Available on: http://gjetr.ep-journals.org/ E-ISSN: 3051-3782, P-ISSN: 3051-3774

Global Journal of Engineering and Technology Review
Volume: 02 Issue: 10 October 2026

# Optimized Deep Convolutional Networks for MRI Brain Tumor Classification and Segmentation: A Systematic Review

**Maher Talal Al-Asaady**
Department of Network Technologies and Intelligent Systems Engineering, Polytechnic College – Mosul, Northern Technical University, Mosul 41001, Iraq

**Published Date: October 26, 2026** **Article DOI : 10.65150/EP-gjetr/V2E10/2026-04**

**ABSTRACT:** Magnetic Resonance Imaging (MRI) is the primary modality for automated brain tumor detection, classification, and segmentation. While conventional Convolutional Neural Networks (CNNs) form the foundation of medical image analysis, they suffer from high computational overhead, local optima traps during hyperparameter tuning, and localized receptive fields that miss global anatomical context. This review systematically evaluates how algorithmic optimizations and hybrid architectures overcome these limitations. Through a multi-stage screening of 1,323 Scopus records published from 2021 to 2027 (including early-access publications), we analyze and compare 20 representative benchmark studies across four domains: (1) bio-inspired meta-heuristics for hyperparameter tuning and feature selection; (2) structural modifications integrating multiscale attention, dual-encoder U-Net and V-Net topologies, and wavelet compression for edge deployment; (3) hybrid frameworks pairing CNN extractors with Support Vector Machines, Long Short-Term Memory networks, Quantum Classifiers, and Graph Convolutional Networks; and (4) fusions of CNNs with Vision Transformers and 3D Mamba state-space models to capture long-range dependencies. Critically, our comparative study shows a complexity-performance compromise: sometimes heavy ensembles and hybrid classifiers fail to perform as well as streamlined architectures because of the saturation of the benchmark studies on the public 2D datasets. Finally, clinical translation will involve prioritizing cross-dataset generalization, linear-scaling efficiency, and quantitatively validated Explainable Artificial Intelligence (XAI).

**KEYWORDS:** Brain tumor detection, convolutional neural networks, magnetic resonance imaging, meta-heuristic optimization.

## INTRODUCTION

Pattern recognition and biometric identification systems have seen a significant transformation in the last decade from heuristic methods to intelligent systems [1-2]. Brain tumor detection and segmentation are important tasks in medical informatics, specifically in neuro-oncology, where automated systems based on Magnetic Resonance Imaging (MRI) scans are crucial for providing valuable diagnostic information to facilitate early clinical intervention and treatment planning [3-4]. Although Convolutional Neural Networks (CNNs) are the foundation of automated medical image analysis, the complexity of multi-modal MRI data structures and computations are not believed to be achievable by CNN for effective analysis [5-7]. The standard models usually have a high computational cost, sensitive to vanishing gradients, and have a large tendency to fall into local minima when training [8-9]. Conventional CNNs are also hardwired, and have limited local receptive field structures, which hinders their modeling of long-range spatial relationships and global anatomical context for accurate tumor boundary delineation and stratification into multiple classes [10-11].

In response to these fundamental constraints, modern studies in applied informatics have now turned into creating algorithmic optimization strategies and hybrid deep learning architectures. For high dimensional feature selection and weight training of neural networks, the use of swarm intelligence and meta-heuristic optimization algorithms has long been known to work well in a variety of medical data classification problems [12-14]. Recently, a few studies have applied bio-inspired metaheuristics to CNN hyperparameter tuning and deep feature selection, such as genetic optimizations [3], hybrid Aquila and Harris hawks optimizers [8] and whale optimization and zebra optimization frameworks [9], [15] to help standard CNN networks find their way out of local optima traps [16]. At the same time, specific architectural changes have been transplanted into 2D and 3D encoder-decoder architectures such as the U-Net, V-Net, and VGG19, to better extract the spatial hierarchy and minimize background noise in the image [5-7], [10] in the field of 2D/3D image processing. Furthermore, attention-based wavelet compression methods have enabled deploying lightweight CNNs to the resource-constrained edge devices [19].

In addition to tuning the standalone convolutional structures, the most revolutionary recent development is the blending of CNNs with different computational paradigms. By combining CNN backbones with different traditional machine learning methods like Support Vector Machines and ensemble trees, widely used in oncological classification [2], [20] or by embedding quantum enhanced methods, it is possible to create more solid decision boundaries for complex MRI feature sets [21-23]. For sequential dependencies and/or topological relationships, researchers have successfully integrated convolutional layers with Long Short-Term Memory (LSTM) networks [24] and Graph Convolutional Networks (GCNs) [25]. Moreover, CNNs and ViTs [26-27] and state-space models [28] when combined directly alleviate the limited receptive field of traditional

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

convolutions. The presented hybrid transformer pipelines combine the self-attention scheme with dual-path skip connections to fuse local textural knowledge with global spatial knowledge [11], [28-29].

In the present review, those recent paradigm shifts are systematically analyzed, using a selective sample of recent literature. This manuscript evaluates the performance of the most successful meta-heuristic based optimization algorithms, structural attention mechanisms, and multi-paradigmatic hybrid systems to address key fundamental challenges of improving the accuracy, efficiency and explainability scores of automated MRI brain tumor diagnosis [21], [26], [28].

The rest of this paper is structured as follows. The systematic search strategy, eligibility criteria, screening process, and multi-stage screening are described in Section II. In Section III, optimized deep convolutional networks are explored, which involves tuning with the aid of meta-heuristic algorithms, as well as structural attention modifications. Section IV discusses hybrid deep learning architectures combining CNNs with traditional classifiers, graph networks, Vision Transformers, and state space models. Section V considers open challenges, such as computational overhead, dataset saturation or explainability, in the critical discussion. Last but not least, the main results are summarized and research prospects are suggested in Section VI.

## METHODOLOGY

A systematic multi-stage screening process was implemented to make sure that literature was selected in a rigorous, transparent, repeatable way for this review. The main aim of this methodology was to identify recent, high impact works that assert algorithmic optimization techniques and hybrid deep learning architectures in the use of CNNs for brain tumor analysis in MRI.

### A. Information Source and Search Strategy

A systematic literature search was performed only by using Scopus database. Scopus was chosen because it indexes the best computing and medical informatics journals (including IEEE, Elsevier, Springer and Nature) and has a very stringent peer-review criteria for indexing. The search was limited to 2021-2027 (including early access and volume assigned articles for 2026 and 2027) so that the review represents the latest state-of-the-art developments.

The search question was designed in a manner that employs boolean operators, which intersect four main conceptual areas: the clinical pathology, the imaging modality, the deep learning model, and the methodology used to enhance the deep learning model. The search string used in the search of both article titles, abstracts and keywords was: ("brain tumor") AND ("MRI" OR "magnetic resonance imaging") AND ("convolutional neural network" OR "CNN") AND ("hybrid" OR "meta-heuristic" OR "optimization").

### B. Eligibility Criteria

In order to have the scope of the study focused and exclude marginal studies from the initial retrieval set, strict inclusion and exclusion criteria were applied during the screening process.

*1) Inclusion Criteria:* Peer-reviewed journal articles (n = 734) and conference proceedings (n = 589) published only in English were eligible for inclusion. The studies needed to employ single-modal or multimodal brain MRI scans for automated tumor detection, multi-class classification, or volumetric segmentation [6, 21]. In addition, the research had to explicitly employ algorithmic optimization (e.g. bio-inspired meta-heuristics for hyperparameter tuning or feature selection) [3], [8-9], structural attention or residual modifications [7], [10], or multi-paradigm hybrid architectures that integrate CNNs with traditional classifiers, graph networks, or Vision Transformers [23], [25], [28].

*2) Exclusion Criteria:* Studies assessing non-MRI neuroimaging techniques (e.g., Computed Tomography or Positron Emission Tomography) and non-cranial disease were not included. Standard, non-modified CNN baselines that did not include novel optimization techniques or hybrid CNN design, book chapters, editorial notes, retracted articles, and non-peer-reviewed preprints were excluded.

### C. Multi-Stage Screening and Selection Process

The literature selection was based on a three-stage funnel approach as described below:

*1) Stage 1 (Initial Identification and Temporal Distribution):* Using the search string and language filters as well as document-type filters, the initial search resulted in a corpus of 1,323 records at stage 1 (Initial Identification and Temporal Distribution). As shown in Figure 1, there was a significant increase in research activity over the time period analyzed, from 39 articles published in 2021 to 218 articles in 2024, 426 articles in 2025, 413 articles in 2026 and 5 early-access articles in 2027.

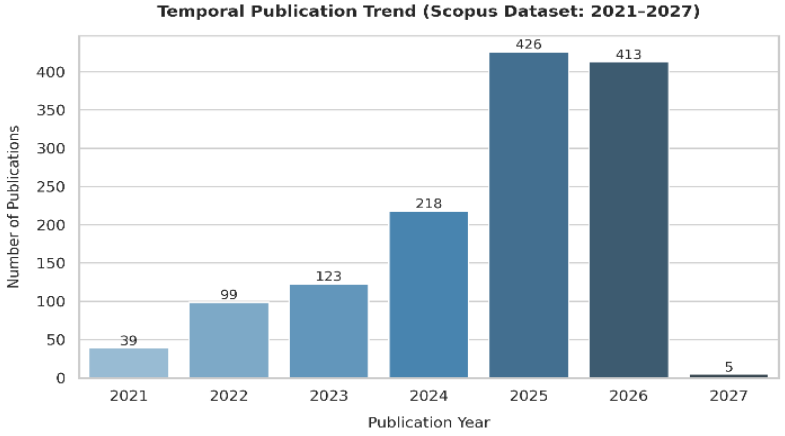



**Figure 1. The temporal trend of Scopus-indexed articles and conference papers published between 2021 and 2027 on the topic of hybrid and optimized CNNs for MRI brain tumor analysis are displayed (n = 1,323).**

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

*2) Stage 2 (Programmatic Domain Categorization):* The 1323 records in the titles were selected for programmatic keyword filtering in Stage 2 (Programmatic Domain Categorization), and 766 records were rejected because they did not satisfy the core methodological criteria. At this stage 557 different candidate papers were located that explicitly addressed the target methodologies. The algorithmic and meta-heuristic optimization keyword matched 175 records while the architectural attention and U-Net modifications keyword matched 193 records, the hybrid traditional machine learning or graph formulations keyword matched 99 records, and the hybrid Vision Transformer or state space architectures keyword matched 170 records (with the multi-paradigm studies distributed across the four thematic domains).

*3) Stage 3 (Priority Ranking and Final Selection):* The categorized subsets were then ranked and selected for a representative corpus of 20 benchmark studies from the search in Scopus (with 9 foundational domain references for the methodological context). Some methodologies have been eliminated from different categories and the titles have been removed to ensure structural diversity. The entire multi-stage selection process is outlined in the PRISMA 2020 diagram in Figure 2. Additionally, the structural taxonomy covering all the studied works in the four methodological domains is provided, in Figure 3. In order to compare these top five different articles for each domain, the following algorithms were selected for analysis: Algorithmic Optimization: [3], [8], [9], [15], [16]; Architectural Optimization: [6], [7], [10], [19], [26]; Hybrid CNNs with Traditional, Sequential and Graph models: [21], [22], [23], [24], [25]; Hybrid CNNs with Vision Transformers and State Space models: [4], [11], [27], [28], [29].

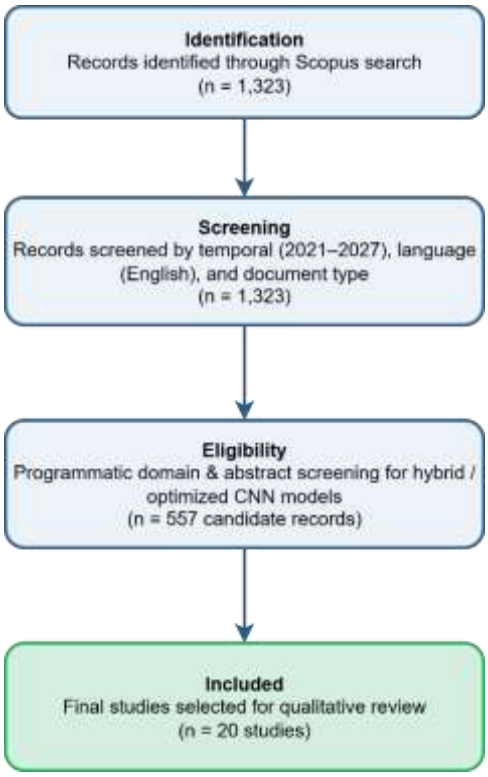



**Figure 2. PRISMA 2020 systematic screening flowchart.**

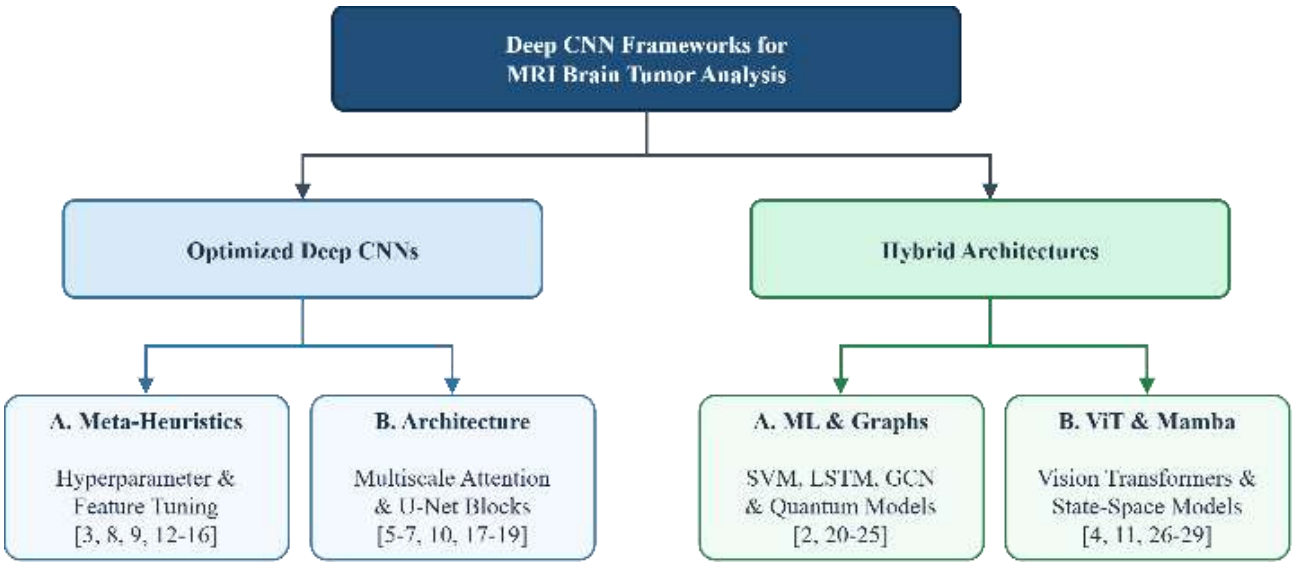



**Figure 3. Methodological taxonomy of optimized and hybrid deep convolutional networks for MRI brain tumor analysis.**

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Corresponding Author: Maher Talal Al-Asaady

**OPTIMIZED DEEP CONVOLUTIONAL NETWORKS**

In this context, while the CNN is the main foundation for neuro-oncological image analysis, conventional CNN implementations often struggle with sub-optimal parameter initialization, feature redundancy, and limited spatial receptive fields. These vulnerabilities have been recently tackled in two different directions: algorithmic optimization using bio-inspired meta-heuristics and structural optimization using multiscale attention and hybrid encoder-decoder topologies.

**A. Algorithmic Optimization via Meta-Heuristics**

In algorithmic optimization, the structural design of the network is separated from the search space of the network parameters, and hyperparameter optimization and deep feature selection are global mathematical optimization problems [12]. Because manual or grid-search tuning of learning rates, filter counts, dropout rates, and batch sizes is computationally prohibitive and prone to local optima, researchers have increasingly deployed evolutionary and swarm-intelligence algorithms.

To address label scarcity in glioma grading, Ravali et al. [3] proposed a multi-stage framework combining SimCLR-based self-supervised representation learning and Deep Embedded Clustering with a weighted ensemble of EfficientNet-B7, ResNet-50, and DenseNet-121. While the standalone EfficientNet-B7 achieved 88% to 90% classification accuracy and the unoptimized ensemble reached 93%, fine-tuning the hyperparameters via a Differential Evolution-optimized Genetic Algorithm (GA) elevated classification accuracy by an additional 3% to 5% while reducing training time by 15% [3]. Similarly, Winarno and Harjoko [9] integrated the Whale Optimization Algorithm (WOA) with adaptive training callbacks (ReduceLROnPlateau and early stopping) to automate CNN hyperparameter selection across glioma, meningioma, pituitary, and non-tumor MRI scans. The baseline CNN has an overall accuracy of 99.1%, increasing to 99.4% with callbacks and peaking at 99.6% (with a mean precision of 99.4% and F1-score of 99.2%) when combined with WOA [9].

On the other hand, if the hyperparameter space is high-dimensional, standalone meta-heuristics often exhibit a trade-off between globally exploring the search space and locally exploiting good solutions [13, 14]. To address this issue, Kumar et al. [8] proposed a hybrid Aquila Optimizer and Harris Hawks Optimization (AO-HHO) algorithm, which was evaluated using a publicly available dataset of 7,023 brain MRI images. The AO-HHO-tuned CNN significantly outperformed the standalone Particle Swarm Optimization (PSO), GA and WOA baselines which have achieved an accuracy range of 78% to 83% on the same experimental split [8] across the globe.The accuracy of the AO-HHO-tuned CNN is significantly higher than the standalone Particle Swarm Optimization (PSO), GA and WOA baselines which has remained within a range of 78% to 83% accuracy, when tested globally on the same experimental split [8]. Moreover, the hybrid AO-HHO converged in 77.85 seconds while conventional baseline optimizers converged in more than 300 seconds [8].

Beyond hyperparameter tuning, bio-inspired algorithms are actively utilized to optimize spatial segmentation centroids and select discriminative deep features [12]. Vikhe and Shrivastava [16] introduced a White-Headed Timber Optimization-based Deep CNN (WHT-based Deep CNN) featuring an eight-layer convolutional classifier and adaptive parallel learning. The model performed the best on the BraTS 2019 dataset with 97.46% accuracy, 97.22% sensitivity and 97.89% specificity, by using probability theory and optimizing centroids of tumor regions before feature extraction [16]. Expanding this paradigm to end-to-end pipelines, Salhi and Hassan [15] employed the Zebra Optimization Algorithm (ZOA) within a preprocessing-intensive framework incorporating N4 bias field correction, Z-score normalization, and Contrast-Limited Adaptive Histogram Equalization (CLAHE). The most relevant features were selected from the representation learned by a U-Net Transformer (UNETR) encoder using a deep feature selection algorithm, and then the classification step was carried out by a Variational Quantum Classifier (VQC). ZOA also tuned the hyperparameters of the model and deep feature selection on the representation before classification with a Dice Similarity Coefficient (DSC) of 97.1%, an Intersection over Union (IoU) of 94.8% and a classification accuracy of 98.7% [15].

A careful comparison of the above studies, however, will reveal that there are some methodological differences, even though all of them provide empirical measures. This is clearly evident from the difference in performance of standard optimizers like WOA that can be seen as 99.6% in [9] and less than 83% in [8] that might be due to data partitioning artifacts and not algorithmic improvements.

To illustrate the working of these combined methodologies, we describe an optimized and hybrid deep convolutional network diagnostic sequence for analysis of brain tumors in MRI images in the entire end-to-end pipeline of Figure 4. The architecture operates across four integrated stages: (1) image preprocessing and frequency-domain compression via N4 bias correction, CLAHE normalization, and Haar parametric wavelet transform [15], [19], [21], [28]; (2) hierarchical feature extraction and spatial representation through multiscale U-Net and V-Net encoder-decoder branches or Vision Transformer blocks [6], [8], [11], [26], [29]; (3) an iterative meta-heuristic optimization feedback loop utilizing algorithms such as Genetic Algorithms, Whale Optimization, and Aquila-Harris Hawks to dynamically tune hyperparameters and select optimal features [3], [8], [9], [15], [16]; and (4) final diagnostic stratification, producing multi-class classifications (via SVM, LSTM, or graph operators [23]–[25]), volumetric tumor segmentation masks, and quantitative Explainable AI (XAI) attribution maps [10], [21].

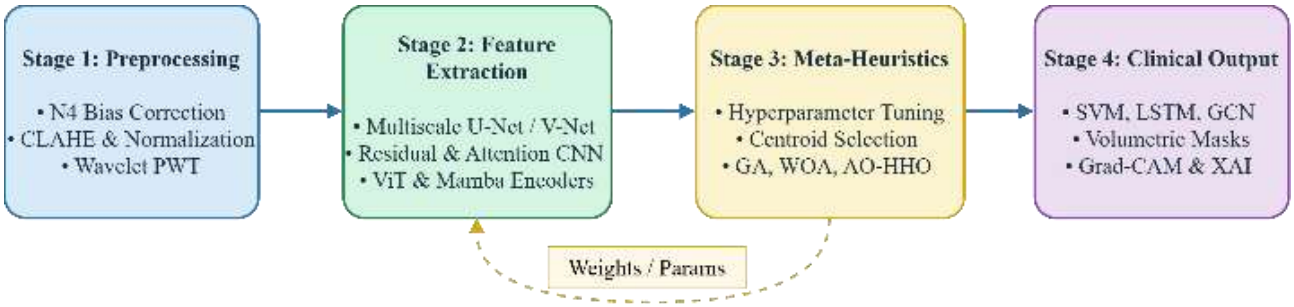



**Figure 4. End-to-end diagnostic pipeline illustrating the sequential interaction.**

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

**B. Architectural Optimization and Attention Mechanisms**

Our architectural optimization directly optimizes the internal topology of the convolutions without the need of external meta-heuristic search loops to find the best spatial hierarchy representation, boundary delineation and computational efficiency [5], [17], [18]. The key structural advancements are related to dual-encoder fusion, multi-scale attention and frequency domain compression.

To integrate both fine tumor boundaries and anatomical context, single-encoder CNNs are limited to volumetric and multimodal segmentation, which are based on the basic segmentation approaches of U-Net for multimodal MRI [5]. Zou et al. [6] proposed an AE-UVNet dual-encoder segmentation architecture for addressing the limitations of single-encoder, which was tested on BraTS 2020. An attention enhancement module (AE-M) links structured tri-skip fusion with a V-Net inspired branch, which uses larger receptive fields, to effectively bridge the cross-scale and cross-modality interactions between different subregions of the diffuse tumor [6]. On the other hand, Aslan and Özüpak [7] showed that simple architectures can also gain competitive performance with the help of transfer learning with complex multi-branch architectures. Their hybrid network was trained directly on a U-Net decoder with a standard skip connection, significantly cutting down the computational burden from large attention-based networks [7] and finalizing the performance on the annotated FLAIR MR imaging data of the lower grade glioma (LGG) from The Cancer Imaging Archive (TCIA) with a Dice score of 92.20%, an IoU of 85.53%, and an Area Under the Curve (AUC) of 95.90%.

Recent architectures in multi-class classification of CNNs aim at parallelizing multi-scale feature fusion and channel attention modules to replace sequential, single-scale layers, adapting to the architectural advancements that has been successful for one-dimensional biomedical signals classification [17], [18]. Prayogo et al. [10] suggested HayabusaNet, an architecture consisting of parallel multiscale branches that have different kernel sizes and gradually increasing channel widths, along with depthwise separable convolutions. Enhanced by a hybrid attention module and validated using Gradient-weighted Class Activation Mapping (Grad-CAM and Grad-CAM++), HayabusaNet achieved 98.86% accuracy and a 98.81% F1-score on the 7,023-image Kaggle multiclass dataset, alongside 98.70% accuracy on the 3,000-image binary Br35H benchmark, while maintaining lower floating-point operations (FLOPs) and parameter counts than standard deep CNNs [10].

To push architectural efficiency toward point-of-care edge deployment, K and M [19] integrated a Haar-initialized Parametric Wavelet Transform (PWT) layer directly ahead of a custom lightweight CNN and multi-head self-attention module. By decomposing MRI inputs into frequency subbands and stacking only the approximation and diagonal detail components, the PWT layer eliminates spatial redundancy before convolutional processing. This frequency domain optimization resulted in a classification accuracy of 95.88% while using less memory than MobileNetV2 (93.10%) and MobileNetV3Small (94.80%) to achieve real-time performance on the Raspberry Pi 5 edge device [19]. Finally, addressing cross-dataset domain shifts and boundary errors in segmentation, Tehsin et al. [26] introduced an attention-guided multi-branch encoder combining CNNs, Transformers, and Mamba state-space blocks with a dual-branch attention decoder and gated multiscale fusion. Evaluated across the Kaggle Brain MRI, Figshare, and BraTS 2020 datasets, the framework achieved mean Dice scores of 91.47%, 90.02%, and 89.14% (and IoU scores of 84.61%, 82.70%, and 81.66%), respectively, while maintaining an Explanation Stability Coefficient (ESC) of up to 0.86 and a cross-dataset transfer Dice score of 88.76% without domain adaptation [26]. To summarize and compare the structural and empirical differences across these ten studies, Table 1 provides a comparative overview of their optimization paradigms, datasets, and reported benchmarks.

**Table 1. Summary of Optimized CNN Methodologies for MRI Brain Tumor Analysis**

| Ref. | Optimization Strategy | Base Architecture | Dataset(s) | Key Performance |
|---|---|---|---|---|
| [3] | Differential Evolution Genetic Algorithm (GA) | SimCLR + Ensemble (EfficientNet-B7, ResNet-50, DenseNet-121) | Glioma MRI Dataset | 96% to 98% Acc (+3% to 5% over ensemble); 15% training time reduction |
| [8] | Hybrid Aquila Optimizer and Harris Hawks Optimization | Custom CNN | Public Brain MRI (7,023 images; 4 classes) | Outperformed PSO, GA, and WOA (78% to 83%); 77.85 s training time (vs. >300 s) |
| [9] | Whale Optimization Algorithm (WOA) + Callbacks | Deep CNN | Multiclass Brain MRI (4 classes) | 99.60% Acc; 99.40% Prec; 99.00% Rec; 99.20% F1-Score |
| [16] | White-Headed Timber (WHT) Centroid Optimization | 8-Layer Deep CNN + Deep Flow Segmentation | BraTS 2019 | 97.46% Acc; 97.22% Sens; 97.89% Spec |
| [15] | Zebra Optimization Algorithm (ZOA) | UNETR + Variational Quantum Classifier (VQC) | Preprocessed Brain MRI (CLAHE/N4) | 97.10% DSC; 94.80% IoU (Seg); 98.70% Acc (Class) |
| [6] | Dual-Encoder Attention + Structured Tri-Skip Fusion | AE-UVNet (Residual U-Net + V-Net branches) | BraTS 2020 | Improved structural coherence and boundary delineation across subregions |
| [7] | Pre-Trained Encoder-Decoder Transfer Fusion | Hybrid VGG19 Encoder + U-Net Decoder | TCIA LGG FLAIR Dataset | 92.20% DSC; 85.53% IoU; 95.90% AUC |

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

| Ref. | Optimization Strategy | Base Architecture | Dataset(s) | Key Performance |
|---|---|---|---|---|
| [10] | Depthwise Separable Multiscale Fusion + Attention | HayabusaNet (Parallel multiscale CNN branches) | Kaggle (7,023) & Br35H (3,000) | 98.86% Acc / 98.81% F1 (Kaggle); 98.70% Acc (Br35H) |
| [19] | Haar Parametric Wavelet Transform + Self-Attention | Lightweight Custom CNN (Raspberry Pi 5 Edge) | Brain MRI Dataset | 95.88% Acc (outperforming MobileNetV2 at 93.1% and V3Small at 94.8%) |
| [26] | Gated Multiscale Fusion + Dual-Branch Attention | Hybrid CNN-Transformer-Mamba Network | Kaggle, Figshare, & BraTS 2020 | 91.47%, 90.02%, & 89.14% DSC; 88.76% cross-dataset transfer DSC |

**HYBRID DEEP LEARNING ARCHITECTURES**

Pure convnets are limited in the number of linear heads for local classification and local receptive fields, but optimizations via algorithms and structures can bring some dramatic improvements. It has been noticed that Standard Convolutional Kernels work well for modelling local spatial hierarchies and edges of textural features in images, but are not adequate to model long-range anatomical dependencies, non-linear feature boundaries, and the topological relationship between distant brain regions [4], [25]. To overcome these structural barriers, contemporary research on neuro-oncology has turned to multi-paradigm hybrid architectures, such as the combination of CNNs and traditional machine learning classifiers, sequential networks, graph networks [21] and global self-attention Vision Transformers or state-space models [5].

**A. Integration with Traditional Classifiers, Sequential Models, and Graph Networks**

A primary hybridization strategy decouples spatial representation learning from decision boundary formulation, bridging the gap between shallow feature extraction models and deep neural architectures [2]. In these pipelines, pre-trained or custom CNNs serve strictly as high-dimensional feature extractors, while downstream traditional machine learning classifiers, recurrent units, or graph operators execute the final diagnostic stratification.

In order to test the generalization ability of non-linear kernel classifiers over the standard neural network output layer, Omran et al. [23] performed a controlled comparative study for 7023 contrast enhanced MRI scans from Kaggle, belonging to four classes, using a VGG19 feature extractor and a standard SoftMax layer or a Support Vector Machine (SVM) with a Radial Basis Function kernel (SVM-RBF). Clearly, while the VGG19-SoftMax baseline achieved 99.0% training accuracy and 98.4% validation accuracy, the hybrid VGG19-SVM (RBF) model recorded 96.2% validation accuracy and 97.8% testing accuracy [23]. Although the authors noted that the SVM-RBF head reduced the training-to-validation gap and improved nonlinear class separation when deployed within an AWS cloud compression workflow, these findings critically demonstrate that replacing SoftMax with an SVM does not universally elevate raw validation accuracy [23].

To extract the inter-slice sequential dependencies prior to SVM classification, Bérété et al. [24] proposed a hybrid paradigm with five different CNN backbones (VGG16, VGG19, ResNet50, ResNet152, and EfficientNetB3) and LSTM networks and an SVM classifier. The ResNet152 + LSTM + SVM configuration (tuned using RMSprop) reached 99.0% accuracy, precision, recall and F1 score in all metrics [24]. Hussain et al. [21] developed an explainable multi-class framework called NeuroBlend-3, an extension of the recurrent-convolutional fusion. After enhancement by CLAHE and five-fold spatial augmentation, the authors tested six feature extractors (HRNet, VGG16, VGG19, ResNet50, ResNet101, and CNN-LSTM), compressed the representations using Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE), and classified with ensemble classifiers (XGBoost, AdaBoost, Bagging, and a custom Tree Selection and Stacking Ensemble-based Random Forest, or TSRF). The best pipeline was an integrated CNN-LSTM feature extraction, RFE dimensionality reduction and the TSRF classifier, with clinical auditing using Grad-CAM, Grad-CAM++, and Local Interpretable Model-Agnostic Explanations (LIME) [21].

In addition to classical ensembles, new paradigms combine quantum machine learning (QML) and topological graph reasoning with visual backbones. Rawas and AlSaeed [22] explored hybrid quantum-classical architectures simulated using Qiskit and PennyLane, which encode preprocessed Brain Tumor MRI features into quantum states for classification using Variational Quantum Circuits (VQCs) and a Quantum Kernel SVM. The Quantum Kernel SVM had just eight trainable parameters, while the VQC required only twice as many as classical Multi-Layer Perceptron (MLP) and CNN baselines, but the accuracy of its diagnosis (89.1%) and F1 score (88.5%) were significantly smaller than for classical deep hybrids, evidencing the representational limitations of simulated low-qubit circuits [22] today. On the other hand, while preserving visual granularity, Elhamzi and Mokni [25] created BrainGraphNet-ViT to model complex anatomical topologies. This dual-stream architecture handles the processing of MRI scans in parallel, using a hierarchical visual extractor (ViT, EfficientNet-B0, ResNet50, and VGG19) and a GCN based on a Graph Transformer operator. BrainGraphNet-ViT incorporated local features by textural features and graph-based structural reasoning to nodes, and thus the peak multi-class accuracy reached 99.31% [25].

**B. Fusion with Vision Transformers and State-Space Models**

Recently, the most significant architecture change in MRI segmentation and classification literature is the combination of CNNs with Vision Transformers (ViTs) and structured state-space sequence models (SSMs). There is a direct remedy to the problem of receptive field, as observed in the comparative study of Mukherjee et al. [4] of different approaches across the BraTS benchmarks (2012 - 2025): from purely convolutional encoders (like standard U-Net) to hybrid transformer topologies (like Swin UNETR and AXUNet) that introduce global self-attention across volumetric scans.

In multi-class tumor classification, Kumar Napa et al. [28] introduced BRAIN-META, a modular meta-ensemble pipeline evaluated on CLAHE-enhanced 2D MRI scans. The authors fused ten distinct pre-trained CNN architectures (DenseNet121, DenseNet169, DenseNet201, MobileNet, MobileNetV2, EfficientNetB0, EfficientNetB1, EfficientNetB4, InceptionV3, and Xception) with Vision Transformer blocks to capture both

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

localized textures and global dependencies. The validation SoftMax probabilities from all ten CNN-ViT hybrids were concatenated into stacked feature vectors to train second-level meta-learners, achieving 97.10% accuracy with XGBoost and 97.03% with Logistic Regression [28]. However, from a computational efficiency perspective, training ten concurrent CNN-ViT backbones yields lower overall accuracy than streamlined single-backbone models (such as [10] or [25]), illustrating the diminishing returns of extreme ensemble stacking.

On the other hand, hybrid CNN-Transformer architectures focus on incorporating transformer encoders into the U-net skip-connection structure to maintain the accuracy of edges in segmentation of tumors. Liu et al. [11] introduced the Enhanced Feature-Based Vision Patch Transformer Network (EF-VPT-Net), which consists of an optimized vision patch transformer encoder and a positional-embedding convolutional U-net decoder, with a novel adaptive block-size layer that allows the network to process high-resolution images and report segmentation accuracy and sensitivity of up to 99.1% on two public datasets. Nguyen-Tat et al. [29] introduced DoubleBlock-ViT to address standard 3D multimodal benchmarks (T1, T1Gd, T2 and T2-FLAIR) under more challenging volumetric metrics. This hybrid 3D U-Net integrates a MaxViT-based 3D DoubleBlock-ViT encoder for global context with a CNN-based Dual-Path Fusion Block along the skip connections to preserve fine spatial margins. Maintaining a compact footprint of only 7.8 million parameters, DoubleBlock-ViT achieved Dice Similarity Coefficients (DSC) of 80.11% for Enhancing Tumor (ET), 86.60% for Tumor Core (TC), and 91.20% for Whole Tumor (WT) on BraTS 2020, and improved to 87.82% (ET), 91.61% (TC), and 92.31% (WT) on BraTS 2021 [29].

Despite the modeling power of ViTs, their quadratic self-attention complexity scales poorly on high-resolution 3D MRI volumes. To capture long-range dependencies with linear computational scaling, Rahim et al. [27] bridged transformers and state-space models on the BraTS 2021 and BraTS 2023 datasets. The authors introduced two hybrid design pathways: SwinUNETR+ (incorporating modality-aware attention, multi-scale windows, and residual refinement) and Mamba-based hybrids, including MambaUNETR (employing tri-directional 3D Mamba blocks in the encoder) and SwinMamba (combining shallow Swin transformer stages with deep Mamba state-space blocks) [27]. On the BraTS 2023 validation set, MambaUNETR and SwinMamba achieved mean Dice scores of 89.30% (0.893) and 89.40% (0.894), respectively, while an ensemble of SwinUNETR+, MambaUNETR, and SwinMamba reached 89.80% validation DSC and 89.70% test DSC [27]. Furthermore, in cross-dataset evaluation on BraTS 2021, MambaUNETR achieved a 90.20% DSC and SwinMamba recorded an F1-score of 91.85%, demonstrating that hybridizing shallow window-based transformers with deep state-space models offers a highly scalable alternative to pure transformer encoders [27]. Table 2 summarizes and compares the architectural paradigms, fusion mechanisms, datasets, and empirical performance metrics across the ten surveyed hybrid studies.

**Table 2. Summary of Hybrid Deep Learning Architectures for MRI Brain Tumor Analysis**

| Ref. | Hybrid Paradigm | Architecture & Fusion | Dataset(s) | Key Performance |
|---|---|---|---|---|
| [21] | CNN-RNN + Feature Selection + Ensemble ML | NeuroBlend-3: CNN-LSTM + PCA/RFE + Tree Selection Random Forest + Grad-CAM/LIME | Augmented Multiclass Brain MRI (CLAHE) | Superior multi-class performance across metrics |
| [22] | Classical Preprocessing + Quantum ML | Hybrid QML: Variational Quantum Circuits (VQC) & Quantum Kernel SVM | Brain Tumor MRI, ChestX-ray14, & CT | 89.1% Prec; 88.50% F1 (8 parameters); 2x faster VQC convergence |
| [23] | Pre-Trained CNN + Kernel Classifier | VGG19 + SVM (RBF) vs. VGG19 + SoftMax (AWS cloud pipeline) | Kaggle Contrast MRI (7,023 images; 4 classes) | 97.80% Test / 96.20% Val Acc (VGG19-SVM) vs. 98.40% Val Acc (SoftMax) |
| [25] | Dual-Stream CNN-ViT + Graph Network | BrainGraphNet-ViT: ViT, EfficientNet-B0, ResNet50, & VGG19 + Graph Transformer GCN | Brain Tumor MRI Dataset (Multiclass) | 99.31% Peak Classification Accuracy |
| [24] | Deep CNN + Recurrent Network + SVM | ResNet152 + LSTM + SVM (alongside VGG16/19, ResNet50, & EfficientNetB3) | Augmented Multiclass Brain MRI | 99.00% Acc, 99.00% Prec, 99.00% Rec, & 99.00% F1 (RMSprop) |
| [4] | CNN-to-Transformer Benchmark | Systematic evaluation of U-Net vs. Swin UNETR and AXUNet hybrid topologies | BraTS Benchmarks (2012 to 2025 GBM) | Superior global context and boundary recovery in hybrid transformers |
| [11] | Vision Patch Transformer + CNN U-Net | EF-VPT-Net: Vision Patch Transformer Encoder + Positional-Embedding U-Net Decoder | Two Public Brain MRI Datasets | Up to 99.10% pixel-level segmentation accuracy and sensitivity |
| [28] | Multi-CNN-ViT Meta-Ensemble | BRAIN-META: 10 CNN-ViT hybrids stacked into XGBoost & Logistic Regression | Multiclass 2D Brain MRI (4 classes; CLAHE) | 97.10% Acc (XGBoost meta-learner); 97.03% Acc (Logistic Regression) |
| [29] | 3D MaxViT Encoder + Dual-Path Skip Fusion | DoubleBlock-ViT: 3D Hybrid U-Net with Dual-Path Fusion Blocks (7.8M params) | BraTS 2020 & BraTS 2021 (3D MRI) | BraTS2020: 80.11% ET, 86.60% TC, 91.20% WT DSC; BraTS2021: 87.82% ET, 91.61% TC, 92.31% WT |

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

| Ref. | Hybrid Paradigm | Architecture & Fusion | Dataset(s) | Key Performance |
|---|---|---|---|---|
| [27] | 3D Swin Transformer + Mamba State-Space | SwinUNETR+, MambaUNETR, & SwinMamba (Tri-directional 3D Mamba + Swin) | BraTS 2023 & BraTS 2021 (3D MRI) | BraTS2023: 89.70% Test DSC (Ensemble); BraTS2021: 90.20% DSC (MambaUNETR) & 91.85% F1 |

**CRITICAL DISCUSSION AND OPEN CHALLENGES**

While the surveyed literature demonstrates that meta-heuristic optimization, attention-guided structural modifications, and multi-paradigm hybrid architectures achieve near-ceiling benchmark metrics in MRI brain tumor analysis, a rigorous comparative analysis reveals significant methodological vulnerabilities that currently impede clinical translation.

**A. The Complexity-Performance Paradox and Computational Overhead**

A critical weakness across recent neuro-oncology informatics is the tendency toward architectural over-engineering without proportional diagnostic gains. Several frameworks construct massive multi-backbone pipelines that incur severe computational overhead during training and inference. For instance, BRAIN-META [28] trains ten separate CNN-ViT hybrid architectures and stacks their outputs into an XGBoost meta-learner to achieve 97.10% classification accuracy on a four-class 2D MRI benchmark. This heavy ten-model ensemble outperforms the single-backbone, light-weight models evaluated on the same public set of 7,023 images, like HayabusaNet [10] (98.86% accuracy with lightweight depthwise separable convolutions) and dual-stream BrainGraphNet-ViT (99.31% accuracy). Likewise, in volumetric and 2D segmentation, the multi-branch networks combining CNNs, Transformers, and Mamba blocks [26] yield mean Dice score ranging from 89.14% to 91.47% whereas a structurally simple VGG19-UNet encoder-decoder [7] obtains a Dice score of 92.20% on lower-grade glioma FLAIR scans without using multi-branch attention modules, which suffer from parameter bloat issues.

In addition, although meta-heuristic algorithms can readily automate hyperparameter tuning and feature selection [3], they are often not reported to be as compute intensive as they actually are [14]. For instance, Kumar et al. [8] report a decrease in final model training time, from more than 300 seconds for baseline optimizers to 77.85 seconds for AO-HHO, while Ravali et al. [3] report a 15% decrease using genetic optimization. But the population-based evolutionary and swarm algorithms have their inherent drawback – that is to say that the deep network is assessed over many candidate agents and many iterations [12]. For deployment in resource-limited clinical environments, these frameworks need to be deployed in parameter-efficient ways, such as efficient channel attention modules [18] or linear-scaling 3D Mamba state-space hybrids [27] or frequency-compressed wavelet CNNs that can perform real-time inference on edge hardware like the Raspberry Pi 5 [19].

**B. Benchmark Saturation, Metric Inconsistencies, and Generalization**

This can be observed in the fact that the studies surveyed made use of a limited number of heavily preprocessed public datasets, namely the Kaggle 4-class dataset (7,023 images), Br35H, Figshare and the Brats series [6] (31 patients), [8] (200 patients), [10] (280 patients), [26] (292 patients), [27] (128 patients), and [29] (280 patients). Since the current 2D classification baselines are small and clearly distinguishable, the performance of the baseline CNNs is often saturated and consequently different baselines are reported in different studies. For instance, according to Winarno and Harjoko [9] the unoptimized CNN baseline achieves an accuracy of 99.1% and with WOA tuning the accuracy is 99.6%. Conversely, on the same 4-class task, Kumar et al. [8] report that standard WOA, PSO, and Genetic Algorithms stagnate between 78% and 83% accuracy before applying their hybrid AO-HHO method. Likewise, Omran et al. [23] showed that a basic VGG19-SoftMax baseline outperforms their proposed model, a hybrid VGG19-SVM (RBF) classifier with 98.4% and 97.8% (validation and test) accuracy, respectively, indicating that hybridizing a CNN with a kernel classifier does not necessarily yield a higher raw accuracy when the underlying dataset is close to saturation.

One of the prevalent methodological challenges in segmentation is the use of different evaluation measures. Frequently, 3D multimodal studies report Dice Similarity Coefficients (DSC) (from 80.11% to 92.31%) for each sub-region of interest (ROI) defined on BraTS benchmarks: Enhancing Tumor (ET), Tumor Core (TC) and Whole Tumor (WT) [27]. In contrast, studies that report only pixel-level segmentation accuracy (such as the 99.1% accuracy reported in [11]) obscure boundary errors because background non-tumor pixels overwhelmingly dominate MRI slices. Moreover, external cross-dataset validation remains rare; among the surveyed corpus, only Tehsin et al. [26] and Rahim et al. [27] explicitly quantified zero-shot cross-dataset transfer (achieving 88.76% DSC across 2D datasets [26] and 90.20% DSC when transferring from BraTS 2023 to BraTS 2021 [27]). In addition, emerging paradigms such as Variational Quantum Classifiers (VQCs) and Quantum Kernel SVMs [15] remain restricted to classical software simulators (such as Qiskit and PennyLane [22]), leaving their resilience against real-world quantum hardware decoherence and clinical scanner noise unverified.

**C. From Post Hoc Visualizations to Quantitative Explainable AI (XAI)**

One of the major challenges to the use of hybrid deep networks in neuro-oncology is their black box property. However, in order to build diagnostic trust, various recent frameworks have been developed that incorporate Explainable AI (XAI) mechanisms such as Grad-CAM [10] and Grad-CAM++ [19], LIME [21] and Integrated Gradients [27]. Most of the implementations are however based on the post hoc qualitative saliency heatmaps [10] exclusively. A visual examination of an MRI slice chosen by hand from a heatmap does not ensure that the trained model is using a legitimate pathological feature when applied to an entire cohort of patients, and the model is not guaranteed to be stable when run with a minor change in the input.

A critical open challenge for applied informatics is transitioning from qualitative saliency maps to rigorous, quantitative explainability metrics and ante hoc (inherently interpretable) architectures [4], [26]. Tehsin et al. [26] have taken a step further from visual heatmaps by mathematically auditing their hybrid CNN-Transformer-Mamba network using an Attention-Ground Truth Alignment Score (AGAS) of 88.71%, an Explanation Coverage Ratio (ECR) of 90.54%, an Attention Sparsity Index (ASI) of 78.61%, and an Explanation Stability Coefficient (ESC) of up to 0.86. Future architectures need to be similarly coupled with standardized quantitative alignment benchmarks to meet the clinical and regulatory auditing requirements, alongside the topological and causal reasoning (such as the graph-based structural pathways in BrainGraphNet-ViT [25]).

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Ultimately, direct quantitative rankings across these surveyed studies remain methodologically constrained by the absence of standardized multi-institutional datasets and uniform evaluation metrics.

## CONCLUSION

This review systematically analyzed and compared recent advancements in algorithmic optimization and hybrid deep convolutional architectures for automated MRI brain tumor detection, classification, and segmentation. A preliminary Scopus corpus of 1323 documents published between 2021 and 2027 was used as a basis and 20 representative benchmark studies (plus 9 methodological papers) were critically analysed in four thematic areas. The literature proves that bio-inspired meta-heuristic algorithms are effective in solving local optima and automating hyperparameter tuning and deep feature selection. At the same time, structural optimizations based on dual-encoder U-Net and V-Net pathways, multiscale inception and depthwise separable convolutions, efficient channel attention and parametric wavelet compression substantially improve the quality of boundary extraction, while enabling the deployment of the model on the edge. In addition, the fusion of CNNs with sequential LSTMs, SVMs and Graph Convolutional Networks, or with global self-attention Vision Transformers and linear-scaling 3D Mamba state-space models has proven to be an effective way to span the divide between local textural extraction and long-range anatomical context.

Our empirical successes notwithstanding, our critical analysis reveals that more complex structures – such as multiple concurrent CNN-ViT backbones or switching out SoftMax layers for SVMs on saturated 2D benchmarks, often fail to deliver significant improvements in diagnostic value over simpler, better-regularized structures. As an outlook, the applied informatics research community needs to move from marginal accuracy improvements of single-institution public datasets to real-time clinical hardware toward addressing three practical research challenges: (1) Designing computationally efficient hybrid architectures (CNN-Mamba, wavelet-compressed networks) suitable for real-time clinical hardware that can be scaled to thousands of patients; (2) Establishing multi-institutional, cross-dataset validation protocols, using standardized volumetric metrics (DSC, HD95) as a standard for comparison; and (3) Replacing qualitative post hoc heatmaps with a quantitative framework to drive a more inherently interpretable XAI framework. Solving those problems is critical for the successful adoption of experimental deep learning models to robust, consistent and automated clinical decision support tools in neuro-oncology.

## REFERENCES


1) A. Khalaf and M. Al-Asaady, "Developed trends in human identification based on behavioral biometrics," *Int. J. Bus. Inf. Commun. Technol.*, vol. 1, no. 1, Sep. 2026, doi: 10.68092/ijbict.01.01.06.
2) S. Q. Hasan, "Shallow model and deep learning model for features extraction of images," *NTU J. Eng. Technol.*, vol. 2, no. 3, Nov. 2023, doi: 10.56286/ntujet.v2i3.449.
3) P. Ravali, P. C. S. Reddy, and P. Praveen, "An efficient neuro-framework for brain tumor classification using a CNN-based self-supervised learning approach with genetic optimizations," *Curr. Neuropharmacol.*, vol. 24, no. 10, pp. 1550-1564, 2026, doi: 10.2174/011570159X378103250806214613.
4) K. Mukherjee, A. Babhulkar, M. V. Munot, and R. C. Jaiswal, "Deep learning for glioblastoma segmentation: A survey from CNNs to transformers," in *Lect. Notes Netw. Syst.*, vol. 2127, 2027, pp. 11-22, doi: 10.1007/978-3-032-34772-5_2.
5) M. T. Ghazal, "Robust U-Net-based approach for accurate brain tumor segmentation using multimodal MRI data," *NTU J. Eng. Technol.*, vol. 2, no. 3, Nov. 2023, doi: 10.56286/ntujet.v2i3.692.
6) H. Zou, A. Anil, and M. Qiao, "AE-UVNet: An attention enhanced hybrid architecture using U-Net and V-Net for multimodal brain tumour segmentation," in *Lect. Notes Comput. Sci.*, vol. 16754, 2027, pp. 199-214, doi: 10.1007/978-981-92-2487-6_13.
7) E. Aslan and Y. Özüpak, "A hybrid deep learning framework based on VGG19 and U-Net for accurate brain tumor segmentation in MRI images," *Magn. Reson. Lett.*, vol. 6, no. 4, Art. no. 200275, 2026, doi: 10.1016/j.mrl.2026.200275.
8) M. Kumar, N. Mohd, G. Shivam, A. Goyal, D. Parashar, and R. Khan, "Hybrid Aquila optimizer-Harris Hawks optimization for CNN hyperparameter tuning in brain tumor classification," *Sci. Rep.*, vol. 16, no. 1, Art. no. 12799, 2026, doi: 10.1038/s41598-026-43329-7.
9) Winarno and A. Harjoko, "Enhancing deep learning model using Whale Optimization Algorithm on brain tumor MRI," *J. Electron. Electromed. Eng. Med. Inform.*, vol. 8, no. 1, pp. 136-152, 2026, doi: 10.35882/jeeemi.v8i1.941.
10) R. D. Prayogo, S. A. Karimah, and H. Nambo, "HayabusaNet: Hybrid attention-based multiscale fusion CNN for accurate and efficient brain tumor classification in MRI scans," *Expert Syst. Appl.*, vol. 331, Art. no. 133135, 2026, doi: 10.1016/j.eswa.2026.133135.
11) J. Liu, U. A. Bhatti, J. Zhang, Y. Zhang, and M. Huang, "EF-VPT-Net: Enhanced feature-based vision patch transformer network for accurate brain tumor segmentation in magnetic resonance imaging," *IEEE J. Biomed. Health Inform.*, vol. 30, no. 9, pp. 7938-7951, 2026, doi: 10.1109/JBHI.2025.3526976.
12) L. Brezočnik, I. Fister, and V. Podgorelec, "Swarm intelligence algorithms for feature selection: A review," *Appl. Sci.*, vol. 8, no. 9, p. 1521, Sep. 2018, doi: 10.3390/app8091521.
13) M. T. Al-Asaady, T. N. M. Aris, N. M. Sharef, and H. Hamdan, "COA-MLP: Training neural networks using the Coyote Optimization Algorithm for medical data classification," in *Proc. 7th Int. Conf. Interdiscip. Comput. Sci. Eng. (ICICSE)*, IEEE, Nov. 2025, pp. 29-36, doi: 10.1109/ICICSE67247.2025.11390755.
14) M. Al-Asaady, T. N. M. Aris, N. M. Sharef, and H. Hamdan, "Recent advances on meta-heuristic algorithms for training multilayer perceptron neural network," *JOIV: Int. J. Inform. Vis.*, vol. 9, no. 2, p. 658, 2025, doi: 10.62527/joiv.9.2.3109.
15) A. Salhi and M. M. Hassan, "Enhanced brain tumor segmentation and classification based on quantum-classical attention-guided hybrid framework with Zebra Optimization Algorithm," *J. Signal Process. Syst.*, vol. 98, no. 1, Art. no. 14, 2026, doi: 10.1007/s11265-025-01973-8.
16) N. B. Vikhe and M. Shrivastava, "White Headed Timber Optimization based deep learning model for brain tumor prediction," *Biomed. Mater. Devices*, vol. 4, no. 2, pp. 2072-2096, 2026, doi: 10.1007/s44174-025-00293-6.


License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Corresponding Author: Maher Talal Al-Asaady 

17) M. T. Al-Asaady, T. N. M. Aris, and A. D. Khalaf, "Multi-scale Inception 1D-CNN for patient-independent congestive heart failure detection from ECG signals," in *Int. Conf. Cybersecur., Data Sci., Mach. Learn. (ICCDM-2026)*, Kuala Lumpur: Springer, 2026.
18) M. Al-Asaady, M. Rashid, and M. Mahmood, "ECA-ECGNet: An Efficient Channel Attention-Enhanced 1D CNN for ECG Arrhythmia Detection", Journal of Electronics, Electromedical Engineering, and Medical Informatics, vol. 8, no. 4, pp. 1609-1624, Oct. 2026.
19) A. N. K and N. A. K. M, "Haar-initialized parametric wavelet compression with attention-driven lightweight CNN for brain tumor classification on edge devices," *Biomed. Phys. Eng. Express*, vol. 12, no. 1, 2026, doi: 10.1088/2057-1976/ae3760.
20) R. H. M. Ameen, N. M. Basheer, and A. K. Younis, "A survey: Breast cancer classification by using machine learning techniques," *NTU J. Eng. Technol.*, vol. 2, no. 1, May 2023, doi: 10.56286/ntujet.v2i1.367.
21) M. I. Hussain, S. H. Chowdhury, M. Mamun, and M. M. Hossain, "NeuroBlend-3: Hybrid deep and machine learning framework with explainable AI for multi-class brain tumor detection using MRI scans," *Medinformatics*, vol. 3, no. 1, pp. 56-66, 2026, doi: 10.47852/bonviewMEDIN52026540.
22) S. Rawas and D. AlSaeed, "Quantum enhanced machine learning for medical image analysis: A hybrid approach," *Appl. Comput. Inform.*, pp. 1-16, 2026, doi: 10.1108/ACI-10-2025-0448.
23) H. M. Omran, K. Ibrahim, G. T. Abdel-Jaber, and A.-N. Sharkawy, "Brain tumor classification from MRI images using hybrid deep learning approaches: VGG19 with SoftMax and SVM classifiers," *Int. J. Robot. Control Syst.*, vol. 6, no. 1, pp. 16-35, 2026, doi: 10.31763/ijrcs.v6i1.2260.
24) M. Bérété, A. Echtioui, L. Sellami, and A. A. Ben Hmida, "Hybrid CNN + LSTM + SVM model for advanced brain tumor classification," *Netw. Model. Anal. Health Inform. Bioinform.*, vol. 15, no. 1, Art. no. 43, 2026, doi: 10.1007/s13721-025-00718-8.
25) W. Elhamzi and R. Mokni, "BrainGraphNet-ViT model: A hybrid model combining vision transformers and graph convolutional networks for brain tumor diagnosis," *Sci. Rep.*, vol. 16, no. 1, Art. no. 16772, 2026, doi: 10.1038/s41598-026-46803-4.
26) S. Tehsin, I. M. Nasir, and R. Damaševičius, "Explainable brain tumor segmentation via attention-guided hybrid CNN-Transformer-Mamba network," *Knowl.-Based Syst.*, vol. 343, Art. no. 116000, 2026, doi: 10.1016/j.knosys.2026.116000.
27) I. Rahim, N. R. Arpa, S. Sarkar, M. M. I. Nibir, and R. Khan, "Bridging transformers and state-space models: An efficient hybrid framework for 3D brain tumor segmentation," *Biomed. Signal Process. Control*, vol. 122, Art. no. 110406, 2026, doi: 10.1016/j.bspc.2026.110406.
28) K. Kumar Napa, S. Murugan, J. S. Murugan, and A. Jayanthi, "BRAIN-META: A reproducible CNN-vision transformer meta-ensemble pipeline for explainable brain tumor classification," *MethodsX*, vol. 16, Art. no. 103769, 2026, doi: 10.1016/j.mex.2025.103769.
29) T. B. Nguyen-Tat, L. Q. Truong, K. Q. Truong, and T. H. Ngo, "DoubleBlock-ViT: A MaxViT-based enhancement and dual-path skip connections for brain tumor segmentation in MRI scans," *Comput. Methods Programs Biomed.*, vol. 274, Art. no. 109165, 2026, doi: 10.1016/j.cmpb.2025.109165.


License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Corresponding Author: Maher Talal Al-Asaady