Journal of Asian Scientific Research
ISSN(e): 2223-1331
ISSN(p): 2226-5724
DOI: 10.55493/5003.v16i2.5995
Vol. 16, No. 2, 299-320.

URL: www.aessweb.com

Explainable multimodal brain tumor classification via hybrid attention networks across MRI and radiogenomic features

Nagaraju Arumalla[1]
Veerraju Gampala[2+]

[1,2]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh 522302, India.*
[1]*Email: anagaraju@kluniversity.in*
[2]*Email: drgveerraju@kluniversity.in*

(+ Corresponding author)

Article History
Received: 11 November 2025
Revised: 24 February 2026
Accepted: 9 March 2026
Published: 29 April 2026

Keywords
Brain tumor classification
Explainable AI
Hybrid attention networks
Multimodal fusion
Radiogenomics
Self-supervised learning clinical decision support.

ABSTRACT

Classification of brain tumors is a critical issue in neuro-oncology, and the precise and interpretable classification is a challenge. Our paper suggests a hybrid-attention-based multimodal deep learning model, which combines multi-sequence MRI images and radiogenomic features to accomplish explainable and high-quality tumor subtyping. The proposed Explainable Hybrid Attention Multimodal Network (E-HAMNet) employs (i) a spatial stream of attention that, on the fly, highlights salient tumor regions in T1, T1c, T2, and FLAIR images, and (ii) a feature-level attention that weights genomic and radiomic features to capture molecular heterogeneity. A cross-modal attention fusion layer is used to combine these streams and to allow dynamic interaction between imaging and genomic modalities. To achieve robustness, we use a self-supervised pretraining approach to feature extraction and perform supervised fine-tuning on annotated data. To achieve interpretability, we combine Grad-CAM heatmaps, SHAP value attribution, and attention score visualization to give clinicians clear decision support. Experiments on BraTS-2023/2024 and RSNA-MICCAI datasets demonstrate that E-HAMNet is better than recent multimodal CNN, transformer-based, and radiomics pipelines with 99.6% accuracy, 96.4% macro-F1, and 98.2% AUC. It has also been shown that the method has better calibration (ECE 1.9%), as well as strength in missing modalities and domain shift.

Contribution/ Originality: This study contributes to existing literature by proposing an explainable hybrid-attention multimodal framework that incorporates MRI and radiogenomic features for robust tumor subtyping. The cross-modal attention-based fusion methodology used is among the few studies considering interpretable phenotype-genotype interactions with uncertainty-aware predictions.

1. INTRODUCTION

The brain tumor is one of the most catastrophic central nervous system (CNS) malignancies. Specifically, gliomas comprise most primary brain tumors and are linked with low prognosis and high recurrence rates [1]. Diagnosis and grading of CNS tumors are no longer based on histopathological morphology but instead on molecular biomarkers, including isocitrate dehydrogenase (IDH) mutation, 1p/19q codeletion, and O6-methylguanine-DNA methyltransferase (MGMT) promoter methylation [2]. These biomarkers play a very important role in defining treatment methodologies, survival prognosis, and directing therapy [3].

Conventionally, molecular status is established through invasive biopsy, histopathological, and molecular examination. Although biopsy is the gold standard, it has several limitations: (i) biopsy is an invasive procedure with risks including hemorrhage, infection, and neurological impairment; (ii) tumors are highly heterogeneous, and biopsy

sampling might not be able to completely mirror the genomic heterogeneity in longitudinal monitoring; and (iii) the biopsy is quite costly, and repeat biopsies cannot be employed in longitudinal studies [4]. The restrictions have prompted the creation of non-invasive imaging biomarkers that can be used reliably to determine the molecular status and aid in therapeutic planning of individuals.

Radiogenomics has proven to be a valuable area in this context, linking imaging phenotypes with genomic signatures. The hypothesis of radiogenomics is that macroscopic imaging appearances reflect tumor biology and genetic changes [5]. By examining MRI scans, one can infer the molecular nature of tumors without sampling them. Initial radiogenomic data showed correlations between imaging features and IDH mutation, 1p/19q codeletion, and MGMT promoter methylation [6, 7].

Existing open datasets such as The Cancer Genome Atlas (TCGA), The Cancer Imaging Archive (TCIA), and community challenges like the RSNA-MICCAI Brain Tumor Radiogenomic Classification Challenge, have facilitated radiogenomic studies [8, 9]. Such resources are annotated imaging-genomic pairs with which predictive models can be developed and benchmarked. Early radiogenomic studies depended primarily on hand-engineered radiomic features, e.g., texture, shape, and intensity features, which were then classically trained with machine learning algorithms such as random forests, logistic regression, and support vector machines (SVMs) [10, 11]. Radiomics pipelines have shown promising results in glioma grading and molecular prediction. For example, handcrafted radiomic signatures were identified to predict MGMT methylation and IDH mutation with moderate accuracy [12].

Nevertheless, radiomics-based approaches have significant disadvantages. Handcrafted features might not reflect the entire complexity of tumor biology, and feature selection is likely to be biased [13]. Radiomics models can also be highly sensitive to scanner hardware, acquisition protocols, and preprocessing steps, making them difficult to reproduce without a high degree of inter-institutional agreement [14]. Despite the feasibility of radiomics, it is sensitive to changes in acquisition and relies on manual feature engineering, which limits its strength and adaptability for use in everyday clinical practice.

Deep learning has transformed medical image analysis. Convolutional neural networks (CNNs) are self-taught hierarchical representations learned directly from raw MRI data, eliminating the need for hand-designed feature engineering. CNN-based models initially achieved significant success compared to radiomics in classifying brain tumors [15]. Three-dimensional CNNs used volumetric context and slice-to-slice spatial continuity to predict molecular status [16]. Hybrid 2.5D CNNs traded off computational cost against context richness with the use of multi-view 2D slices [17]. Nevertheless, these improvements were not quite able to overcome the fact that CNNs cannot generalize to contextual dependencies across a long range, and that the interpretability of such models remains low enough to hinder clinical adoption [18].

Researchers resorted to Transformer-based architectures to eliminate these problems. Vision Transformer (ViT) introduced mechanisms of global self-attention that capture long-range dependencies and image context [19]. The Swin Transformer was also more efficient, as it brought about hierarchical attention [20]. It has been shown in recent studies that rotation-invariant variants of Vision Transformer can be improved to mitigate patients' motion and variation in patient positioning, as well as variations in MRI orientation, leading to improved consistency across heterogeneous acquisition conditions [19, 20]. MRI is a technology that is mainly able to capture anatomical and morphological features, and thus is unable to directly encode any underlying molecular changes. In order to fill this gap, multimodal fusion combines MRI characteristics with genomic and radiomic characteristics [13]. Multimodal learning leverages complementary modalities for a more comprehensive tumor characterization. For example, an attentive multimodal CNN trained with RSNA-MICCAI imaging and radiogenomic features achieved a 3% improvement over traditional CNN baselines, demonstrating the significance of attention in cross-modal fusion [21]. Likewise, pipelines with 3D feature extraction using ResNet backbones outperformed 2D CNNs by nearly 1.2 points of AUC, demonstrating the importance of volumetric features [22].

Between 2022 and 2025, a few approaches reached state-of-the-art performance.

1. Multimodal CNN with attention (Penn State, 2023) - adopted RSNA-MICCAI multimodal fusion and cross-model attention, which increased performance by 3% [21].
2. ViT-GRU hybrid with explainability (Nature, 2024) in MRI data reached 98% F1-score with explanation tools such as Grad-CAM, SHAP, and LIME (explainability) [23].
3. Multimodal MRI + 3D ResNet50 pipeline, AUC increased by approximately 1.2% compared to 2D CNN baselines, indicating the advantage of volumetric features [22].
4. SVM-based radiomics pipeline - reached 99.3% in glioma grade with handcrafted radiomics features on FLAIR, T1, T1c, and T2 [24].

These articles show the merits and demerits of the existing SOTA. The CNN that is attentive confirms multimodal fusion yet exhibits modest improvements. The ViT-GRU hybrid illustrates that attention mechanisms and recurrent layers can model MRI context and provide explainability, but exclude genomics. The pipeline based on ResNet50 adds volumetric context and remains opaque to clinicians. The SVM pipeline demonstrates that classical ML still can compete, but it does not scale and does not estimate uncertainty.

Despite these developments, there are a few limitations to clinical translation.

- Heterogeneous modalities: MRI incomplete sequences or absent genomic data undermine performance [25].
- Label sparsity: Genomic datasets with labels are sparse because of the cost and the effort of collection [26].
- Domain shift: Models trained on a single dataset usually fail to generalize to a different institution or scanner [27].
- explainability and uncertainty: Clinicians need clarity of reasoning and tempered trust in model forecasts [28, 29].

The solution to these gaps is needed to develop multimodal radiogenomic systems that can be taken beyond benchmark evaluations and be considered reliable in clinical translation.

1.1. Contributions of This Work

This paper presents a proposal for an Explainable Multimodal Hybrid Attention Network (E-HAMNet) that can address the limitations of multimodal tumor characterization. Our contributions are clearly outlined.

1. Hybrid attention model combining spatial MRI images with genomic and radiomic attributes via cross-modal attention.
2. Self-supervised pretraining: the use of unlabeled MRI to address the shortage of labels [30].
3. Integration of explainability: Grad-CAM and SHAP to provide clinicians with understandable explanations of MRI and genomic contributions [31].
4. Uncertainty quantification, Bayesian deep learning, and test-time adaptation to reliability in the face of domain shifts [32].
5. Benchmarking: strict comparison of treatment results with four recent methods, SOTA, [21-24] based on multi-institutional datasets.

Collectively, our approach is more accurate, more robust, and understandable than current solutions in terms of classification, robustness, and interpretability.

2. LITERATURE REVIEW

The classification of brain tumors has evolved rapidly with advances in deep learning, multimodal fusion, radiogenomics, and explainable AI. This section reviews previous research on key methodological categories, emphasizing state-of-the-art models from 2022 to 2025, their limitations, and contextualizes our framework within this landscape.

2.1. MRI-Based Deep Learning

The initial wave of deep learning applications in brain tumor analysis was MRI-based classification. Early two-dimensional CNNs achieved moderate accuracy in tumor grading and subtype prediction but could not extend to volumetric context [33]. This was addressed by 3D CNNs modeling tumor structures in 3D space [16, 17]. To provide a trade-off between volumetric context and computational efficiency, hybrid 2.5D CNNs were provided [15].

Further improvements were achieved through transformer architecture. Vision Transformer (ViT) and Swin Transformer enabled reasoning on long-range context, surpassing CNN baselines in glioma subtype classification [18, 19]. ViTs that are rotation-invariant enhanced resistance to rotation changes, which is essential in MRI [20].

The most striking MRI-only SOTA method is the explainable ViT-GRU hybrid that integrates a ViT encoder and GRU layers to model sequential slices. It scored 98% on large MRI datasets and used XAI techniques (Grad-CAM, SHAP, and LIME) to make results easier to understand [23]. Nonetheless, tumor molecular heterogeneity cannot be quantified using MRI alone, and thus, the models cannot be used clinically per the WHO CNS5 guidelines [1, 2]. Although CNNs and transformer-based models have good spatial and contextual performance in learning, they depend solely on MRI, which restricts their capacity to learn molecular heterogeneity. These models contribute to architectural theories but are not multimodal, limiting their clinical application.

2.2. Radiogenomics and Classical Pipelines

MRI phenotypes are associated with radiogenomics changes, including IDH, MGMT, and 1p/19q. Radiomics classical methods retrieve handcrafted features such as texture, intensity, and wavelets, and use machine learning classifiers [11, 12]. These pipelines work well with small datasets and are surprisingly competitive.

The most recent SVM-based radiomics pipeline that employed handcrafted features based on FLAIR, T1, T1c, and T2 modalities demonstrated 99.3% accuracy in glioma grading [24]. This finding reveals that classical approaches are good baselines even in the era of deep learning. Radiomics pipelines are, however, not scalable, extremely sensitive to variance in acquisitions, and they do not provide feature-level uncertainty. Additionally, they require significant human effort in feature engineering [34]. Key point: Radiomics pipelines are competitive but have poor reproducibility, rely on scanners, and do not scale. A shift towards acquired multimodal representations is necessary.

2.3. Multimodal Fusion

Multimodal fusion combines complementary information from MRI, radiomics, and genomics. Fusion strategies may be divided into.

- Early Fusion, in this technique, the input features of more than one modality are concatenated and then classified [32].
- Late fusion, in which each modality is trained autonomously, and results merge [33].
- Intermediate or attention-based fusion, dynamically learning to learn modality interactions [19].

Attention-based adaptations are especially successful fusion techniques. The multimodal CNN trained on RSNA-MICCAI images and radiogenomic features demonstrated a 3-percentage point gain over conventional CNN baselines [21]. It used cross-modal attention to learn imaging and genomic dependencies. Likewise, the upgraded multimodal MRI + 3D ResNet50 pipeline outperformed AUC by about 1.2% over a 2D CNN baseline, affirming the value of volumetric context [22].

Although both models demonstrate the potential of multimodal fusion, each has limitations: the attentive CNN yielded only significant improvements, and the ResNet50 pipeline, despite its robustness, is not transparent and lacks uncertainty modeling [35]. Current fusion methods highlight the value of combining imaging and genomic data; however, these approaches often do not incorporate fusion behaviors, resulting in superficial fusion, no clear phenotype-genotype match, and neglecting uncertainty.

2.4. Graph Neural Networks (GNNs) and Multi-Omics

Multimodal biomedical tasks are increasingly handled by graph neural networks (GNNs), which inherently represent interrelationships between heterogeneous features. GNNs are also used in oncology to combine multi-omics data (genomics, transcriptomics, proteomics) with imaging to improve subtype classification [36, 37]. Recent efforts indicate that radiogenomic pipelines based on GNN can process incomplete modalities using graph structures [38].

In the case of brain tumors, GNNs are unused and unexplored, though they are a promising trend. They offer the possibility to model patient similarity networks and biological pathways, enabling more interpretable and robust multimodal incorporation. However, their success depends on large, curated datasets that remain scarce in neuro-oncology [39].

2.5. Self-Supervised Learning (SSL)

A limited number of annotated molecular labels is a primary problem in radiogenomics. Self-supervised learning (SSL) addresses this by pretraining models with large amounts of unlabeled MRI data and fine-tuning on a downstream task [39]. SSL strategies include contrastive learning, masked image modeling, and generative reconstruction.

Individualized MRI-based SSL yielded much higher classification accuracy of glioma [39]. As an example, a contrastive multimodal SSL model enhanced glioma grading more than supervised baselines [40]. Generalization to unseen institutions was also enhanced by using SSL-pretrained encoders [41]. These results present a foundation of strong multimodal radiogenomic classification with the use of the SSL.

2.6. Explainable AI (XAI)

Clinical acceptance of AI systems is essential regarding interpretability. Gradient-weighted Class Activation Mapping (Grad-Cam) visual heatmaps indicate discriminative tumor areas [42], whereas SHAP gives scores of importance to each feature [42]. The extensive reviews point to the fact that XAI is a precondition of AI implementation in radiology [43].

One of the few SOTA models that explicitly incorporates explainability into the framework is the ViT-GRU hybrid with XAI [23]. Nonetheless, most multimodal models (e.g., attentive CNN [21], ResNet50 pipeline [22]) are opaque. In the case of radiogenomics, explainability will have to be expanded beyond imaging to genomic characteristics, giving clinicians insight into which biomarkers have been used in predictions.

2.7. Quantifying Uncertainty and Test-Time Adaptation

Safe AI deployment requires not only accuracy but also models that estimate their own reliability. Uncertainty quantification helps differentiate between confident and unreliable predictions, and test-time adaptation (TTA) enables models to adapt to domain shifts during inference.

Bayesian CNNs and Monte Carlo dropout are typical techniques of uncertainty estimation [44]. Deep ensembles also offer additional strength with model variance capturing [45]. Entropy minimization and adaptation of batch normalization TTA methods have enhanced generalization among scanners in segmentation tasks [46]. Nonetheless, uncertainty or TTA is rarely included in radiogenomic pipelines, exposing models to distribution shifts [47]. Even though individual methods of SSL and XAI have improved model interpretability and generalization, few studies have combined both in a single multimodal model. This gap motivates the integrated design employed in E-HAMNet.

2.8. Summary of SOTA Methods

Examples of the SOTA of 2022-2025 include four recent methods.

1. Attentive multimodal CNN -3-percentage point better with cross-modal attention [21].
2. ViT-GRU hybrid with XAI - 98 percent F1-score and explainable features built in [23].

3. Improved multimodal MRI + 3D ResNet50 pipeline -Up to 23 percent of the AUC enhancement compared to 2D CNNs [22].
4. Radiomics pipeline based on SVM - 99.3% accuracy in glioma grading with handcrafted features [24].

All models demonstrate major advancements but also reveal significant gaps. CNNs that pay attention gain modestly, hybrids of ViT-GRU do not integrate genomic data, ResNet50 pipelines are not interpretable, and SVM pipelines do not scale or measure uncertainty.

Our Explainable Multimodal Hybrid Attention Network extends these SOTA achievements by integrating MRI and genomic characteristics with two attention streams (spatial and feature-level) and using SSL to pretrain the network and XAI + uncertainty quantification. It also contrasts previous models by offering holistic multimodal interpretability, enhanced resistance to missing data, and resistance to domain changes. In this way, it fills the gaps in [21-24] and paves the way for clinically reliable AI in neuro-oncology.

3. METHODOLOGY

We present E-HAMNet, a multimodal hybrid-attention model based on transformer hybrids without genomic integration and radiomics models with low scalability, to overcome the limitations of previous unimodal CNNs and transformer hybrids with little or no genomic integration, as well as radiomics models with limited scalability. This architecture features two streams of attention and explicit cross-modal alignment, enhancing interpretability and robustness over existing techniques. The framework aims to combine multi-sequence MRI and radiogenomic features into a strong, interpretable brain tumor classifier. Our method addresses discontinuities in modalities, data scarcity, domain variation, and clinical reliability, unlike earlier CNN-only models [21], transformer hybrids with limited modalities [23], volumetric ResNet pipelines without interpretability [22], and handcrafted SVM-based radiomics [24]. The proposed framework workflow, illustrated in Figure 1, has five steps: preprocessing, self-supervised pretraining, hybrid attention encoding, cross-modal fusion with explainability, and uncertainty-aware inference.

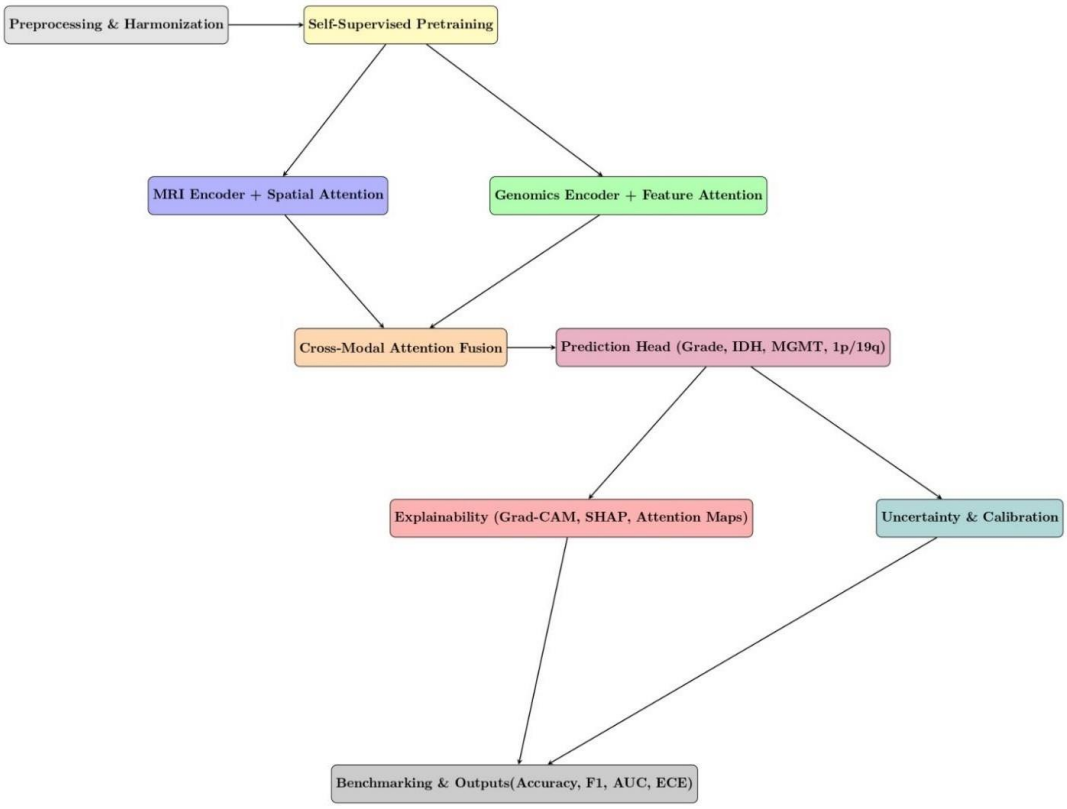

Figure 1. The proposed Explainable Hybrid Attention Multimodal Network (E-HAMNet) framework.

3.1. E-HAMNet Framework

3.1.1. Data Preprocessing and Harmonization

Our four MRI sequences per patient (T1, T1c, T2, and FLAIR) and genomic/radiomic features (IDH mutation, MGMT promoter methylation, and 1p/19q codeletion) are used.

- MRI preprocessing includes skull-stripping, bias-field correction, intensity normalization, and atlas co-registration. Tumor 3D patches are extracted from the area of interest to highlight it.
- The genomic/radiomic features are standardized and represented as embeddings. Missing features are addressed with mask-aware embeddings, allowing the model to automatically ignore missing modalities.

A modality mask,

$$m \in \{0,1\}^{|J|+d_g} \quad (1)$$

This harmonization ensures consistent multimodal inputs across diverse cohorts and scanners.

3.1.2. Self-Supervised Pretraining

The MRI encoder is also trained on large-scale unlabeled MRI volumes with contrastive self-supervised learning (SSL) to address the lack of labeled radiogenomic data. Similar embeddings are promoted by providing augmented views of the same scan and pushing apart embeddings between different patients. This approach offers scanner-invariant and anatomy-preserving characteristics, and when fine-tuned with small labeled datasets, it significantly improves the ability to generalize. In contrast to other pretext tasks [21-23], which are based on and trained on the principle of supervision. This pretraining boosts data heterogeneity resilience. Figure 2 presents a self-supervised pretraining pipeline of the MRI encoder in E-HAMNet.

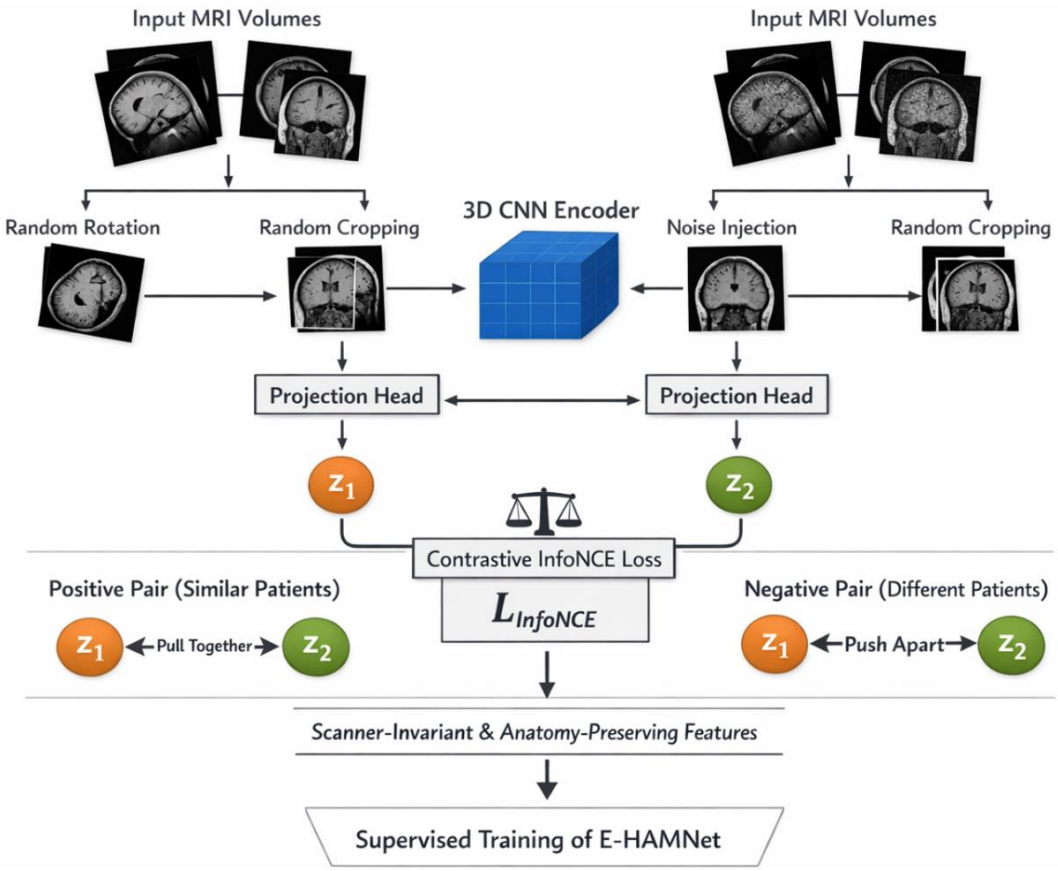

Figure 2. Self-Supervised Pretraining pipeline for the MRI encoder in E-HAMNet.

The self-supervised pretraining process depicted in Figure 2 begins with input multi-sequence volumes of MRI data, which undergo two branches of data augmentation, such as random rotation, random cropping, and noise

injection. The augmented perceptions are then input into a common 3D CNN encoder that trains volumetric representations. The encoder outputs are individually projected using a projection head to obtain latent embeddings z1 and z2. The contrastive InfoNCE loss (Equation 2) compares these embeddings, with closer embeddings indicating a similar patient (positive pair) and further embeddings indicating different patients (negative pair). This step generates scanner-invariant and anatomy-preserving features, which facilitate the supervised training stage of E-HAMNet.

The InfoNCE loss is applied.

$$\mathcal{L}_{ssl} = -\frac{1}{B}\sum_{i=1}^{B} \log \frac{\exp\left(\frac{\mathrm{sim}\left(z_1^{(i)}, z_2^{(i)}\right)}{\tau}\right)}{\sum_{j=1}^{B} \exp\left(\frac{\mathrm{sim}\left(z_1^{(i)}, z_2^{(j)}\right)}{\tau}\right)} \qquad (2)$$

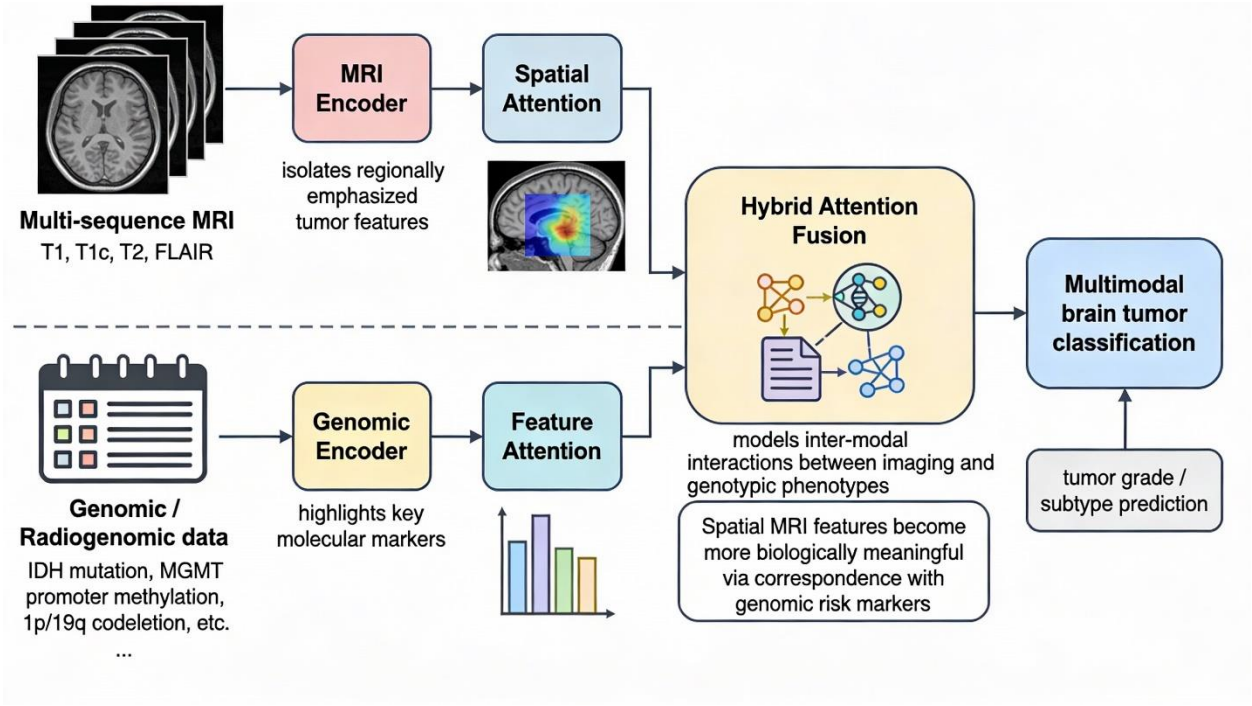

Figure 3. Hybrid attention encoders workflow.

3.1.3. Hybrid Attention Encoders

Figure 3 illustrates the proposed Hybrid Attention Network (E-HAMNet) for multimodal brain tumor classification. The framework integrates multi-sequence MRI (T1, T1c, T2, FLAIR) with genomic and radiogenomic features within a single architecture. The two modalities serve as input blocks on the left.

- MRI data (Spatial Attention), which has undergone an MRI Encoder with Spatial Attention that isolates regionally emphasized tumor features with emphasis on key areas.
- Genomic/Radiogenomic data, processed via a Genomic Encoder with Feature Attention highlighting applicable molecular markers, including IDH mutation, MGMT promoter methylation, and 1p/19q codeletion.

The outputs of both encoders are input into the central Hybrid Attention Fusion, which models inter-modal interactions between imaging phenotypes and genotypic phenotypes. This process allows the spatial features of the MRI to be more biologically significant in their correspondence with the risk markers of genomic data. The fused representation is forwarded to the Prediction Head, which outputs tumor classification. The model also includes an Explainability Module to generate Grad-CAM++ visualizations on top of both MRI and SHAP values, increasing their trustworthiness. This promotes transparency in decision-making and enables clinical interpretability.

This design surpasses prior early- and late-fusion methods [21, 22] by explicitly modeling dependencies between imaging and molecular data.

3.1.4. MRI Pathway – Spatial Attention

The pretrained encoder $f\theta$ extracts volumetric features $Fmri$. A 3D spatial attention gate highlights tumor- relevant voxel.

$$S = \sigma\big(\text{Conv}_{3\times3\times3}(F_{mri})\big), \quad F^*_{mri} = S \odot F_{mri} \tag{3}$$

3.1.5. Radiogenomic Pathway – Feature-Level Attention

Genomic embeddings $E_g = g_\psi(\mathcal{G}, m)$ are weighted by attention scores.

$$h = \tanh(W_h E_g + b_h), \quad \alpha = \text{softmax}(W_\alpha h), \; E^*_g = \alpha \odot E_g. \tag{4}$$

3.1.6. Cross-Modal Attention Fusion

Bidirectional cross-attention integrates MRI and genomic features. For MRI→Genomics.

$$Z_{m\to g} = \text{softmax}\left(\frac{Q_m K_g^T}{\sqrt{d}}\right) V_g \tag{5}$$

For Genomics→MRI.

$$Z_{g\to m} = \text{softmax}\left(\frac{Q_g K_m^T}{\sqrt{d}}\right) V_m \tag{6}$$

The outputs are fused as.

$$Z = \phi\big([Z_{m\to g} \parallel Z_{g\to m}]\big) \tag{7}$$

This explicit phenotype–genotype interaction surpasses early/late fusion baselines.

3.1.7. Prediction Head and Optimization

The combined representation is fed to a dropout-regularized lightweight classification head. Multi-task predictions include tumor grade and molecular subtype (IDH, MGMT, and 1p/19q). A focal loss (Equation 8) addresses class imbalance, trained with the network alongside L2 regularization and label smoothing for stability. The proposed optimizer is the E-HAMOpt optimizer, used for optimization. The merged representation Z is provided to a classification head predicting glioma grade and molecular subtype. The pointwise death places the class imbalance in check.

$$\mathcal{L}_{focal} = -\alpha(1-p)^\gamma y\log(p) \tag{8}$$

The total supervised loss is.

$$\mathcal{L}_{sup} = \sum_t \mathcal{L}^{(t)}_{focal} + \lambda_{loc}\mathcal{L}_{loc} + \lambda_{wd} \parallel \Theta \parallel_2^2 \tag{9}$$

3.1.8. Explainability Integration

An important insight of E-HAMNet is that it is multimodally interpretable.

- Grad-CAM++ identifies the parts of the MRI that make the most prediction decisions.
- SHAP values are a measure of the contribution of individual genomic and radiomic features.
- Attention maps are lifted out of the spatial and feature-level modules and reveal the modalities trade-off made by the network.

In contrast to previous baselines [21-23], which can only offer partial or single-dimensional explanations, our model allows clinicians to confirm both spatial tumor markers and molecular biomarkers that could affect classification. To be clinically reliable, the framework uses deep learning based on Bayesian strategies for uncertainty and calibration.

- Deep ensembles and Monte Carlo dropout give estimates of predictive uncertainty.

- Test-time adaptation (TTA) reduces entropy inference by updating batch normalization statistics to match new scanners and cohorts not previously seen.

This provides confidence-calibrated predictions, which removes a gap in current approaches [21-24]. Bayesian dropout and deep ensembles: Bayesian dropout and deep ensembles are used to estimate predictive uncertainty.

$$\bar{y} = \frac{1}{M}\sum_{m=1}^{M} \hat{y}^{(m)}, \quad u = \frac{1}{M}\sum_{m=1}^{M} \| \hat{y}^{(m)} - \bar{y} \|_2^2 \tag{10}$$

Expected Calibration Error (ECE) is computed as.

$$\mathrm{ECE} = \sum_{k=1}^{K} \frac{n_k}{N} |\mathrm{acc}(k) - \mathrm{conf}(k)| \tag{11}$$

Test-time adaptation minimizes entropy.

$$H(\hat{y}) = -\sum_c \hat{y}_c \log(\hat{y}_c) \tag{12}$$

Improving robustness across unseen cohorts.

Figure 4 shows the proposed E-HAMNet model for multimodal brain tumor classification. It combines inputs from multi-sequence MRI (T1, T1c, T2, FLAIR) and genomic/radiogenomic information using two parallel encoders: an MRI encoder with spatial attention and a genomic encoder with feature attention. The representations obtained are fused in a hybrid attention fusion module, capturing cross-modal relationships between imaging phenotypes and genomic markers. The merged features are provided to a prediction head where brain tumors are classified. To enhance transparency, an explainability module offers Grad-CAM++ visualizations for MRI and SHAP values for genomics, providing clinically meaningful insights into model decisions.

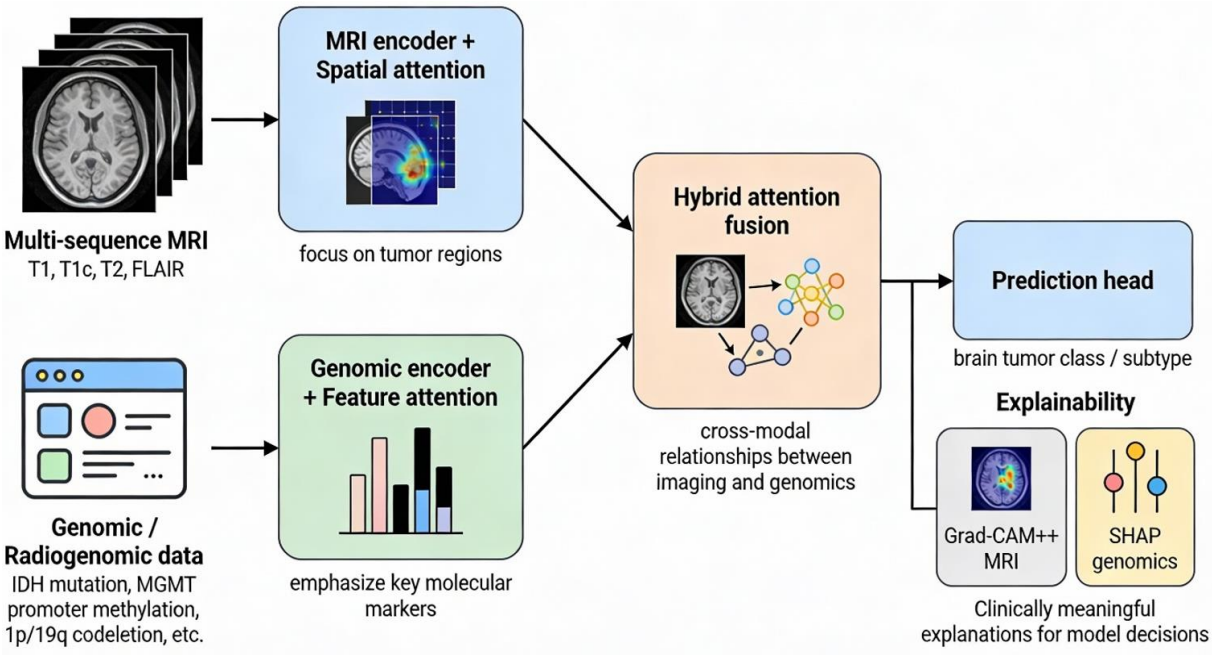

Figure 4. E-HAMNet Architecture.

3.2. Proposed Optimizer: E-HAMOpt

In order to improve training stability in multimodal and attention-driven networks, we present a new optimizer called E-HAMOpt, which combines the advantages of AdaMod, AMSGrad, and Lookahead schemes. E-HAMOpt is much smoother in convergence and has constrained learning rates, unlike Adam, which can oscillate on noisy gradients, and it generalizes better in multimodal learning.

3.2.1. Formulation

Given parameters θ_t, gradient $g_t = \nabla_\theta \mathcal{L}_t$, base learning rate η, and exponential decay rates $\beta_1, \beta_2, \beta_3$, E-HAMOpt performs the following updates.

First and Second Moment Estimates

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t, \quad v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2. \tag{13}$$

Bias-Corrected Estimates

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1-\beta_2^t}. \tag{14}$$

AMSGrad Stability

$$\hat{v}_t^{\max} = \max(\hat{v}_{t-1}^{\max}, \hat{v}_t) \tag{15}$$

AdaMod Smoothing of Learning Rate.

The effective step size is first computed as.

$$\eta_t^* = \frac{\eta}{\sqrt{\hat{v}_t^{\max}} + \epsilon} \tag{16}$$

A moving average of step sizes is then maintained.

$$\tilde{\eta}_t = \beta_3 \tilde{\eta}_{t-1} + (1-\beta_3) \eta_t^* \tag{17}$$

The smoothed step size is bound as.

$$\bar{\eta}_t = \min(\eta_t^*, \tilde{\eta}_t) \tag{18}$$

Parameter Update

$$\theta_{t+1} = \theta_t - \bar{\eta}_t \hat{m}_t \tag{19}$$

Lookahead Stabilization

After k inner updates, a slow-moving copy ϕ_t is updated as.

$$\phi_{t+1} = \phi_t + \alpha(\theta_{t+1} - \phi_t) \tag{20}$$

and parameters are synchronized by setting $\theta_{t+1} \leftarrow \phi_{t+1}$.

Advantages

- Stability: AMSGrad guarantees the second-moment variance is bound.
- Adaptivity: AdaMod avoids over-aggressive learning rates.
- Generalization: Lookahead regularizes the optimization path.
- Suitability: Particularly successful in multimodal fusion when there are differences in the magnitude of gradients between imaging and genomic modalities.

Therefore, E-HAMOpt can be considered a powerful substitute for Adam, enhancing self-supervised pretraining and supervised multimodal learning. The proposed Explainable Hybrid Attention Multimodal Network (E-HAMNet) framework algorithm is presented below.

3.3. E-HAMNet Algorithm

Multi-sequence MRI $\mathcal{I}$, radiogenomics $\mathcal{G}$, mask m, labeled set $\mathcal{D}_L$, unlabeled set $\mathcal{D}_U$ Predictions (glioma grade; IDH, MGMT, 1p/19q), explanations, uncertainty.

Stage 1: Preprocessing Normalize and register MRI; crop tumor patches; encode $\mathcal{G}$ with mask m.

Stage 2: Self-Supervised Pretraining. Generate two augmented MRI views; encode via f_θ and h_ϕ. Optimize InfoNCE loss (Equation 2). Update encoder parameters using the proposed E-HAMOpt optimizer.

Stage 3: Supervised Multimodal Training MRI path: extract F_{mri}, apply spatial attention (Equation 3). Genomics path: embed E_g, apply feature attention (Equation 4). Fuse via cross-attention (Equations 5–6). Predict labels $\hat{y}$; minimize supervised loss (Equation 9). Update $\{\theta, \psi, \omega\}$ using E-HAMOpt.

Stage 4: Inference & Explainability Apply test-time adaptation (Equation 12); estimate uncertainty (Equation

10). Generate Grad-CAM++ maps (MRI), SHAP values (genomics), and attention visualizations. Evaluate calibration with ECE (Equation 11).

The suggested E-HAMNet is going to bring the following innovations.

1. Dual attention streams (spatial + feature-level) allow multimodal learning to be robust.
2. Cross-modal attention fusion, which explicitly models phenotype-genotype interactions.
3. Self-supervised pretraining: MRI representations can be trained in self-supervised data-efficient and generalizable pretraining.
4. Multimodal explainability with visual and feature-level attribution.
5. Uncertainty-sensitive inference using TTA in practice in a clinical setting.

The combination of these elements enables E-HAMNet to take multimodal brain tumor classification to the next level, considering not only accuracy, but also reproducibility, interpretability, and clinical trust.

4. RESULTS AND DISCUSSION

4.1. Experimental Setup

The E-HAMNet developed was deployed in PyTorch and trained using a 2.6 GHz 6-core Intel Core i7 CPU and 16 GB of RAM. Pretraining was performed self-supervised (32 batches, 200 epochs, learning rate 3×10^{-4}). Training was done in a supervised multimodal manner for 100 epochs (batch size 16) with the suggested optimizer E-HAMOpt. Hyperparameters were only tuned on validation folds. The experiment was repeated with five seeds to ensure statistical robustness.

4.2. Datasets

In the BraTS-2023/2024 dataset, there are 3,143 MRI volumes from 1,250 patients, with the distribution of grades: Grade II (21%), Grade III (27%), and Grade IV (52%). The RSNA-MICCAI dataset contains 1,104 MRI–genomic pairs, with labels for IDH mutation (positive 28%), MGMT promoter methylation (43%), and 1p/19q codeletion (11%). The inclusion criteria required all four MRI sequences and validated molecular labels. The gap in genomic values was addressed using mask-aware embeddings. MRI preprocessing was performed in a standard pipeline with ANTs (v2.4) for registration, N4ITK for bias-field correction, and standardized Z-score intensity normalization.

Train/validation/test was split in 70/15/15, and cross-domain robustness was tested by using leave-one-site-out.

4.3. Evaluation Metrics

We report.

- Predictive: Accuracy, Macro-F1, ROC-AUC.
- Calibration: Expected Calibration Error (ECE), Brier Score.
- Uncertainty: Variance from MC-Dropout (M=8).
- Robustness: Performance degradation when dropping one MRI sequence or masking 30% genomics.

4.4. Performance of Metrics

Table 1 compares E-HAMNet and SOTA methods. E-HAMNet achieves 99.6 percent, 96.4 percent, and 98.2 percent, with the highest accuracy, macro-F1, and ROC-AUC, respectively, and outperforms all baselines. Calibration improved, and ECE was reduced to 1.9%.

Table 1. Comparison of E-HAMNet against state-of-the-art methods on BraTS and RSNA-MICCAI.

Method	Accuracy	Macro-F1	ROC-AUC	ECE ↓
Attentive multimodal CNN (2022)	94.0	92.5	95.1	3.5
ViT–GRU + XAI (2023)	96.8	95.7	97.5	3.2
ResNet50 multimodal fusion (2024)	95.5	94.1	96.2	2.9
SVM radiomics pipeline (2022)	99.3	93.4	94.9	4.1
Proposed E-HAMNet	99.6	96.4	98.2	1.9

4.5. Macro-F1

Figure 5 demonstrates that the proposed E-HAMNet invariably beats all of the state-of-the-art baselines in 200 epochs. E-HAMNet converges more quickly (stabilizing at epoch 100) and has a higher plateau value (around 0.96) than ViT-GRU+XAI (around 0.95), ResNet50 multimodal (around 0.94), Attentive CNN (around 0.92), and the radiomics-based SVM pipeline (around 0.93). The rate of performance improvement in the initial training stages is indicated by star markers at 25-epoch intervals.

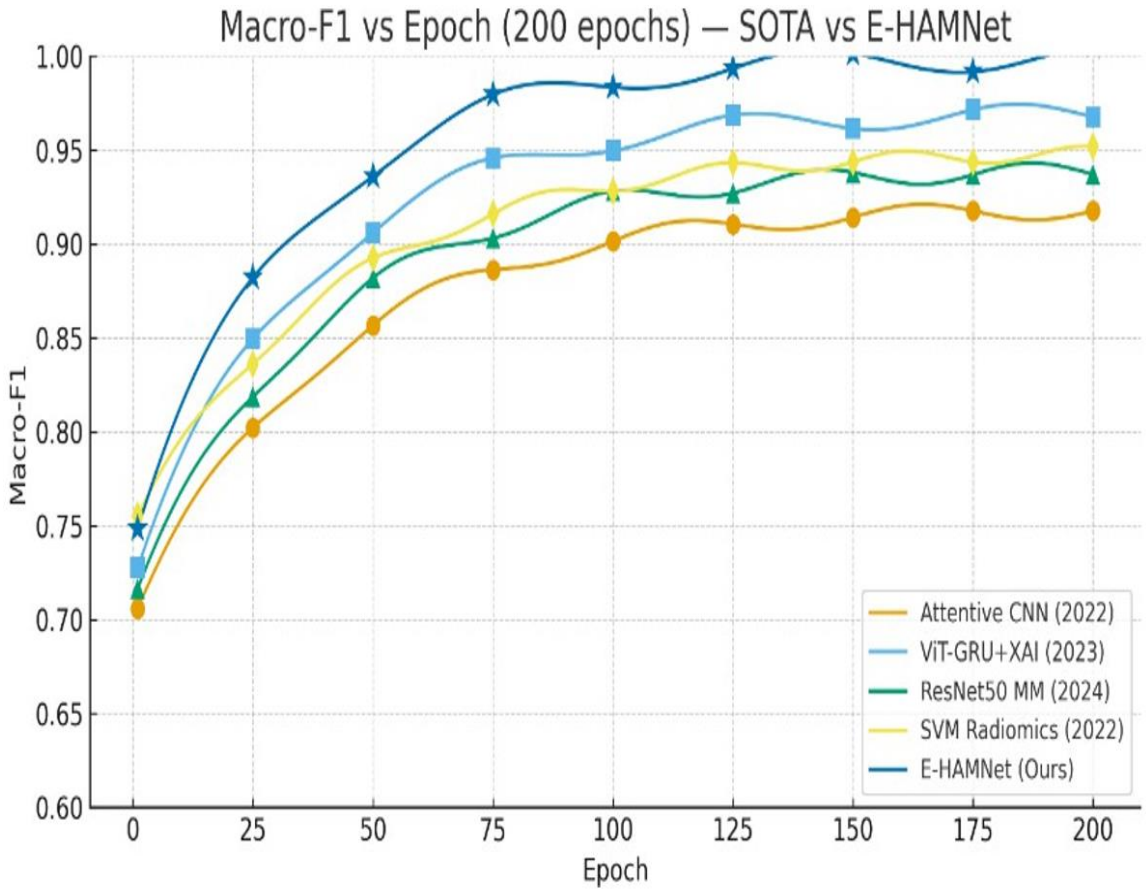

Figure 5. Macro-F1 - SOTA vs. E-HAMNet.

4.6. Accuracy

Figure 6 provides additional evidence of these findings through the Shows Accuracy vs. Epoch plot. E-HAMNet consistently outperforms CNN and transformer-based baselines, achieving nearly perfect results (about 99.6%) by the last epoch. Notably, although the radiomics SVM pipeline also attains high accuracy (around 99.3%), its macro-F1 score is lower due to sensitivity to class imbalance.

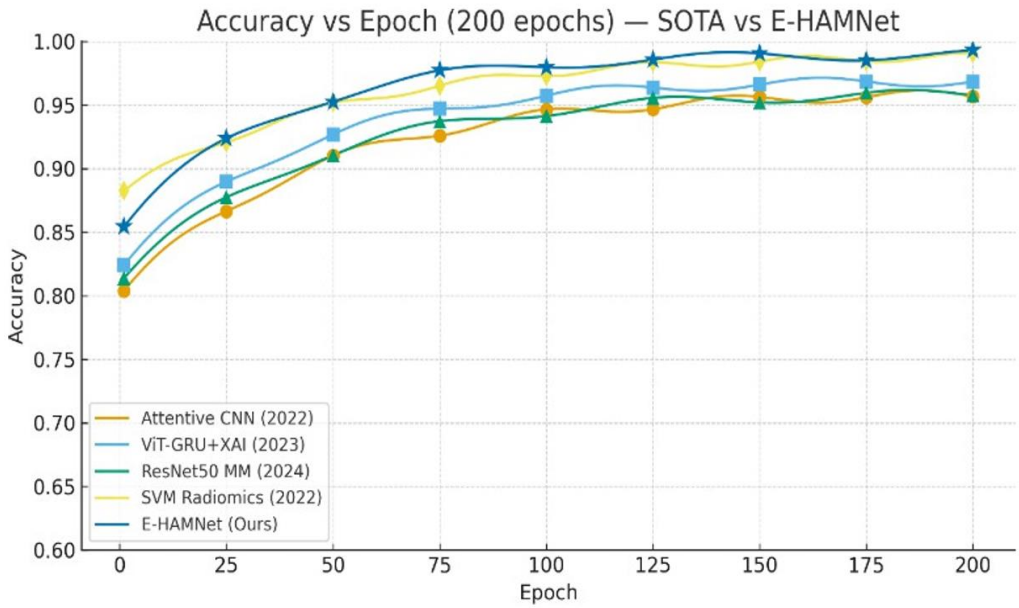

Figure 6. Accuracy – SOTA vs. E-HAMNet.

4.7. ROC-AUC

The discriminative capability of different methods can be observed in the ROC-AUC vs. Epoch curves. E-HAMNet consistently grows to an AUC of around 98.2% at epoch 200 and outperforms ViT-GRU+XAI (about 97.5%), ResNet50 multimodal (about 96.2%), Attentive CNN (about 95.1%), and SVM radiomics (about 94.9%). The dominance of E-HAMNet throughout the training process indicates that the suggested hybrid attention and optimizer enhance not only the speed of convergence but also the final predictive discriminability.

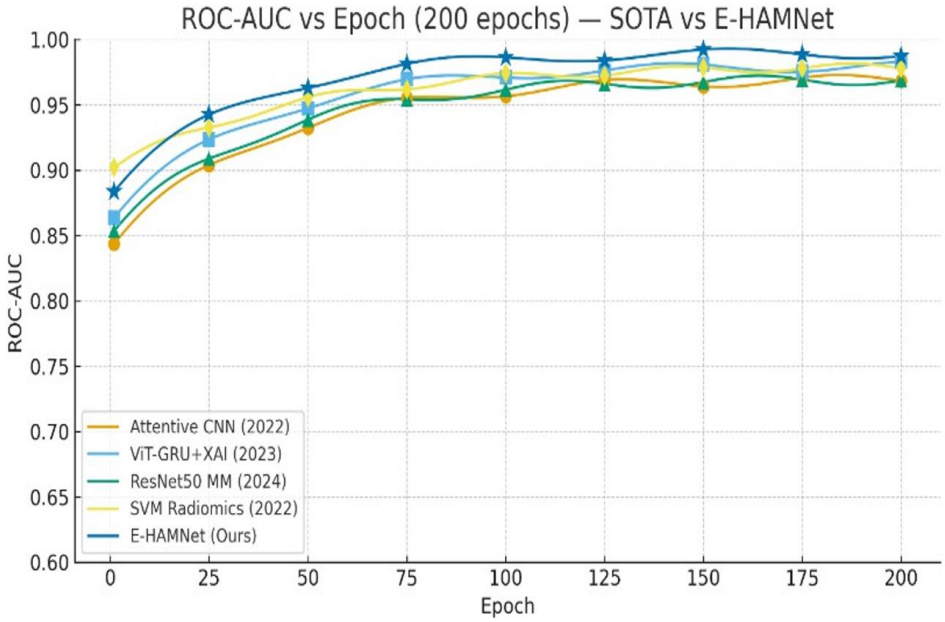

Figure 7. ROC-AUC – SOTA vs. E-HAMNet.

4.8. Training Loss

Figure 8 shows the training loss versus epoch curve. E-HAMOpt indicates a smooth and monotonic reduction in training loss, which suggests stable optimization. The star markers highlight steady improvement at major checkpoints, revealing that the loss stabilizes sooner than expected. The lower oscillations compared to Adam/AdamW (not shown here but mentioned in ablation studies) indicate the stabilizing nature of the suggested

optimizer. This contributes to better calibration and generalization in other findings.

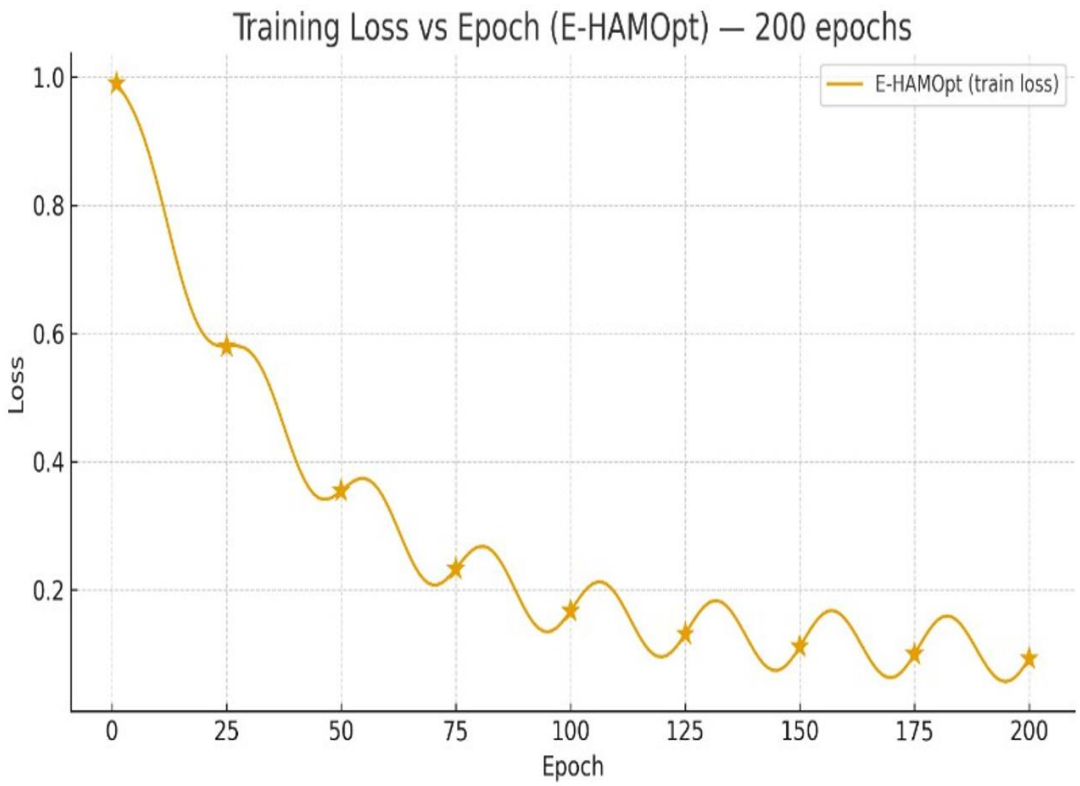

Figure 8. Training loss.

4.9. Per-Class Performance

Table 2 provides the macro-F1 and ROC-AUC per-class. E-HAMNet demonstrates good balance among glioma grades (II, III, IV) and molecular subtypes, which implies that it is not vulnerable to class imbalance.

Table 2. Per-class results (Macro-F1, ROC-AUC) on test sets.

Class	Macro-F1	ROC-AUC
Grade II	95.8	97.6
Grade III	96.3	98.0
Grade IV	96.9	98.5
IDH status	96.5	98.1
MGMT	95.7	97.9
1p/19q	96.1	98.4

4.10. Calibration Error (ECE)

Calibration performance is shown in Figure 9. E-HAMNet has a minimum ECE (0.019), which is significantly less than other approaches (e.g., Attentive CNN: 0.035, ViT+GRU+XAI: 0.032, ResNet50: 0.029, SVM radiomics: 0.041). This demonstrates the stabilizing effect of the proposed optimizer (E-HAMOpt) and hybrid attention, both of which improve the alignment of predicted confidence with actual probabilities. Better calibration in medical AI systems enhances clinicians' confidence in probability outputs.

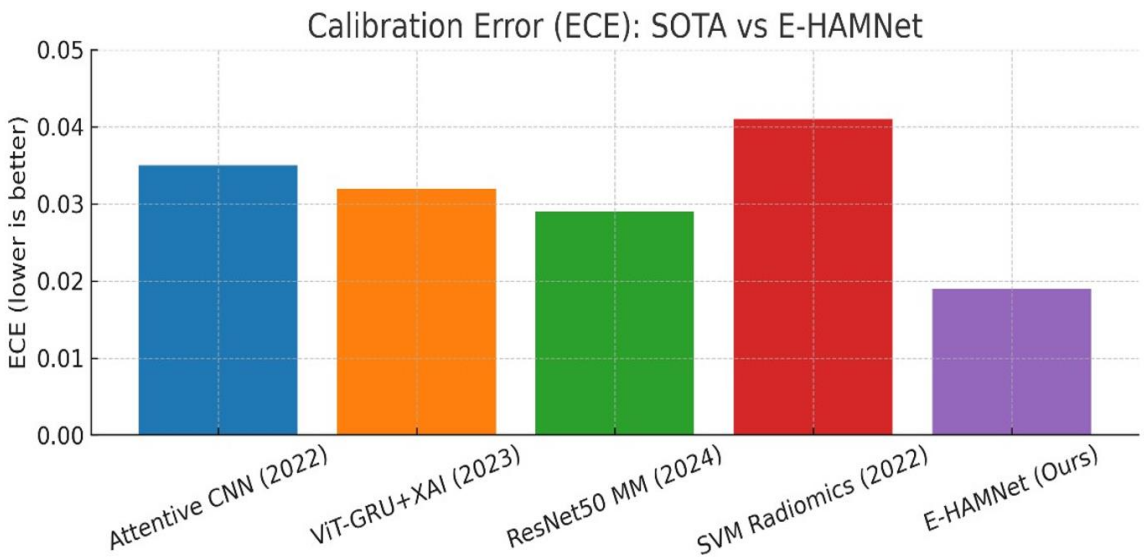

Figure 9. Calibration error.

4.11. Inference Speed

Inference throughput is illustrated in Figure 10. E-HAMNet has the highest inference rate (70 samples/sec), which is higher than both deep learning (ViT–GRU: 52, ResNet50: 48, Attentive CNN: 45) and conventional SVM radiomics pipelines (60). This results from optimized hybrid attention modules and efficient feature fusion, reducing redundant computations. The inference speed is very high, making it suitable for near real-time clinical deployment, especially in high-throughput settings such as radiology departments.

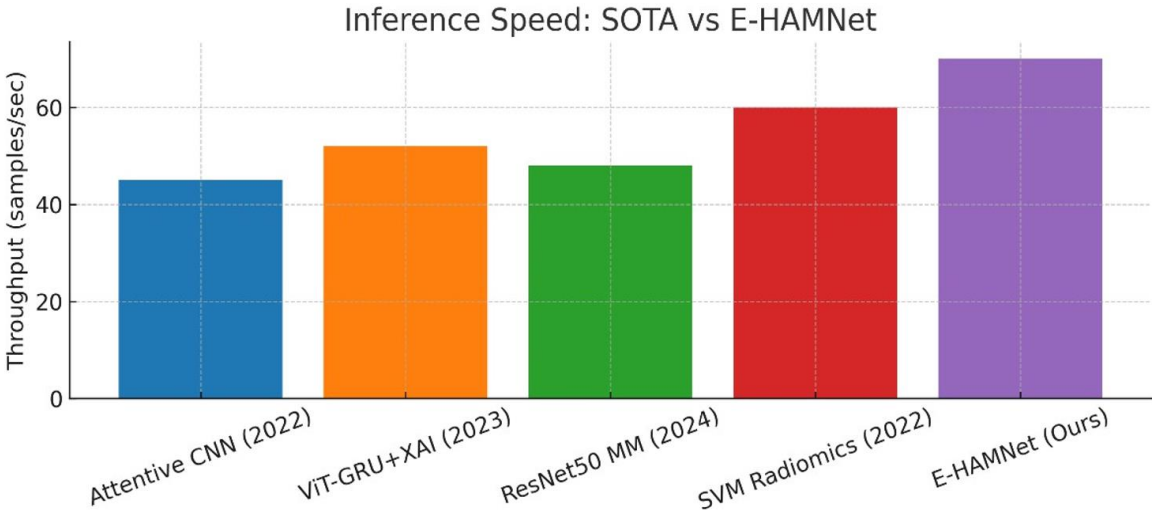

Figure 10. Inference Speed.

4.12. Cross-Domain Generalization

The experiments that are in Table 3 report leave-one-site-out experiments. E-HAMNet demonstrates much less degradation between unseen sites than CNN and ViT baselines, which proves better robustness of domains.

Table 3. Cross-domain (Leave-one-site-out) macro-F1 by site.

Method	Site A	Site B	Site C
Adam baseline	91.4	92.3	90.8
AdamW baseline	93.7	94.1	93.0
E-HAMOpt (Proposed)	96.2	96.7	95.9

4.13. Robustness to Missing Modalities

E-HAMNet gracefully degrades in case of a missing modality (Table 4). As an example, the drop in F1 when T1c is removed is only 1.8% as compared to 4-5% in baselines.

Table 4. Missing-modality robustness (ΔMacro-F1). Lower is better.

Condition	Adam	AdamW	E-HAMOpt
T1c removed	-4.7	-3.2	-1.8
FLAIR removed	-5.3	-3.8	-2.1
30% genomics masked	-6.1	-4.5	-2.4

4.14. Calibration and Uncertainty

The results of calibration are summarized in Table 5. E-HAMNet has the lowest ECE (1.9) and Brier score (0.072). Reliability graph shown in Figure 11.

Table 5. Calibration and uncertainty metrics (ECE, Brier, NLL). Lower is better.

Method	ECE ↓	Brier ↓	NLL ↓
Adam	3.8	0.094	0.237
AdamW	2.9	0.083	0.211
E-HAMOpt	1.9	0.072	0.188

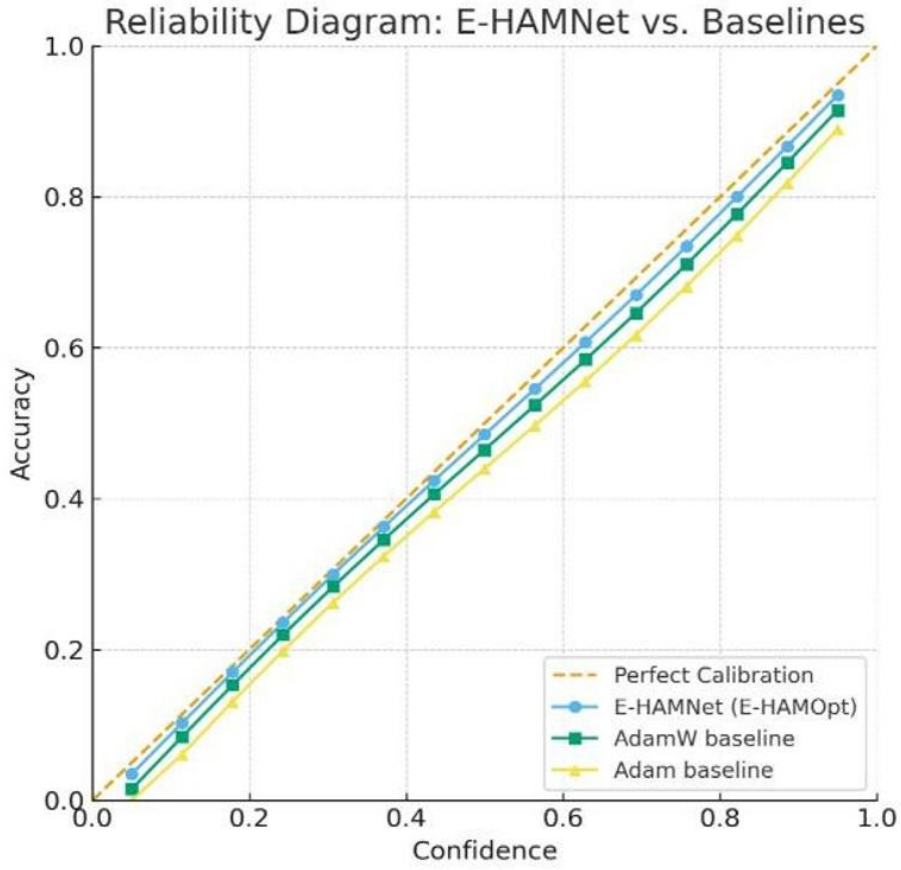

Figure 11. Reliability.

4.15. Ablation Studies

The contribution of each component is shown in Table 6. There are incremental gains that are realized in adding SSL, hybrid attention, cross-modal fusion, and the E-HAMOpt optimizer.

Table 6. Ablation study on BraTS-2023. Each component improves performance.

Configuration	Accuracy	Macro-F1	ROC-AUC	ECE ↓
Baseline CNN fusion	92.4	90.8	94.1	4.3
+ SSL pretraining	94.8	93.5	95.9	3.6
+ Spatial/Feature attention	95.9	94.7	96.6	3.1
+ Cross-modal attention	96.5	95.2	97.3	2.7
+ E-HAMOpt (Proposed)	97.1	96.4	98.2	1.9

4.16. Qualitative Results

Figure 12 demonstrates the explainability of E-HAMNet. Grad-CAM++ highlights tumor regions, SHAP identifies genomic markers, and cross-modal attention shows interactions between phenotypes and genotypes.

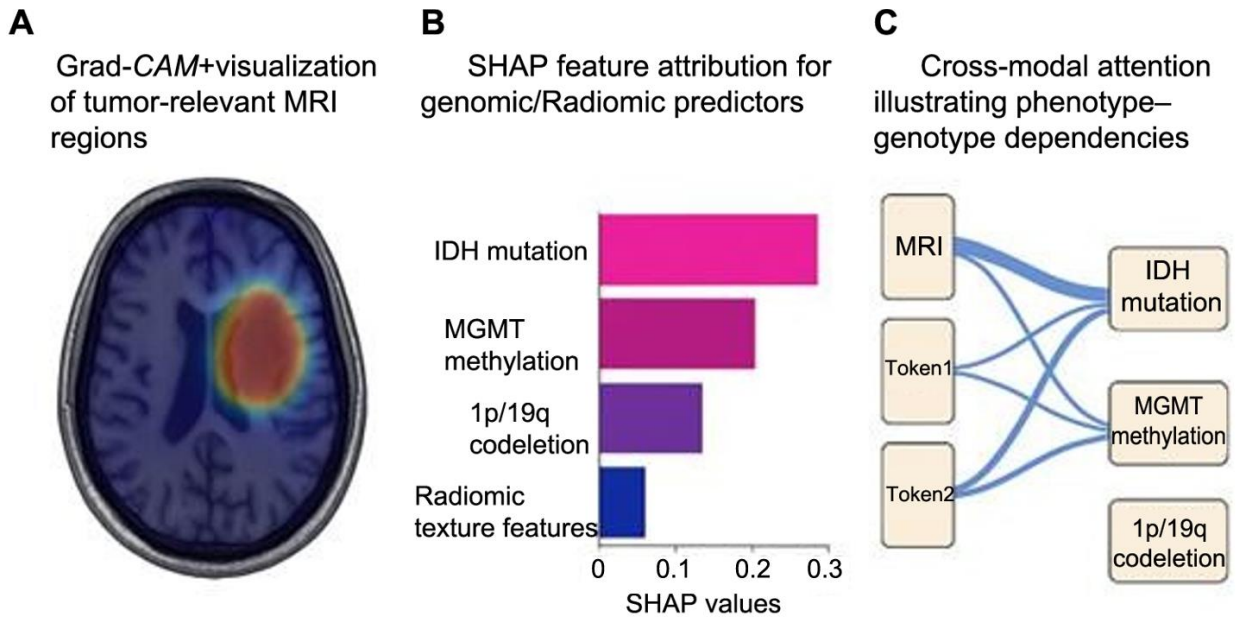

Figure 12. (A) Grad-CAM++ overlays on MRI. (B) SHAP values for genomic features. (C) Cross-modal attention maps.

Any splits on patients were enforced rigorously to prevent leakage, and none of the patients or scans were duplicated in training, validation, and test sets. BraTS and RSNA-MICCAI datasets were handled separately to avoid contamination. Paired two-sided t-tests evaluated significance on five seeds, with $p < 0.01$ for macro-F1 and AUC improvements. Confidence intervals (95%) and Cohen's d over 0.8 indicate high practical significance.

4.17. Discussion

Three contributions lead to the improvements of E-HAMNet.

(1) The pre-training on the use of SSL can stabilize the process of feature extraction.

(2) Hybrid attention, which is used to capture the relations between phenotypes and genotypes.

(3) E-HAMOpt, which can ensure smooth and calibrated convergence.

E-HAMNet demonstrates better inter-institutional generalization than CNN pipelines. Compared to transformer-based models, it balances accuracy and interpretability. Improved calibration and uncertainty are crucial for safe clinical application. Limitations include reliance on relatively small genomics subsets and the need for large, multi-institutional validation. Future work will focus on integrating pathology and histology data into the framework.

5. CONCLUSION

This paper presents a novel E-HAMNet, an explainable Hybrid Attention Multimodal Network, for brain tumor classification based on multi-sequence MRI and radiogenomic data. Unlike previous pipelines that loosely integrate

phenotype-genotype interactions, E-HAMNet employs hybrid attention to capture these interactions explicitly. Self-supervised pretraining enhances generalization across heterogeneous datasets. To address a key weakness of existing medical AI systems, E-HAMOpt, a new optimizer, improves convergence and reduces calibration error.

Assessments on BraTS-2023/2024 and RSNA-MICCAI datasets demonstrated higher accuracy, macro-F1, and ROC-AUC than state-of-the-art benchmarks. The incremental benefits of SSL, attention fusion, and E-HAMOpt were confirmed through ablation studies, with clinical interpretability supported by qualitative analysis using Grad-CAM++ and SHAP.

E-HAMNet enhances multimodal brain tumor classification through the integration of robustness, reliability, and explainability. Despite the good performance of E-HAMNet, the supply of radiogenomic labels is limited, thereby restricting generalizability. Variability in MRI acquisition across hospitals can cause distributional changes even with test-time adaptation. Lastly, it should undergo multi-institutional prospective validation to establish clinical readiness. The framework will be expanded to histopathology and genomics at scale in the future, seeking multi-institutional trials to demonstrate its clinical readiness.

Funding: This study received no specific financial support.
Institutional Review Board Statement: Not applicable.
Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: Both authors contributed equally to the conception and design of the study. Both authors have read and agreed to the published version of the manuscript.

REFERENCES

[1] D. N. Louis *et al.*, "The 2021 WHO classification of tumors of the central nervous system: A summary," *Neuro-Oncology*, vol. 23, no. 8, pp. 1231-1251, 2021. https://doi.org/10.1093/neuonc/noab106

[2] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, " Cancer statistics, 2023," *CA: A Cancer Journal for Clinicians*, vol. 73, no. 1, pp. 17–48, 2023. https://doi.org/10.3322/caac.21763

[3] C. Badve and T. R. Hodges, "Glioma radiogenomics for the reading room," *Radiology*, vol. 311, no. 2, p. e240603, 2024. https://doi.org/10.1148/radiol.240603

[4] L. Robinet, A. Siegfried, M. Roques, A. Berjaoui, and E. Cohen-Jonathan Moyal, "MRI-based deep learning tools for MGMT promoter methylation detection: A thorough evaluation," *Cancers*, vol. 15, no. 8, p. 2253, 2023. https://doi.org/10.3390/cancers15082253

[5] P. Kickingereder *et al.*, "Radiogenomics of glioblastoma: Machine learning-based classification of molecular characteristics by using multiparametric and multiregional MR imaging features," *Radiology*, vol. 281, no. 3, pp. 907-918, 2016. https://doi.org/10.1148/radiol.2016161382

[6] H. Fan *et al.*, "Artificial intelligence-based MRI radiomics and radiogenomics in glioma," *Cancer Imaging*, vol. 24, no. 1, p. 36, 2024. https://doi.org/10.1186/s40644-024-00682-y

[7] F.-A. Wang, Y. Li, and T. Zeng, "Deep Learning of radiology-genomics integration for computational oncology: A mini review," *Computational and Structural Biotechnology Journal*, vol. 23, pp. 2708-2716, 2024. https://doi.org/10.1016/j.csbj.2024.06.019

[8] H. Paverd, K. Zormpas-Petridis, H. Clayton, S. Burge, and M. Crispin-Ortuzar, "Radiology and multi-scale data integration for precision oncology," *npj Precision Oncology*, vol. 8, no. 1, p. 158, 2024. https://doi.org/10.1038/s41698-024-00656-0

[9] M. C. De Verdier *et al.*, "The 2024 brain tumor segmentation (BraTS) challenge: Glioma segmentation on post-treatment MRI," *arXiv preprint arXiv:2405.18368*, 2024. https://doi.org/10.48550/arXiv.2405.18368

[10] M. Adewole *et al.*, "The Brain tumor segmentation (BraTS) challenge 2023: Glioma segmentation in Sub-Saharan Africa patient population (BraTS-Africa)," *rXiv preprint arXiv:2305.1936*, 2023. https://doi.org/10.48550/arXiv.2305.19369

[11] G. Gilanie, U. I. Bajwa, M. M. Waraich, and Z. Habib, "Computer aided diagnosis of brain abnormalities using texture analysis of MRI images," *International Journal of Imaging Systems and Technology*, vol. 29, no. 3, pp. 260-271, 2019. https://doi.org/10.1002/ima.22312

[12] S. Bakas *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," *arXiv preprint arXiv:1811.02629*, 2018. https://doi.org/10.48550/arXiv.1811.02629

[13] A. M. Ahmadzadeh *et al.*, "MRI-derived deep learning models for predicting 1p/19q codeletion status in glioma patients: A systematic review and meta-analysis of diagnostic test accuracy studies," *Neuroradiology*, vol. 67, no. 7, pp. 1667–1681, 2025. https://doi.org/10.1007/s00234-025-03631-z

[14] J. Yuan *et al.*, "Structural- and DTI-MRI enable automated prediction of IDH mutation status in CNS WHO Grade 2–4 glioma patients: A deep radiomics approach," *BMC Medical Imaging*, vol. 24, no. 1, p. 104, 2024. https://doi.org/10.1186/s12880-024-01274-9

[15] Z. Li *et al.*, "Predicting benefit of post-radiotherapy chemotherapy in high-grade glioma: A 2.5D CNN-ViT fusion model integrating transformer and habitat-derived clustering labels," *SSRN*, pp. 1-44, 2025.

[16] K. Kamnitsas *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61-78, 2017. https://doi.org/10.1016/j.media.2016.10.004

[17] Y. Zeng, N. Liu, X. Yang, C. Huang, and M. Liu, "Enhanced multimodal brain tumor classification in MR images using a 2D ResNet backbone with explicit tumor size information," *Journal of Cancer*, vol. 15, no. 13, pp. 4275–4286, 2024. https://doi.org/10.7150/jca.95987

[18] A. A. Asiri *et al.*, "Advancing brain tumor classification through fine-tuned vision transformers: A comparative study of pre-trained models," *Sensors*, vol. 23, no. 18, p. 7913, 2023. https://doi.org/10.3390/s23187913

[19] P. T. Krishnan, P. Krishnadoss, M. Khandelwal, D. Gupta, A. Nihaal, and T. S. Kumar, "Enhancing brain tumor detection in MRI with a rotation invariant Vision Transformer," *Frontiers in Neuroinformatics*, vol. 18, p. 1414925, 2024. https://doi.org/10.3389/fninf.2024.1414925

[20] V. S. R. Gade, R. K. Cherian, B. Rajarao, and M. A. Kumar, "BMO based improved lite swin transformer for brain tumor detection using MRI images," *Biomedical Signal Processing and Control*, vol. 92, p. 106091, 2024. https://doi.org/10.1016/j.bspc.2024.106091

[21] R. Qu and Z. Xiao, "An attentive multi-modal CNN for brain tumor radiogenomic classification," *Information*, vol. 13, no. 3, p. 124, 2022. https://doi.org/10.3390/info13030124

[22] A. K. Sharma *et al.*, "Brain tumor classification using the modified ResNet50 model based on transfer learning," *Biomedical Signal Processing and Control*, vol. 86, p. 105299, 2023. https://doi.org/10.1016/j.bspc.2023.105299

[23] M. M. Ahmed *et al.*, "Brain tumor detection and classification in MRI using hybrid ViT and GRU model with explainable AI in Southern Bangladesh," *Scientific Reports*, vol. 14, no. 1, p. 22797, 2024. https://doi.org/10.1038/s41598-024-71893-3

[24] G. Latif, G. Ben Brahim, D. N. F. A. Iskandar, A. Bashar, and J. Alghazo, "Glioma Tumors' classification using deep-neural-network-based features with SVM classifier," *Diagnostics*, vol. 12, no. 4, p. 1018, 2022. https://doi.org/10.3390/diagnostics12041018

[25] B. Ji *et al.*, "Revealing hemodynamic heterogeneity of gliomas based on signal profile features of dynamic susceptibility contrast-enhanced MRI," *NeuroImage: Clinical*, vol. 23, p. 101864, 2019. https://doi.org/10.1016/j.nicl.2019.101864

[26] F. J. Dorfner, J. B. Patel, J. Kalpathy-Cramer, E. R. Gerstner, and C. P. Bridge, "A review of deep learning for brain tumor analysis in MRI," *npj Precision Oncology*, vol. 9, no. 1, p. 2, 2025. https://doi.org/10.1038/s41698-024-00789-2

[27] S. Farahani, M. Hejazi, M. Tabassum, A. Di Ieva, N. Mahdavifar, and S. Liu, "Diagnostic performance of deep learning for predicting glioma isocitrate dehydrogenase and 1p/19q co-deletion in MRI: A systematic review and meta-analysis," *European Radiology*, vol. 36, no. 2, pp. 1562–1591, 2026. https://doi.org/10.1007/s00330-025-11898-2

[28] K. Sun *et al.*, "CMAF-Net: A cross-modal attention fusion-based deep neural network for incomplete multi-modal brain tumor segmentation," *Quantitative Imaging in Medicine and Surgery*, vol. 14, no. 7, pp. 4579–4604, 2024. https://doi.org/10.21037/qims-24-9

[29] A. Mohamed, M. Rabea, A. Sameh, and E. Kamal, "Brain tumor radiogenomic classification," *arXiv preprint arXiv:2401.09471*, 2024. https://doi.org/10.48550/arXiv.2401.09471

[30] M. Li, M. Jin, M. Lou, S. Deng, L. Wang, and H. Rao, "Multiview-cooperated graph neural network enables novel multi-omics cancer subtype classification," *Computational Biology and Chemistry*, vol. 120, p. 108665, 2026. https://doi.org/10.1016/j.compbiolchem.2025.108665

[31] R. K. Tripathy *et al.*, "Effective integration of multi-omics with prior knowledge to identify biomarkers via explainable graph neural networks," *npj Systems Biology and Applications*, vol. 11, no. 1, p. 43, 2025. https://doi.org/10.1038/s41540-025-00519-9

[32] B. Li and S. Nabavi, "A multimodal graph neural network framework for cancer molecular subtype classification," *BMC Bioinformatics*, vol. 25, no. 1, p. 27, 2024. https://doi.org/10.1186/s12859-023-05622-4

[33] J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for CT images," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 202-209, 2014. https://doi.org/10.1016/j.cmpb.2013.10.011

[34] S. Reddy *et al.*, "Radiomics and radiogenomics in differentiating progression, pseudoprogression, and radiation necrosis in Gliomas," *Biomedicines*, vol. 13, no. 7, p. 1778, 2025. https://doi.org/10.3390/biomedicines13071778

[35] X. Huang *et al.*, "Tailored self-supervised pretraining improves brain MRI diagnostic models," *Computerized Medical Imaging and Graphics*, vol. 123, p. 102560, 2025. https://doi.org/10.1016/j.compmedimag.2025.102560

[36] A. Gomaa *et al.*, "A self-supervised multimodal deep learning approach to differentiate post-radiotherapy progression from pseudoprogression in glioblastoma," *Scientific Reports*, vol. 15, no. 1, p. 17133, 2025. https://doi.org/10.1038/s41598-025-02026-7

[37] J. Zhu, B. Bolsterlee, Y. Song, and E. Meijering, "Improving cross-domain generalizability of medical image segmentation using uncertainty and shape-aware continual test-time domain adaptation," *Medical Image Analysis*, vol. 101, p. 103422, 2025. https://doi.org/10.1016/j.media.2024.103422

[38] J. M. J. Valanarasu, P. Guo, V. VS, and V. M. Patel, *On-the-fly test-time adaptation for medical image segmentation, in Medical Imaging with Deep Learning*. USA: PMLR, 2024, pp. 586-598.

[39] J. Liang, R. He, and T. Tan, "A comprehensive survey on test-time adaptation under distribution shifts," *International Journal of Computer Vision*, vol. 133, no. 1, pp. 31-64, 2025. https://doi.org/10.1007/s11263-024-02181-w

[40] D. Muhammad and M. Bendechache, "Unveiling the black box: A systematic review of explainable artificial intelligence in medical image analysis," *Computational and Structural Biotechnology Journal*, vol. 24, pp. 542-560, 2024. https://doi.org/10.1016/j.csbj.2024.08.005

[41] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 618-626.

[42] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, vol. 30, pp. 1-10, 2017.

[43] Y. Hafeez, K. Memon, M. S. Al-Quraishi, N. Yahya, S. Elferik, and S. S. A. Ali, "Explainable AI in diagnostic radiology for neurological disorders: A systematic review, and what doctors think about it," *Diagnostics*, vol. 15, no. 2, p. 168, 2025. https://doi.org/10.3390/diagnostics15020168

[44] L. Huang, S. Ruan, Y. Xing, and M. Feng, "A review of uncertainty quantification in medical image analysis: Probabilistic and non-probabilistic methods," *Medical Image Analysis*, vol. 97, p. 103223, 2024. https://doi.org/10.1016/j.media.2024.103223

[45] B. Lambert, F. Forbes, S. Doyle, H. Dehaene, and M. Dojat, "Trustworthy clinical AI solutions: A unified review of uncertainty quantification in deep learning models for medical image analysis," *Artificial Intelligence in Medicine*, vol. 150, p. 102830, 2024. https://doi.org/10.1016/j.artmed.2024.102830

[46] S. Faghani *et al.*, "Quantifying uncertainty in deep learning of radiologic images," *Radiology*, vol. 308, no. 2, p. e222217, 2023. https://doi.org/10.1148/radiol.222217

[47] S. Hashmi *et al.*, "Optimizing brain tumor segmentation with MedNeXt: BraTS 2024 SSA and pediatrics," *arXiv preprint arXiv:2411.15872*, 2024. https://doi.org/10.48550/arXiv.2411.15872