Metaheuristic Optimization Review

♦ ISSN: 3066-280X Vol: 06, No. 01, 2026 ♦ PP. 103–117

Article DOI: https://doi.org/10.54216/MOR.060108

REVIEW ARTICLE

Deep Learning-Based Classification of Brain Tumors from Magnetic Resonance Imaging Scans Using a Convolutional Neural Network Model

Karim Eldreny[1,*]

[1] Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura 35111, Egypt

Email: Ch2000186@dhiet.edu.eg

Received: December 31, 2025 Revised: February 04, 2026 Accepted: April 01, 2026 ⋆ Corresponding author

ABSTRACT

Brain tumors are serious neurological conditions that require accurate and timely classification to support medical evaluation and treatment planning. This project presents a deep learning-based system for classifying brain Magnetic Resonance Imaging (MRI) scans into four categories: glioma, meningioma, pituitary tumor, and no tumor. The proposed system uses a Convolutional Neural Network (CNN) trained on a balanced dataset of 7,200 MRI images collected from publicly available sources. The images were preprocessed through RGB conversion, resizing, tensor transformation, and normalization to ensure consistent input for model training and testing. The trained model achieved an overall classification accuracy of 94.31% on a held-out test set of 1,600 MRI images, demonstrating strong performance in multi-class brain tumor classification. A Streamlit-based web application was also developed to allow users to upload MRI images and view the predicted class, confidence score, and probability distribution across the four categories. The system is intended for educational and research purposes only and should not replace professional medical diagnosis, clinical judgment, or radiological evaluation.

Keywords: Brain Tumor Classification; Magnetic Resonance Imaging; Deep Learning; Convolutional Neural Network; Medical Image Classification

1. INTRODUCTION

Brain tumor diagnosis represents one of the most critical challenges in contemporary medical imaging because it directly affects clinical decision-making, treatment planning, surgical intervention, radiotherapy design, and long-term patient monitoring. Brain tumors may appear with highly variable anatomical locations, tissue characteristics, growth patterns, and visual boundaries, which makes their interpretation from medical images a complex task even for experienced specialists. Magnetic Resonance Imaging (MRI) has become a major imaging modality for brain tumor assessment because it provides high soft-tissue contrast and can reveal structural abnormalities within the brain with greater anatomical clarity than many other imaging techniques. However, the interpretation of MRI scans remains dependent on expert evaluation, visual experience, image quality, tumor appearance, and the availability of trained radiologists. These factors have encouraged extensive research into artificial intelligence (AI), machine learning (ML), and deep learning (DL) techniques that can support automated brain tumor detection and classification from MRI images [1].

The increasing use of computational intelligence in medical imaging is mainly driven by the need for accurate, repeatable, and time-efficient diagnostic support systems. Traditional clinical interpretation requires careful visual inspection of

MRI slices, comparison between anatomical regions, and recognition of abnormal tissue patterns. Although human expertise remains essential, manual interpretation may become difficult when the tumor is small, poorly contrasted, irregularly shaped, or visually similar to surrounding healthy tissue. Automated classification methods can assist by extracting quantitative patterns from images and mapping them to diagnostic categories in a consistent manner. In this context, brain tumor classification has become an important research direction because it attempts not only to detect the presence of abnormal tissue, but also to assign MRI images to clinically meaningful tumor classes using computational models [2].

Deep learning has significantly changed the methodological landscape of brain MRI analysis. Unlike earlier ML approaches, which often depended on manually designed texture descriptors, intensity statistics, segmentation masks, or handcrafted shape features, DL models can learn hierarchical image representations directly from pixel-level data. Convolutional Neural Networks (CNNs) are especially influential in this field because their convolutional filters can capture low-level image patterns such as edges and local intensity transitions, while deeper layers can represent more abstract visual structures related to tumor morphology. This ability makes CNN-based systems suitable for distinguishing between tumor types that may differ in location, boundary definition, internal texture, surrounding edema, and signal intensity distribution. As a result, DL-based MRI classification has become a central research direction in computer-aided diagnosis and medical image analysis [3].

A major motivation behind automated brain tumor classification is the clinical and technical difficulty caused by tumor heterogeneity. Brain tumors do not follow a single visual pattern across all patients or MRI acquisitions. Some tumors may appear as clearly defined masses, while others may have diffuse, infiltrative, or irregular borders. In addition, image appearance can vary according to scanner settings, acquisition protocols, slice orientation, contrast enhancement, and preprocessing procedures. These variations can reduce the reliability of simple image-processing pipelines and increase the need for robust learning models that generalize across diverse MRI appearances. Recent studies have therefore investigated improved DL architectures, transfer learning strategies, hybrid models, and optimized classification pipelines to enhance the reliability of tumor recognition under variable imaging conditions [4].

Despite the strong progress achieved by CNN-based models, brain tumor classification from MRI remains an open and active research problem. One important challenge is that high accuracy on a specific dataset does not always guarantee reliable performance on images collected from different hospitals, scanners, or patient populations. Dataset bias, class imbalance, limited external validation, and possible overlap between training and testing samples can lead to optimistic results that may not fully reflect real-world performance. Therefore, review studies in this area must examine not only reported accuracy values, but also dataset composition, preprocessing strategies, model evaluation protocols, validation design, and the transparency of the experimental workflow. A model that performs well under controlled conditions may still require broader validation before it can be considered suitable for clinical decision-support environments [5].

Transfer learning has become a common strategy for improving MRI classification performance, particularly when available medical datasets are not large enough to train deep architectures from the beginning. In transfer learning, a model that has already learned general image features from a large dataset is adapted to a medical imaging task by modifying and retraining its final layers. This approach can reduce training time and improve feature extraction, especially when the target dataset contains a limited number of labeled MRI images. However, the difference between natural images and medical images must be considered carefully because features learned from general image datasets may not always correspond directly to the subtle anatomical and pathological patterns present in MRI scans. Consequently, transfer learning should be evaluated with rigorous testing and careful interpretation rather than being treated as a universal solution [6].

In addition to CNN and transfer-learning models, recent research has expanded toward hybrid architectures and advanced AI systems that combine multiple learning mechanisms. Hybrid models may integrate convolutional feature extraction with recurrent structures, attention mechanisms, transformer-based components, or optimization procedures to improve classification capability. These approaches aim to capture complementary information from MRI images, strengthen feature representation, and reduce misclassification between visually similar tumor types. Nevertheless, increasing architectural complexity can also introduce challenges related to computational cost, interpretability, reproducibility, and fair comparison with simpler baseline models. For this reason, the evaluation of hybrid systems should consider not only predictive performance but also model transparency, implementation feasibility, and suitability for medical-image workflows [7].

Explainability is another essential direction in AI-based brain tumor classification. Medical imaging systems cannot be assessed only by their final prediction labels, because clinical users also need to understand why a model produced a particular output and whether the decision appears visually consistent with meaningful tumor regions. Explainable AI methods are therefore increasingly used to highlight image regions that contribute to model predictions, support visual verification, and improve trust in automated systems. In MRI-based brain tumor analysis, explainability can help determine whether the model is focusing on tumor tissue, surrounding anatomical structures, image artifacts, or irrelevant background regions. This is particularly important because a highly accurate model may still learn undesirable shortcuts if the dataset contains hidden biases or inconsistent acquisition patterns [8].

The literature also indicates that performance improvement in brain tumor classification depends on the interaction between several methodological components rather than on model architecture alone. Dataset quality, image preprocessing, augmentation strategy, feature extraction method, classifier design, hyperparameter selection, validation procedure, and evaluation metrics all influence final classification performance. Some recent studies have explored image enhancement, model refinement, and optimized learning pipelines to improve the separability of tumor classes in MRI images.

These directions are important because brain tumor classes may contain overlapping visual characteristics, and classification errors can occur when tumors share similar shapes, locations, or intensity distributions. A strong review must therefore discuss how different methods attempt to reduce these errors and whether their reported improvements are supported by reliable experimental design [9].

The present review focuses on AI-based brain tumor classification from MRI images, with particular attention to ML and DL approaches, CNN-based systems, transfer learning, hybrid models, optimization-based improvements, and explainability-oriented methods. The purpose is to provide an academically structured discussion of how automated MRI classification systems have evolved, what methodological strengths they offer, and what limitations still restrict their translation into dependable medical-support tools. Since this work is prepared as a review-oriented manuscript, the emphasis is placed on conceptual analysis, comparison of methodological trends, discussion of existing limitations, and identification of future research directions rather than on presenting a new experimental pipeline. By synthesizing recent developments in the field, this review aims to clarify the role of intelligent MRI classification systems in supporting brain tumor analysis while maintaining the essential distinction between computational assistance and professional medical diagnosis [10].

2. LITERATURE REVIEW

Research on MRI–based brain tumor classification has moved through several methodological stages, beginning with conventional image-processing pipelines and progressing toward deep learning models capable of learning discriminative representations directly from medical images. Early automated systems usually depended on a sequence of preprocessing, segmentation, handcrafted feature extraction, feature selection, and classical classification. These approaches were useful because they converted visual MRI patterns into measurable numerical descriptors, such as intensity distributions, texture statistics, shape measurements, and region-based features. However, their success was highly dependent on the quality of the selected features and on the ability of the preprocessing stage to preserve clinically meaningful tumor information. Since brain tumors can vary widely in location, size, boundary clarity, tissue signal, and internal structure, manually designed features may fail to capture the full complexity of tumor appearance. This limitation encouraged the adoption of learning-based systems that can automatically identify relevant patterns from MRI data with less reliance on handcrafted descriptors [11].

The transition from handcrafted features to convolutional learning marked a major development in the field. Convolutional Neural Networks (CNNs) are particularly suitable for MRI analysis because they process images through local receptive fields, allowing the model to identify spatially meaningful features such as edges, boundaries, intensity gradients, and textural variations. As information passes through deeper layers, the network gradually transforms low-level visual patterns into more abstract representations that can support tumor classification. In brain MRI classification, this hierarchical feature learning is important because different tumor classes may differ not only in local appearance but also in anatomical position and global structure. For example, some tumor types may appear as well-defined masses, whereas others may be more diffuse or heterogeneous. CNN–based systems can therefore provide stronger representation capacity than many traditional pipelines, especially when trained and evaluated on sufficiently organized MRI datasets [12].

A central issue in the reviewed literature is the quality and composition of the datasets used for model training and testing. MRI classification performance is strongly affected by class balance, image resolution, tumor visibility, scanner variability, preprocessing consistency, and the separation between training and testing samples. If a dataset contains duplicated or visually near-identical images across different splits, the resulting performance may be artificially high because the model may encounter test images that are highly similar to training data. Similarly, if the dataset is imbalanced, the model may achieve high overall accuracy while performing poorly on minority classes. These issues are especially important in medical imaging, where reliable generalization is more valuable than high performance on a narrow benchmark. For this reason, recent literature increasingly emphasizes careful data preparation, transparent reporting of dataset structure, and the use of evaluation metrics beyond simple accuracy [13].

Deep neural network refinement has also received considerable attention because MRI tumor classification requires both high sensitivity to pathological patterns and robustness against irrelevant image variations. CNN models may be improved through changes in layer depth, kernel configuration, activation functions, pooling strategy, dropout rate, normalization, and optimizer settings. These design choices can influence how effectively the model extracts features and how well it generalizes to unseen images. Regularization methods are especially important because medical datasets are often smaller than large natural-image datasets, creating a risk that the network may memorize training examples rather than learning general tumor characteristics. Therefore, the literature commonly employs dropout, data augmentation, batch normalization, learning-rate scheduling, and validation-based model selection to reduce overfitting. Such refinements support more reliable MRI classification, but they must be reported clearly to ensure that performance comparisons between studies remain meaningful [14].

Residual learning introduced another important direction for MRI classification because deeper CNN architectures can extract richer features but may also suffer from training degradation when depth increases. Residual connections allow information to pass across layers more effectively, which can improve gradient flow and support stable training of deeper networks. In the context of brain tumor MRI classification, residual architectures are useful because they can combine low-level anatomical features with high-level tumor-related representations. This is particularly relevant when the classification task involves visually similar categories, where subtle differences in texture, boundary structure, or spatial distribution may be necessary for correct prediction. Nevertheless, deeper networks are not automatically superior. Their performance depends on dataset size, augmentation quality,

parameter tuning, and the ability of the model to avoid learning dataset-specific artifacts. Thus, residual CNNs should be understood as powerful tools whose value depends on rigorous experimental design rather than architectural complexity alone [15].

Transfer learning is another dominant strategy in MRI–based brain tumor classification. Because medical image datasets are often limited in size, researchers frequently use models pretrained on large image collections and then adapt them to brain MRI classification by replacing or fine-tuning the final layers. This approach can reduce training time and improve convergence because the early layers of pretrained networks often learn general visual structures such as edges, corners, and texture primitives. However, transfer learning must be interpreted carefully in medical imaging because natural images and MRI scans have very different visual characteristics and semantic meanings. A feature that is useful for classifying natural objects may not directly correspond to medically meaningful MRI patterns. Therefore, transfer learning can be highly effective when properly adapted, but its success depends on fine-tuning strategy, dataset similarity, preprocessing compatibility, and validation rigor [16].

Recent studies have also examined transfer learning from the perspective of efficiency and practical deployment. In medical-image systems, accuracy alone is not the only criterion for success. A model may achieve strong classification performance but still be unsuitable for practical use if it requires excessive memory, long inference time, or specialized hardware. This is especially relevant for educational tools, web applications, or low-resource medical environments where computational resources may be limited. Efficient transfer-learning pipelines attempt to balance model complexity with usability by selecting architectures that provide acceptable performance while remaining feasible for deployment. Such work is important because brain tumor classification systems are increasingly expected to operate not only as experimental models but also as accessible tools that can present results quickly and clearly to users. This has encouraged comparative evaluation of pretrained architectures according to accuracy, number of parameters, inference speed, and general implementation feasibility [17].

Hybrid deep learning models have emerged as a response to the limitations of relying on one architecture alone. These models may combine CNN feature extraction with ensemble learning, transformer-based attention, recurrent components, or other classification layers in order to capture complementary aspects of MRI data. The motivation behind hybridization is that tumor appearance may require both local and global interpretation. Local features can describe boundaries, textures, and intensity transitions, while global features can reflect anatomical position and broader spatial relationships. Hybrid systems may therefore improve classification performance when tumor categories overlap visually or when a single model architecture fails to capture all relevant information. However, hybrid models also introduce methodological challenges because they are more complex, harder to reproduce, and often more computationally expensive. Their reported improvements should therefore be assessed together with their transparency, implementation cost, and comparison against simpler baseline models [18].

Attention-based and vision-transformer approaches have further expanded the methodological landscape of MRI tumor classification. Unlike conventional CNNs, which mainly process local neighborhoods through convolutional filters, transformer-based models can represent relationships between distant image regions through attention mechanisms. This can be useful in MRI classification because tumor interpretation may depend on global anatomical context as well as local lesion appearance. For example, the relationship between a lesion and surrounding brain structures may contribute to distinguishing tumor types. Nevertheless, transformer-based systems often require large amounts of data or strong pretraining to reach stable performance, and they can be computationally demanding. For this reason, many recent approaches combine convolutional layers with attention or transformer modules rather than replacing CNNs entirely. Such combinations seek to preserve the local sensitivity of CNNs while adding broader contextual modeling through attention mechanisms [19].

Explainability has become a necessary component of modern AI–based brain tumor classification because medical users require more than a final class label. A model that produces a prediction without showing the basis of its decision may be difficult to trust, especially in high-stakes medical contexts. Explainable AI methods such as Local Interpretable Model-Agnostic Explanations (LIME) and Gradient-weighted Class Activation Mapping (Grad-CAM) are commonly used to visualize which image regions contribute to a model output. In MRI classification, these tools can help determine whether the model is focusing on tumor tissue, surrounding anatomical structures, or irrelevant artifacts. However, explanation maps should be interpreted as supportive evidence rather than definitive clinical proof. A highlighted region does not automatically confirm diagnostic correctness, and a visually plausible explanation does not replace expert radiological assessment. Therefore, explainability strengthens transparency but must be combined with careful validation, class-wise error analysis, and responsible medical framing [20].

Overall, the literature demonstrates that MRI–based brain tumor classification has advanced from handcrafted feature extraction toward CNNs, residual networks, transfer learning, hybrid architectures, attention-based models, and explainable AI frameworks. The field has achieved strong progress, but several limitations remain consistent across existing studies. Reported performance may be affected by dataset size, class balance, preprocessing differences, lack of external validation, and insufficient reporting of class-wise errors. In addition, many models are evaluated on benchmark datasets that may not fully represent real clinical variability across scanners, institutions, acquisition protocols, and patient populations. For a review-oriented study, these issues are central because they determine how reported results should be interpreted. The current body of work supports the value of AI-assisted MRI classification, but it also indicates that reliable deployment requires robust validation, transparent reporting, explainable outputs, and a clear distinction between computational classification and professional medical diagnosis.

Table 1 summarizes the methodological directions of selected studies related to MRI–based tumor classification, medical image analysis, explainable artificial intelligence, transfer

learning, segmentation, radiomics, and optimization–assisted model design. The table is designed to support the review discussion by showing how different methodological strategies contribute to automated medical image classification and how they relate to brain tumor MRI analysis.

Table 1. Methodological comparison of selected MRI–based and artificial intelligence studies relevant to tumor classification.

Ref.	Study direction	Methodological strategy	Relevance to the present review
[21]	Brain tumor MRI classification	Generative AI with deep–learning classification	Represents the use of generative AI to improve MRI–based tumor classification by increasing data diversity or supporting representation learning, which is relevant to discussions of augmentation, limited medical datasets, and generalization.
[22]	Multi–class brain tumor classification	AI–based hybrid framework using BO–DenseXGB	Provides an example of a hybrid intelligent classification pipeline, showing how deep feature extraction and advanced decision models can be combined for multi–class tumor recognition from MRI images.
[23]	Multi–institutional neuro–oncology classification	Comparative machine–learning and deep–learning evaluation	Highlights the importance of testing AI models across multiple institutions, which is essential for assessing robustness against scanner variation, protocol differences, and dataset–specific bias.
[24]	Multi–class brain tumor MRI classification	Hybrid deep–learning framework with explainable AI	Supports the review discussion on combining predictive performance with interpretability, especially when MRI classification systems are expected to provide transparent and clinically understandable outputs.
[25]	Brain tumor MRI classification overview	Comparative analysis of state–of–the–art algorithms	Provides a broad methodological reference for comparing modern algorithms used in MRI–based tumor classification, including trends in deep learning, hybrid systems, and performance evaluation.
[26]	Multiparametric MRI lesion classification	AI–based classification using multiple MRI sequences	Demonstrates the value of multiparametric MRI inputs, which is relevant to future brain tumor studies that may move beyond single–slice or single–sequence classification toward richer imaging representations.
[27]	Brain tumor MRI detection	ResNet50 with Grad–CAM explainability	Illustrates the integration of CNN–based detection with visual explanation maps, supporting the argument that MRI classification models should provide interpretable evidence rather than prediction labels alone.
[28]	Brain tumor detection and classification	Integrated AI–based classification workflow	Shows the movement toward complete AI pipelines that combine image input, classification, and user–oriented output, which is important for practical educational and decision–support applications.
[29]	Low–grade glioma MRI segmentation	AI–based brain MRI segmentation for diagnosis and planning	Emphasizes the methodological connection between segmentation and classification, since accurate tumor localization can improve feature extraction, spatial interpretation, and treatment–planning relevance.
[30]	MRI–based brain tumor classification	Explainable AI–driven deep–learning model	Supports the review focus on explainability as a necessary component of medical AI systems, particularly when model outputs must be interpreted by clinicians or educational users.
[31]	Brain tumor grade classification	CNN model supported by Genetic Algorithm optimization	Demonstrates an optimization–assisted CNN methodology, where GA can contribute to architecture or parameter selection for MRI–based tumor grading and classification.
[32]	Primary brain tumor multi–classification	Vision Transformer and D–CNN models with XAI	Represents attention–based and convolutional modeling for MRI tumor classification, showing how local CNN features and global transformer representations can be compared or combined.
[33]	Brain tumor MRI detection	AI–based review of advancements and challenges	Provides methodological context for discussing the progress and limitations of MRI–based AI detection systems, including challenges related to reliability, validation, and clinical translation.
[34]	Glioma MRI classification	Lightweight CNN with explainable AI	Highlights the importance of efficient model design, showing that lightweight CNN architectures can support MRI classification while reducing computational requirements for deployment.
[35]	Brain tumor MRI classification	Two–phase fine–tuned Xception model with explainable AI	Supports the discussion of staged transfer–learning strategies, where fine–tuning and explainability are combined to improve classification reliability and output transparency.
[36]	Glioma tumor grading	Radiomics–based machine–learning classification	Shows how radiomic feature extraction remains relevant for glioma grading, especially when classification is linked to structured imaging biomarkers and updated tumor classification criteria.
[37]	Brain tumor detection and classification	Deep learning and transfer learning	Provides methodological support for the role of transfer learning in MRI tumor analysis, particularly when pretrained networks are adapted to limited medical image datasets.
[38]	Brain tumor MRI classification	Fine–tuned transfer learning with explainable AI	Reinforces the recent trend of combining pretrained architectures with explanation methods, supporting model transparency and more responsible interpretation of classification outputs.
[39]	MRI texture–feature classification	Machine–learning classification of tumor and peri–tumor texture features	Demonstrates the importance of peri–tumor information and texture analysis, which is relevant to brain MRI classification because surrounding tissue patterns may improve class discrimination.
[40]	MRI–based cancer classification	AI classification under different molecular subtype reference standards	Highlights that label definitions and reference standards can strongly affect AI classification outcomes, supporting the need for careful ground–truth construction in MRI tumor studies.

Notes: AI: Artificial Intelligence; MRI: Magnetic Resonance Imaging; CNN: Convolutional Neural Network; D–CNN: Deep Convolutional Neural Network; XAI: Explainable Artificial Intelligence; Grad–CAM: Gradient–weighted Class Activation Mapping; GA: Genetic Algorithm; BO–DenseXGB: Bayesian Optimization–Dense Extreme Gradient Boosting; ViT: Vision Transformer.

The methodological literature on MRI–based tumor classification shows that recent studies are no longer limited to direct CNN classification alone, but increasingly integrate generative learning, explainability, transfer learning, segmentation, radiomics, optimization–assisted modeling, and multi–institutional validation. One recent direction investigates the role of generative artificial intelligence in improving brain tumor MRI classification by addressing the limited diversity of medical image datasets. This direction is important because MRI datasets are often smaller, less diverse, and more difficult to annotate than general computer–vision datasets. Generative models may help increase the effective variability of the training distribution, support augmentation, and reduce overfitting when the generated samples are realistic and clinically meaningful. However, the methodological value of this direction depends on whether synthetic or enhanced samples improve performance on real unseen MRI images rather than only improving internal validation results. Therefore, generative AI should be interpreted as a potentially useful support mechanism for data enrichment, but not as a substitute for high–quality clinical data, careful validation, and transparent reporting of training and testing separation [21].

Another methodological direction focuses on hybrid AI–based classification frameworks for multi–class brain tumor MRI recognition. Hybrid frameworks are designed to combine complementary modeling strengths, such as deep feature extraction and advanced classification layers, rather than depending on a single model family. This approach is relevant because brain tumor classes may differ in anatomical position, texture, boundary sharpness, internal heterogeneity, and surrounding tissue response. A single classifier may not capture all of these variations equally well, whereas a hybrid system can potentially improve discriminative ability by combining learned representations with stronger decision mechanisms. Nevertheless, hybrid models must be evaluated cautiously because increased complexity may reduce reproducibility and make it difficult to identify which component is responsible for the observed improvement. A strong hybrid methodology should therefore include transparent architecture description, fair comparison with simpler baselines, class–wise analysis, and validation on unseen data [22].

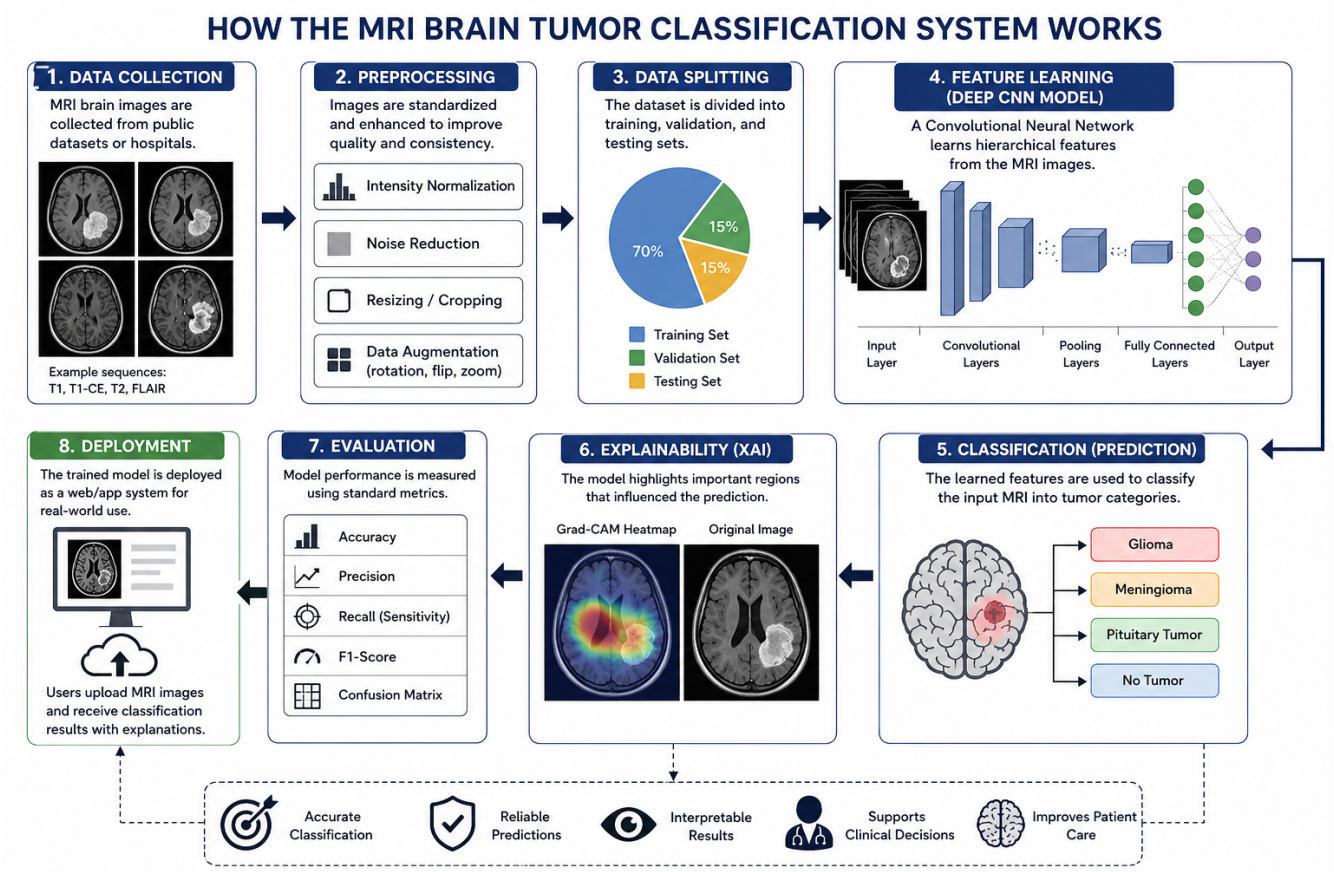

Figure 1. General workflow of the MRI–based brain tumor classification system. The diagram presents the complete processing pipeline, starting from MRI image collection and preprocessing, followed by dataset splitting, CNN–based feature learning, tumor class prediction, explainability analysis, performance evaluation, and final deployment.

Figure 1 illustrates the complete workflow of the MRI–based brain tumor classification system. The process begins with the collection of MRI brain images, which may be obtained from public datasets, hospital archives, or clinical repositories. These images represent the raw input of the classification pipeline and may include different MRI sequences, image resolutions, contrast levels, and anatomical views. Since MRI data can vary significantly according to scanner type, acquisition protocol, and image quality, the collected images must be standardized before they are used by the learning model.

The preprocessing stage is responsible for preparing the MRI images in a consistent format suitable for model training and inference. This stage may include intensity normalization, noise reduction, resizing, cropping, and data augmentation. Intensity normalization reduces differences in brightness and contrast between images, while resizing ensures that all input images have the same spatial dimensions. Data augmentation can improve generalization by exposing the model to controlled variations of the training images, such as rotation, flipping, or zooming. These preprocessing operations help the model focus on tumor–related visual patterns rather than irrelevant variations caused by image acquisition or formatting.

After preprocessing, the dataset is divided into training, validation, and testing subsets. The training set is used to optimize the model parameters, the validation set is used to monitor learning behavior and support model selection, and the testing set is reserved for final performance evaluation on unseen images. This separation is essential because it reduces the risk of biased evaluation and provides a more reliable estimate of how the model may behave on new MRI images. A well–designed split also helps prevent data leakage, which can otherwise produce overly optimistic classification results.

The core classification component shown in Figure 1 is the CNN–based feature learning stage. In this stage, the preprocessed MRI images are passed through convolutional layers, pooling layers, and fully connected layers. The early convolutional layers learn basic visual patterns such as edges, intensity transitions, and local textures, while deeper layers

extract more complex features related to tumor shape, location, and structural abnormality. These learned features are then used by the classification layer to assign the input MRI image to one of the target classes, such as glioma, meningioma, pituitary tumor, or no tumor.

The diagram also includes an explainability stage, which is important for improving the transparency of medical AI systems. Explainability methods, such as Grad–CAM, can highlight the image regions that contributed most strongly to the model prediction. This allows researchers and users to inspect whether the model is focusing on meaningful tumor–related regions or on irrelevant background areas. Although explainability does not replace professional radiological interpretation, it provides useful visual support for understanding model behavior.

The evaluation stage measures the performance of the trained model using standard metrics such as accuracy, precision, recall, F1–score, and the confusion matrix. These metrics are necessary because overall accuracy alone may not fully describe model behavior, especially in multi–class classification problems where some tumor types may be easier to identify than others. Finally, the deployment stage shows how the trained model can be integrated into a web application or software system, allowing users to upload MRI images and receive classification outputs. In this review context, the deployed system should be treated as an educational and research–support tool rather than a substitute for expert medical diagnosis.

2.1 Deep Learning and Transfer Learning for MRI–Based Brain Tumor Classification

This subsection can discuss the evolution from conventional machine–learning methods to deep learning models, with emphasis on CNN architectures, pretrained networks, transfer learning, and fine–tuning strategies. It should explain why CNNs became dominant in MRI–based brain tumor classification, how convolutional layers learn hierarchical image representations, and why transfer learning is commonly used when medical datasets are limited. This part can also discuss the strengths and weaknesses of pretrained models, especially the difference between natural image pretraining and medical MRI interpretation. The subsection should be large because it forms the technical foundation of the review and connects directly to most MRI classification studies.

2.2 Hybrid, Optimization–Assisted, and Explainable Artificial Intelligence Models

This subsection can focus on more advanced methodological directions, including hybrid deep learning frameworks, attention–based models, transformer–inspired architectures, optimization–assisted CNN design, and explainable artificial intelligence. It can explain how hybrid models combine different computational components to improve feature representation and classification reliability. It can also discuss how optimization methods, such as Genetic Algorithm (GA), may support model selection, parameter tuning, or architecture design. The explainability part should discuss Grad–CAM, LIME, attention maps, and the importance of showing which MRI regions influence the model prediction. This subsection is important because it moves beyond simple classification accuracy and addresses transparency, interpretability, and model trust.

2.3 Dataset Quality, Validation Strategy, and Clinical Applicability

This subsection can discuss the methodological and practical limitations of MRI–based brain tumor classification studies. It should cover dataset size, class balance, duplicate removal, data leakage prevention, preprocessing consistency, multi–institutional validation, external testing, and class–wise evaluation metrics. This part should also explain why high accuracy on a benchmark dataset does not automatically mean clinical readiness. It can include discussion of accuracy, precision, recall, F1–score, confusion matrices, and the need to distinguish between computational classification and professional medical diagnosis. This subsection is essential because it gives the review a critical academic tone rather than simply summarizing model performance.

3. DISCUSSION

The reviewed body of work shows that MRI–based brain tumor classification has developed into a mature but still evolving research direction within medical artificial intelligence. The field has moved beyond the early question of whether artificial intelligence can identify tumor patterns in MRI images and has shifted toward more demanding questions related to model reliability, interpretability, robustness, generalization, and practical usability. This progression is important because medical image classification is not only a technical pattern–recognition problem. It is also connected to clinical uncertainty, image acquisition variability, annotation quality, dataset representativeness, and the need for transparent communication of model outputs. Therefore, the discussion of MRI–based brain tumor classification must consider the entire methodological chain, starting from data collection and preprocessing and extending to model evaluation, explainability, and deployment.

A central observation from the reviewed literature is that deep learning, particularly CNN–based modeling, has become the dominant approach for brain tumor MRI classification because of its ability to learn visual features directly from image data. This advantage is especially relevant in brain tumor analysis, where the visual characteristics of tumors are highly variable. Tumors may differ in anatomical location, shape, internal texture, boundary clarity, contrast enhancement, surrounding edema, and relationship to adjacent brain structures. Traditional handcrafted features can describe some of these properties, but they may not fully capture the complex spatial and textural representations required for reliable classification. CNN models address this limitation by learning hierarchical features, where early layers detect local visual primitives and deeper layers combine these features into higher–level tumor representations. However, this representational strength also introduces a challenge: the learned features are not always directly interpretable, and the model may learn dataset–specific shortcuts if the data are not carefully curated.

The role of dataset quality is one of the most critical issues in this research area. The reliability of any MRI–based classification model depends strongly on the structure, diversity, and integrity of the dataset used for training and

testing. If the dataset is small, imbalanced, duplicated, or collected from a narrow imaging environment, the resulting model may achieve strong internal performance while failing to generalize to external data. This problem is particularly important in medical imaging because MRI scans vary widely across institutions, scanners, acquisition protocols, field strengths, imaging sequences, and preprocessing procedures. A model trained on one dataset may unintentionally learn technical properties associated with a specific dataset rather than tumor–specific characteristics. For this reason, future studies should place greater emphasis on independent testing, multi–institutional validation, and transparent reporting of data splitting. Without these safeguards, reported performance values may overestimate the actual reliability of the system.

Another important issue is the distinction between classification performance and clinical usefulness. A model may achieve high accuracy on a benchmark dataset, but this does not automatically mean that it is ready for clinical deployment. Brain tumor diagnosis is a complex process that may require multiple MRI sequences, patient history, neurological examination, histopathology, molecular markers, and expert radiological interpretation. An image–level classification model can support this process by identifying visual patterns and suggesting a possible class, but it cannot replace the broader diagnostic reasoning performed by medical professionals. This distinction must be maintained carefully in review papers and applied systems because overstatement of artificial intelligence capabilities can create unrealistic expectations. In this context, MRI–based classification systems should be framed as research tools, educational systems, or decision–support components unless they have undergone rigorous clinical validation.

The reviewed studies also show that transfer learning remains highly influential because medical datasets are usually smaller than the datasets required for training deep networks from the beginning. Transfer learning provides a practical solution by adapting pretrained models to MRI classification tasks. This can improve convergence, reduce training requirements, and support strong performance even when the available labeled data are limited. Nevertheless, transfer learning has limitations that must be discussed with care. Most pretrained models are originally trained on natural images, which differ substantially from MRI scans in visual structure, texture distribution, semantic meaning, and diagnostic relevance. Therefore, transfer learning should not be interpreted as a direct transfer of medical knowledge. Instead, it should be understood as a computational initialization strategy that may provide useful low–level visual filters but still requires domain–specific adaptation, careful preprocessing, and reliable validation.

Hybrid models represent another major direction in the literature. Their main advantage is that they attempt to combine complementary strengths from different computational methods. For example, CNNs can extract local spatial features, transformer–based components can model global relationships, ensemble systems can reduce dependence on a single model, and optimization methods can support architecture or parameter selection. This direction is methodologically attractive because brain tumor MRI classification often requires both local lesion analysis and broader anatomical context. However, hybrid models can also become difficult to reproduce and interpret. When a system contains several interconnected components, it may be unclear whether improved performance comes from genuinely better representation learning, stronger classification boundaries, heavier parameterization, or dataset–specific tuning. Therefore, the value of hybrid models should be assessed not only through accuracy but also through transparency, computational feasibility, reproducibility, and comparison with simpler alternatives.

Explainability is one of the strongest themes emerging from recent MRI classification research. In medical artificial intelligence, prediction alone is not sufficient. Users need to understand why a model produced a given output and whether the decision appears consistent with meaningful anatomical or pathological regions. Explainability methods such as Grad–CAM and related visualization strategies can help by highlighting areas that contributed to the model decision. This is particularly valuable for brain tumor MRI classification because it allows researchers to inspect whether the model focuses on the tumor region, peri–tumor tissue, normal brain structures, or irrelevant image artifacts. However, explainability should not be treated as a complete solution to the interpretability problem. A heatmap does not prove that the model is medically correct, and an apparently plausible explanation can still accompany an incorrect prediction. Thus, explainability should be used as part of a broader validation framework that also includes quantitative metrics, class–wise analysis, expert inspection, and external testing.

The workflow shown in Figure 1 reflects the need to consider MRI classification as a complete pipeline rather than as an isolated model. The process begins with MRI image collection and continues through preprocessing, dataset splitting, CNN–based feature learning, classification, explainability, evaluation, and deployment. Each stage can influence the final output. Preprocessing choices such as resizing, normalization, cropping, and augmentation affect what information is preserved and how the model perceives the input images. Dataset splitting determines whether the evaluation is reliable or biased by data leakage. Feature learning determines how the model transforms MRI patterns into discriminative representations. Classification produces the predicted label, while explainability and evaluation help determine whether the prediction is reliable and interpretable. Deployment then introduces another layer of responsibility, because the way results are presented to users can influence how the system is understood and trusted.

Performance evaluation requires particular attention in multi–class brain tumor classification. Overall accuracy is useful, but it can be insufficient when the classification task includes several tumor types with different levels of visual difficulty. A model may perform very well on clearly distinguishable classes while struggling with heterogeneous or visually overlapping tumor categories. Therefore, precision, recall, F1–score, and confusion matrix analysis are necessary to understand class–specific behavior. Recall is especially important when the objective is to identify all cases of a particular tumor class, while precision is important when evaluating how often the predicted class is correct. The confusion matrix is valu-

able because it reveals which classes are commonly confused with each other. In review studies, the use of class–wise metrics is essential because it prevents a single aggregate score from hiding important weaknesses in model behavior.

The reviewed literature also suggests that radiomics and classical ML remain relevant even though DL dominates many recent studies. Radiomics can provide structured features related to tumor intensity, shape, texture, and heterogeneity, which may be useful for grading or clinically oriented analysis. Classical ML models can also be valuable when datasets are small, when interpretability is prioritized, or when features are carefully engineered from medically meaningful regions. Rather than viewing radiomics, ML, and DL as competing approaches, it is more accurate to view them as complementary methodological families. DL is powerful for automatic feature learning, radiomics is useful for structured quantitative analysis, and ML can provide efficient classification when informative features are available. Future systems may benefit from combining these approaches in a careful and transparent way.

A major limitation across the field is the frequent lack of external validation. Many studies report strong results on public or internally curated datasets, but fewer studies test their models across independent institutions or clinically diverse cohorts. This limitation reduces confidence in generalization. External validation is difficult because it requires access to well–annotated medical images from different sources, but it is essential for assessing real–world reliability. Brain tumor MRI classification models should be tested across different scanners, image qualities, sequences, patient populations, and acquisition protocols. This would help determine whether the model has learned tumor–specific features or has simply adapted to the visual characteristics of a particular dataset. Without external validation, model performance should be interpreted as evidence of benchmark success rather than evidence of clinical readiness.

Another limitation concerns the use of two–dimensional images instead of full volumetric MRI studies. Many classification pipelines use individual slices, which simplifies model design and reduces computational requirements. However, brain tumors are three–dimensional structures, and important diagnostic information may exist across adjacent slices or across multiple MRI sequences. Slice–based classification may miss spatial continuity, tumor volume, and relationships between the lesion and surrounding anatomical structures. Three–dimensional CNNs, multi–sequence models, and volumetric transformer approaches could address this limitation by incorporating richer spatial context. However, these methods require more computational resources, larger datasets, and more careful preprocessing. Therefore, the future of MRI–based classification will likely involve a balance between model complexity, data availability, and practical deployment requirements.

The deployment of MRI–based classification systems also raises ethical and practical considerations. If the output is presented too confidently, users may mistake a computational prediction for a confirmed diagnosis. This is especially risky when systems are made available through web applications or educational interfaces. Responsible deployment requires clear disclaimers, uncertainty communication, and careful wording that distinguishes model prediction from medical diagnosis. Users should be informed that the system is intended to support research, education, or preliminary interpretation, not to replace a radiologist or clinical team. Confidence scores should also be explained carefully because a high softmax probability reflects model certainty under its learned distribution, not clinical certainty. Ethical deployment therefore requires both technical reliability and responsible communication.

Overall, the discussion indicates that MRI–based brain tumor classification is a promising but methodologically sensitive field. The strongest current directions include CNN–based feature learning, transfer learning, hybrid modeling, explainability, lightweight architectures, radiomics, segmentation support, and multi–institutional validation. However, the field still faces limitations related to dataset bias, limited external validation, inconsistent reporting, model interpretability, computational complexity, and clinical translation. The most reliable future systems will likely be those that combine high predictive performance with transparent methodology, robust validation, interpretable outputs, and careful communication of limitations. In this sense, progress in MRI–based tumor classification should not be measured only by higher accuracy values, but by the development of systems that are scientifically reproducible, clinically cautious, and practically usable.

4. CONCLUSION

This review presented an expanded discussion of artificial intelligence methods for MRI–based brain tumor classification, with emphasis on CNN models, transfer learning, hybrid architectures, explainable artificial intelligence, radiomics, segmentation–assisted analysis, optimization–supported design, and deployment–oriented workflows. The reviewed literature demonstrates that automated MRI classification has advanced considerably and that modern AI models can learn meaningful visual representations from brain MRI images. These systems have shown strong potential for supporting tumor recognition, educational analysis, and research–oriented decision support. However, the review also shows that the field remains constrained by important methodological challenges that must be addressed before such systems can be considered broadly reliable.

The most important conclusion is that model performance cannot be evaluated through accuracy alone. Brain tumor MRI classification is a multi–class medical imaging problem in which different tumor categories may vary substantially in visual complexity. Therefore, reliable evaluation requires class–wise precision, recall, F1–score, and confusion matrix analysis. These metrics help reveal whether a model performs consistently across all classes or whether it performs strongly only on visually easier categories. This distinction is essential because a high aggregate score may hide clinically meaningful weaknesses. Future studies should therefore report detailed evaluation results and avoid relying only on a single global performance number.

A second conclusion is that data quality and validation design are as important as model architecture. Deep learning models can learn powerful representations, but they can also learn biases, artifacts, or dataset–specific patterns when the training

data are limited or poorly controlled. Reliable classification requires balanced datasets, careful preprocessing, duplicate removal, prevention of data leakage, and independent testing. External validation across institutions and imaging protocols should become a standard requirement rather than an optional extension. Without such validation, reported results should be interpreted as benchmark performance rather than evidence of real–world clinical readiness.

A third conclusion is that explainability is no longer an optional feature in medical AI systems. Since brain tumor classification affects a high–risk medical domain, users need transparent outputs that support inspection and interpretation. Explainability tools can help identify whether the model focuses on tumor–related regions, but they should be understood as supportive visual aids rather than definitive clinical evidence. The most responsible systems will combine explainability with rigorous quantitative evaluation, expert review, uncertainty communication, and clear limitations. In this way, explainable AI can improve transparency while avoiding the false impression that a visual heatmap is equivalent to a medical diagnosis.

The review also concludes that future research should move toward richer and more clinically realistic modeling. Multi–sequence MRI inputs, volumetric analysis, three–dimensional architectures, segmentation–guided classification, radiomics–deep learning integration, and lightweight deployable models are all promising directions. These approaches can improve the ability of AI systems to represent tumor structure, surrounding tissue, and anatomical context. However, increased complexity must be accompanied by reproducibility, computational feasibility, and careful validation. A complex model is not necessarily better unless it provides measurable, interpretable, and generalizable improvement over simpler methods.

Finally, MRI–based brain tumor classification should be framed as a supportive technology rather than an autonomous diagnostic replacement. AI systems can assist in image analysis, accelerate research workflows, provide educational explanations, and support preliminary classification. However, final medical interpretation must remain the responsibility of qualified healthcare professionals who can integrate imaging findings with clinical history, neurological examination, pathology, molecular information, and treatment context. The future value of AI in brain tumor MRI analysis will depend not only on achieving higher classification performance, but also on building systems that are transparent, validated, ethical, and aligned with real medical practice

Multi–institutional evaluation is one of the most important methodological requirements for medical AI because it tests whether a model can generalize beyond the dataset on which it was developed. MRI images collected from different institutions may vary in scanner type, acquisition protocol, field strength, slice thickness, contrast settings, preprocessing routine, and annotation practice. These variations can strongly influence model behavior because a network may learn institution–specific visual patterns instead of tumor–specific pathology. For this reason, AI models that perform well on a single internal dataset may not necessarily achieve the same performance when tested on images from other hospitals or imaging centers. Multi–institutional studies are therefore valuable because they expose classification systems to more realistic variability and provide stronger evidence for robustness. In brain tumor classification, this methodological direction is especially important because generalization is a necessary step before any model can be considered reliable for broader medical use [23].

Hybrid deep learning combined with explainable AI represents a further methodological advancement in MRI–based brain tumor classification. The purpose of this direction is not only to improve predictive performance, but also to make the prediction process more interpretable for users. In medical imaging, a classification model should ideally provide more than a final label because clinicians and researchers need to understand whether the model is focusing on tumor–related regions, normal anatomical structures, or irrelevant image artifacts. Explainability methods can therefore support visual inspection of model attention and help identify possible failure cases. When hybrid deep learning is combined with explainability, the resulting framework can offer both strong feature representation and a more transparent interpretation of the classification decision. However, explainability should still be treated as supportive evidence rather than definitive proof of diagnostic correctness, because visual explanation maps may be affected by model architecture, layer selection, and preprocessing choices [24].

A broad comparative view of state–of–the–art algorithms is methodologically useful because it helps clarify which model families are commonly applied to brain tumor MRI classification and how they differ in their strengths and limitations. Comparative algorithmic studies can include conventional ML methods, CNN architectures, transfer learning models, ensemble systems, attention–based networks, and hybrid frameworks. Such comparisons are valuable because the same dataset may produce different results depending on preprocessing, augmentation, optimization, architecture depth, and evaluation metrics. They also help identify whether improvements are caused by genuinely better representation learning or by methodological factors such as dataset simplicity or weak validation design. For a review paper, this type of reference supports a balanced discussion because it places individual model results within the wider development of the field rather than treating each reported accuracy as an isolated achievement [25].

Multiparametric MRI classification provides an important methodological lesson for brain tumor analysis because it shows that richer imaging inputs may improve the representation of disease characteristics. A single MRI image or sequence may capture only part of the pathological information, while multiple MRI sequences can provide complementary details about tissue composition, enhancement behavior, edema, vascularity, and lesion structure. Although some brain tumor classification studies rely on two–dimensional images or single–sequence inputs, future systems may benefit from multi–sequence learning strategies that use T1–weighted, T2–weighted, contrast–enhanced, or FLAIR information together. The relevance of this direction is that tumor classes may become easier to distinguish when the model receives more complete imaging context. However, multiparametric learning also requires careful alignment, preprocessing, and handling of missing sequences, which makes the methodology

more demanding than single–image classification [26].

CNN–based detection supported by Grad–CAM explainability is methodologically important because it connects high–level prediction with visual evidence. CNN models can produce strong classification results, but their internal decision processes are difficult to interpret directly. Grad–CAM addresses this limitation by generating heatmaps that highlight image regions contributing strongly to the model output. In MRI–based brain tumor detection and classification, this is useful because it allows researchers to inspect whether the model is attending to the tumor region or to irrelevant background structures. This type of explanation is particularly valuable in medical AI because a model may appear accurate while still relying on non–clinical shortcuts. Nevertheless, Grad–CAM maps should be interpreted carefully because they are not equivalent to expert segmentation masks or radiological annotations. They provide approximate visual evidence of model attention, not a definitive clinical explanation [27].

Integrated AI workflows for brain tumor detection and classification are relevant because real applications require more than model training. A complete workflow may include image upload, preprocessing, model inference, class probability estimation, visualization, result interpretation, and user communication. This is methodologically important because even a high–performing model can be difficult to use if the surrounding system is poorly designed. In practical MRI classification environments, users need clear outputs, confidence interpretation, and appropriate warnings about the limitations of automated prediction. Integrated systems therefore connect algorithmic performance with usability, accessibility, and responsible communication. For review purposes, this direction shows that the evaluation of AI–based brain tumor classification should include not only the classifier but also the pipeline through which image data are processed and results are delivered to the user [28].

Segmentation–based AI research contributes strongly to brain tumor classification because localization and classification are closely connected. Segmentation identifies the spatial extent of the tumor, while classification assigns a category or diagnostic label. When tumor localization is accurate, classification models can focus more directly on pathological tissue and may avoid being influenced by irrelevant background regions. This can improve feature extraction, support volumetric analysis, and provide more clinically meaningful interpretation. In low–grade glioma analysis, segmentation is particularly important because tumor boundaries may be subtle, diffuse, or difficult to distinguish from surrounding tissue. However, segmentation requires detailed annotations, and errors in segmentation can propagate into downstream classification. Therefore, segmentation–assisted classification is powerful but methodologically more complex than direct image–level classification [29].

Explainable deep learning models for MRI–based brain tumor classification emphasize the need for transparent AI in medical image analysis. A model that produces a class label without justification may be scientifically incomplete and clinically difficult to trust. Explainable AI attempts to address this issue by providing visual or analytical evidence of the decision process. This is especially important in brain tumor MRI classification because misclassification may occur when tumor types share similar imaging features or when image quality is low. Explainability can help researchers examine whether the model uses meaningful tumor features, but it must be combined with quantitative validation, confusion matrix analysis, and expert review. The methodological value of explainable models therefore lies not only in making the system more understandable, but also in supporting error analysis and identifying weaknesses in learned representations [30].

Optimization–assisted CNN design is another important methodological route in MRI–based tumor grading and classification. In this direction, deep learning is combined with an optimization method to improve model configuration, parameter selection, feature selection, or architecture design. The use of Genetic Algorithm (GA) in this context is relevant because GA can search through possible configurations and select better solutions according to a defined objective function. This can be useful when manual tuning is difficult or when the model contains many design choices. In MRI brain tumor analysis, optimization–assisted CNNs may help improve performance by selecting more effective network structures or hyperparameters. However, such approaches must clearly describe what is being optimized and how the search process is validated, because optimization can also increase the risk of overfitting if it is repeatedly guided by limited validation data [31].

Attention–based modeling and transformer–inspired classification have become important because they provide a mechanism for capturing long–range relationships in medical images. CNNs are strong at local feature extraction, but transformer–based methods can model relationships between distant image regions using attention. In brain tumor MRI classification, this can be useful because tumor interpretation may depend not only on lesion texture but also on anatomical position and surrounding structures. The comparison between vision transformer models and deep convolutional models is therefore methodologically valuable because it clarifies whether global context modeling improves classification beyond local convolutional features. Explainable analysis within this setting is also important because attention maps and explanation tools can help inspect what the model has learned. However, transformer–based systems often require large datasets or strong pretraining, so their benefits must be evaluated against computational cost and data requirements [32].

Review work on AI–based brain tumor detection provides methodological context by summarizing the general progress, challenges, and limitations of automated MRI analysis. Such studies are useful because they show that the field has advanced significantly, but still faces recurring problems related to dataset quality, lack of external validation, limited interpretability, and inconsistent evaluation metrics. They also help distinguish between detection, classification, segmentation, grading, and diagnosis, which are often incorrectly treated as equivalent. From a methodological perspective, this distinction is essential because each task requires different data labels, model outputs, evaluation metrics, and clinical interpretation. A detection model may only identify whether a tumor is present, while a classification model assigns a tumor

category, and a grading model estimates severity. Therefore, broad AI reviews help frame brain tumor MRI classification within the larger medical image analysis pipeline [33].

Lightweight CNN architectures address the practical need for efficient MRI classification models. While large deep networks may achieve strong performance, they can require substantial memory, processing power, and inference time. This creates barriers for deployment in web applications, low–resource environments, educational tools, or clinical settings without advanced hardware. Lightweight models attempt to reduce computational cost while preserving classification accuracy. This direction is methodologically important because practical AI systems must balance predictive performance with usability and accessibility. In glioma MRI classification, lightweight CNNs are especially relevant because glioma appearance can be heterogeneous and difficult to classify, yet efficient models are still needed for scalable use. The integration of explainable AI with lightweight models further strengthens their value by making compact systems more transparent and easier to inspect [34].

Two–phase fine–tuning strategies represent a structured approach to transfer learning in MRI classification. Instead of adapting a pretrained model in a single step, a two–phase method may first train selected layers and then fine–tune deeper layers more carefully. This can help the model preserve general image features while gradually adapting to the specific characteristics of brain MRI data. Such staged training is useful when the target dataset is not large enough to support full training from the beginning, but still requires domain–specific adaptation. The methodological strength of this approach is that it can improve stability and reduce overfitting compared with uncontrolled fine–tuning. When combined with explainable AI, two–phase fine–tuning can also provide a more transparent classification workflow by allowing researchers to evaluate both predictive performance and visual attention patterns [35].

Radiomics–based machine learning remains relevant even in the deep learning era because it provides structured and interpretable quantitative features from medical images. In glioma grading, radiomics can describe tumor texture, shape, intensity heterogeneity, and spatial relationships in a way that may relate to biological aggressiveness. This is different from end–to–end deep learning, where features are learned automatically but are often difficult to interpret. Radiomics pipelines can therefore support clinically meaningful analysis when features are carefully extracted and validated. However, radiomics is sensitive to segmentation accuracy, scanner differences, preprocessing choices, and feature selection. In brain tumor studies, radiomics is methodologically important because it bridges imaging biomarkers and machine learning, but it requires strict standardization to avoid unstable or non–reproducible results [36].

Deep learning and transfer learning approaches for brain tumor detection and classification demonstrate the continued importance of pretrained models in medical imaging. Transfer learning is attractive because it allows researchers to adapt models that have already learned general visual features, reducing the need for very large MRI datasets. In brain tumor MRI analysis, this can improve convergence and support classification when labeled medical data are limited. However, pretrained models must be adapted carefully because MRI images differ significantly from natural images in structure, contrast, and semantic meaning. The methodological relevance of this direction is that transfer learning offers a practical compromise between training efficiency and classification performance, but its success depends on preprocessing consistency, fine–tuning strategy, dataset quality, and evaluation design [37].

Fine–tuned transfer learning combined with explainable AI reflects one of the most current methodological trends in MRI–based brain tumor classification. This approach attempts to combine the benefits of pretrained models, domain–specific adaptation, and decision transparency. Fine–tuning allows the model to adjust its learned features to MRI data, while explainability helps inspect whether the adapted model is focusing on medically meaningful regions. This is valuable because a model may achieve high accuracy but still rely on biased or irrelevant visual cues. By combining transfer learning with explanation methods, researchers can produce classification systems that are more transparent and more suitable for review–level discussion. However, such systems still require independent validation and careful reporting because explanation outputs do not guarantee clinical correctness by themselves [38].

Texture–feature classification from tumor and peri–tumor regions provides an important reminder that information surrounding the visible lesion may also be diagnostically meaningful. In MRI, peri–tumor tissue can contain edema, infiltration, vascular response, or structural distortion, depending on the tumor type and biological behavior. Classical ML approaches that extract texture features from both tumor and peri–tumor regions can therefore capture patterns that may not be represented by the lesion core alone. This concept is relevant to brain tumor classification because some tumor categories may be distinguished by how they interact with surrounding tissue rather than only by internal texture. Methodologically, this supports the idea that future deep learning systems may benefit from spatially aware inputs, segmentation masks, or region–specific feature extraction rather than treating the entire image as a uniform input [39].

The definition of ground truth and reference standards is a critical methodological issue in AI–based MRI classification. A model learns from the labels provided during training, so inconsistent or clinically ambiguous labels can directly affect model behavior. Differences in reference standards may lead to different classification outcomes, even when the same imaging data are used. This issue is highly relevant to brain tumor MRI analysis because tumor diagnosis may depend on histopathology, molecular markers, radiological findings, and clinical context. If datasets define labels using different criteria, model comparisons become difficult and reported results may not reflect the same diagnostic target. Therefore, careful label construction, transparent reference standards, and consistent class definitions are essential for reliable AI evaluation in medical imaging [40].

Taken together, these methodological references show that MRI–based tumor classification is developing into a multi–component research field that includes data preparation, image enhancement, deep representation learning, optimization, segmentation, radiomics, transfer learning, explainability, and

deployment–oriented design. The strongest methodological trend is the movement from isolated classification models toward complete, interpretable, and more generalizable AI pipelines. However, the reviewed studies also show that high classification performance must be interpreted with caution unless it is supported by strong validation, transparent data handling, class–wise evaluation, and clinically meaningful explanation. For this reason, the methodological discussion in this review treats automated MRI classification as a promising support tool while maintaining that final medical interpretation must remain grounded in expert clinical and radiological assessment.

REFERENCES

[1] Gopal S. Tandel, Antonella Balestrieri, Tanay Jujaray, Narender N. Khanna, Luca Saba, and Jasjit S. Suri. Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm. *Computers in Biology and Medicine*, 122:103804, 2020.

[2] Ahmad Saleh, Rozana Sukaik, and Samy S. Abu-Naser. Brain tumor classification using deep learning. In *2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech)*, pages 131–136. IEEE, 2020.

[3] Hapsari Peni Agustin Tjahyaningtijas, Dewinda Julianensi Rumala, Cucun Very Angkoso, Nurul Zainal Fanani, Joan Santoso, Anggraini Dwi Sensusiati, Peter MA Van Ooijen, IKE Ketut Eddy Purnama, and Mauridhi Hery Purnomo. Brain tumor classification in MRI images using en-CNN. *International Journal of Intelligent Engineering and Systems*, 14(4):437–451, 2021.

[4] Evangelia I. Zacharaki, Sumei Wang, Sanjeev Chawla, Dong Soo Yoo, Ronald Wolf, Elias R. Melhem, and Christos Davatzikos. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magnetic Resonance in Med*, 62(6):1609–1618, December 2009.

[5] Yuki Wong, Eileen Lee Ming Su, Che Fai Yeong, William Holderbaum, and Chenguang Yang. Brain tumor classification using MRI images and deep learning techniques. *PloS one*, 20(5):e0322624, 2025.

[6] Shuvashis Sarker. Transfer Learning and Explainable AI for Brain Tumor Classification: A Study Using MRI Data from Bangladesh. In *2024 6th International Conference on Sustainable Technologies for Industry 5.0 (STI)*, pages 1–6. IEEE, 2024.

[7] Md Mahfuz Ahmed, Md Maruf Hossain, Md Rakibul Islam, Md Shahin Ali, Abdullah Al Noman Nafi, Md Faisal Ahmed, Kazi Mowdud Ahmed, Md Sipon Miah, Md Mahbubur Rahman, and Mingbo Niu. Brain tumor detection and classification in MRI using hybrid ViT and GRU model with explainable AI in Southern Bangladesh. *Scientific reports*, 14(1):22797, 2024.

[8] Md. Ariful Islam, M. F. Mridha, Mejdl Safran, Sultan Alfarhood, and Md. Mohsin Kabir. Revolutionizing Brain Tumor Detection Using Explainable AI in MRI Images. *NMR in Biomedicine*, 38(3):e70001, March 2025.

[9] Abdullah A. Asiri, Toufique Ahmed Soomro, Ahmed Ali Shah, Ganna Pogrebna, Muhammad Irfan, and Saeed Alqahtani. Optimized brain tumor detection: a dual-module approach for mri image enhancement and tumor classification. *IEEE access*, 12:42868–42887, 2024.

[10] Sanjukta Chakraborty and Dilip Kumar Banerjee. A review of brain cancer detection and classification using artificial intelligence and machine learning. *Journal of Artificial Intelligence and Systems*, 6(1):146–178, 2024.

[11] M. Thachayani and Sneha Kurian. AI based classification framework for cancer detection using brain MRI images. In *2021 International conference on system, computation, automation and networking (ICSCAN)*, pages 1–4. IEEE, 2021.

[12] Diponkor Bala, Mohammad Anwarul Islam, Mohammad Iqbal Hossain, Mohammed Mynuddin, Mohammad Alamgir Hossain, and Md Shamim Hossain. Automated brain tumor classification system using convolutional neural networks from mri images. In *2022 International Conference on Engineering and Emerging Technologies (ICEET)*, pages 1–6. IEEE, 2022.

[13] L. Reddy, Muniyandy Elangovan, M. Vamsikrishna, and Ch Ravindra. Brain Tumor Detection and Classification Using Deep Learning Models on MRI Scans. *EAI Endorsed Transactions on Pervasive Health & Technology*, 10(1), 2024.

[14] Asma Alshuhail, Arastu Thakur, R Chandramma, T R Mahesh, Ahlam Almusharraf, V Vinoth Kumar, and Surbhi Bhatia Khan. Refining neural network algorithms for accurate brain tumor classification in MRI imagery. *BMC Med Imaging*, 24(1):118, May 2024.

[15] Sarah Ali Abdelaziz Ismael, Ammar Mohammed, and Hesham Hefny. An enhanced deep learning approach for brain cancer MRI images classification using residual networks. *Artificial intelligence in medicine*, 102:101779, 2020.

[16] Chetana Srinivas, Nandini Prasad K. S., Mohammed Zakariah, Yousef Ajmi Alothaibi, Kamran Shaukat, B. Partibane, and Halifa Awal. Deep Transfer Learning Approaches in Performance Analysis of Brain Tumor Classification Using MRI Images. *Journal of Healthcare Engineering*, 2022:1–17, March 2022.

[17] Nadia Shamshad, Danish Sarwr, Ahmad Almogren, Kiran Saleem, Alia Munawar, Ateeq Ur Rehman, and Salil Bharany. Enhancing brain tumor classification by a comprehensive study on transfer learning techniques and model efficiency using MRI datasets. *IEEe Access*, 12:100407–100418, 2024.

[18] Shaimaa E. Nassar, Ibrahim Yasser, Hanan M. Amer, and Mohamed A. Mohamed. A robust MRI-based brain tumor classification via a hybrid deep learning technique. *J Supercomput*, 80(2):2403–2427, January 2024.

[19] Shahriar Hossain, Amitabha Chakrabarty, Thippa Reddy Gadekallu, Mamoun Alazab, and Md Jalil Piran. Vision transformers, ensemble model, and transfer learning leveraging explainable AI for brain tumor detection and classification. *IEEE Journal of Biomedical and Health Informatics*, 28(3):1261–1272, 2023.

[20] Yoon Han Chel and Lin Lih Poh. Brain tumor classification in MRI: Insights from LIME and Grad-CAM explainable AI techniques. *IEEE Access*, 2025.

[21] Soun Marina. Improving diagnostic accuracy of brain tumor MRI classification using generative AI and deep learning techniques. *Babylonian Journal of Artificial Intelligence*, 2025:55–63, 2025.

[22] Monika Sachdeva and Alok Kumar Singh Kushwaha. AI-based intelligent hybrid framework (BO-DenseXGB) for multi-classification of brain tumor using MRI. *Image and Vision Computing*, 154:105417, 2025.

[23] Girish Bathla, Durjoy Deb Dhruba, Neetu Soni, Yanan Liu, Nicholas B. Larson, Blake A. Kassmeyer, Suyash Mohan, Douglas Roberts-Wolfe, Saima Rathore, and Nam H. Le. AI-based classification of three common malignant tumors in neuro-oncology: A multi-institutional comparison of machine learning and deep learning methods. *Journal of neuroradiology*, 51(3):258–264, 2024.

[24] Hamid R. Alsanad. Hybrid Deep Learning Framework with Explainable AI for Multi-Class Brain Tumor Classification Using MRI Images. *Anbar Journal of Engineering Sciences*, 17(1):192–203, 2026.

[25] Joice J. Anish and D. Ajitha. Exploring the state-of-the-art algorithms for brain tumor classification using MRI data. *IEEE Access*, 2025.

[26] Mehmet U. Dalmiş, Albert Gubern-Mérida, Suzan Vreemann, Peter Bult, Nico Karssemeijer, Ritse Mann, and Jonas Teuwen. Artificial intelligence–based classification of breast lesions imaged with a multiparametric breast MRI protocol with ultrafast DCE-MRI, T2, and DWI. *Investigative radiology*, 54(6):325–332, 2019.

[27] Mohamed Musthafa M, Mahesh T. R, Vinoth Kumar V, and Suresh Guluwadi. Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50. *BMC Med Imaging*, 24(1):107, May 2024.

[28] V. Yamuna, R. V. S. Praveen, R. Sathya, M. Dhivva, R. Lidiya, and P. Sowmiya. Integrating AI for improved brain tumor detection and classification. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)*, pages 1603–1609. IEEE, 2024.

[29] Shah Foysal Hossain, Md Al Amin, Irin Akter Liza, Shahriar Ahmed, Md Musa Haque, Md Azharul Islam, and Sarmin Akter. AI-based brain MRI segmentation for early diagnosis and treatment planning of low-grade gliomas in the USA. *British Journal of Nursing Studies*, 3(2):37–55, 2023.

[30] Vinayaka R. Srinivas and Ramasubramanian Parvathi. Explainable AI-driven MRI-based brain tumor classification: a novel deep learning approach. *Frontiers in Artificial Intelligence*, 8:1700214, 2025.

[31] Amin Kabir Anaraki, Moosa Ayati, and Foad Kazemi. Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *biocybernetics and biomedical engineering*, 39(1):63–74, 2019.

[32] Hiba Mzoughi, Ines Njeh, Mohamed BenSlima, Nouha Farhat, and Chokri Mhiri. Vision transformers (ViT) and deep convolutional neural network (D-CNN)-based models for MRI brain primary tumors images multi-classification supported by explainable artificial intelligence (XAI). *Vis Comput*, 41(4):2123–2142, March 2025.

[33] Sahar Gull and Shahzad Akbar. Artificial intelligence in brain tumor detection through MRI scans: advancements and challenges. *Artificial intelligence and internet of things*, pages 241–276, 2021.

[34] Md Rahman, Mustavi Masum, Khan Hasib, M. Mridha, Sultan Alfarhood, Mejdl Safran, and Dunren Che. GliomaCNN: an effective lightweight CNN model in assessment of classifying brain tumor from magnetic resonance images using explainable AI. *Computer Modeling in Engineering & Sciences*, 140(3):2425, 2024.

[35] Arsalan Khan, Sugyani Rani Panda, Mansha Gupta, Md Hasnain Raza, Soumya Snigdha Mohapatra, and Debendra Muduli. A Two-Phase Fine Tuned Xception Model with Explainable AI for Brain Tumor Classification in MRI Images. In *2025 International Conference on Intelligent and Cloud Computing (ICoICC)*, pages 1–6. IEEE, 2025.

[36] Farzan Moodi, Fereshteh Khodadadi Shoushtari, Delaram J. Ghadimi, Gelareh Valizadeh, Ehsan Khormali, Hanieh Mobarak Salari, Mohammad Amin Dabbagh Ohadi, Yalda Nilipour, Amin Jahanbakhshi, and Hamidreza Saligheh Rad. Glioma Tumor Grading Using Radiomics on Conventional MRI : A Comparative Study of WHO 2021 and WHO 2016 Classification of Central Nervous Tumors. *Magnetic Resonance Imaging*, 60(3):923–938, September 2024.

[37] Faris Rustom, Ezekiel Moroze, Pedram Parva, Haluk Ogmen, and Arash Yazdanbakhsh. Deep learning and transfer learning for brain tumor detection and classification. *Biology Methods and Protocols*, 9(1):bpae080, 2024.

[38] Essam H. Houssein, Amr M. Gamal, Eman M. G. Younis, and Ebtsam Mohamed. Explainable artificial intelligence for brain tumor classification via fine-tuned transfer learning. *Discov Artif Intell*, 6(1):306, March 2026.

[39] Lal Hussain, Pauline Huang, Tony Nguyen, Kashif J. Lone, Amjad Ali, Muhammad Salman Khan, Haifang Li, Doug Young Suh, and Tim Q. Duong. Machine learning classification of texture features of MRI breast tumor and peri-tumor of combined pre- and early treatment predicts pathologic complete response. *BioMed Eng OnLine*, 20(1):63, June 2021.

[40] Yu Ji, Heather M. Whitney, Hui Li, Peifang Liu, Maryellen L. Giger, and Xuening Zhang. Differences in Molecular Subtype Reference Standards Impact AI-based Breast Cancer Classification with Dynamic Contrast-enhanced MRI. *Radiology*, 307(1):e220984, April 2023.