Improvement of CNN model using integration of multimodal MRI sequences and multilevel fusion to enhance performance for brain tumor classification

Ade Umar Ramadhan [a,1,*], Shofwatul Uyun [a,2]

[a] Informatics Department, Faculty of Science and Technology, Universitas Islam Negeri Sunan Kalijaga, Yogyakarta, Indonesia
[1] 23206051003@student.uin-suka.ac.id; [2] 19820511200604200 2@uin-suka.ac.id
* corresponding author

ARTICLE INFO

Article history
Received February 5, 2025
Revised November 11, 2025
Accepted December 18, 2025
Available online May 31, 2026

Keywords
Brain tumor
Multilevel fusion
Feature-level fusion
Data-level fusion
Ensemble learning

ABSTRACT

The prevalence of brain tumors has been increasing annually, and headaches, a common initial symptom, represent the most common manifestation. However, there is a paucity of research on effective methods of assessing brain tumors. This study proposes a novel approach by introducing various modality fusion techniques based on their fusion levels, which are then categorized into four groups: single-modal, data-level fusion, feature-level fusion, and multilevel fusion. A total of 51 combinations are designed to evaluate the efficacy of these fusion techniques and modality configurations. The experiments used a BraTS2021, which comprises four magnetic resonance imaging (MRI) sequences (flair, t1, t1ce and t2). Initially, the image was pre-processed, encompassing data selection, conversion, and normalization. Subsequently, it was input into a 13-layer CNN architecture for feature extraction. Classification was facilitated by a soft voting method in ensemble learning, incorporating support vector machine (SVM), k-nearest neighbor (KNN), logistic regression, random forest, and decision tree algorithms. The predictive efficacy of the model was rigorously assessed through a comprehensive suite of metrics, prominently featuring accuracy, AUCROC, AUCPR, Cohen's Kappa, and MCC. The results indicate that multilevel fusion exhibits optimal performance, with an average accuracy of 95.84%, followed by feature-level fusion and data-level fusion, at 95.12% and 94.77%, respectively. The optimal fusion technique was identified as the combination with the FF configuration (1,2),3,4), producing an accuracy of 96.62%. The best-model combination proposed exhibited an accuracy difference of nearly 6% from the baseline model, underscoring the efficacy of the proposed approach. These empirical results establish a robust baseline for future investigations into sophisticated fusion architectures across hierarchical integration levels.

1. Introduction

Brain tumors are the most prevalent fatal malignancy, exhibiting high morbidity and mortality rates in 2021 [1]. In the same year, brain tumors accounted for a disproportionate number of cancer-related deaths, with two thirds of patients dying at least five years after diagnosis. In 2023, that figure rose to 76.54% [2]. The incidence of brain tumors is approximately 7 per 100,000 people, headaches being the initial presenting symptom in up to 50% of cases [3]. Accordingly, rapid and precise diagnostic delineation of brain tumors is critical for optimizing therapeutic interventions and improving patient

prognosis [4] and requires the expertise of skilled radiologists [5]. Advances in contemporary medical imaging technology have led to substantial progress in the field of diagnostics and the detection of dangerous pathologies, particularly brain tumors [6]. Among the various imaging modalities available, magnetic resonance imaging (MRI) and computed tomography (CT) examinations are the most clinically advanced diagnostic techniques [7]. However, the results of MRI imaging show superiority in providing textural and morphological information on tumors compared to CT scans [8]. Furthermore, the MRI modality allows for a more comprehensive evaluation of the dimensions, morphological characteristics, and specific location of brain tissue [9].

Computer-assisted diagnosis (CAD) systems have become a very important tool in improving the accuracy and efficiency of medical diagnosis. In the context of brain tumors, CAD systems have been shown to help radiologists analyze MRI images more quickly and thoroughly, thereby reducing their manual workload and minimizing the risk of misdiagnosis due to human fatigue. Research indicates that the implementation of deep learning-based approaches can significantly enhance diagnostic precision, with recent models achieving classification accuracies exceeding 98% and outperforming traditional diagnostic methods [10]. In addition, AI-based concurrent reading systems have been shown to streamline the radiological workflow, effectively reducing reading times while maintaining high sensitivity even for less-experienced radiologists [11]. Implementing CAD systems in the diagnosis of brain tumors has shown encouraging results, with accuracy rates that reach 95% in specific cases [12]. A study conducted in 15 hospitals demonstrated that the incorporation of deep learning-based CAD systems can improve diagnosis sensitivity by 89% and specificity by 91% [13].

Despite the obvious potential of CAD to facilitate brain tumor diagnosis, its methodology is frequently restricted to the use of unimodal data, such as MRI or CT images, without considering multimodal data integration. This methodological ludicrity is due to the inability to harness the vast data repositories present within the diverse modalities of medical data, a situation that can potentially result in inaccurate diagnoses or less precise results [14]. Prior studies have demonstrated that the synergy of multimodal data and AI technologies provides a robust framework for medical diagnosis, offering a more comprehensive and accurate understanding of pathological conditions compared to unimodal approaches [15]. In routine radiological workflows, brain MRI interpretation typically begins with a rapid visual assessment followed by a more detailed evaluation across multiple sequences. However, diagnostic uncertainty frequently arises when certain sequences provide limited soft-tissue contrast or when findings appear subtle or ambiguous across modalities [16]. To address this challenge, our binary classification model is positioned as a clinically meaningful “second reader” or triage tool that automatically identifies potentially abnormal scans during the pre-reading stage [17]. By prioritizing suspicious cases, the model can assist radiologists in managing high workloads, reducing cognitive burden, and improving consistency in decision-making, particularly in settings with limited staffing or time-critical diagnostic demands. This explicit integration into existing radiologist workflows underscores the practical clinical value of our approach beyond methodological performance alone [18].

The efficacy of data fusion performed on augmentation techniques for brain tumor diagnosis using MobileNetV1 and MobileNetV2 was confirmed through the analysis of the area under the curve (AUC) metric derived from the receiver operating characteristic (ROC) analysis [19]. Research conducted on brain tumor classification for feature-level fusion using concatenation with principal component analysis (PCA) selection on three unique MRI datasets successfully outperformed other existing single models [20]. Furthermore, decision-level fusion using soft voting methods on ensemble learning models consisting of multiple algorithms with a case study of disease classification using medical images showed significantly better results than various single models [21],[22]. However, optimally combining information remains a challenge; single-level approaches often struggle with feature alignment, whereas recent frameworks employing multilevel contrastive information fusion have shown significant promise in managing the radio genomic complexities of brain tumors [23].

While recent studies utilizing the BraTS2021 dataset have explored multiple fusion strategies, most approaches still face challenges in effectively integrating both spatial and semantic information across

modalities [24]. State-of-the-art attention-based multimodal networks and transformer-based models are highly effective in modeling long-range dependencies, yet they demand substantial computational resources and typically require large training datasets to converge reliably [25]. Deep ensemble fusion methods, on the other hand, often process modalities independently and fuse them only at the decision level, which risks overlooking important cross-modal spatial relationships embedded in the raw MRI sequences [26]. In contrast to these limitations, this study introduces a unified Multilevel Fusion strategy that performs fusion at both the data level, capturing low-level spatial texture, and the feature level, encoding high-level semantic abstractions. By hierarchically integrating complementary information from multiple MRI sequences within a single framework, the proposed approach enhances cross-modal correlation modeling while maintaining a balanced trade-off between computational efficiency and diagnostic accuracy, outperforming conventional single-stage fusion architectures.

To address these limitations, we propose a novel learning model for brain tumor classification that employs data fusion at multiple levels to improve the model. This approach utilizes a combination of 51 fusion techniques, which encompasses various modalities, to achieve enhanced classification accuracy. Specifically, the model uses four MRI image sequences as modalities, utilizing ensemble learning to achieve optimal performance. The principal contributions of this study are delineated as follows:

- A learning model that utilizes 51 combinations of fusion techniques from four MRI sequences (flair, t1, t1ce, and t2) in a multimodal context.
- Ensemble learning classification employs soft vote as decision-level fusion with five machine learning algorithms (support vector machine (SVM), k-nearest neighbor (KNN), logistic regression, random forest, and decision tree).
- Fusion techniques include data-level fusion, feature-level fusion, and multilevel fusion, using 51 combinations of input from MRI sequences.
- The experimental results of the 51 combinations of fusion techniques were evaluated for their performance in terms of accuracy, AUC of the ROC curve, AUC of the precision recall curve, cohen's Kappa, and Matthew's correlation coefficient (MCC).

The scientific article is structured as follows: Section 2 discusses related research, Section 3 presents methods and materials, Section 4 discusses results and discussion, and Section 5 provides research conclusions.

2. Related Works

The quality of the data set is a fundamental factor in the production of reliable and accurate analysis [27]. For instance, Nehru and Prabhu utilized a comprehensive multimodal MRI dataset to evaluate the performance of automated segmentation and localization models [28]. At the same time, Peng and Sun developed a segmentation method using datasets from three sources. BraTS19 (335 images), BraTS20 (369 images) and FeTS21 (336 images), each of which integrated five sequences [29]. In a similar vein, Usha et al. explored classification with a multimodal dataset consisting of 561 unimodal images from Kaggle and 170 MRI-CT combination sets from MedHarvard [30]. In a similar study, Ullah et al. implemented classification using BraTS2020 and BraTS2021 data sets with four standard sequences, covering data from 369 and 1251 patients, respectively [31]. Maqsood et al. integrated data sets from Figshare and BraTS 2018, comprising a total of 3,064 T1-weighted contrast-enhanced images, which were categorized into three classes: meningiomas (708 images), gliomas (1,426 images) and pituitary (930 images) [32]. Taking into account the novelty and suitability of the proposed method, the BraTS2021 dataset was used in this study, as previously indicated by research findings.

Raw MRI data necessitates rigorous pre-processing investigation to mitigate artifacts and intensity inhomogeneity, as the presence of noise can severely compromise the integrity of downstream anomaly detection algorithms [33]. The preprocessing stage is crucial to address issues such as data redundancy, missing values, and data inconsistency [34]. In this regard, numerous researchers have employed various preprocessing methodologies, including the following. Nehru and Prabhu implemented a robust

preprocessing pipeline, incorporating intensity normalization and image resizing, to optimize multimodal MRI inputs for their Hybrid Res2-UNext segmentation model [28]. Usha et al. implemented a series of techniques, including elimination of salt and pepper noise, standardization of the image dimension to 227×227 pixels, and data stratification with a 70: 30 ratio for training and validation purposes [30]. Ullah et al. implemented a different approach through the manual selection of brain tissue layers in the data of each patient, resulting in the transformation of the BraTS2020 and BraTS2021 datasets into 110,700 and 24,200 images, respectively [31]. Maqsood et al. optimized data quality using the linear contrast stretching method to improve the accuracy of the analysis [32]. The pre-processing stage, involving the manual selection of images from the BraTS dataset, offers the distinct advantage of determining the representation and quantity of data in the class. The BraTS data set has undergone a series of image enhancements to ensure that image selection is more suitable for preprocessing in the proposed model.

Feature extraction is a pivotal phase in the learning pipeline, where advanced techniques such as feature-enhanced deep learning with soft attention are increasingly employed to highlight salient tumor regions [35]. In this context, various researchers have implemented advanced feature extraction methods: Nehru and Prabhu applied a Hybrid Res2-UNext architecture to effectively capture multi-scale features for precise tumor delineation [28], while Peng and Sun developed an auto-weight dilated convolution unit (AD unit) approach that utilizes multi-scale convolution feature maps for channel separation feature extraction [29]. Usha et al. adopted the Tumnet technique structured with an architecture of five convolution layers and three pooling layers, integrated with the ReLU activation function [30]. A more sophisticated approach was demonstrated by Ullah et al. through the implementation of Grey Wolf method that has been optimized using Jaya algorithm [31], while Maqsood et al. proposed the development of MobileNetV2 deep Convolutional Neural Network (CNN) architecture as a feature extraction method [32]. The feature extraction in our research uses a combination of convolution layers, pooling, and several other layers, as has been done in previous research.

The selection of an appropriate classification architecture is a fundamental aspect of predictive modeling, particularly as recent advancements demonstrate that integrating global and efficient channel attention mechanisms into models like EfficientNetV2 significantly enhances decision-making accuracy in MRI-based diagnosis [36]. For example, Nehru and Prabhu optimized a deep learning framework incorporating residual connections and attention mechanisms to handle complex tumor boundaries effectively [28]. Meanwhile, Peng and Sun developed a more complex architecture by integrating three different residual units to build Block-R1, Block-R2, and Block-R3 neural networks [29]. Usha et al. demonstrated a different approach through the implementation of a Tumnet model structured with three fully connected layers for the classification process [30]. Ullah et al. proposed the development of an optimized ResNet-50 architecture for the classification of multiclass brain tumors [31], while Maqsood et al. adopted a multiclass support vector machine (M-SVM) approach to classify three different categories [32]. Given the satisfactory outcomes achieved by previous research using multiple calcifiers, our proposed method uses ensemble learning.

In the most recent developments of 2023-2024, the landscape of brain tumor classification has expanded to include Vision Transformers (ViT) and attention-based mechanisms which excel at capturing global contexts [37]. For instance, Tabatabaei et al. [38] introduced a pyramidal attention-based network that integrates hierarchical features to enhance interpretability. Similarly, hybrid approaches like the Trans-IMSM model [39] and mask-attention frameworks [40] have demonstrated that fusing spatial-language-vision information can significantly improve boundary delineation. However, while these transformer-based architectures offer powerful long-range dependency modeling, they typically require massive datasets to avoid overfitting and entail substantial computational costs compared to CNN-based solutions. Consequently, for datasets with moderate sizes like BraTS2021 subsets, optimized CNN architectures with multilevel fusion strategies often provide a more balanced trade-off between computational efficiency and diagnostic accuracy.

3. Method

3.1. Dataset

This study uses the BraTS2021 dataset, which comprises multiparametric MRI (mpMRI) scan data of brain tumors. This dataset was released at a top-level academic conference facilitated by the Medical Image Computing and Computer Assisted Interventions (MICCAI) society [41]. The data set is sourced from multiple institutions but remains within standard clinical conditions and has undergone various pre-processing steps carried out by the organizers [29]. The BraTS2021 data set comprises four distinct sequences derived from MRI scans: flair, t1, t1ce and flair. Representative samples of the BraTS2021 dataset, captured at the same object and time for each sequence, are depicted in Fig. 1.

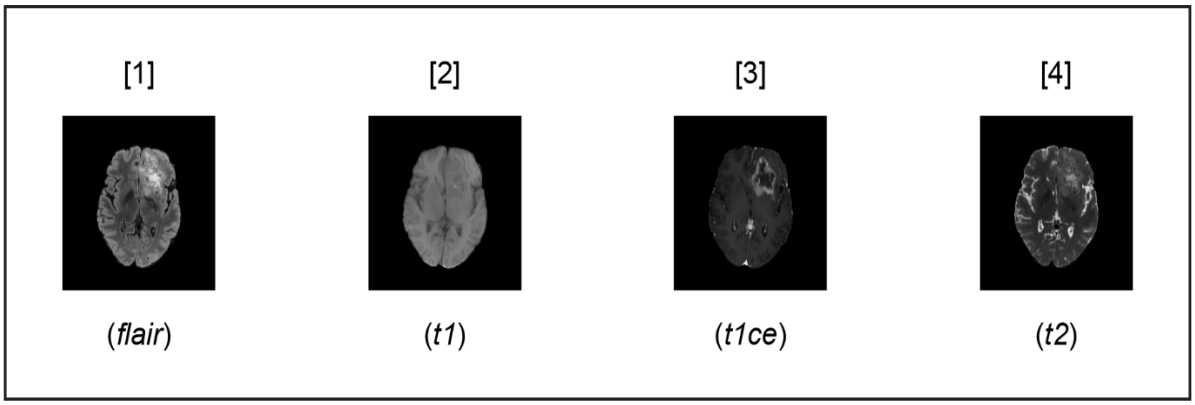

Fig. 1. A sample of the each sequence of MRI

Each sequence represents data from 1,666 patients, comprising a total of 79,921 images, resulting in a comprehensive BraTS2021 dataset of 319,456 images. The data set is meticulously classified into two distinct classes: tumor and non-tumor. The quantitative distribution of the data across these classes is elucidated in Table 1.

Table 1. The raw dataset (before pre-processing)

MRI sequences	Tumor	Non-tumor	Image totals	Patient counts
flair	60,697	19,224	79,921	1,666
t1	60,697	19,224	79,921	1,666
t1ce	60,697	19,224	79,921	1,666
t2	60,697	19,224	79,921	1,666

3.2. Proposed Method

Multilevel fusion is the core novelty of this study, designed as a sequential two-stage integration process that unifies the benefits of early and late fusion [42]. Instead of treating fusion as a single event, we structured it hierarchically. First, Data-Level Stage, selected raw MRI sequences (e.g., T1 and T1ce) are stacked channel-wise to form a composite input tensor (240×240×N). This allows the initial CNN layers to learn pixel-level correlations and local texture dependencies immediately [43]. Second, Feature-Level Stage, The latent feature vectors extracted from this composite input are then concatenated with feature vectors derived independently from other modalities (e.g., FLAIR or T2). This second stage ensures that high-level semantic information from distinct anatomical views is preserved [44]. By combining these stages, the model architecture effectively captures both the intra-modal spatial details and the inter-modal semantic context before the final soft-voting classification. The theoretical justification for prioritizing multilevel fusion lies in its ability to address tumor heterogeneity, a critical challenge in neurological imaging where tumor boundaries and sub-regions (e.g., edema vs. necrotic core) manifest differently across MRI sequences [45]. While data-level fusion preserves fine-grained spatial details essential for texture analysis, it is often susceptible to noise [46]. Conversely, feature-level

fusion captures high-level semantic representations but risks losing morphological subtleties due to pooling operations [47]. Multilevel fusion bridges this gap by creating complementary spatial–semantic representations [48]. It allows the network to learn low-level inter-modal correlations (e.g., verifying a boundary seen in T2 with T1ce) while simultaneously integrating high-level abstractions [49]. This architecture effectively exploits cross-sequence redundancy, ensuring that vital diagnostic features missed at one level are captured at another, thereby enhancing the model's robustness against the complex variations found in gliomas [50]. The overall schema of the model utilized in this research is depicted in Fig. 2.

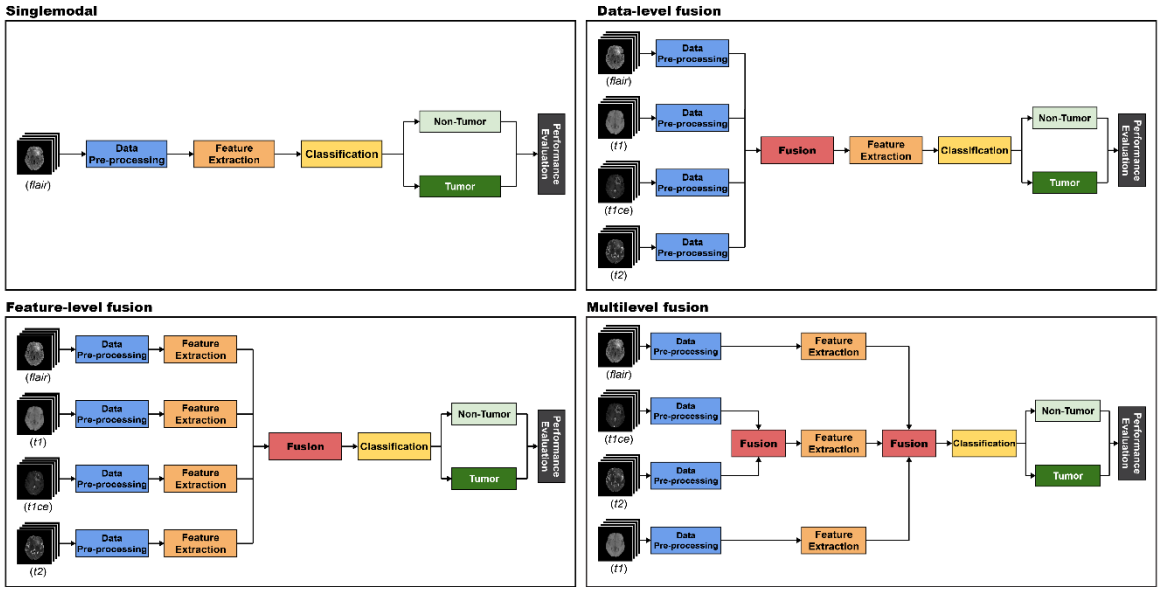

Fig. 2. The proposed models are utilized in single-modal and multimodal with varying levels of fusion settings

Initially, a baseline model is designed, devoid of any fusion technique (single-modal), which serves as a reference point to assess the enhancement in performance of the model when equipped with fusion techniques. Each MRI sequence is executed in the model sequentially. One of the baseline model combinations is depicted in Fig. 2 of the first sequence. We then constructed a model with multimodal input using the data-level fusion technique, which was implemented as follows. The data-level fusion strategy was systematically implemented across varying sequence configurations, using combinations of MRI sequences ranging from two to four modalities, producing a total of 11 combinations. The resultant data-level fusion combination is illustrated in the second sequence of Fig. 2. Subsequently, we designed a model with multimodal inputs using the feature-level fusion technique. The feature-level fusion process was executed in an analogous manner, employing combinations of MRI sequences ranging from two to four modalities, yielding a total of 11 combinations. One such feature-level fusion combination is depicted in the third sequence of Fig. 2.

Finally, we propose a model design with multimodal input using the multilevel fusion technique, which is a combination of data-level fusion and feature-level fusion techniques. Data-level fusion and feature-level fusion were executed in an alternating sequence, utilizing combinations of MRI sequences ranging from 3 to 4, producing a total of 25 combinations. One of the feature-level fusion combinations is illustrated in the fourth sequence of Fig. 2.

3.2.1. Pre-processing

The preliminary processing stage is performed prior to the feature extraction stage or the fusion process. In this research, three distinct preprocessing stages are identified: data selection, conversion, and normalization, as illustrated in Fig. 3.

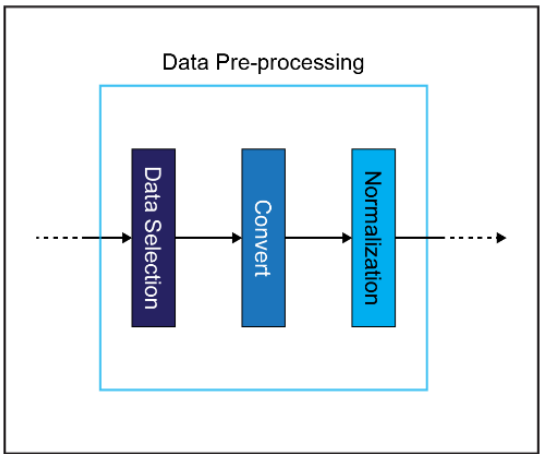

Fig. 3. Pre-processing stage

- Data Selection

Data selection was performed to create a representative and computationally manageable dataset for the extensive 51-fusion combination experiments. From the original cohort, a subset of 200 patients was randomly selected to ensure diverse demographic representation while reducing the computational overhead. From each patient's MRI volume, we extracted the central 50 slices, excluding the peripheral slices (approx. top and bottom 25%). This 'central-slice strategy' was rigorously chosen to maximize the Tumor-to-Background Ratio (TBR) [51]. In neuro-oncology imaging, peripheral slices often contain only healthy tissue or skull artifacts (non-informative background), which can introduce severe class imbalance and bias the model towards false negatives [52]. Although this 2D slice-based approach inherently reduces the global 3D volumetric context compared to full 3D-CNNs, it offers a necessary trade-off for significantly lowers memory consumption and allows for the testing of complex fusion architectures without the prohibitive cost of 3D processing [53]. Furthermore, extracting a continuous block of 50 slices preserves local inter-slice spatial continuity, ensuring that the most critical tumor features are retained for classification. Dataset after data selection detailed in Table 2.

Table 2. The subsequent data selection process

MRI sequences	Tumor	Non-tumor	Image totals	Patient counts
flair	5,000	5,000	10,000	200
t1	5,000	5,000	10,000	200
t1ce	5,000	5,000	10,000	200
t2	5,000	5,000	10,000	200

- Convert

Data conversion involves the conversion of image data in the form of portable network graphics (PNG) into a numpy array, with the objective of reducing computational burden during the feature extraction and classification stages [54]. Subsequent to conversion, each MRI sequence currently has a matrix of 240×240×10,000.

- Normalization

The final stage of preprocessing is normalization, a process that has been shown to improve model stability and accelerate convergence during training [55]. Normalization involves adjusting the scale of image values from 0-255 to a scale of 0-1, using a normalization factor of 1/255. The specific normalization formula employed is outlined in Equation (1).

$$Norm. = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

3.2.2. Feature Reaction

The research uses a 13-layer CNN architecture specifically engineered for high-dimensional feature extraction. The dimensions of the image entered into the input layer determine the number of modalities to be used. Specifically, the number of modalities employed is determined by the number of images in the input. For example, a combination of two images uses a configuration of 240×240×2, a combination of three images employs a configuration of 240×240×3, and a combination of four images utilizes a configuration of 240×240×4. Conversely, if the input consists of a single image, without any combinations of images, it employs a configuration of 240×240×1. A more detailed exposition of the layer architecture, type, activation, and number of parameters used for feature extraction in a combination of MRI sequences can be found in Table 3.

Table 3. Learnable parameters for the extraction of features from the fusion of data levels (flair, t1, t1ce, t2)

Layers	Name	Type	Activations	Total learned
1	InputImage 240×240×4[a] images	Input Image	240×240×4	-
2	Conv_1 32 3×3×3 convolution, stride 1×1, pad 'same'	Convolution	240×240×32	1,184
3	MaxpooI_1 2×2 max pooIing, stride 1×1, padding 'same'	Max Pooling	120×120×32	-
4	Conv_2 64 3×3×32 convoIution, stride 1×1, padding 'same'	Convolution	120×120×64	18,496
5	MaxpooI_2 2×2 max pooIing, Stride 1×1, padding 'same'	Max Pooling	60×60×64	-
6	Conv_3 128 3×3×64 convoIution with stride 1×1 and padding 'same'	Convolution	60×60×128	73,856
7	MaxpooI_3 2×2 max pooIing, Stride 1×1, padding 'same'	Max Pooling	30×30×128	-
8	Conv_4 256 3×3×128 convoIution, stride 1×1, padding 'same'	Convolution	30×30×256	295,168
9	MaxpooI_4 2×2 max pooIing, stride 1×1, padding 'same'	Max Pooling	15×15×256	-
10	GAP_1	Global Average Pooling 2D	1×1×256	388,704
11	Concat_1	Concatenate	1×1×256	-
12	Batch-norm_1 Batch Normalization and 256 channels	Batch Normal form	1×1×256	1,024
13	Flatten_1	Flatten	1×1×256	-

[a] Image size of data-level fusion (*flair, t1, t1ce, t2*) model

The convolution layer is responsible for extracting local features such as edges, textures, and patterns. The max-pooling layer serves to reduce dimensionality while preventing overfitting. The global average pooling (GAP) layer is useful for reducing dimensionality by averaging data at the channel level (when using data-level fusion) while preventing overfitting at the vector level. The concatenate layer plays a role in the combination of features in vector form. The batch normalization layer is instrumental in accelerating convergence and regularization. The Flatten layer is a crucial component that transforms multidimensional tensors into a flat structure, devoid of any specific operations.

3.2.3. Fusion

In this research, three types of fusion are used: data-level fusion, feature-level fusion, and multilevel fusion. The fusion method utilized in data-level fusion is fourth-dimensional concatenation, where the fourth dimension is incorporated into the channel representation. This happens because the first dimension signifies the amount of data, the second dimension indicates the image height, and the third dimension denotes the image width. The number of channels generated by concatenation will correspond to the modality of the MRI sequences used. The GAP layer plays a very important role here, because it is responsible for providing a data representation of several channels that have been combined using concatenation at the data level.

The fusion method used at the feature-data level is the concatenation of the feature extraction results in the form of vectors. The vector is generated from the pooling of all channels by the GAP layer. The number of features that will be output is proportional to the data representation generated by the GAP layer because the data representation will be combined in the concatenation layer.

Multilevel fusion will only occur if both the GAP and Concatenation layers have more than one input, as the GAP layer will pool more than one channel, which makes the data representation different from the input data. Meanwhile, if the concatenation layer only has one input, there will be no features that can be combined.

3.2.4. Classification

In this research, soft voting is chosen as a parameter in determining the results of the proposed ensemble learning. Soft voting serves to average the probability values generated by machine learning classifiers with SVM, KNN, logistic regression, random forest, and decision tree methods. The hyperparameters of each method are adjusted on the basis of the experimental results to optimize the performance of the model. Soft voting was chosen because it is more robust than hard voting, which provides prediction results only with a majority vote on the prediction in the form of classes generated from each method. The classifier used can be seen in Fig. 4.

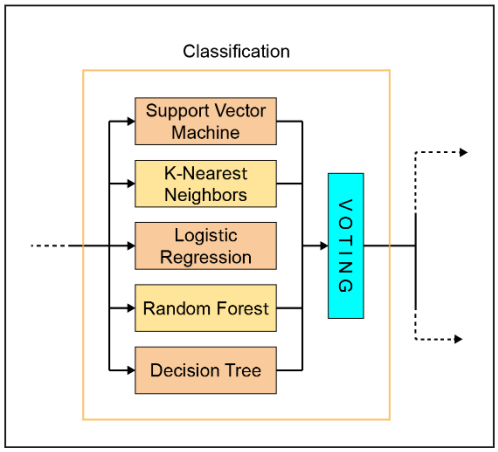

Fig. 4. Proposed ensemble learning classifier

3.2.5. Performance Evaluation

The efficacy of the trained model was quantified through fundamental classification metrics derived from indicators issued during the classification process. The metrics used to evaluate performance in this study are accuracy, AUC of the ROC curve, AUC of the precision recall curve, Cohen's Kappa, and MCC.

Accuracy (Acc) is the ratio of correct predictions to the total of all predictions; however, it is suitable only for balanced class distributions. Higher Acc values indicate better model performance [56]. The accuracy metric is calculated using the (2) formula.

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

The second metric is the AUC of the ROC curve (AUCROC), which serves to quantify the model's discriminative capability in distinguishing between target classes, even in instances of imbalanced class distribution. An ideal model has a value close to 1, while a random model has a value close to 0.5 [57]. The AUCROC is derived by plotting the values of the true positive rate (TPR) (3) and the false positive rate (FPR) (4). Formula (5) is used to calculate the AUCROC metric.

$$TPR = \frac{TP}{TP+FN} \tag{3}$$

$$FPR = \frac{FP}{FP+TN} \tag{4}$$

$$AUC_{ROC} = \sum_{i=1}^{n-1} \frac{(TPR[i+1]+TPR[i]}{2}.(FPR[i+1] - FPR[i]) \tag{5}$$

The third metric is the AUC of the Precision-Recall curve (AUCPR), which is used to assess the positive classes generated during classification without considering the negative classes. This metric is also suitable for use on imbalanced data distributions, because in imbalanced datasets the proportion of positive classes appears less frequently than that of negative classes. The model is better or perfect if the value is close to 1, while the model is more random if the value is close to 0.5 [57]. The AUCPR is derived by plotting the precision (6) and recall (7) values. The AUCPR metric is calculated using the formula (8).

$$Precision\ (P)\ = \frac{TP}{TP+FP} \tag{6}$$

$$Recall\ (R) = \frac{TP}{TP+FN} \tag{7}$$

$$AUC_{PR} = \sum_{i=1}^{n-1} \frac{(P[i+1]+P[i])}{2}.(R[i+1] - R[i]) \tag{8}$$

The fourth metric, Cohen's Kappa, is used to calculate the level of agreement between the predictions and the actual data, considering the probability of agreement that occurs by chance. The model shows perfect agreement if the value is closer to 1, no agreement if the value is closer to 0, and no agreement and coincidence if the value is closer to −1 [58]. The calculation of Cohen's Kappa involves the utilization of two values: the observed agreement value (P_o) and the agreement value that is expected to occur by chance (P_e). To derive P_o, one must refer to Equations (9) and (10), whereas P_e can be found using Equations (9) and (11). The final step in determining Cohen's Kappa metric (12).

$$N = TP + TN + FP + FN \tag{9}$$

$$P_o = \frac{TP+TN}{N} \tag{10}$$

$$P_e = \frac{(TP+FP)(TP+FN)+(TN+FN)(TN+FP)}{N^2} \tag{11}$$

$$K = \frac{P_o - P_e}{1-P_e} \tag{12}$$

The final metric, known as MCC, serves to determine the correlation between the prediction results and the actual data. A perfect correlation is indicated by values that approximate 1. In the absence of correlation, values closer to 0 indicate a random relationship. On the contrary, negative values do not imply correlation, resulting in erroneous predictions [56]. The MCC is determined using formula (13) below:

$$MCC = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)TN+FN)}} \tag{13}$$

4. Results and Discussion

The 51 combinations of MRI sequences were trained using the KFold cross-validation method with a K value of 10 folds. The training data was evaluated by testing the prediction results against the actual data. The evaluation results produced a confusion matrix organized on the basis of the basic indicators of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). These basic indicators are used to evaluate performance in the form of accuracy, AUCROC, AUCPR, Cohen's Kappa, and MCC matrices. The metric values generated for each training validation result using the modalities of the combinations of MRI sequences can be seen in Table 4.

Combinations using data-level fusion techniques will be denoted by "DF", feature-level fusion techniques by "FF", while combinations using multilevel fusion techniques will use the notation of both. The modality of the MRI sequences used will be denoted using numbers. Flair sequences are denoted by "1", t1 sequences by "2", t1ce sequences by "3", and t2 sequences by "4".

To illustrate, in the event that multilevel fusion is applied with the configuration of sequences t1ce and t2 as data-level fusion, the result of the data-level fusion will be combined at the feature level with sequences flair and t1, as demonstrated in Fig. 2 of the fourth sequence. This is expressed as FF(DF). Sequence 5 is expressed as FF(DF (3,4),1,2).

Table 4. Mean metrics for 10-fold of each combination

Combination Number	Fusion Configuration	Fusion Level	*Acc*	AUC_{ROC}	AUC_{PR}	Cohen's Kappa	MCC
1	1	Singlemodal	94.62%	0.9902	0.9904	0.8924	0.8926
2	2	Singlemodal	90.95%	0.9737	0.9731	0.8190	0.8201
3	3	Singlemodal	90.85%	0.9679	0.9688	0.8170	0.8170
4	4	Singlemodal	92.30%	0.9779	0.9782	0.8460	0.8461
5	DF(1,2)	Data-level	95.75%	0.9931	0.9931	0.9150	0.9151
6	DF(1,3)	Data-level	95.47%	0.9922	0.9923	0.9094	0.9096
7	DF(1,4)	Data-level	95.10%	0.9911	0.9912	0.9020	0.9021
8	DF(2,3)	Data-level	93.17%	0.9816	0.9819	0.8634	0.8637
9	DF(2,4)	Data-level	93.81%	0.9855	0.9855	0.8762	0.8764
10	DF(3,4)	Data-level	93.15%	0.9818	0.9822	0.8630	0.8630
11	DF(1,2,3)	Data-level	95.45%	0.9926	0.9927	0.9090	0.9091
12	DF(2,3,4)	Data-level	94.22%	0.9878	0.9881	0.8844	0.8846
13	DF(3,4,1)	Data-level	95.08%	0.9921	0.9923	0.9016	0.9018
14	DF(4,1,2)	Data-level	95.98%	0.9943	0.9944	0.9196	0.9197
15	DF(1,2,3,4)	Data-level	95.32%	0.9928	0.9929	0.9064	0.9067
16	FF(1,2)	Feature-level	94.87%	0.9924	0.9925	0.8974	0.8978
17	FF(1,3)	Feature-level	95.68%	0.9926	0.9928	0.9136	0.9137
18	FF(1,4)	Feature-level	95.44%	0.9926	0.9925	0.9088	0.9090
19	FF(2,3)	Feature-level	93.76%	0.9857	0.9857	0.8752	0.8755
20	FF(2,4)	Feature-level	94.38%	0.9872	0.9869	0.8876	0.8879
21	FF(3,4)	Feature-level	93.28%	0.9834	0.9834	0.8656	0.8656
22	FF(1,2,3)	Feature-level	95.90%	0.9939	0.9939	0.9180	0.9181
23	FF(2,3,4)	Feature-level	94.88%	0.9893	0.9891	0.8976	0.8977
24	FF(3,4,1)	Feature-level	95.82%	0.9935	0.9936	0.9164	0.9165
25	FF(4,1,2)	Feature-level	95.97%	0.9940	0.9940	0.9194	0.9197
26	FF(1,2,3,4)	Feature-level	96.35%	0.9955	0.9955	0.9270	0.9271
27	FF(DF(1,2),3)	Multilevel	95.63%	0.9934	0.9934	0.9126	0.9127

Combination Number	Fusion Configuration	Fusion Level	*Acc*	AUC_{ROC}	AUC_{PR}	Cohen's Kappa	MCC
28	FF(DF(1,2),4)	Multilevel	95.79%	0.9938	0.9938	0.9158	0.9160
29	FF(DF(1,3),2)	Multilevel	95.92%	0.9941	0.9942	0.9184	0.9185
30	FF(DF(1,3),4)	Multilevel	95.81%	0.9937	0.9937	0.9162	0.9163
31	FF(DF(1,4),2)	Multilevel	96.06%	0.9942	0.9944	0.9212	0.9214
32	FF(DF(1,4),3)	Multilevel	94.92%	0.9915	0.9916	0.8984	0.8986
33	FF(DF(2,3),1)	Multilevel	95.90%	0.9943	0.9944	0.9180	0.9182
34	FF(DF(2,3),4)	Multilevel	94.66%	0.9896	0.9898	0.8932	0.8935
35	FF(DF(2,4),1)	Multilevel	95.85%	0.9938	0.9939	0.9170	0.9172
36	FF(DF(2,4),3)	Multilevel	94.91%	0.9899	0.9898	0.8982	0.8983
37	FF(DF(3,4),1)	Multilevel	95.32%	0.9920	0.9919	0.9064	0.9067
38	FF(DF(3,4),2)	Multilevel	94.75%	0.9891	0.9892	0.8950	0.8952
39	FF(DF(1,2),3,4)	Multilevel	96.62%	0.9955	0.9955	0.9324	0.9325
40	FF(DF(1,3),2,4)	Multilevel	96.46%	0.9956	0.9956	0.9292	0.9294
41	FF(DF(1,4),2,3)	Multilevel	96.29%	0.9946	0.9946	0.9258	0.9259
42	FF(DF(2,3),1,4)	Multilevel	96.35%	0.9950	0.9949	0.9270	0.9272
43	FF(DF(2,4),1,3)	Multilevel	96.34%	0.9948	0.9949	0.9268	0.9269
44	FF(DF(3,4),1,2)	Multilevel	96.37%	0.9947	0.9946	0.9274	0.9276
45	FF(DF(1,2,3),4)	Multilevel	96.00%	0.9944	0.9945	0.9200	0.9202
46	FF(DF(2,3,4),1)	Multilevel	95.95%	0.9942	0.9943	0.9190	0.9192
47	FF(DF(3,4,1),2)	Multilevel	96.09%	0.9943	0.9945	0.9218	0.9221
48	FF(DF(4,1,2),3)	Multilevel	95.79%	0.9937	0.9937	0.9158	0.9160
49	FF(DF(1,2),DF(3,4))	Multilevel	96.60%	0.9956	0.9957	0.9320	0.9321
50	FF(DF(1,3),DF(2,4))	Multilevel	96.07%	0.9946	0.9946	0.9214	0.9215
51	FF(DF(1,4),DF(2,3))	Multilevel	95.55%	0.9931	0.9932	0.9110	0.9114

The most effective model, which has an accuracy of 96.62%, has been determined to be combination number 39 with the configuration of the fusion technique using FF(DF(1,2),3,4). This configuration involves performing data-level fusion on the flair and t1 modalities and subsequently performing feature fusion of the results with additional modalities t1ce and t2. However, to provide a comprehensive assessment of the results, the performance evaluation metrics for each combination are calculated using the averaging method, focusing on the attributes of the fusion level. As illustrated in Fig. 5, the validated average accuracy and loss graphs for each fusion level are presented. It should be noted that the loss value serves as a complement to support accuracy, with the results of the accuracy and loss values produced by each fusion level consistently being inversely proportional. Combining multilevel MRI sequences with multilevel fusion demonstrates the superior performance compared to other methods. This superiority is evident in the highest average accuracy and the lowest loss values recorded, which are 95.84% and 0.1371, respectively. The graph illustrates that the most effective subsequent combination of MRI sequences is level fusion using feature level, which achieves an accuracy value of 95.12%, followed by feature level at 94.77%. In contrast, the single-modal model without the fusion technique exhibits the least optimal performance.

As illustrated in Fig. 6, the average ROC curve and the precision-recall curve are presented at each fusion level. The AUC value is a quantitative metric that assesses the performance of the model. A higher AUC value indicates a more effective model. In this graph, the combination of MRI sequences with multilevel fusion levels demonstrates the most superior performance compared to the other models. This superiority is reflected in the highest average AUC generated by the ROC curve and the precision-recall curve, which are 0.9938 and 0.9939, respectively. The graph also demonstrates that the combination of

MRI sequences with level fusion using data-level and/or feature-level is significantly more effective than single-modal models without any fusion.

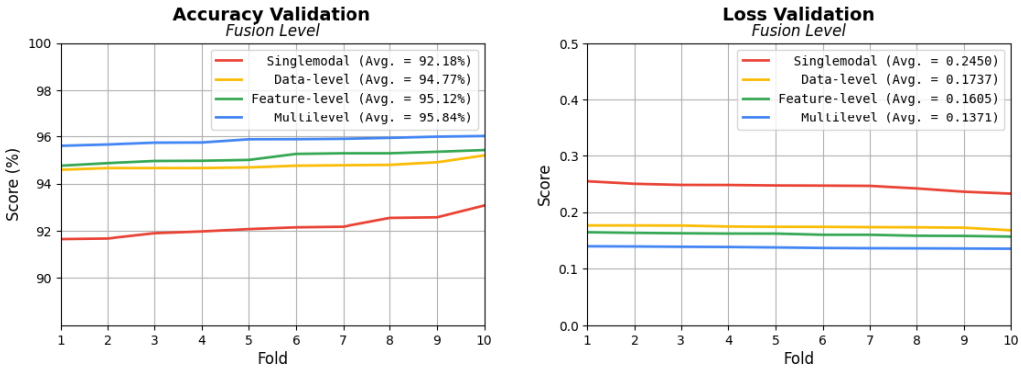

Fig. 5. Grafik akurasi dan loss pada tiap level fusion

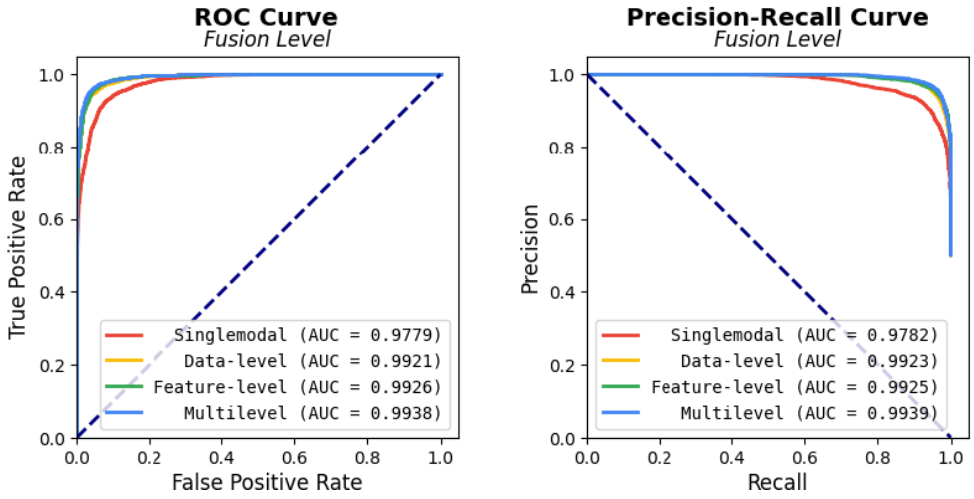

Fig. 6. Grafik ROC curve dan Precision-Recall curve pada tiap level fusion

As illustrated in Table 5, the mean values of Cohen's Kappa and MCC are reported at each fusion level. Combining multilevel MRI sequences with multilevel fusion, when assessed from both metrics, exhibits the most superior performance compared to other combinations. This is evidenced by the mean Cohen's Kappa and MCC values, which are 0.9229 and 0.9230, respectively. The subsequent optimal configuration of the MRI sequences is identified as feature-level fusion, resulting in Cohen's Kappa and MCC values of 0.9088 and 0.9090, respectively. This is followed by data-level fusion, which yields Cohen's Kappa and MCC values of 0.8898 and 0.8899. In particular, the model exhibiting the lowest performance is the single-modal configuration without fusion techniques.

Table 5. Average Cohen's Kappa amd MCC of each fusion level

Fusion Level	Cohen's Kappa	MCC
Singlemodal	0.8436	0.8439
Data-level	0.8898	0.8899
Feature-level	0.9088	0.9090
Multilevel	0.9229	0.9230

A multitude of extensive investigations have been undertaken that address the classification of brain tumors through the utilization of various data preprocessing methodologies. Among these, the approach used by [31], which involves the exclusive selection of data from the central slice, has demonstrated superior performance compared to other methods that do not employ this technique. The use of multimodal approaches has consistently yielded results that surpass those of models based on unimodal techniques [28]–[32]. The empirical results of this investigation substantiate that that incorporating a

greater number of modalities and features consistently produces superior results, as evidenced by the top 10 combinations of merging techniques, which collectively utilize four modalities and four features (see Table 4). Consistent with the findings reported in [23], which asserts that fusion methodologies invariably possess inherent limitations at any given level, our experimental outcomes depicted in Fig. 5 and Fig. 6 demonstrate that the mean outcomes of data-level and feature-level fusion fall short of those attained through multilevel fusion. Table 6 presents a comprehensive comparative benchmarking of the obtained results against of the present study with those of previous research efforts that focused on the classification of multimodal brain tumors.

Table 6. Performance comparison with the current methodologies for multimodal classification of brain tumors

Researcher	Dataset	Pre-processing	Methodology	Result
Nehru and Prabhu [28]	BraTS 2021	-	Hybrid Res2-UNext	92% TC, 93% WT, 92% ET
Peng and Sun [29]	BraTS19 dan FeTS21	Randomization clipping, rotation, intensity enhancement, mirroring, and normalization	AD unit multiscale convolution, 3 residual neural net blocks	WT 90% , TC 80%, and ET 76%
Usha et al. [30]	Kaggle for unimodal MRI and CT, MedHarvard for multimodal MRI-CT	Salt-and-pepper noise removal, resize, and split data	Tumnet, 3 layer FCL	Unimodal 96% and multimodal 98%
Ullah et al. [31]	BraTS2020 and BraTS2021	Data selection on middle portion of MRI slice	Optimized Grey Wolf, ResNet-50	98%
Maqsood et al. [32]	BraTS2018	Linear contrast streching	MobileNetV2 deep CNN, M-SVM	97.47%
Curent study	**BraTS2021**	**Data selection, convert, and normalization**	**13-layered CNN, Multimodal MRI sequences & multilevel fusion**	**Multilevel (avg.) 95.84% and multilevel (highest) 96.62%**

5. Conclusion

This study has made a comparison of models in terms of the use of modalities and fusion techniques within the domain of brain tumor classification utilizing multiparametric MRI sequences from the BraTS2021 dataset, with a configuration of 51 model combinations. The model combinations were merged into four fusion techniques, namely single-modal, data-level, feature-level, and multilevel fusion. Dataset preprocessing is done in three stages, namely data selection, conversion, and normalization. The feature extraction is designed based on a CNN architecture with 13 layers. Classification uses a soft-voting method in ensemble learning, which consists of SVM, KNN, logistic regression, random forest, and decision tree algorithms. The experimental results on the accuracy metrics, AUCROC, AUCPR, Cohen's Kappa, and MCC consistently show that multilevel fusion has the most effective performance, followed by the feature level, the data level, and then the single modal as the baseline. The optimal combination of models, identified by rigorous experimentation, is the configuration FF(DF(1,2),3,4). A critical avenue for future research lies in investigating the application of sophisticated methods to modality fusion at various levels.

Declarations

Author contribution. AUR executed the experimental design and drafted the original manuscript, SU supervised the experiment, and provided editorial feedback on the manuscript

Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.

Conflict of interest. The authors unequivocally declare that they have no competing financial or non-financial interest.

Additional information. Supplementary information accompanying this manuscript is available online.

Data and Software Availability Statements

The research employed third-party data from https://www.synapse.org/ that had previously been compiled under the joint organization of the Radiological Society of North America (RSNA), the American Society of Neuroradiology (ASNR), and the Medical Image Computing and Computer Assisted Interventions (MICCAI) society.

References

[1] K. D. Miller *et al.*, "Brain and other central nervous system tumor statistics, 2021," *CA. Cancer J. Clin.*, vol. 71, no. 5, pp. 381–406, Sep. 2021, doi: 10.3322/caac.21693.

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

[3] L. R. Schaff and I. K. Mellinghoff, "Glioblastoma and Other Primary Brain Malignancies in Adults: A Review," *JAMA*, vol. 329, no. 7, pp. 574–587, Feb. 2023, doi: 10.1001/jama.2023.0023.

[4] M. Arabahmadi, R. Farahbakhsh, and J. Rezazadeh, "Deep Learning for Smart Healthcare, A Survey on Brain Tumor Detection from Medical Imaging," *Sensors*, vol. 22, no. 5, p. 1960, Mar. 2022, doi: 10.3390/s22051960.

[5] B. S. Negi, R. Chauhan, R. Rawat, Y. Chanti, and G. Kaur, "Brain Tumor Detection Using integrated approach of FCM & CNN," in *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, Apr. 2024, pp. 1–5, doi: 10.1109/I2CT61223.2024.10543848.

[6] B. Abhisheka, S. K. Biswas, B. Purkayastha, D. Das, and A. Escargueil, "Recent trend in medical imaging modalities and their applications in disease diagnosis: a review," *Multimed. Tools Appl.*, vol. 83, no. 14, pp. 43035–43070, Oct. 2023, doi: 10.1007/s11042-023-17326-1.

[7] M. C. Florkow, K. Willemsen, V. V Mascarenhas, E. H. G. Oei, M. van Stralen, and P. R. Seevinck, "Magnetic resonance imaging versus computed tomography for Three-Dimensional bone imaging of musculoskeletal pathologies: a review," *J. Magn. Reson. Imaging*, vol. 56, no. 1, pp. 11–34, 2022, doi: 10.1002/jmri.28067.

[8] M. C. Florkow, K. Willemsen, V. V. Mascarenhas, E. H. G. Oei, M. van Stralen, and P. R. Seevinck, "Magnetic Resonance Imaging Versus Computed Tomography for Three-Dimensional Bone Imaging of Musculoskeletal Pathologies: A Review," *J. Magn. Reson. Imaging*, vol. 56, no. 1, pp. 11–34, Jul. 2022, doi: 10.1002/jmri.28067.

[9] A. M. Gab Allah, A. M. Sarhan, and N. M. Elshennawy, "Edge U-Net: Brain tumor segmentation using MRI based on deep U-Net model with boundary information," *Expert Syst. Appl.*, vol. 213, p. 118833, Mar. 2023, doi: 10.1016/j.eswa.2022.118833.

[10] H. Sadr, M. Nazari, S. Yousefzadeh-Chabok, H. Emami, R. Rabiei, and A. Ashraf, "Enhancing brain tumor classification in MRI images: A deep learning-based approach for accurate diagnosis," *Image Vis. Comput.*, vol. 159, p. 105555, Jun. 2025, doi: 10.1016/j.imavis.2025.105555.

[11] S. Labus *et al.*, "A concurrent, deep learning–based computer-aided detection system for prostate multiparametric MRI: a performance study involving experienced and less-experienced radiologists," *Eur. Radiol.*, vol. 33, no. 1, pp. 64–76, Jul. 2022, doi: 10.1007/s00330-022-08978-y.

[12] M. S. I. Khan *et al.*, "Accurate brain tumor detection using deep convolutional neural network," *Comput. Struct. Biotechnol. J.*, vol. 20, pp. 4733–4745, Jan. 2022, doi: 10.1016/j.csbj.2022.08.039.

[13] P. Xue *et al.*, "Unassisted Clinicians Versus Deep Learning–Assisted Clinicians in Image-Based Cancer Diagnostics: Systematic Review With Meta-analysis," *J. Med. Internet Res.*, vol. 25, p. e43832, Mar. 2023, doi: 10.2196/43832.

[14] S. R. Stahlschmidt, B. Ulfenborg, and J. Synnergren, "Multimodal deep learning for biomedical data fusion: a review," *Brief. Bioinform.*, vol. 23, no. 2, p. bbab569, Mar. 2022, doi: 10.1093/bib/bbab569.

[15] X. Xu *et al.*, "A Comprehensive Review on Synergy of Multi-Modal Data and AI Technologies in Medical Diagnosis," *Bioengineering*, vol. 11, no. 3, p. 219, Feb. 2024, doi: 10.3390/bioengineering11030219.

[16] S. Lee *et al.*, "Ensemble learning-based radiomics with multi-sequence magnetic resonance imaging for benign and malignant soft tissue tumor differentiation," *PLoS One*, vol. 18, no. 5, p. e0286417, May 2023, doi: 10.1371/journal.pone.0286417.

[17] E. Mahmoud *et al.*, "MU-Glioma Post: A comprehensive dataset of automated MR multi-sequence segmentation and clinical features," *Sci. Data*, vol. 12, no. 1, p. 1847, Nov. 2025, doi: 10.1038/s41597-025-06011-7.

[18] E. C. Yilmaz *et al.*, "Evaluating deep learning and radiologist performance in volumetric prostate cancer analysis with biparametric MRI and histopathologically mapped slides," *Abdom. Radiol.*, vol. 50, no. 6, pp. 2732–2744, Dec. 2024, doi: 10.1007/s00261-024-04734-6.

[19] A. Hekmat, Z. Zhang, S. Ur Rehman Khan, I. Shad, and O. Bilal, "An attention-fused architecture for brain tumor diagnosis," *Biomed. Signal Process. Control*, vol. 101, p. 107221, 2025, doi: 10.1016/j.bspc.2024.107221.

[20] N. F. Aurna, M. A. Yousuf, K. A. Taher, A. K. M. Azad, and M. A. Moni, "A classification of MRI brain tumor based on two stage feature level ensemble of deep CNN models," *Comput. Biol. Med.*, vol. 146, p. 105539, 2022, doi: 10.1016/j.compbiomed.2022.105539.

[21] A. Srinivas and J. P. Mosiganti, "A brain stroke detection model using soft voting based ensemble machine learning classifier," *Meas. Sensors*, vol. 29, p. 100871, 2023, doi: 10.1016/j.measen.2023.100871.

[22] S. 'Uyun, L. Choridah, S. Riyadi, and A. U. Ramadhan, "Convolutional Layer-Based Feature Extraction in an Ensemble Machine Learning Model for Breast Cancer Classification," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 12. pp. 418–425, 2024, doi: 10.14569/IJACSA.2024.0151244.

[23] Asmita and P. Mittal, "Brain Tumor Radiogenomic Classification using multilevel contrastive information fusion," *Procedia Comput. Sci.*, vol. 259, pp. 260–268, 2025, doi: 10.1016/j.procs.2025.03.327.

[24] H. Liu, J. Huang, Q. Li, X. Guan, and M. Tseng, "A deep convolutional neural network for the automatic segmentation of glioblastoma brain tumor: Joint spatial pyramid module and attention mechanism network," *Artif. Intell. Med.*, vol. 148, p. 102776, Feb. 2024, doi: 10.1016/j.artmed.2024.102776.

[25] G. Zheng *et al.*, "An attention-based multi-modal MRI fusion model for major depressive disorder diagnosis," *J. Neural Eng.*, vol. 20, no. 6, p. 066005, Dec. 2023, doi: 10.1088/1741-2552/ad038c.

[26] E. K. Jadoon, F. G. Khan, S. Shah, A. Khan, and M. ElAffendi, "Deep Learning-Based Multi-Modal Ensemble Classification Approach for Human Breast Cancer Prognosis," *IEEE Access*, vol. 11, pp. 85760–85769, 2023, doi: 10.1109/ACCESS.2023.3304242.

[27] M. M. M, M. T. R, V. K. V, and S. Guluwadi, "Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50," *BMC Med. Imaging*, vol. 24, no. 1, p. 107, May 2024, doi: 10.1186/s12880-024-01292-7.

[28] V. Nehru and V. Prabhu, "Automated Multimodal Brain Tumor Segmentation and Localization in MRI Images Using Hybrid Res2-UNeXt," *J. Electr. Eng. Technol.*, vol. 19, no. 5, pp. 3485–3497, Jul. 2024, doi: 10.1007/s42835-023-01779-3.

[29] Y. Peng and J. Sun, "The multimodal MRI brain tumor segmentation based on AD-Net," *Biomed. Signal Process. Control*, vol. 80, p. 104336, 2023, doi: 10.1016/j.bspc.2022.104336.

[30] M. P. Usha, G. Kannan, and M. Ramamoorthy, "Multimodal Brain Tumor Classification Using Convolutional Tumnet Architecture," *Behav. Neurol.*, vol. 2024, no. 1, p. 4678554, Jan. 2024, doi: 10.1155/2024/4678554.

[31] M. S. Ullah, M. A. Khan, N. A. Almujally, M. Alhaisoni, T. Akram, and M. Shabaz, "BrainNet: a fusion assisted novel optimal framework of residual blocks and stacked autoencoders for multimodal brain tumor classification," *Sci. Rep.*, vol. 14, no. 1, p. 5895, Mar. 2024, doi: 10.1038/s41598-024-56657-3.

[32] S. Maqsood, R. Damaševičius, and R. Maskeliūnas, "Multi-Modal Brain Tumor Detection Using Deep Neural Network and Multiclass SVM," *Medicina (B. Aires).*, vol. 58, no. 8, p. 1090, Aug. 2022, doi: 10.3390/medicina58081090.

[33] S. Srivastava, M. Arfat, S. Pachar, R. Krishna Yellapragada, and P. Jyotiyana, "Pre-Processing Investigation For Brain Abnormality Detection And Analysis Through MRI of Brain," in *2023 1st International Conference on Innovations in High Speed Communication and Signal Processing (IHCSP)*, Mar. 2023, pp. 418–422, doi: 10.1109/IHCSP56702.2023.10127204.

[34] S. De Sutter, J. Wuts, W. Geens, A.-M. Vanbinst, J. Duerinck, and J. Vandemeulebroucke, "Modality redundancy for MRI-based glioblastoma segmentation," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 19, no. 10, pp. 2101–2109, 2024, doi: 10.1007/s11548-024-03238-4.

[35] B. C. Mohanty, P. K. Subudhi, R. Dash, and B. Mohanty, "Feature-enhanced deep learning technique with soft attention for MRI-based brain tumor classification," *Int. J. Inf. Technol.*, vol. 16, no. 3, pp. 1617–1626, Mar. 2024, doi: 10.1007/s41870-023-01701-0.

[36] I. Pacal, O. Celik, B. Bayram, and A. Cunha, "Enhancing EfficientNetv2 with global and efficient channel attention mechanisms for accurate MRI-Based brain tumor classification," *Cluster Comput.*, vol. 27, no. 8, pp. 11187–11212, Nov. 2024, doi: 10.1007/s10586-024-04532-1.

[37] S. Aburass, O. Dorgham, J. Al Shaqsi, M. Abu Rumman, and O. Al-Kadi, "Vision Transformers in Medical Imaging: a Comprehensive Review of Advancements and Applications Across Multiple Diseases," *J. Imaging Informatics Med.*, vol. 38, no. 6, pp. 3928–3971, Mar. 2025, doi: 10.1007/s10278-025-01481-y.

[38] S. Tabatabaei, K. Rezaee, and M. Zhu, "Attention transformer mechanism and fusion-based deep learning architecture for MRI brain tumor classification system," *Biomed. Signal Process. Control*, vol. 86, p. 105119, Sep. 2023, doi: 10.1016/j.bspc.2023.105119.

[39] V. Durairaj and P. Uthirapathy, "Interactive Multi-scale Fusion: Advancing Brain Tumor Detection Through Trans-IMSM Model," *J. Imaging Informatics Med.*, vol. 38, no. 2, pp. 757–774, Aug. 2024, doi: 10.1007/s10278-024-01222-7.

[40] G. Kong, C. Wu, Z. Zhang, C. Yin, and D. Qin, "M3: using mask-attention and multi-scale for multi-modal brain MRI classification," *Front. Neuroinform.*, vol. 18, p. 1403732, Jul. 2024, doi: 10.3389/fninf.2024.1403732.

[41] L. Liu and K. Xia, "BTIS-Net: Efficient 3D U-Net for Brain Tumor Image Segmentation," *IEEE Access*, vol. 12, pp. 133392–133405, 2024, doi: 10.1109/ACCESS.2024.3460797.

[42] L. M. Pereira, A. Salazar, and L. Vergara, "A Comparative Analysis of Early and Late Fusion for the Multimodal Two-Class Problem," *IEEE Access*, vol. 11, pp. 84283–84300, 2023, doi: 10.1109/ACCESS.2023.3296098.

[43] A. He, T. Li, Y. Wu, K. Zou, and H. Fu, "FRCNet: Frequency and Region Consistency for Semi-supervised Medical Image Segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2024, pp. 305–315, doi: 10.1007/978-3-031-72111-3_29.

[44] Y. Jiang *et al.*, "Anatomical Invariance Modeling and Semantic Alignment for Self-supervised Learning in 3D Medical Image Analysis," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2023, pp. 15813–15823, doi: 10.1109/ICCV51070.2023.01453.

[45] S. Li, Y. Dai, J. Chen, F. Yan, and Y. Yang, "MRI-based habitat imaging in cancer treatment: current technology, applications, and challenges," *Cancer Imaging*, vol. 24, no. 1, p. 107, Aug. 2024, doi: 10.1186/s40644-024-00758-9.

[46] M.-T. Tran, H.-J. Yang, S.-H. Kim, and G.-S. Lee, "Prediction of Survival of Glioblastoma Patients Using Local Spatial Relationships and Global Structure Awareness in FLAIR MRI Brain Images," *IEEE Access*, vol. 11, pp. 37437–37449, 2023, doi: 10.1109/ACCESS.2023.3266771.

[47] Y. Su, J. Cheng, C. Zhong, C. Jiang, J. Ye, and J. He, "Accurate polyp segmentation through enhancing feature fusion and boosting boundary performance," *Neurocomputing*, vol. 545, p. 126233, Aug. 2023, doi: 10.1016/j.neucom.2023.126233.

[48] Z. Zhu, M. Sun, G. Qi, Y. Li, X. Gao, and Y. Liu, "Sparse Dynamic Volume TransUNet with multi-level edge fusion for brain tumor segmentation," *Comput. Biol. Med.*, vol. 172, p. 108284, Apr. 2024, doi: 10.1016/j.compbiomed.2024.108284.

[49] Z. Xiong, "ResSAXU-Net for multimodal brain tumor segmentation from brain MRI," *Sci. Rep.*, vol. 15, no. 1, p. 24179, Jul. 2025, doi: 10.1038/s41598-025-09539-1.

[50] Z. Li, W. Silamu, Y. Ma, and Y. Li, "DualTrans: A Novel Glioma Segmentation Framework Based on a Dual-Path Encoder Network and Multi-View Dynamic Fusion Model," *Appl. Sci.*, vol. 14, no. 11, p. 4834, Jun. 2024, doi: 10.3390/app14114834.

[51] I. A. Abbasi, M. Alshehri, and Y. AlQahtani, "Tumor-specific PET tracer imaging and contrast-enhanced Mri based tumor volume differences inspection of glioblastoma patients," *Sci. Rep.*, vol. 15, no. 1, p. 30011, Aug. 2025, doi: 10.1038/s41598-025-15185-4.

[52] P. L. Y. Tang, E. A. H. Warnert, and M. Smits, "Conventional and Advanced MRI in Neuro-Oncology," in *Advanced Imaging and Therapy in Neuro-Oncology*, Cham: Springer Nature Switzerland, 2024, pp. 9–30, doi: 10.1007/978-3-031-59341-3_2.

[53] F. Raab, W. Malloni, S. Wein, M. W. Greenlee, and E. W. Lang, "Investigation of an efficient multi-modal convolutional neural network for multiple sclerosis lesion detection," *Sci. Rep.*, vol. 13, no. 1, p. 21154, Nov. 2023, doi: 10.1038/s41598-023-48578-4.

[54] S. Fanton and W. H. Thompson, "NetPlotBrain : A Python package for visualizing networks and brains," *Netw. Neurosci.*, vol. 7, no. 2, pp. 461–477, Jun. 2023, doi: 10.1162/netn_a_00313.

[55] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao, "Normalization Techniques in Training DNNs: Methodology, Analysis and Application," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 10173–10196, Aug. 2023, doi: 10.1109/TPAMI.2023.3250241.

[56] D. Chicco, N. Tötsch, and G. Jurman, "The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation," *BioData Min.*, vol. 14, no. 1, p. 13, Feb. 2021, doi: 10.1186/s13040-021-00244-z.

[57] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Min.*, vol. 16, no. 1, p. 4, Feb. 2023, doi: 10.1186/s13040-023-00322-4.

[58] D. Chicco, M. J. Warrens, and G. Jurman, "The Matthews Correlation Coefficient (MCC) is More Informative Than Cohen's Kappa and Brier Score in Binary Classification Assessment," *IEEE Access*, vol. 9, pp. 78368–78381, 2021, doi: 10.1109/ACCESS.2021.3084050.