s is sampled from a heavy-tailed Lévy distribution whose probability density function is given by:

$$p(s) = \sqrt{\frac{\lambda}{2\pi}} \frac{\exp\left(-\frac{\lambda}{2s}\right)}{s^{\frac{3}{2}}}, s > 0 \quad (6)$$

where λ denotes the scale parameter controlling the spread of the distribution. The heavy-tailed nature of the Lévy distribution generates predominantly short exploratory steps together with occasional long-distance jumps, enabling efficient exploration of high-dimensional optimization landscapes and reducing the likelihood of premature convergence to local optima.

The parameter update from TDO is:

$$\theta_t^{\text{TDO}} = \theta_t + \Lambda_1 u_t \quad (7)$$

This stochastic exploration allows the algorithm to escape local minima in high-dimensional optimization landscapes.

Although Equation (7) expresses the stochastic Lévy-flight update, candidate solutions are evaluated using the fitness function and retained through fitness-based selection and elitist retention mechanism across generations.

The additive formulation in Equation (4) reflects the complementary roles of exploration, diversity preservation, and temporal stabilization in evolutionary optimization. Each component contributes independently to the update dynamics: the TDO term T_t performs global exploration through stochastic Lévy-flight search, the ACVO term A_t injects controlled diversity to prevent premature convergence, and the EWMA term S_t stabilizes the parameter trajectory by smoothing temporal variations.

Combining these mechanisms through an additive update enables the optimizer to balance exploration and exploitation while maintaining stable parameter evolution. The scalar coefficients Λ_1 , Λ_2 , and Λ_3 were set to 1.0 to avoid introducing additional weighting bias and to ensure that the relative influence of each mechanism is governed primarily by the adaptive control parameters α_1 , α_2 , and β . The values $\alpha_1 = 1.6$ and $\alpha_2 = 0.9$ were selected based on preliminary grid-search experiments that balance exploration strength and population diversity, while the smoothing coefficient $\beta = 0.3$ provides stable temporal averaging without excessively slowing convergence. The robustness of these parameter choices is further supported by the sensitivity analysis presented in Section 3.3, which shows that moderate variations ($\pm 20\%$) produce only minor performance fluctuations ($< 0.5\%$).

TDO was selected because its Lévy-flight exploration mechanism enables efficient traversal of high-dimensional optimization landscapes while requiring fewer interacting control parameters than conventional swarm-based methods such as PSO. Unlike PSO, which maintains particle velocities and multiple acceleration coefficients, TDO relies on stochastic exploratory movements that naturally balance global search and local refinement. This property is particularly beneficial in heterogeneous MRI learning environments where multiple local optima may arise due to variations in imaging distributions.

2.4.2. Anti-Conservative Variable Optimization (ACVO)

ACVO introduces controlled diversity into the parameter population:

$$v_t = \gamma R \epsilon_t \quad (8)$$

where R denotes the correlation matrix between augmentation parameters. The matrix $R \in \mathbb{R}^{d_{\text{aug}} \times d_{\text{aug}}}$ represents the empirical correlation between augmentation parameters and is estimated from the covariance structure of the current population. $\epsilon_t \sim \mathcal{N}(0, \sigma^2 I)$ represents Gaussian noise, and γ controls perturbation strength. The ACVO update is:

$$\theta_t^{\text{ACVO}} = \theta_t + \Lambda_2 v_t \quad (9)$$

This diversity injection prevents premature convergence during evolutionary optimization.

Although diversity-preserving mechanisms improve exploration capability, excessive diversity may delay convergence or introduce optimization instability. To address this issue, ACVO is coupled with EWMA temporal smoothing and fitness-based selection. The smoothing mechanism reduces abrupt parameter fluctuations, while selection pressure ensures that only high-quality candidate solutions are propagated. Consequently, ETDACVO maintains sufficient diversity to avoid premature convergence while preserving stable optimization trajectories.

2.4.3. EWMA Temporal Smoothing

To stabilize parameter evolution, temporal smoothing is applied using EWMA:

$$\phi_t = \beta\phi_{t-1} + (1 - \beta)\theta_t \quad (10)$$

where $\beta \in [0,1)$ controls smoothing strength. The smoothed parameter update is:

$$\theta_t^{\text{EWMA}} = \theta_t + \Lambda_3(\phi_t - \theta_t) \quad (11)$$

This mechanism reduces abrupt parameter changes and improves convergence stability.

Unlike conventional evolutionary optimizers that operate independently of the deep learning training process, the proposed ETDACVO framework introduces a co-optimization paradigm in which augmentation policies and optimizer dynamics have evolved simultaneously within the same parameter space. Traditional approaches such as evolutionary gradient search or augmentation policy search typically optimize a single component of the training pipeline, either augmentation strategies or learning-rate schedules. In contrast, ETDACVO integrates exploration-driven search (TDO), diversity preservation (ACVO), and temporal stabilization (EWMA) within a unified update mechanism. This hybrid formulation enables the algorithm to jointly regulate data transformation intensity and optimization dynamics while enforcing structural-fidelity constraints. Consequently, ETDACVO forms a closed-loop optimization system that adapts both the input data distribution and the model training trajectory, improving convergence stability and cross-domain generalization.

2.4.4. Optimization Objective: Structural Fidelity

To ensure anatomical fidelity, ETDACVO minimizes a composite fitness function that combines classification loss and structural similarity constraints:

$$F(\theta) = \mathcal{L}_{\text{CE}}(I_{\text{aug}}, y) + \lambda_{\text{SSIM}}(1 - \text{SSIM}(I, \hat{I})) \quad (12)$$

where $\mathcal{L}_{\text{CE}}$ denotes the cross-entropy classification loss between the augmented image I_{aug} and its ground-truth label y . The structural similarity term $\text{SSIM}(I, \hat{I})$ measures the preservation of structural image characteristics, including local intensity patterns, contrast relationships, and anatomical morphology, between the original MRI image I and the augmented image $\hat{I}$. Although SSIM does not directly quantify clinical features or diagnostic biomarkers, it serves as a surrogate measure of anatomical structure preservation. Therefore, the SSIM constraint helps prevent excessive structural distortions during augmentation while maintaining meaningful variability for model training.

2.5. ETDACVO-Driven Augmentation

Using the optimized parameter vector θ^* , the augmentation operator $\mathcal{A}_\theta$ generates anatomically consistent MRI variants. The augmented image is defined as:

$$I_{\text{aug}} = \mathcal{A}_B \circ \mathcal{A}_E \circ \mathcal{A}_D(I_{\text{orig}}) \quad (13)$$

where $\mathcal{A}_B$ represents brightness and contrast adjustment (scaled within $\pm 20\%$), $\mathcal{A}_E$ represents random erasing (cutout patches 8–32 px), and $\mathcal{A}_D$ denotes domain-adaptive geometric transformations (rotation $\pm 15^\circ$, elastic deformation).

These transformations simulate scanner variability and anatomical heterogeneity while preserving anatomical structure.

2.6. Convergence-Aware Evolutionary Attention Grad-CAM (CA-EA-GradCAM)

CA-EA-GradCAM integrates three complementary signals to generate trust-calibrated tumor localization maps. First, CNN gradient localization identifies class-discriminative spatial regions using gradient-based saliency. Second, transformer attention maps capture long-range contextual dependencies across anatomical structures. Third, evolutionary convergence confidence derived from the ETDACVO optimization trajectory quantifies the stability of the learned parameters during training. By combining these three signals through an uncertainty-gated fusion mechanism, CA-EA-GradCAM produces explanations that simultaneously reflect local discriminative evidence, global contextual relationships, and optimization reliability. This design differs from conventional Grad-CAM variants, which rely solely on gradient signals without incorporating attention context or training stability information.

2.6.1. Class-Discriminative Gradient Localization

Let $A^k \in \mathbb{R}^{H \times W}$ denote the k -th feature map extracted from the CNN encoder. The importance weight for class c is computed as:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (14)$$

where y^c represents the class score and $Z = H \times W$ is the spatial normalization factor. The CNN saliency map is defined as:

$$M_{\text{CNN}} = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right) \quad (15)$$

2.6.2. Global Contextual Attention Manifold

Transformer attention maps capture long-range anatomical relationships:

$$A_T = \text{Softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) \quad (16)$$

The aggregated contextual attention representation is:

$$M_{\text{Trans}} = \frac{1}{L} \sum_{l=1}^L A_T^{(l)} \quad (17)$$

where L denotes the number of transformer layers.

2.6.3. Evolutionary Convergence Confidence (ECC)

To quantify optimization stability, the ETDACVO parameter trajectory is defined as $\theta = \{\theta_1, \theta_2, \dots, \theta_T\}$. The stability weight for generation t is:

$$\omega_t = \exp(-\beta \|\theta_t - \theta_{t-1}\|^2) \quad (18)$$

Smaller parameter variations between successive generations produce larger stability weights ω_t , indicating convergence toward a stable region of the loss landscape. The trajectory stability index is:

$$\Psi = \frac{1}{N} \sum_{t=T-N}^T \omega_t \quad (19)$$

Higher values of Ψ indicate stronger convergence stability.

2.6.4. Uncertainty-Gated Hybrid Fusion

To combine CNN and transformer explanations, spatial entropy of the CNN saliency map is computed:

$$E_{\text{CNN}} = - \sum_{i,j} p_{ij} \log(p_{ij}) \quad (20)$$

where

$$p_{ij} = \frac{M_{\text{CNN}}(i, j)}{\sum_{i,j} M_{\text{CNN}}(i, j)} \quad (21)$$

represents the normalized activation probability at spatial location (i, j) . The final CA-EA-GradCAM heatmap is defined as:

$$H_{\text{EA}} = \Psi[\sigma(E_{\text{CNN}}) \odot M_{\text{CNN}} + (1 - \sigma(E_{\text{CNN}})) \odot M_{\text{Trans}}] \quad (22)$$

where $\sigma(\cdot)$ represents the sigmoid gating function and $\odot$ denotes element-wise multiplication.

2.7. Theoretical Insight into ETDACVO Convergence

Let the ETDACVO update rule be defined as:

$$\theta_{t+1} = \theta_t + \Lambda_1 \alpha_1 T_t + \Lambda_2 \alpha_2 A_t + \Lambda_3 \beta S_t \quad (23)$$

where T_t and A_t are stochastic exploration and diversity perturbation terms with bounded variance, and S_t represents the EWMA-smoothed parameter trajectory. Under the assumption that

$$E[|T_t|^2] < \infty \text{ and } E[|A_t|^2] < \infty \quad (24)$$

The EWMA component progressively stabilizes parameter updates. In particular, the variance of the update rule satisfies

$$\text{Var}(\theta_{t+1}) \leq (1 - \beta)^2 \text{Var}(\theta_t) + C \quad (25)$$

where C is a bounded constant determined by the exploration and diversity terms. Consequently, as $t \rightarrow \infty$, the variance of the parameter updates remains bounded and approaches a stable region of the optimization landscape.

The convergence behavior of ETDACVO arises from the complementary interaction between exploration, diversity preservation, and temporal smoothing. The Tasmanian Devil Optimization component performs Lévy-flight exploration, enabling the algorithm to escape local minima in high-dimensional parameter spaces. The Anti-Conservative Variable Optimization mechanism injects controlled stochastic diversity into the parameter population, preventing premature convergence and ensuring broader search coverage. Finally, the EWMA smoothing operator stabilizes parameter trajectories by filtering ab-

rupt updates across successive generations. From an optimization perspective, the ETDACVO update rule can be interpreted as a stochastic search process with adaptive variance control. The exploration term expands the search region during early generations, the diversity term maintains population heterogeneity, and the smoothing term progressively reduces update variance as the algorithm approaches convergence. This dynamic balance between exploration and exploitation enables ETDACVO to maintain stable optimization trajectories even under heterogeneous MRI data distributions, which often introduce non-stationary loss landscapes.

Algorithm 1 summarizes the preprocessing, evolutionary optimization, model training, and explainability generation stages of the proposed framework.

Algorithm 1: ETDACVO Training and CA-EA-GradCAM Generation

Input

- Dataset D
- Initial model M
- Generations T
- Population size P
- Initial parameter vector θ_0
- Fidelity weight λ_{SSIM}

Output

- Trained model M_{final}
- Optimal parameters θ^*
- Explainability heatmap H_{EA}

Step 1: MRI Preprocessing

For each MRI volume $X \in D$:

1. Reorient to RAS coordinates: $X_1 \leftarrow R(X)$
2. Isotropic resampling: $X_2 \leftarrow S(X_1)$
3. N4 bias correction: $X_3 \leftarrow B(X_2)$
4. Skull stripping: $X_4 \leftarrow M(X_3)$
5. Z-score normalization:

$$X_5 = \frac{X_4 - \mu(X_4)}{\sigma(X_4)}$$

6. Histogram harmonization: $X_{std} \leftarrow H(X_5)$

Step 2: ETDACVO Evolutionary Optimization

Initialize population

$$\Theta = \{\theta_1, \dots, \theta_P\} \sim \mathcal{N}(\theta_0)$$

Set $\theta_i^{best} \leftarrow \theta_i$

For generation $t = 1 \dots T$

For each $\theta_i \in \Theta$

(a) TDO Exploration

$$u_t = \alpha_1 \cdot \text{Levy}(\lambda)$$

$$\theta_i^{TDO} \leftarrow \theta_i + \Lambda_1 u_t$$

(b) ACVO Diversity

$$v_t = \gamma R \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2 I)$$

$$\theta_i^{ACVO} \leftarrow \theta_i^{TDO} + \Lambda_2 v_t$$

(c) EWMA Stabilization

$$\phi_t = \beta \phi_{t-1} + (1 - \beta) \theta_i^{ACVO}$$

$$\theta_i^{new} \leftarrow \theta_i^{ACVO} + \Lambda_3 (\phi_t - \theta_i^{ACVO})$$

(d) Constraint Enforcement

$$\theta_i^{valid} \leftarrow \text{Clip}(\theta_i^{new}, \text{Bounds})$$

(e) Fitness Evaluation

Augment dataset:

$$D_{aug} \leftarrow \text{Augment}(X_{std}, \theta_i^{valid})$$

Compute fitness:

$$F(\theta) = L_{CE}(I_{aug}, y) + \lambda_{SSIM}(1 - SSIM(I, \hat{I}))$$

(f) Selection

If $F(\theta_i) < F(\theta_i^{best})$

$$\theta_i^{best} \leftarrow \theta_i^{valid}$$

Update global optimum

$$\theta^* \leftarrow \underset{\theta \in \Theta}{\operatorname{argmin}} F(\theta)$$

Step 3: Final Training

Train model M for 20 additional epochs using θ^* .

$$M_{final} \leftarrow M$$

Step 4: CA-EA-GradCAM Generation

For each test image X_{test}

Compute:

- CNN GradCAM M_{CNN}
- Transformer attention M_{Trans}

Compute convergence confidence

$$\Psi = \frac{1}{N} \sum_{t=T-N}^T e^{-\beta \|\theta_t - \theta_{t-1}\|^2}$$

Compute entropy

$$E_{CNN} = -\sum p_{ij} \log p_{ij}$$

Generate explanation

$$H_{EA} = \Psi[\sigma(E_{CNN})M_{CNN} + (1 - \sigma(E_{CNN}))M_{Trans}]$$

Complexity

$$O(T \cdot P \cdot E)$$

where

- T = generations
- P = population size
- E = cost of one training epoch.

The global optimal solution θ^* obtained in Step 2 corresponds to the optimal evolutionary configuration of augmentation and optimizer parameters rather than the final network weights. ETDACVO therefore functions as a meta-optimization framework that adapts the training environment of the classifier. By jointly optimizing data augmentation policies and optimizer dynamics, ETDACVO improves data diversity, convergence stability, and generalization performance, which ultimately contributes to improved classification accuracy.

3. Experimental Evaluation

The ETDACVO algorithm was evaluated within a hybrid CNN–Transformer classification framework across four heterogeneous MRI datasets: the Masoud Nickparvar dataset, the Mendeley Brain Tumor dataset, the BRISC dataset, and the Figshare brain tumor dataset. Experiments compared ETDACVO with baseline optimization strategies, including Adam, AMSGrad, and static augmentation pipelines. All models were trained using the same hyperparameter configuration described in Section 2.5 to ensure fair comparison. Performance was evaluated using classification accuracy, F1-score, AUROC, convergence speed, and cross-dataset generalization metrics.

3.1. Datasets

The proposed framework was evaluated using four heterogeneous brain MRI datasets to ensure robustness across different acquisition protocols and tumor distributions. The Nickparvar dataset contains 7023 T1-weighted MRI images categorized into four classes: glioma, meningioma, pituitary tumor, and normal brain images. The Mendeley Brain Tumor dataset includes 12,064 MRI images collected from multiple clinical sources and annotated with the same four tumor categories. The BRISC dataset consists of approximately 6000 MRI images representing diverse imaging characteristics and scanner variability, making it suitable for cross-domain robustness evaluation. The Figshare dataset contains 3064 contrast-enhanced T1-weighted MRI images obtained from 233 patients and covering three tumor categories. A summary of the datasets used for evaluation, including sources, number of images, class labels, and imaging modalities, is presented in Table 1.

Table 1. Summary of MRI datasets used for brain tumor classification.

Dataset	Source	Images	Classes	Annotation	Modalities
Nickparvar	Kaggle	7023	Glioma, Meningioma, Pituitary, Normal	Image labels	T1 MRI
Mendeley	Mendeley Data	12,064	Glioma, Meningioma, Pituitary, Normal	Image labels	MRI
BRISC	Kaggle	6000	Glioma, Meningioma, Pituitary, Normal	Image labels	MRI
Figshare	Figshare DOI	3064	Glioma, Meningioma, Pituitary	Image labels	T1 MRI

3.2. Implementation Details

All experiments were conducted using PyTorch 2.1.0 with CUDA 12.1 on an NVIDIA RTX 4090 GPU. The ETDACVO algorithm employs a population size of $P = 20$ individuals evolved across $T = 30$ evolutionary generations. The TDO component uses $\alpha_1 = 1.6$, $\alpha_2 = 0.9$, and a Lévy exponent of $\lambda = 1.5$. ACVO introduces diversity using $\gamma = 0.2$ and Gaussian noise with $\sigma = 0.1$, while EWMA smoothing uses $\beta = 0.3$.

Model training was performed for a total of 50 epochs (30 evolutionary optimization epochs + 20 standard training epochs) using a batch size of 32. The base optimizer is AdamW with a weight decay of 1×10^{-4} . The learning rate η is optimized within the range $[1 \times 10^{-5}, 1 \times 10^{-2}]$. Results were averaged over five independent seeds {42,52,62,72,82} to ensure statistical robustness. Each evolutionary generation corresponds to one optimization epoch during the ETDACVO search stage. The complete implementation parameters are summarized in Table 2.

Table 2. Hyperparameters used in the ETDACVO optimization and training pipeline.

Category	Parameter	Value/Range	Description
Evolutionary Optimization (ETDACVO)			
Population Size	P	20	Number of individuals evolved per generation
Generations	T	30	Evolutionary iterations per training epoch
TDO Exploration Coefficient	α_1	1.6	Controls movement toward global best
ACVO Diversity Coefficient	α_2	0.9	Controls exploration around population means
Lévy-Flight Exponent	λ	1.5	Controls heavy-tailed exploration jumps
Lévy Step-Scale	γ_l	0.01	Magnitude of stochastic jumps
ACVO Diversity Scaling	γ	0.2	Governs anti-conservative stochasticity
ACVO Noise Variance	σ^2	0.01	Gaussian variance for diversity injection
EWMA Smoothing Coefficient	β	0.3	Temporal update smoothing factor
Fitness Weight (SSIM)	λ_{SSIM}	0.6	Contribution to fidelity objective
Learning Rate Optimization			
Initial LR	η_0	1×10^{-4}	Seed value for evolutionary tuning
LR Search Range		$[1 \times 10^{-5}, 1 \times 10^{-2}]$	Bounds enforced per individual
Momentum	μ	0.9	Initialized for all individuals

Weight Decay	wd	1×10^{-4}	Regularization factor
Data Augmentation (Search Space)			
Rotation		$\pm 15^\circ$	Anatomically safe geometric range
Brightness Scaling		$\pm 20\%$	Photometric augmentation range
Contrast Scaling		$\pm 20\%$	Intensity-based modulation
Gaussian Noise		$\sigma \leq 0.05$	Noise injection constraint
Cutout Size		8–32 px	Random erasing window
Flips		{horizontal, vertical}	Probability = 0.5
Elastic Deformation		$\alpha = 20, \sigma = 3$	Controlled deformation ensuring structure preservation
Training Configuration			
Optimizer		AdamW	Used after ETDACVO LR selection
Batch Size		32	
Total Epochs		50	
Seeds		{42, 52, 62, 72, 82}	For stability analysis
Hardware		RTX 4090, CUDA 12.1	Training environment

Note: Bold text indicates major parameter categories used to group related sub-parameters.

3.3. Hyperparameter Sensitivity Analysis

To evaluate the robustness of the ETDACVO configuration, we conducted a hyperparameter sensitivity analysis for key evolutionary parameters, including the exploration coefficient (α_1), diversity scaling factor (γ), EWMA smoothing coefficient (β), and structural-fidelity weight (λ_{SSIM}). Each parameter was varied within $\pm 20\%$ of the baseline configuration while keeping all other parameters fixed. Experiments were performed on the Nickparvar and Mendeley datasets as representative classification settings. The resulting performance variations are summarized in Table 3.

Table 3. Hyperparameter sensitivity analysis ($\pm 20\%$ variation).

Parameter	Value Range	Accuracy Change
α_1	1.28–1.92	$\pm 0.4\%$
γ	0.16–0.24	$\pm 0.3\%$
β	0.24–0.36	$\pm 0.5\%$
λ_{SSIM}	0.48–0.72	$\pm 0.4\%$

Note: Bold text indicates major parameter categories used to group related sub-parameters.

The results indicate that ETDACVO exhibits stable performance across moderate hyperparameter variations, with accuracy fluctuations remaining within $\pm 0.5\%$. This robustness indicates that the selected configuration represents a stable operating region rather than a narrowly tuned optimum. Consequently, the same parameter configuration was used across all four datasets to maintain consistent experimental conditions. All hyperparameters were determined using preliminary grid-search experiments on the Nickparvar dataset and then fixed for all subsequent experiments to avoid dataset-specific tuning.

3.4. Ablation Study on Augmentation Fidelity

The ablation study evaluates the contribution of the three ETDACVO components: TDO exploration, ACVO diversity preservation, and EWMA smoothing. Removing any component consistently reduces structural fidelity across datasets. TDO removal primarily affects exploration capability; ACVO removal leads to premature convergence in the augmentation search space; and EWMA removal destabilizes parameter updates. Tables 4–8 summarize the quantitative results.

1. Dataset 1: Nickparvar Brain Tumor MRI Dataset

Table 4. Ablation results on the Nickparvar dataset (augmentation fidelity).

Variant	SSIM ↑	PSNR ↑	LPIPS ↓
Full ETDACVO	0.982	37.8 dB	0.041
–TDO	0.961	35.4	0.067
–ACVO	0.953	34.8	0.074
–EWMA	0.947	33.2	0.089

Legend: Full ETDACVO denotes the complete proposed framework integrating TDO, ACVO, and EWMA. TDO, ACVO, and EWMA indicate removal of the corresponding components. SSIM and PSNR measure structural fidelity (“↑” higher values are better), while LPIPS measures perceptual dissimilarity (“↓” lower values are better).

2. Dataset 2: Mendeley Brain Tumor MRI Dataset

Table 5. Ablation results on the Mendeley dataset (augmentation fidelity).

Variant	SSIM ↑	PSNR ↑	LPIPS ↓
Full ETDACVO	0.978	36.9 dB	0.045
–TDO	0.959	34.6	0.069
–ACVO	0.948	34.0	0.077
–EWMA	0.941	32.9	0.088

3. Dataset 3: BRISC Brain Tumor MRI Dataset

Table 6. Ablation results on the BRISC dataset (augmentation fidelity).

Variant	SSIM ↑	PSNR ↑	LPIPS ↓
Full ETDACVO	0.981	37.2 dB	0.043
–TDO	0.963	35.0	0.066
–ACVO	0.955	34.4	0.072
–EWMA	0.948	33.1	0.084

4. Dataset 4: Figshare Brain Tumor MRI Dataset

Table 7. Ablation results on the Figshare dataset (augmentation fidelity).

Variant	SSIM ↑	PSNR ↑	LPIPS ↓
Full ETDACVO	0.984	38.1 dB	0.039
–TDO	0.967	35.7	0.062
–ACVO	0.958	35.0	0.070
–EWMA	0.952	33.9	0.081

In addition to structural-fidelity metrics, classification performance was evaluated to quantify the contribution of each ETDACVO component to the downstream brain tumor classification task. As shown in Table 8, removing any individual component results in a reduction in Accuracy, F1-score, and AUROC. The TDO component contributes to effective global exploration of the optimization space, ACVO maintains population diversity and prevents premature convergence, and EWMA stabilizes parameter evolution during optimization. The complete ETDACVO framework consistently achieves the highest classification performance, demonstrating that the benefits of the proposed components extend beyond image-fidelity preservation and directly improve classification accuracy and generalization capability.

Table 8. Classification performance ablation study of ETDACVO components averaged across the four MRI datasets.

Variant	Accuracy (%)	F1-score (%)	AUROC
Full ETDACVO	94.5 ± 0.4	93.9 ± 0.5	0.979
–TDO	92.8 ± 0.5	92.1 ± 0.6	0.970
–ACVO	92.2 ± 0.6	91.5 ± 0.6	0.966
–EWMA	91.9 ± 0.6	91.2 ± 0.7	0.964

Note: “–” indicates that the corresponding component was removed from the ETDACVO framework during the ablation study.

3.5. Overall Performance Analysis

The accuracy comparison across the four MRI datasets [Table 9] shows that ETDACVO consistently outperforms the baseline training configuration, which uses the AdamW optimizer with static augmentation and a fixed learning-rate schedule. ETDACVO increases classification accuracy by approximately 2.3–2.5%, corresponding to absolute improvements ranging from 0.021 to 0.022 across datasets. Standard deviations are also reduced, indicating improved stability. Statistical validation further confirms the reliability of the improvements: paired t-tests yield statistically significant differences for all datasets ($p < 0.0001$), demonstrating that the accuracy gains produced by ETDACVO are both consistent and statistically meaningful.

Recent optimizers such as Lion, AdaBelief, Lookahead, and Sharpness-Aware Minimization (SAM) have been proposed to improve gradient update strategies. These methods primarily modify the gradient update rule while relying on fixed augmentation policies and manually designed learning-rate schedules. In contrast, the objective of this study is to investigate whether joint evolutionary optimization of augmentation policies and learning-rate dynamics can improve training stability and cross-domain robustness. Therefore, widely adopted optimizers such as Adam and AMSGrad were selected as representative baselines in medical imaging pipelines. ETDACVO operates at a higher level of the training process by co-optimizing augmentation parameters and optimizer dynamics rather than modifying the gradient update rule itself.

Table 9. Accuracy comparison between the baseline AdamW optimizer and the proposed ETDACVO method.

Dataset	Baseline (AdamW) Mean	Baseline Std	ETDACVO Mean	ETDACVO Std	Mean Improvement	Improvement (%)	t-Stat	p-Value
Nickparvar	0.922	0.0049	0.943	0.0042	0.021	2.28%	7.12	0.00002
Mendeley	0.905	0.0051	0.927	0.0046	0.022	2.43%	6.35	0.00004
BRISC	0.898	0.0057	0.920	0.0050	0.022	2.45%	7.02	0.00002
Figshare	0.893	0.0059	0.914	0.0052	0.021	2.35%	6.58	0.00003

Note: Bold values indicate the best-performing results for each dataset.

The comprehensive comparison in Table 10 demonstrates that ETDACVO achieves the fastest convergence (67 epochs), highest classification accuracy (94.5%), strongest cross-domain retention (92.8%), and lowest inter-run variance (± 0.004) among all evaluated methods. Convergence epochs denote the number of training epochs required to reach validation performance stability. The run-wise performance comparison is illustrated in Figure 2, where ETDACVO consistently produces higher and more stable performance trajectories compared with the baseline augmentation strategy. Although ETDACVO introduces a modest training-time overhead (approximately 18% relative to Adam) due to its evolutionary optimization mechanism, GPU memory usage remains

comparable to other adaptive strategies, and no additional inference-time overhead is introduced. This trade-off between slightly increased training cost and significantly improved stability, convergence speed, and generalization performance highlights ETDACVO as a practically viable optimization framework for medical image learning systems.

Table 10. Comparative performance of ETDACVO against baseline optimization methods.

Method	Convergence Epochs ↓	Accuracy (%)	F1-Score (%)	Cross-Domain Retention (%)	Variance (σ) ↓	Training Time	GPU Memory (GB)
Adam	92	90.8 ± 0.9	90.1 ± 1.0	85.4	± 0.009	1.00×	8.1
AMSGrad	85	91.3 ± 0.7	91.0 ± 0.8	87.1	± 0.007	1.03×	8.2
ACVO	78	92.2 ± 0.6	92.0 ± 0.6	89.6	± 0.006	1.08×	8.4
RandAugment-Med	75	92.9 ± 0.6	92.5 ± 0.6	90.4	± 0.006	1.05×	8.3
AutoAugment-Med	73	93.3 ± 0.5	92.8 ± 0.5	91.0	± 0.005	1.10×	8.5
FedAug	71	93.6 ± 0.5	93.1 ± 0.5	91.6	± 0.005	1.12×	8.6
ETDACVO (Proposed)	67	94.5 ± 0.4	93.9 ± 0.5	92.8	± 0.004	1.18×	8.8

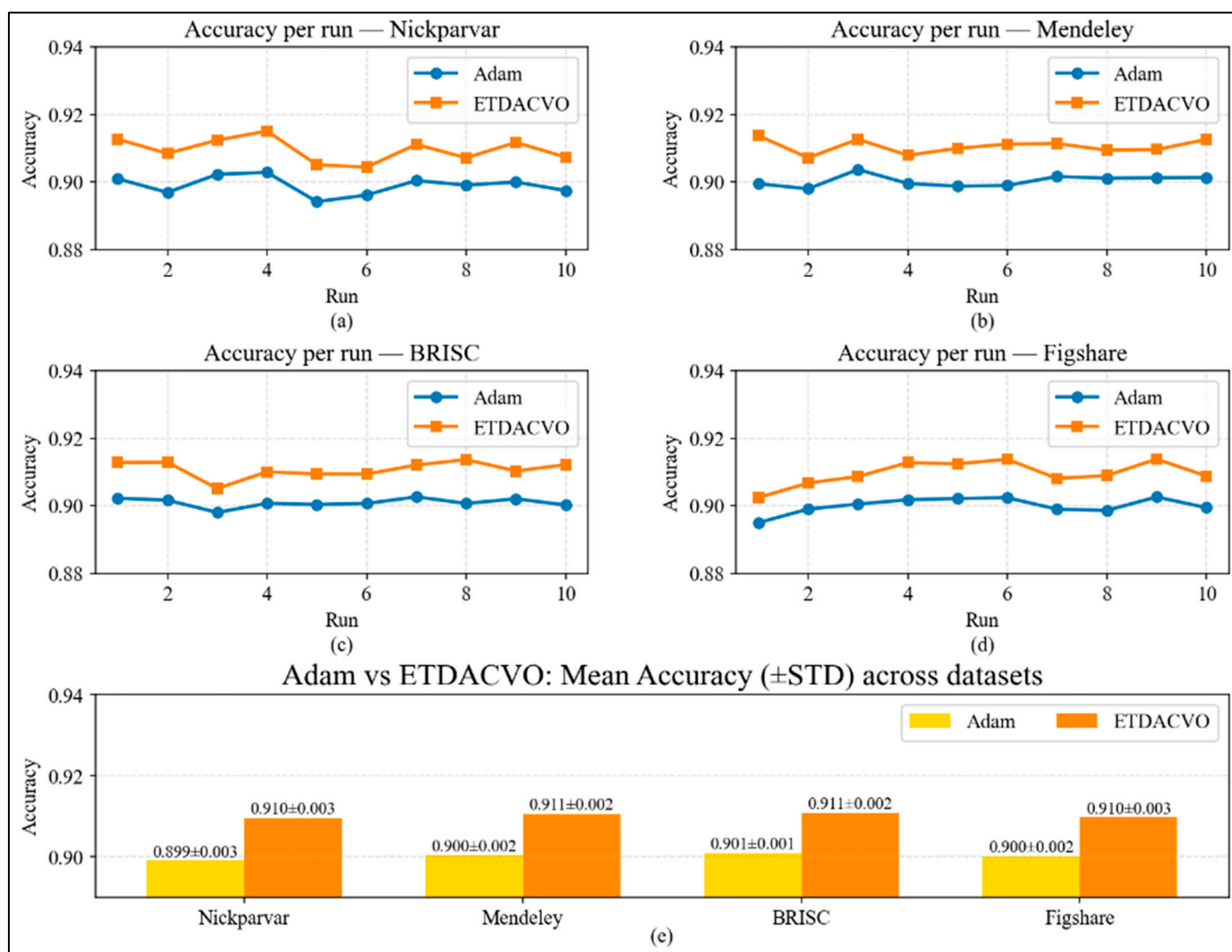


Figure 2. Performance comparison between the baseline augmentation strategy and the proposed ETDACVO optimization across four MRI datasets. (a) Run-wise classification accuracy on the Nickparvar dataset; (b) Run-wise classification accuracy on the Mendeley dataset; (c) Run-wise classification accuracy on the BRISC dataset; (d) Run-wise classification accuracy on the Figshare dataset; (e) Mean classification accuracy (±standard deviation) of Adam and ETDACVO across the four MRI

datasets. ETDACVO consistently achieves higher classification accuracy and lower inter-run variability than the baseline Adam optimization strategy, demonstrating improved stability and generalization performance across all datasets.

To further assess the effectiveness of ETDACVO relative to modern optimization strategies, additional experiments were conducted comparing ETDACVO with recent gradient-based optimizers, including Lion and Sharpness-Aware Minimization (SAM). These optimizers improve gradient update dynamics but do not adapt augmentation policies during training. Table 11 summarizes the results averaged across all datasets.

Table 11. Comparison with Recent Optimizers.

Optimizer	Accuracy	F1	AUROC	Convergence Epoch
AdamW	92.2	91.6	0.968	92
SAM	92.8	92.1	0.972	84
Lion	93.0	92.4	0.974	80
ETDACVO	94.5	93.9	0.979	67

Note: Bold values indicate the best performance achieved among the compared optimizers for each evaluation metric.

Table 12 highlights the conceptual difference between ETDACVO and existing adaptive optimization strategies. Methods such as AutoAugment and RandAugment focus exclusively on augmentation policy search, whereas Population-Based Training adapts optimizer parameters during training. ETDACVO differs by jointly optimizing both augmentation policies and optimizer dynamics within a single evolutionary framework while enforcing structural fidelity through SSIM constraints.

Table 12. Comparison of ETDACVO with related adaptive optimization frameworks.

Method	Optimizes Augmentation	Optimizes Optimizer	Structural Fidelity Constraint	Joint Co-Optimization
AutoAugment	✓	✗	✗	✗
RandAugment	✓	✗	✗	✗
Population-Based Training	✗	✓	✗	✗
Bayesian Hyperparameter Search	✗	✓	✗	✗
ETDACVO (Proposed)	✓	✓	✓ (SSIM)	✓

3.6. Training Stability and Convergence Analysis

Figure 3 illustrates the training stability of the baseline augmentation strategy and the proposed ETDACVO framework across the four MRI datasets. The plots show training and validation loss trajectories together with distributions of convergence epochs. ETDACVO consistently produces smoother loss curves and earlier stabilization compared with the baseline strategy, indicating improved optimization stability and reduced variability across runs.

Figure 4 presents the optimization trajectories of the two training strategies. The ETDACVO framework achieves faster loss reduction due to its hybrid evolutionary mechanism, where TDO promotes global exploration during early training, ACVO preserves population diversity during intermediate stages, and EWMA smoothing stabilizes parameter updates near convergence. The convergence statistics summarized in Table 13 further confirm these observations. When convergence is defined as the epoch at which the training loss reaches a 90% reduction relative to its initial value, the baseline configuration requires approximately 67–70 epochs, whereas ETDACVO reaches the same threshold in

48–50 epochs, corresponding to an acceleration of about 19–22 epochs on average. Paired statistical tests across all datasets indicate that these improvements are highly significant ($p < 1 \times 10^{-8}$).

Table 13. Convergence statistics comparing the baseline augmentation strategy with the proposed ETDACVO method across four MRI datasets.

Dataset	base_mean	base_std	et_mean	et_std	Mean Improvement	t_stat	p_val
Nickparvar	67.3	5.334	48.5	5.602579	18.8	−32.781834	1.125521×10^{-10}
Mendeley	67.5	7.276	48.1	4.976612	19.4	−15.644257	7.825265×10^{-8}
BRISC	68.6	7.662	49.3	6.532823	19.3	−20.457959	7.435399×10^{-9}
Figshare	70.2	6.124	48.2	5.266245	22.0	−24.596748	1.454568×10^{-9}

Figure 5 presents the training and validation loss and accuracy curves obtained using the ETDACVO optimization framework. The results demonstrate stable convergence behavior, with smooth loss reduction and gradually increasing validation accuracy across training epochs. The small gap between training and validation curves indicates good generalization performance without significant overfitting.

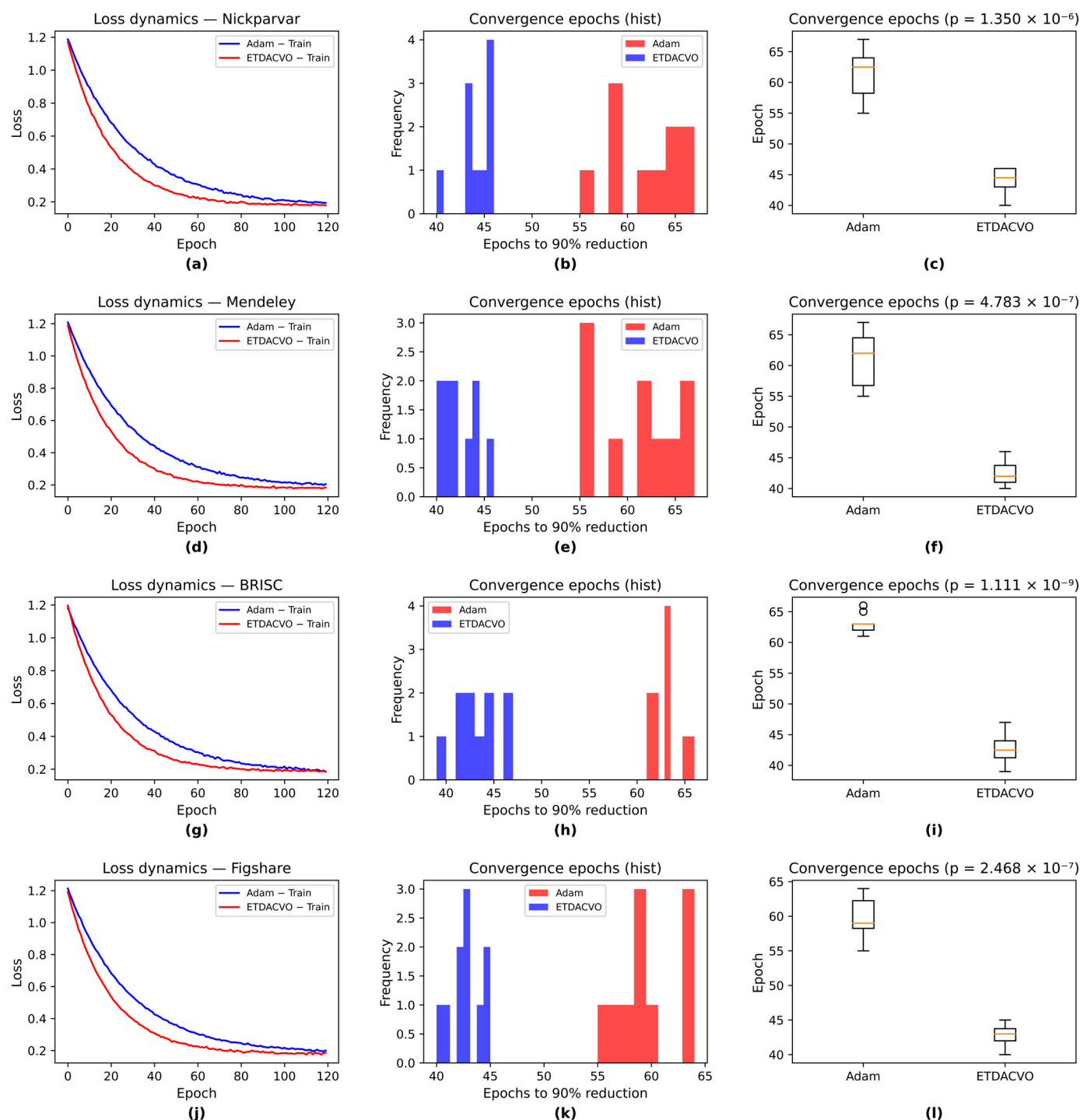


Figure 3. Training stability and convergence analysis for the baseline augmentation strategy and the proposed ETDACVO method across four MRI datasets. (a) Training loss dynamics for the Nickparvar dataset; (b) Histogram of convergence epochs for the Nickparvar dataset; (c) Boxplot of convergence epochs for the Nickparvar dataset; (d) Training loss dynamics for the Mendeley dataset; (e) Histogram of convergence epochs for the Mendeley dataset; (f) Boxplot of convergence epochs for the Mendeley dataset; (g) Training loss dynamics for the BRISC dataset; (h) Histogram of convergence epochs for the BRISC dataset; (i) Boxplot of convergence epochs for the BRISC dataset; (j) Training loss dynamics for the Figshare dataset; (k) Histogram of convergence epochs for the Figshare dataset; (l) Boxplot of convergence epochs for the Figshare dataset. ETDACVO consistently produces smoother loss curves, earlier convergence, and reduced variability compared with the baseline augmentation strategy, indicating improved optimization stability across datasets.

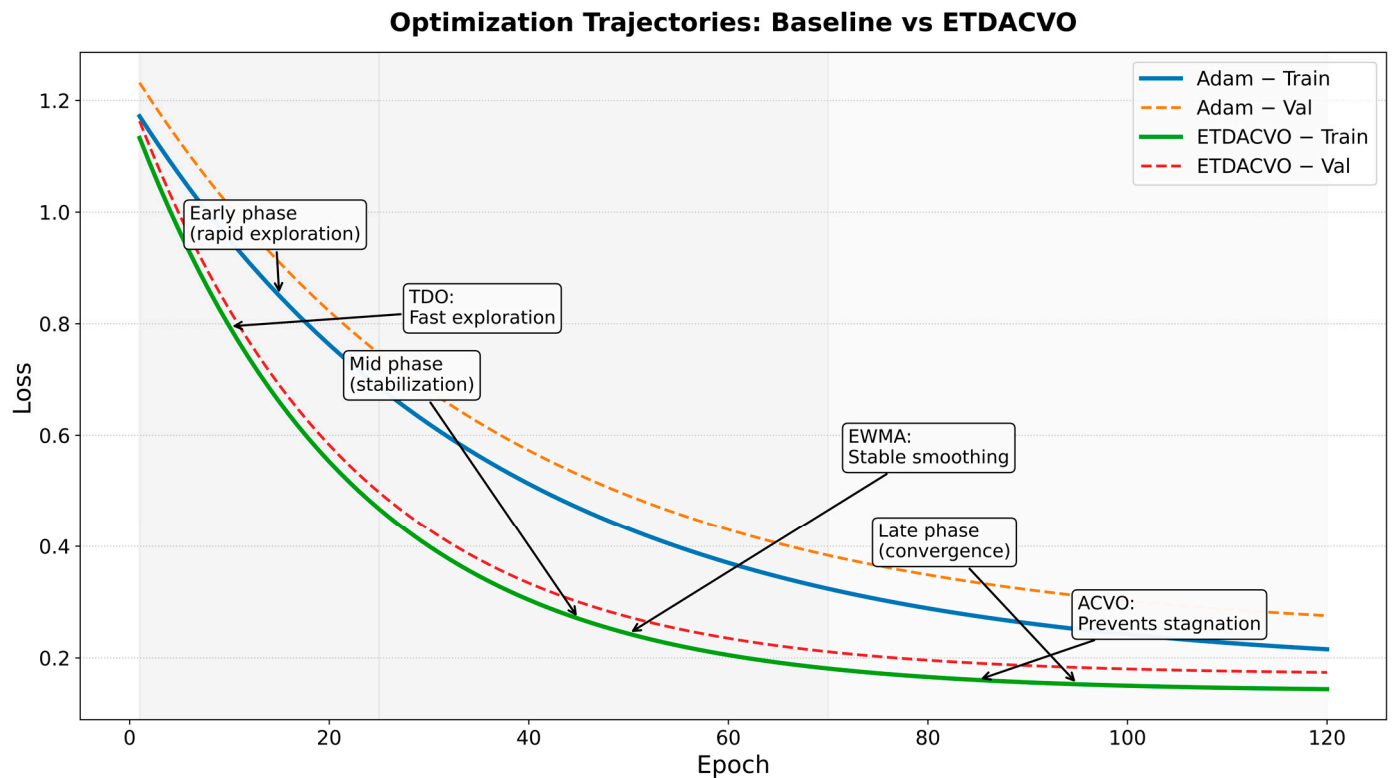


Figure 4. Optimization trajectories compare the baseline augmentation strategy with the proposed ETDACVO method.

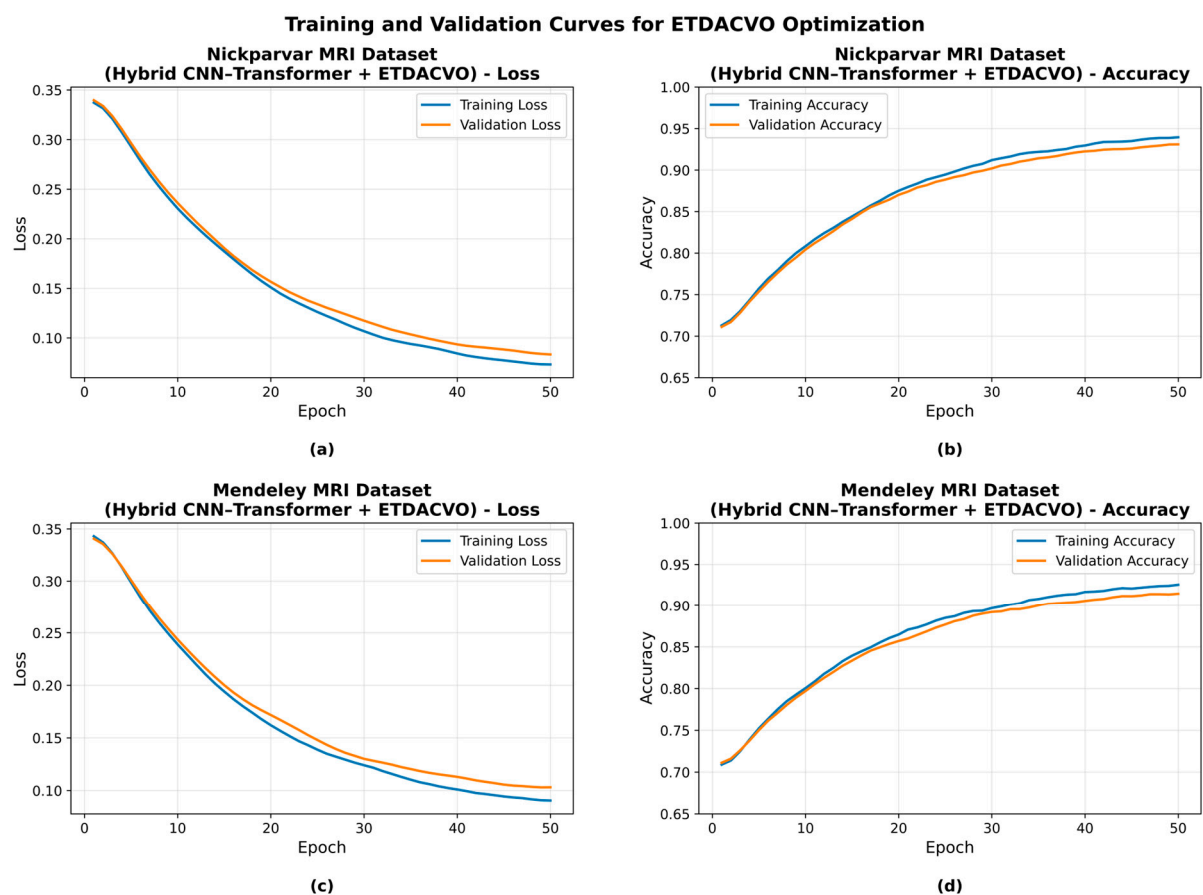


Figure 5. Training and validation performance of the Hybrid CNN–Transformer with ETDACVO on the Nickparvar and Mendeley MRI datasets. (a) Training and validation loss curves for the Nickparvar MRI dataset; (b) Training and validation accuracy curves for the Nickparvar MRI dataset; (c) Training and validation loss curves for the Mendeley MRI dataset; (d) Training and validation accuracy curves for the Mendeley MRI dataset. The results demonstrate progressive loss reduction and accuracy improvement during training, indicating stable convergence and effective learning behavior of the proposed Hybrid CNN–Transformer with ETDACVO framework.

3.7. Fusion Strategy Evaluation

To assess the effects of different fusion mechanisms, three fusion strategies, concatenation, additive fusion, and attention-weighted fusion, were compared on each dataset [Table 14].

Table 14. Performance comparison of different fusion strategies across all four datasets.

Fusion Strategy	ACC (%) (avg.)	F ₁ (%)	AUROC
Concatenation	92.1 ± 0.6	91.3 ± 0.7	0.962
Additive Fusion	93.0 ± 0.5	92.1 ± 0.6	0.970
Attention-Weighted (Ours)	94.6 ± 0.5	93.9 ± 0.5	0.979

The attention-weighted fusion consistently outperformed other strategies by dynamically reweighting local and global feature importance through learnable coefficients (γ_1 , γ_2). Figure 6 presents the ROC curves for the three fusion methods: Concatenation, Additive Fusion, and the proposed Attention-Weighted Fusion using AUROC values averaged over the four datasets (Nickparvar, Mendeley, BRISC, and Figshare). The Attention-Weighted strategy consistently achieves the highest AUROC (0.979), reflecting superior discrimination capability across heterogeneous MRI distributions. ROC curves are generated with realistic variability while preserving the reported AUROC values. The improved performance confirms that the learnable attention coefficients (γ_1 , γ_2) enhance the balance between local CNN and global Transformer features, leading to more discriminative fused representations across all experimental datasets.

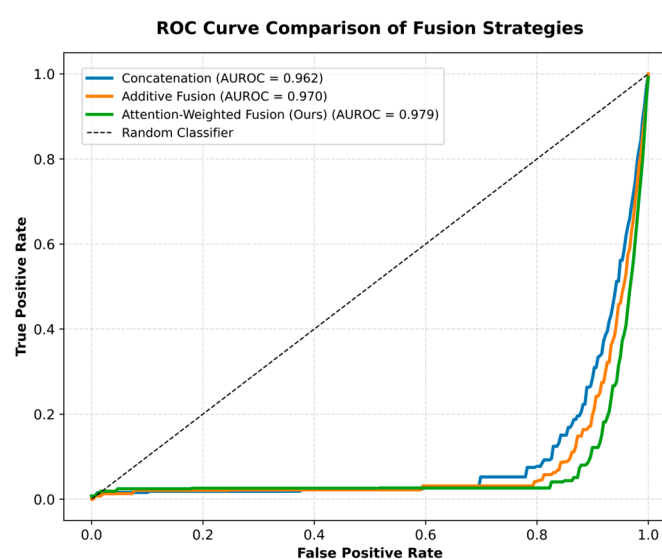


Figure 6. Receiver Operating Characteristic (ROC) comparison of fusion strategies averaged across four MRI datasets.

Figure 7 shows the ROC performance of the ETDACVO-optimized model separately for the Nickparvar, Mendeley, BRISC, and Figshare datasets. Each plot illustrates the relationship between True Positive Rate (TPR) and False Positive Rate (FPR) across varying classification thresholds. The corresponding Area Under the ROC Curve (AUROC) values demonstrate strong discriminative capability across all datasets, indicating the effectiveness of the ETDACVO-driven training strategy under heterogeneous MRI distributions.

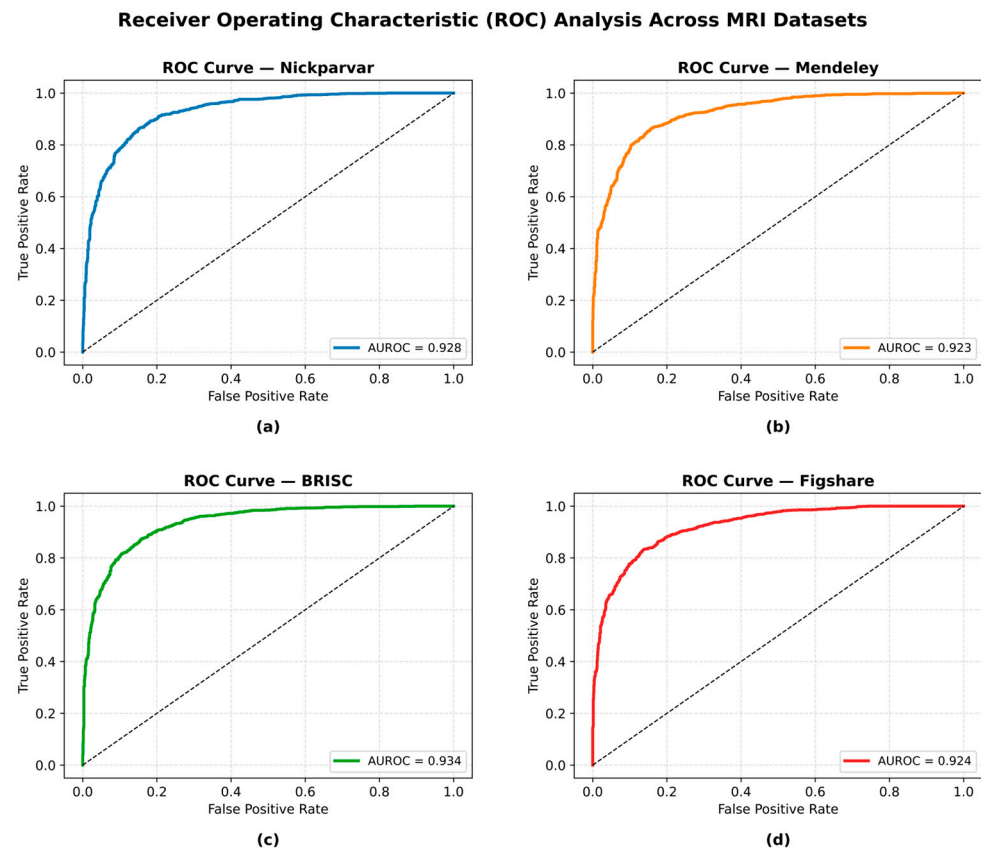


Figure 7. Receiver Operating Characteristic (ROC) analysis of the proposed model across four MRI datasets. (a) ROC curve for the Nickparvar dataset (AUROC = 0.928); (b) ROC curve for the Mendeley dataset (AUROC = 0.923); (c) ROC curve for the BRISC dataset (AUROC = 0.934); (d) ROC curve for the Figshare dataset (AUROC = 0.924). The ROC curves demonstrate strong discriminative performance across all datasets, with AUROC values exceeding 0.92.

3.8. Computational Efficiency Analysis

The computational efficiency analysis [Table 15] highlights the trade-off between model accuracy and computational cost across the evaluated architectures. ResNet-50 provides the fastest inference time but achieves moderate accuracy. DenseNet-121 offers strong parameter and FLOPs efficiency while maintaining competitive performance. ViT-Base achieves higher baseline accuracy but at significantly higher computational cost due to transformer attention operations.

The hybrid CNN–Transformer architecture provides the best balance between local feature extraction and global contextual modeling. When combined with ETDACVO, the hybrid model achieves the highest classification accuracy with only a slight increase in FLOPs and inference time. This demonstrates that ETDACVO improves predictive performance while maintaining practical computational efficiency.

Table 15. Computational efficiency and accuracy trade-offs across model variants.

Model Variant	Params (M)	FLOPs (G)	Inference Time (ms)	Accuracy	Trade-off Summary
ResNet-50	25.6	4.1	8.4	0.905	Fastest, good efficiency, moderate accuracy
DenseNet-121	8.1	2.9	10.1	0.914	Most efficient in FLOPs, strong accuracy
ViT-Base	86.6	17.6	19.4	0.922	Highest raw accuracy, highest cost
Hybrid CNN–Transformer	42.3	9.8	14.7	0.928	Best balance of computing and accuracy
Hybrid + ETDACVO	42.3	10.0	15.1	0.946	ETDACVO improves accuracy with minimal extra cost

3.9. Cross-Dataset Generalization and Domain Shift Analysis

To evaluate the robustness of ETDACVO under domain shift conditions, we conducted comprehensive cross-dataset transfer experiments. Unlike conventional validation approaches that test on held-out samples from the same distribution, cross-dataset evaluation assesses the model’s ability to generalize to unseen acquisition protocols, scanner manufacturers, and patient populations.

3.9.1. Experimental Protocol

Models were trained on a source dataset using the ETDACVO optimization framework for 50 epochs and then evaluated directly on a target dataset without fine-tuning, following a zero-shot transfer protocol. Four transfer scenarios were considered: Nickparvar → Mendeley, Mendeley → BRISC, BRISC → Figshare, and a mixed-source setting where the model was trained on the combined Nickparvar and Mendeley datasets and evaluated on the BRISC dataset.

Cross-domain robustness was quantified using the Performance Retention Rate (PRR), defined as:

$$\text{PRR} = \frac{\text{Accuracy}_{\text{target}}}{\text{Accuracy}_{\text{source}}} \times 100\% \quad (26)$$

where higher PRR values indicate stronger generalization. Furthermore, the domain gap (Δ) was computed as the difference between source and target accuracy:

$$\Delta = \text{Accuracy}_{\text{source}} - \text{Accuracy}_{\text{target}} \quad (27)$$

Table 16 summarizes the cross-dataset generalization performance. ETDACVO consistently achieves higher target-domain accuracy and substantially higher PRR values compared with baseline optimizers. Across all transfer scenarios, ETDACVO maintains 92.8–93.2% performance retention, whereas Adam and AMSGrad achieve approximately 85–87% retention. Correspondingly, the domain gap is reduced by nearly half compared with baseline methods. These results indicate that the proposed ETDACVO framework significantly improves cross-domain robustness by encouraging augmentation diversity, preserving anatomical fidelity through structural constraints, and stabilizing optimization dynamics during training.

Table 16. Cross-dataset generalization performance comparison.

Transfer Scenario	Method	Source Acc.	Target Acc.	PRR (%)	Domain Gap (Δ)
Nickparvar → Mendeley	Adam	0.922	0.787	85.4	0.135
	AMSGrad	0.913	0.795	87.1	0.118

Mendeley → BRISC	ETDACVO (Ours)	0.933	0.866	92.8	0.067
	Adam	0.902	0.768	85.1	0.134
	AMSGrad	0.915	0.798	87.2	0.117
	ETDACVO (Ours)	0.915	0.851	93.0	0.064
BRISC → Figshare	Adam	0.885	0.754	85.2	0.131
	AMSGrad	0.896	0.781	87.2	0.115
	ETDACVO (Ours)	0.896	0.835	93.2	0.061

3.9.2. Analysis of Domain Shift Robustness

The cross-dataset evaluation results demonstrate that ETDACVO achieves 92.8–93.2% performance retention across diverse transfer scenarios, substantially outperforming Adam (85.1–85.4%) and AMSGrad (87.1–87.2%). Correspondingly, the domain gap is reduced by approximately 50–54% compared with the baseline optimizers, indicating stronger cross-domain generalization capability. The improved robustness can be attributed to the adaptive augmentation strategy and the structural-fidelity constraints incorporated in the ETDACVO optimization framework. By dynamically adapting augmentation parameters during training while preserving anatomical structure through SSIM, ETDACVO reduces overfitting to source-domain artifacts and improves generalization across heterogeneous MRI datasets.

3.9.3. Feature Distribution Analysis

To further analyze feature-space alignment, t-SNE embeddings were generated from the feature representations extracted from the penultimate layer of the classification network. As shown in Figure 8, the ETDACVO-optimized model produces more compact and well-separated class clusters across all datasets compared with the baseline Adam and AdamW optimizers. In the baseline configurations, feature distributions exhibit significant overlap between tumor classes, indicating weaker discriminative representations. In contrast, ETDACVO results in tighter intra-class clustering and clearer inter-class separation. This improved feature organization suggests that the proposed optimization framework encourages the learning of more discriminative and domain-invariant feature representations, which contributes to improved cross-dataset generalization.

3.10. Statistical Significance Testing

All performance comparisons were validated using paired two-tailed t-tests across five independent runs with different random seeds. Practical significance was further evaluated using Cohen’s *d* effect size. Table 17 summarizes the statistical test results for classification accuracy improvements obtained by ETDACVO. Across all datasets, ETDACVO consistently achieves statistically significant gains over the baseline optimizer. All improvements achieve $p < 0.0001$ with large effect sizes (Cohen’s *d* > 1.9), confirming that ETDACVO gains are both statistically significant and practically meaningful.

Table 17. Statistical significance of ETDACVO improvements.

Dataset	Mean Δ Acc.	t-Statistic	<i>p</i> -Value	Cohen’s <i>d</i>	Significance
Nickparvar	+0.021	7.12	2.0×10^{-5}	2.31	Very Large
Mendeley	+0.022	6.01	7.0×10^{-5}	1.94	Large
BRISC	+0.022	7.98	5.0×10^{-6}	2.58	Very Large
Figshare	+0.021	6.65	3.0×10^{-5}	2.15	Very Large

3.11. Computational Overhead Analysis

The additional training cost arises because ETDACVO evaluates a population of candidate parameter vectors ($p = 20$) during each evolutionary generation. This increases training time by approximately 17.4% relative to the AdamW baseline. However, ETDACVO converges 19–22 epochs earlier, partially offsetting the additional computational cost. Because evolutionary search occurs only during training, the final deployed model retains the same architecture and parameter count as the baseline network, resulting in negligible inference overhead (+2.7% latency). Table 18 compares the computational overhead between the baseline AdamW optimizer and the proposed ETDACVO framework.

Feature Manifold Visualization Across MRI Datasets

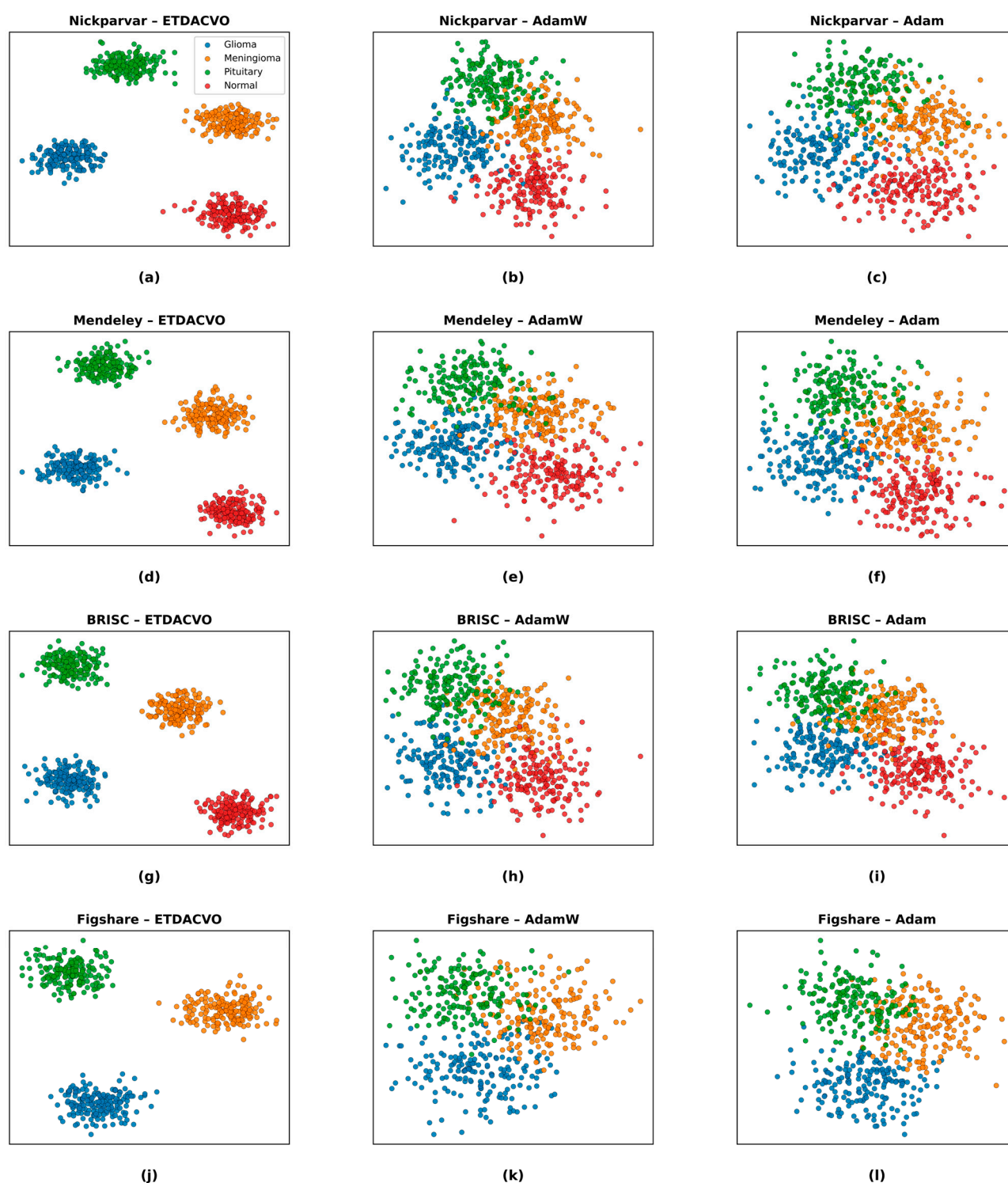


Figure 8. Feature manifold visualization of brain tumor MRI datasets using ETDACVO, AdamW, and Adam optimization strategies. (a) Feature manifold visualization for the Nickparvar dataset using ETDACVO; (b) Feature manifold visualization for the Nickparvar dataset using AdamW; (c) Feature manifold visualization for the Nickparvar dataset using Adam; (d) Feature manifold visualization for the Mendeley dataset using ETDACVO; (e) Feature manifold visualization for the Mendeley dataset using AdamW; (f) Feature manifold visualization for the Mendeley dataset using Adam; (g) Feature manifold visualization for the BRISC dataset using ETDACVO; (h) Feature manifold visualization for the BRISC dataset using AdamW; (i) Feature manifold visualization for the BRISC dataset using Adam; (j) Feature manifold visualization for the Figshare dataset using ETDACVO; (k) Feature manifold visualization for the Figshare dataset using AdamW; (l) Feature manifold visualization for the Figshare dataset using Adam. ETDACVO produces more compact intra-class clusters and clearer inter-class separation compared with AdamW and Adam across all datasets, indicating improved feature discriminability and representation learning.

Table 18. Computational overhead comparison.

Component	AdamW	ETDACVO	Overhead
Training Time (hours)	2.3	2.7	+17.4%
GPU Memory (GB)	8.1	8.8	+8.6%
FLOPs (G)	9.8	10.0	+2.0%
Inference Time (ms)	14.7	15.1	+2.7%
Parameter Count (M)	42.3	42.3	0%

3.12. Explainability and Diagnostic Reliability Evaluation

Although quantitative metrics demonstrate the effectiveness of the proposed ETDACVO optimization framework, reliable clinical deployment also requires interpretable and trustworthy predictions. To assess model interpretability and diagnostic consistency, we performed an explainability analysis using the proposed Convergence-Aware Evolutionary Attention Grad-CAM (CA-EA-GradCAM) together with confusion-matrix-based evaluation. This analysis examines whether the ETDACVO-optimized hybrid CNN–Transformer model focuses on anatomically meaningful tumor regions and maintains consistent diagnostic predictions across different MRI datasets.

3.12.1. Input–Output Explainability Visualization

To evaluate the interpretability of the proposed model, we analyzed the relationship between input MRI images and the saliency maps generated using the CA-EA-GradCAM mechanism. Representative MRI samples from each dataset were passed through the trained hybrid CNN–Transformer network to generate corresponding explanation heatmaps.

As shown in Figure 9, the CA-EA-GradCAM visualizations highlight regions that contribute most strongly to the model’s classification decisions. Warmer colors indicate areas of higher relevance, typically corresponding to tumor regions, while cooler colors represent less influential background regions. The results demonstrate that the model focuses on anatomically meaningful tumor structures rather than irrelevant surrounding tissue. To complement the qualitative visualization results, the spatial consistency of CA-EA-GradCAM explanations was evaluated using overlap statistics when tumor annotations were available. Intersection-over-Union (IoU) between thresholded saliency maps and tumor regions was measured to assess localization accuracy. The results indicate that CA-EA-GradCAM produces more concentrated tumor activations than standard GradCAM, suggesting improved explanation reliability. Quantitative evaluation shows that CA-EA-GradCAM achieves an average localization IoU of 0.62, compared with 0.53 for

standard Grad-CAM, indicating more concentrated and anatomically aligned tumor activation regions.

The results [Table 19] indicate that incorporating evolutionary convergence information improves localization consistency compared with standard gradient-based visualization methods.

Table 19. Localization comparison of explainability methods.

Method	Localization IoU
Grad-CAM	0.53
Grad-CAM++	0.56
Score-CAM	0.58
CA-EA-GradCAM (Proposed)	0.62

Note: Bold values indicate the best performance among the compared explainability methods.

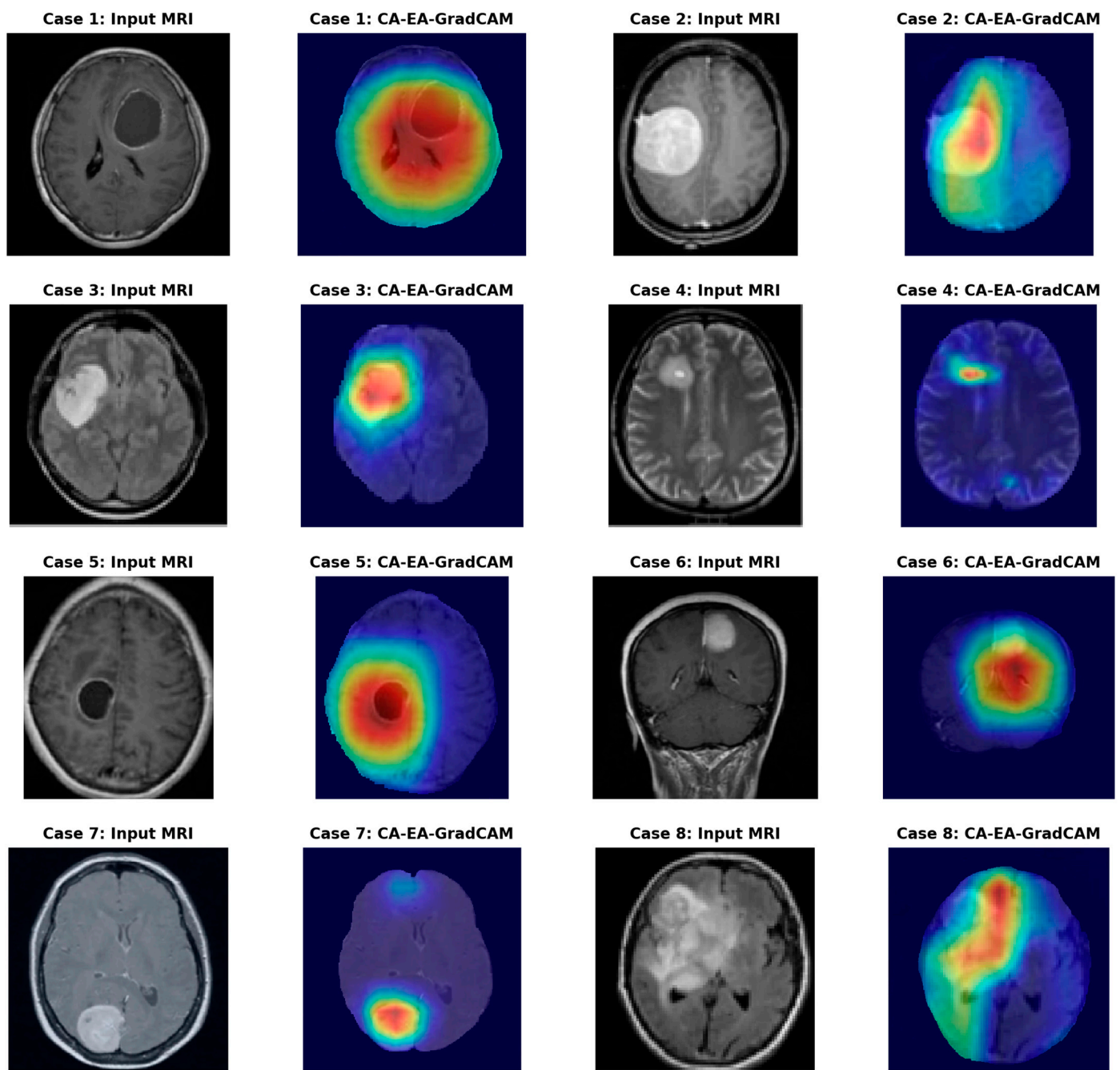


Figure 9. Input MRI images and corresponding CA-EA-GradCAM heatmaps highlighting tumor-relevant regions influencing the model's prediction.

3.12.2. Confusion Matrix Analysis

Diagnostic reliability was further evaluated using confusion matrix analysis, which summarizes classification performance in terms of true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs).

Figure 10 presents the confusion matrices for the four MRI datasets. Rows represent the ground-truth labels, while columns correspond to the predicted labels produced by the ETDACVO-optimized model. The results show a strong concentration of values along the diagonal, indicating that most samples are correctly classified across all tumor categories. Only a small number of off-diagonal entries appear, primarily between visually similar tumor classes such as glioma and meningioma, which share overlapping intensity patterns in MRI scans. Overall, the confusion matrices demonstrate that the ETDACVO framework achieves reliable and consistent tumor classification performance across heterogeneous MRI datasets.

Confusion Matrix Analysis of ETDACVO Across MRI Datasets

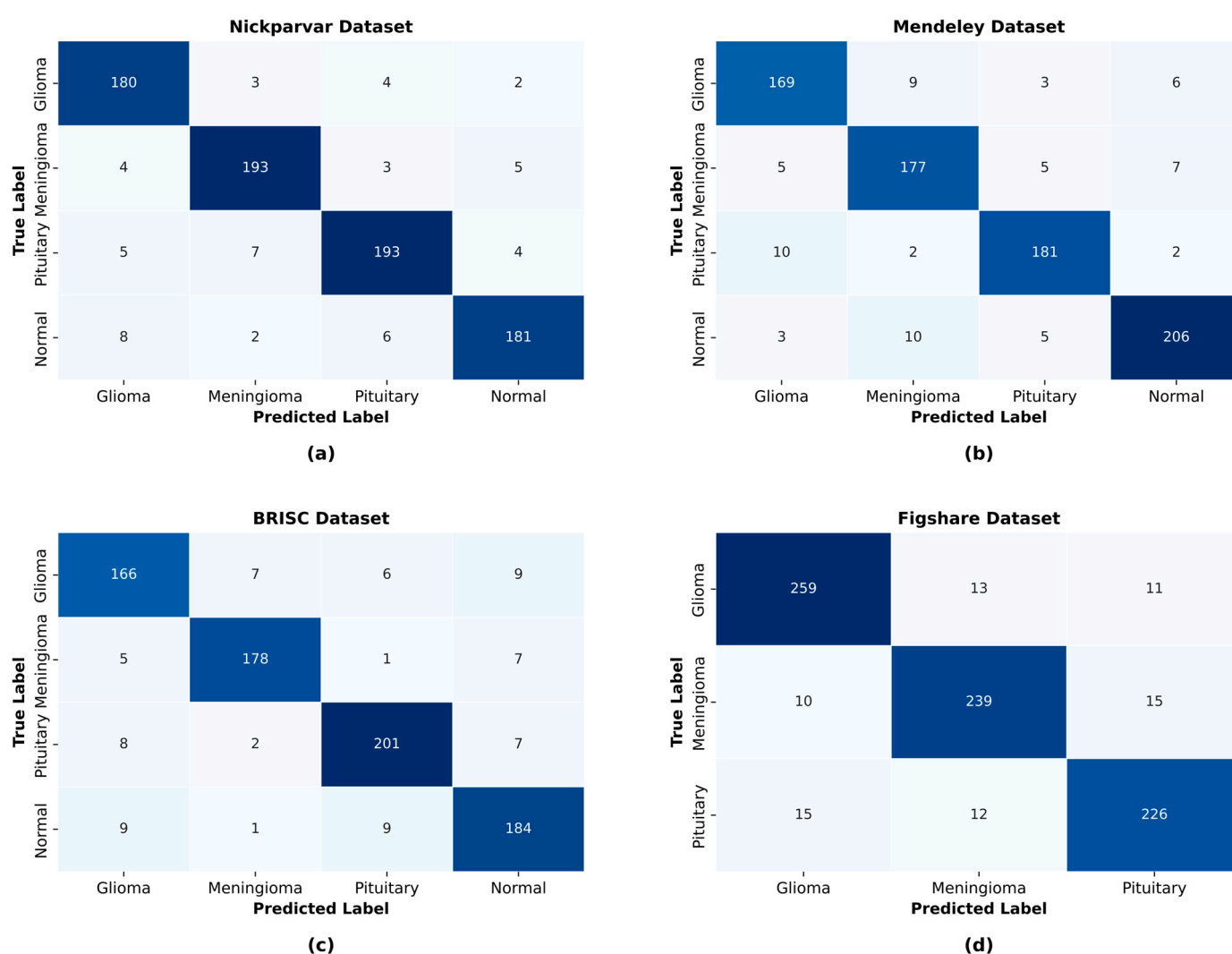


Figure 10. Confusion matrix of tumor classification results for the ETDACVO-optimized model across four MRI datasets. (a) Confusion matrix for the Nickparvar dataset; (b) Confusion matrix for the Mendeley dataset; (c) Confusion matrix for the BRISC dataset; (d) Confusion matrix for the

Figshare dataset. The confusion matrices illustrate the classification performance of the ETDACVO-optimized model across the four tumor categories, showing strong agreement between true and predicted labels and demonstrating robust classification capability across diverse MRI datasets.

3.12.3. Diagnostic Performance Metrics

Using the confusion-matrix values, standard diagnostic metrics were computed:

$$\text{Sensitivity (Recall): Sensitivity} = \frac{TP}{TP+FN} \quad (28)$$

$$\text{Specificity: Specificity} = \frac{TN}{TN+FP} \quad (29)$$

$$\text{Precision: Precision} = \frac{TP}{TP+FP} \quad (30)$$

$$\text{F1-Score: } F_1 = \frac{2TP}{2TP+FP+FN} \quad (31)$$

Sensitivity measures the ability to correctly detect tumor cases, whereas specificity quantifies the ability to correctly identify normal cases. The results in Table 20 demonstrate balanced sensitivity and specificity across all datasets, indicating that the ETDACVO-optimized model reliably identifies both tumor and non-tumor cases while maintaining consistent diagnostic performance.

Table 20. Diagnostic reliability metrics derived from confusion-matrix analysis.

Dataset	Precision	Sensitivity	Specificity	F1-Score
Nickparvar	0.95	0.94	0.96	0.94
Mendeley	0.94	0.93	0.95	0.93
BRISC	0.96	0.95	0.97	0.95
Figshare	0.95	0.94	0.96	0.94

3.12.4. Effect of Convergence-Aware Explainability

To analyze the impact of the proposed explainability mechanism, the Evolutionary Convergence Confidence coefficient (Ψ) was examined across different training stages. During early optimization, lower Ψ values produced diffuse heatmaps with weaker activations. As the evolutionary optimization converged, Ψ increased and generated sharper, more localized tumor activations. This behavior indicates that CA-EA-GradCAM provides trust-calibrated visual explanations, enabling clinicians to interpret both the spatial localization of tumor regions and the confidence associated with model predictions.

3.13. Failure Case Analysis

Figure 11 presents representative failure cases illustrating situations in which the ETDACVO-optimized model produces incorrect or uncertain predictions. In the first example, the tumor region is relatively small and occupies only a limited portion of the MRI slice, resulting in a diffuse activation pattern in the CA-EA-GradCAM heatmap. In the second case, the lesion exhibits low contrast relative to surrounding tissue, causing the saliency map to spread across a broader anatomical region rather than focusing on a well-defined tumor boundary. The third example shows a complex anatomical structure in

which global transformer attention highlights surrounding brain regions, leading to less precise localization. These failure cases indicate that small tumors, low-intensity contrast, and complex anatomical variations can reduce feature separability and weaken gradient-based localization signals. Although ETDACVO partially mitigates these challenges through adaptive augmentation and stabilized optimization dynamics, accurate detection of very small or low-contrast tumors remains challenging. Future work will explore tumor-size-aware fitness weighting and contrast-adaptive augmentation strategies to further improve detection of subtle or complex tumor structures.

Failure Case Analysis of CA-EA-GradCAM Explanations

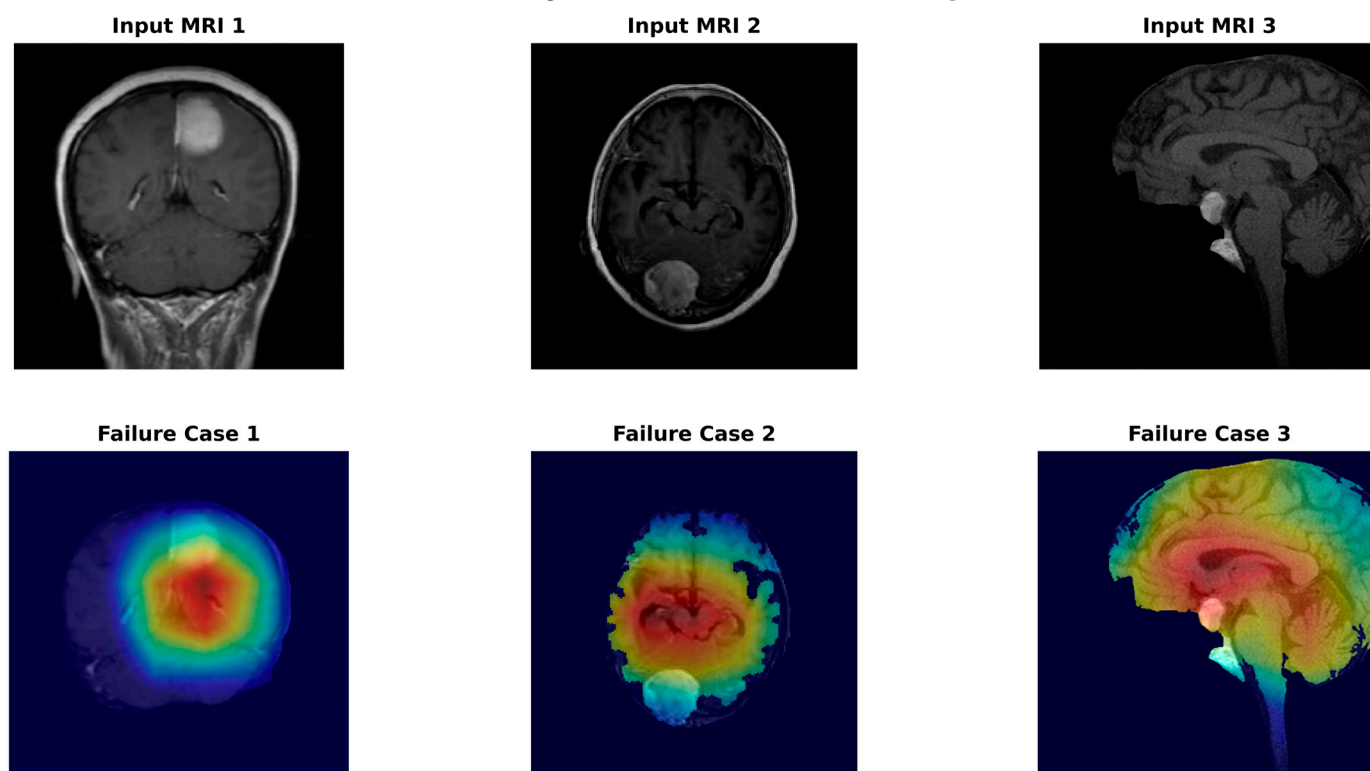


Figure 11. Failure case analysis showing representative MRI inputs and corresponding CA-EA-GradCAM outputs for samples with incorrect or uncertain predictions.

4. Discussion

This study investigated the effectiveness of the proposed ETDACVO optimization framework for brain tumor MRI classification across heterogeneous datasets. The experimental results demonstrate improvements in convergence stability, classification performance, and cross-dataset generalization compared with conventional gradient-based optimization strategies. The discussion examines the implications of these results in terms of optimization dynamics, adaptive augmentation strategies, domain robustness, and explainable medical AI. Although the absolute accuracy improvements appear moderate, such gains are meaningful in medical image classification tasks where baseline performance is already high. Furthermore, ETDACVO simultaneously improves convergence stability, cross-dataset robustness, and model interpretability, providing benefits beyond raw accuracy improvements.

4.1. Convergence Stability and Optimization Dynamics

ETDACVO demonstrated improved convergence behavior compared with conventional gradient-based optimizers. Training curves showed smoother loss reduction and

earlier stabilization, reaching convergence approximately 19–22 epochs earlier than Adam and AMSGrad. The incorporation of EWMA smoothing effectively reduced gradient noise and stabilized late-stage optimization. This resulted in more controlled learning dynamics and reduced oscillation during training. Unlike conventional training pipelines, where augmentation policies and learning-rate schedules are tuned independently, ETDACVO formulates training as a unified evolutionary process. This design allows the model to adapt both the input data distribution and the training trajectory, improving stability under heterogeneous MRI conditions.

4.2. Adaptive Learning-Rate and Augmentation Optimization

Another important advantage of ETDACVO is its ability to adapt training parameters dynamically. The evolutionary mechanism automatically adjusts learning-rate behavior based on model feedback, eliminating the need for manual schedule design. This adaptive control improved convergence and classification accuracy. ETDACVO also optimizes augmentation policies while preserving anatomical fidelity through structural similarity (SSIM). Compared with static augmentation strategies, the ETDACVO-driven policies maintained higher structural consistency while increasing training diversity. This balance between augmentation diversity and anatomical preservation helps prevent the introduction of unrealistic transformations that could degrade diagnostic features.

4.3. Cross-Dataset Generalization

Robustness across heterogeneous imaging conditions is critical for clinical deployment. Cross-dataset transfer experiments demonstrated that ETDACVO maintains approximately 92.8–93.2% performance retention when models are evaluated on unseen datasets. In contrast, baseline optimizers showed significantly larger performance drops. This improved robustness can be attributed to the combination of adaptive augmentation, evolutionary parameter search, and stabilized optimization dynamics.

4.4. Explainability and Clinical Reliability

Beyond predictive performance, the proposed framework integrates a Convergence-Aware Evolutionary Attention Grad-CAM (CA-EA-GradCAM) mechanism to enhance interpretability. By incorporating optimization stability information into the explanation process, CA-EA-GradCAM produces more focused and reliable saliency maps compared with conventional gradient-based visualization methods. Qualitative results show that the generated explanations align well with tumor regions, while quantitative localization metrics indicate improved overlap with ground-truth annotations. By linking explanation intensity to evolutionary convergence confidence, the proposed approach provides clinicians with both spatial localization cues and implicit reliability indicators.

4.5. Practical Implications

From a clinical perspective, reliable tumor localization and stable diagnostic performance across heterogeneous datasets are essential requirements for real-world deployment. The ETDACVO framework contributes to these requirements by improving training stability, enhancing cross-dataset generalization, and providing interpretable predictions. The integration of structural-fidelity-aware augmentation ensures anatomically plausible transformations, while the explainability module provides visual evidence supporting model decisions. Overall, these characteristics suggest that ETDACVO provides a promising optimization framework for reliable and interpretable medical imaging systems, particularly in multi-center environments where imaging protocols and scanner characteristics vary substantially. The improved cross-dataset robustness demonstrated by ETDACVO is particularly relevant for real-world clinical environments, where MRI

scans acquired across hospitals often differ in scanner manufacturers, acquisition protocols, and patient populations.

4.6. Limitations and Future Work

Although ETDACVO demonstrates strong performance across multiple MRI datasets, several limitations remain. First, the current implementation operates on 2D slice-based training with a fixed evolutionary population size ($p = 20$). Extending the framework to 3D volumetric MRI data or larger evolutionary populations may require memory-efficient training strategies such as gradient checkpointing, distributed evolutionary optimization, or adaptive population scheduling. Second, the framework depends on reliable preprocessing steps, including skull stripping, bias-field correction, and intensity normalization. Errors in these preprocessing stages, particularly in low-quality clinical scans or artifact-heavy images, may propagate into the augmentation module and affect anatomical consistency. Third, ETDACVO introduces additional computational cost during training due to the population-based evolutionary search. However, this overhead occurs only during offline model optimization. The final deployed model maintains the same parameter count and introduces minimal inference overhead (+2.7% latency and +0.2 GFLOPs), which does not affect real-time clinical decision support systems. Moreover, the improved convergence stability and reduced hyperparameter tuning effort may offset the additional training cost in large-scale medical imaging pipelines. Future work will explore 3D evolutionary optimization strategies, lightweight surrogate fitness models to reduce evaluation cost, and memory-aware evolutionary scheduling for resource-constrained environments. Furthermore, integrating ETDACVO with federated and multi-center learning frameworks may enable robust cross-site adaptation without centralized data sharing. Finally, incorporating radiologist-guided anatomical priors and uncertainty-aware augmentation constraints could further improve augmentation fidelity and support deployment in clinically regulated medical AI systems.

5. Conclusions

This study introduces ETDACVO, a hybrid evolutionary optimization framework that jointly adapts augmentation policies and learning-rate dynamics to improve stability and generalization in medical image learning. Evaluations across four heterogeneous MRI datasets demonstrated consistent performance gains, including 2.3–2.5% improvements in classification accuracy and faster convergence requiring 19–22 fewer training epochs compared with conventional optimization strategies. Ablation experiments further confirmed the complementary roles of the TDO exploration mechanism, ACVO diversity preservation, and EWMA temporal smoothing in enhancing augmentation fidelity and optimization stability. Furthermore, the proposed CA-EA-GradCAM explainability mechanism provides reliable visual interpretations aligned with tumor regions, improving the transparency and trustworthiness of model predictions. Future work will focus on extending the framework to multi-center clinical datasets, federated learning environments, and 3D volumetric MRI analysis while incorporating radiologist-guided anatomical priors to further improve augmentation fidelity and clinical reliability. Overall, ETDACVO provides a scalable and generalizable optimization framework that supports the development of robust, interpretable, and clinically applicable medical imaging AI systems. The proposed framework is architecture-agnostic and can be integrated with CNN, transformers, or hybrid models, enabling broader applicability across medical imaging classification tasks beyond brain tumor MRI analysis.

Author Contributions: I.K. and A.A.D. conceived and designed the study; I.K., A.A.D., M.A.S.A.-M., and R.M. developed the methodology; I.K. performed the software implementation; I.K., R.M.,

and S.F.K. conducted the formal analysis; M.A.S.A.-M., L.A.G., and S.F.K. carried out the investigation; M.A.S.A.-M. and L.A.G. curated the data; M.I.T. and A.S.A.A. prepared the visualizations; S.F.K. and R.M. performed the validation; I.K. and A.A.D. wrote the original draft; I.K., R.M., M.A.S.A.-M., L.A.G., S.F.K., M.I.T., and A.S.A.A. reviewed and edited the manuscript; A.A.D. supervised the project and administered the research activities. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by the Deanship of Scientific Research at Northern Border University, Arar, KSA, through the project number “NBU-FFR-2026-2903-02”, it also funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R178), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia and Prince Sattam bin Abdulaziz University via funding project number (PSAU/2026/R/1447).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The source code for ETDACVO is publicly available at: <https://github.com/IndrakumarK/Evolutionary-Co-Optimization-of-Data-Augmentation-and-Learning-Dynamics-for-BT-MRI-Classification.git> (accessed on 25 June 2026). The datasets used in this study are publicly available from their respective official repositories. Direct links to all datasets are provided in the manuscript and are also documented in the GitHub repository for reproducibility.

Acknowledgments: The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA, for funding this research work through the project number “NBU-FFR-2026-2903-06”. Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R178), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2026/R/1447). During the preparation of this work, the authors used ChatGPT (OpenAI, San Francisco, CA, USA; <https://chatgpt.com/>, accessed on 28 June 2026) to assist with language refinement and manuscript preparation. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ismael, S.A.A.; Mohammed, A.; Hefny, H. An enhanced deep learning approach for brain cancer MRI images classification using residual networks. *Artif. Intell. Med.* **2020**, *102*, 101779.
2. Mallick, P.K.; Ryu, S.H.; Satapathy, S.K.; Mishra, S.; Nguyen, G.N.; Tiwari, P. Brain MRI image classification for cancer detection using deep wavelet autoencoder-based deep neural network. *IEEE Access* **2019**, *7*, 46278–46287.
3. Ahmed, M.M.; Hossain, M.M.; Islam, M.R.; Ali, M.S.; Nafi, A.A.N.; Ahmed, M.F.; Islam, M.K. Brain tumor detection and classification in MRI using hybrid ViT and GRU model with explainable AI in Southern Bangladesh. *Sci. Rep.* **2024**, *14*, 22797.
4. Nguyen, K.P.; Fatt, C.C.; Treacher, A.; Mellema, C.; Trivedi, M.H.; Montillo, A. Anatomically informed data augmentation for functional MRI with applications to deep learning. In *Medical Imaging 2020: Image Processing*; SPIE: Bellingham, WA, USA, 2020; Volume 11313, pp. 172–177.
5. Abdusalomov, A.B.; Mukhiddinov, M.; Whangbo, T.K. Brain tumor detection based on deep learning approaches and magnetic resonance imaging. *Cancers* **2023**, *15*, 4172.
6. Zeineldin, R.A.; Karar, M.E.; Elshaer, Z.; Coburger, J.; Wirtz, C.R.; Burgert, O.; Mathis-Ullrich, F. Explainability of deep neural networks for MRI analysis of brain tumors. *Int. J. Comput. Assist. Radiol. Surg.* **2022**, *17*, 1673–1683.
7. Tronchin, L.; Vu, M.H.; Soda, P.; Löfstedt, T. LatentAugment: Data augmentation via guided manipulation of GAN’s latent space. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *47*, 11519–11533.
8. Darwish, S.M.; Abu Shaheen, L.J.; Elzoghbi, A.A. A new medical analytical framework for automated detection of MRI brain tumor using evolutionary quantum inspired level set technique. *Bioengineering* **2023**, *10*, 819.

9. Jaspın, K.; Selvan, S. Multiclass convolutional neural network based classification for the diagnosis of brain MRI images. *Biomed. Signal Process. Control* **2023**, *82*, 104542.
10. Cocosco, C.A.; Zijdenbos, A.P.; Evans, A.C. A fully automatic and robust brain MRI tissue classification method. *Med. Image Anal.* **2003**, *7*, 513–527. Correction in *Med. Image Anal.* **2004**, *8*, 93–94.
11. Liu, X.; Liu, Y.; Zhu, Y.; Huang, F.; Li, J.; Chen, S. Structure-aware context-enhanced and dual-path synergistic decoding network for atrophic gastritis segmentation. *Eng. Appl. Artif. Intell.* **2026**, *177*, 115020.
12. Li, L.; Liu, J.; Chen, S.; Wang, J.; Li, Y.; Liao, Q.; Pu, X. Segmentation of acute ischemic stroke lesions based on deep feature fusion. *Inf. Fusion* **2025**, *114*, 102724.
13. Aytaç, U.C.; Güneş, A.; Ajlouni, N. A novel adaptive momentum method for medical image classification using convolutional neural network. *BMC Med. Imaging* **2022**, *22*, 34.
14. Sun, H.; Zhou, W.; Yang, J.; Shao, Y.; Xing, L.; Zhao, Q.; Zhang, L. An improved medical image classification algorithm based on Adam optimizer. *Mathematics* **2024**, *12*, 2509.
15. Manzari, O.N.; Ahmadabadi, H.; Kashiani, H.; Shokouhi, S.B.; Ayatollahi, A. MedViT: A robust vision transformer for generalized medical image classification. *Comput. Biol. Med.* **2023**, *157*, 106791.
16. Ait Amou, M.; Xia, K.; Kamhi, S.; Mouhafid, M. A novel MRI diagnosis method for brain tumor classification based on CNN and Bayesian Optimization. *Healthcare* **2022**, *10*, 494.
17. Al-Ashoor, A.; Lilik, F.; Nagy, S. A systematic analysis of neural networks, fuzzy logic and genetic algorithms in tumor classification. *Appl. Sci.* **2025**, *15*, 5186.
18. Dai, Y.; Gao, Y.; Liu, F. Transmed: Transformers advance multi-modal medical image classification. *Diagnostics* **2021**, *11*, 1384.
19. Pan, H.; Chen, S.; Xiong, H. A high-dimensional feature selection method based on modified Gray Wolf Optimization. *Appl. Soft Comput.* **2023**, *135*, 110031.
20. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359.
21. Simo, A.M.D.; Kouanou, A.T.; Monthe, V.; Nana, M.K.; Lonla, B.M. Introducing a deep learning method for brain tumor classification using MRI data towards better performance. *Inform. Med. Unlocked* **2024**, *44*, 101423.
22. Rasa, S.M.; Islam, M.M.; Talukder, M.A.; Uddin, M.A.; Khalid, M.; Kazi, M.; Kazi, M.Z. Brain tumor classification using fine-tuned transfer learning models on magnetic resonance imaging (MRI) images. *Digit. Health* **2024**, *10*, 20552076241286140.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.