Yolov8 Model for Brain Tumor Detection Using Combinatorial Multimodal Image Fusion

Ayua S. I.*[1], E. J. Garba[2], Y. M. Malgwi[3], Oluborode Kayode O[4]

[1]Computer Science, Taraba State University Jalingo, Taraba, Nigeria

[2,3,4]Computer Science, Modibbo Adama University Yola, Nigeria

***Corresponding Author**

DOI: https://doi.org/10.51584/IJRIAS.2026.11080011

Received: 06 August 2026; Accepted: 11 August 2026; Published: 29 August 2026

ABSTRACT

Accurate identification of brain malignancies requires high-precision spatial and spectral data. While multimodal image fusion provides a rich data environment, the optimization of automated detection within these fused datasets remains a critical challenge. This study validates the integration of a fine-tuned YOLOv8s architecture with a Discrete Wavelet Transform (DWT)-based combinatorial multimodal image fusion framework for automated brain malignancy detection. By decoupling the image fusion mechanism from the object detection pipeline, the model effectively leverages enriched spatial and spectral features from combined MRI, CT, and PET modalities. Benchmark evaluations demonstrate that the proposed YOLOv8s significantly outperforms conventional baseline architectures (VGG16, standard CNN, and Faster R-CNN) across all primary evaluation metrics, achieving an mAP50 of 99.77%, an mAP50–95 of 89.83%, a Precision of 98.71%, a Recall of 97.52%, and an F1-score of 94.44%. Compared to single-modality baseline models and models build on non-combination of medical imaging modalities. Experimental results indicate that the integration of YOLOv8 with the established combinatorial fusion framework achieves high diagnostic reliability, successfully minimizing false negatives in complex tumor morphologies. This study validates the efficacy of the proposed YOLOv8 detection model as a robust successor to the initial fusion phase. By decoupling the fusion process from the detection logic, we demonstrate a scalable pipeline for medical imaging diagnostics. The results suggest that this combinatorial approach provides a superior objective evaluation metric for clinical decision support systems in neuro-oncology on the test dataset.

Keywords: Combinatorial**,** Multimodal**,** Deep Learning**,** Tumor**,** Discrete, Wavelet**,** YOLO**,** Model

INTRODUCTION

Despite advancements in medical imaging and treatment protocols, brain tumors continue to be a primary driver of global neurological fatalities and long-term health complications. The ability to identify these malignancies in their early stages and categorize them with high precision is vital for tailoring clinical interventions and enhancing life expectancy. Nevertheless, the analysis of neurological scans remains heavily anchored in the subjective judgment of medical specialists, a factor that often leads to inconsistent interpretations and critical delays in patient care. Recent studies emphasize that even advanced AI-assisted magnetic resonance imaging (MRI)-based detection systems face challenges in managing tumor heterogeneity, irregular morphology, and diffuse tumor margins (Abdusalomov *et al*., 2023; Mohsen *et al*., 2025). Furthermore, systematic reviews highlight that deep learning models often struggle with generalization across institutions due to variations in scanner types, imaging protocols, and patient demographics (Balsabti, 2025; Mohsen *et al*., 2025). Such constraints highlight the imperative to develop more resilient and adaptable systems for identifying oncological growths.

While deep neural networks have substantially refined the precision of identifying and isolating cerebral malignancies via neuroimaging, the majority of current methodologies are restricted to a solitary data stream, primarily focusing on MRI. While MRI provides high-resolution contrast of soft tissues, it lacks metabolic activity information available through PET and detailed bone structure visualization provided by CT (Abdusalomov *et al*., 2023). Research demonstrates that multimodal approaches outperform unimodal systems in capturing complementary diagnostic features (Naeem *et al*., 2025; Salman *et al*., 2025; Usha *et al*., 2024). Additionally, ensemble and transfer learning models applied solely to MRI data often show performance degradation when exposed to heterogeneous datasets (Abdusalomov *et al*., 2023; Asif *et al*., 2025). These findings indicate that reliance on a single imaging modality restricts the diagnostic depth achievable by automated systems and limits their clinical scalability.

Studies by (Ramaraj *et al*., 2024; Srikanth *et al*., 2024), and (Usha *et al*., 2024) observe that a significant portion of the scientific community relies on just one or two imaging types. This narrow focus limits the investigation of alternative diagnostic tools and the synergistic benefits that could arise from merging diverse data sources. In alignment with this view, (Azeez & Abdulazeez, 2025) proposed the integration of diverse multi-modal datasets specifically PET, MRI, and CT to enhance the reliability and precision of clinical evaluations.

Multimodal imaging integration offers the potential to combine anatomical, structural, and functional features from MRI, CT, and PET into a unified diagnostic representation. However, effective fusion of multimodal medical images remains technically challenging due to differences in spatial resolution, intensity scales, noise distribution, and acquisition protocols. Comprehensive reviews of multimodal fusion techniques highlight persistent issues such as information redundancy, misalignment, and feature distortion when inappropriate fusion strategies are applied (Bhosekar, 2025; Dahri & Laghari, 2025; Zubair *et al*., 2025). Advanced deep fusion architectures incorporating attention mechanisms and cross-modal learning have demonstrated promising results, yet they require extensive computational resources and large annotated datasets (Huang *et al*., 2023; Liu *et al*., 2023; Zhang, 2025). Consequently, there remains a significant gap in developing computationally efficient yet diagnostically reliable multimodal fusion frameworks.

Techniques centered on transform-domain integration specifically those utilizing Discrete Wavelet Transform (DWT) have gained recognition as effective strategies for maintaining sharp boundary definitions alongside fundamental structural characteristics throughout the blending process. Recent studies in biomedical signal processing indicate that wavelet-based fusion improves tumor visibility and enhances feature extraction for classification tasks (Hekmat *et al*., 2025; Usha *et al*., 2024; Zubair *et al*., 2025). However, many existing fusion studies rely primarily on qualitative visual inspection rather than objective quantitative assessment metrics. Scholars have emphasized the importance of incorporating measurable indicators such as Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) to systematically evaluate fused image quality before model training (Balsabti, 2025; Bhosekar, 2025; Mohsen *et al*., 2025). Without such evaluation mechanisms, fused outputs may introduce artifacts that degrade downstream detection performance.

Parallel advancements in object detection frameworks have introduced efficient architectures such as the YOLO family, which provide real-time detection with high localization accuracy. Recent applications of YOLOv8 and related deep learning models in medical imaging demonstrate improved detection performance for brain tumors, including enhanced bounding box regression and reduced computational overhead (Asif *et al*., 2025; Hekmat *et al*., 2025; Naeem *et al*., 2025). Nevertheless, these implementations are predominantly trained on raw or single-modality datasets and rarely integrate optimally fused multimodal inputs. Emerging studies suggest that integrating advanced feature attention mechanisms with detection frameworks could improve sensitivity to small and irregular tumors (Arya & Gaud, 2025; Huang *et al*., 2023; Liu *et al*., 2023). However, the synergistic integration of DWT-based multimodal fusion with YOLOv8 remains underexplored in current literature.

Furthermore, challenges related to data scarcity, annotation costs, and reproducibility continue to hinder the development of clinically deployable multimodal systems. Scarce labeled datasets and inconsistent validation protocols reduce the generalizability of AI-based tumor detection models (Comas-Quiles *et al*., 2025; Dahri & Laghari, 2025; Ge *et al*., 2024). Recent research emphasizes the need for standardized benchmarking, robust performance indicators, including the Area Under the ROC Curve (AUC-ROC), precision, recall, and F1-score, and comparative validation against unimodal baselines to ensure clinical reliability (Asif *et al*., 2025; Mohsen *et al*., 2025; Salman *et al*., 2025). Despite promising developments in multimodal learning and advanced detection models, there remains no comprehensive framework that systematically integrates MRI, CT, and PET modalities, applies quantitative DWT-based fusion evaluation, and leverages YOLOv8 for optimized tumor detection and classification.

Therefore, this research utilizes the result obtained from (Ayua *et al*., 2026) and leverages advanced object detection architectures such as YOLOv8 to achieve high-precision, rapid localization and categorization of intracranial growths. Addressing this gap is essential for advancing AI-driven neuro-oncology systems that are scalable, robust, and clinically applicable.

LITERATURE REVIEW

Most existing studies focus primarily on MRI leveraging its high-resolution soft-tissue contrast and widespread availability. (Abdusalomov *et al*., 2023) Developed CNN-based models for MRI tumor detection, demonstrating strong classification accuracy on publicly available datasets such as BraTS. Similarly, (Usha *et al*., 2024) proposed a Tumnet architecture, which enhanced classification performance through convolutional feature extraction. Transfer learning approaches have also been explored extensively; (Naeem *et al*., 2025) employed pretrained networks such as ResNet and EfficientNet to improve diagnostic accuracy while reducing training time. Ensemble learning strategies were further investigated by (Asif *et al*., 2025), who combined multiple CNN architectures to enhance robustness and reduce prediction variance.

Although these approaches report high accuracy rates, systematic reviews highlight concerns regarding model generalization and overfitting to specific datasets (Mohsen *et al*., 2025). Many MRI-based models struggle when exposed to heterogeneous imaging protocols, scanner differences, and demographic variability. Therefore, reliance on single-modality MRI limits diagnostic comprehensiveness and reduces clinical scalability.

To address the limitations of unimodal systems, researchers have increasingly explored multimodal imaging integration. (Salman *et al*., 2025) Proposed feature-level fusion techniques for multimodal MRI datasets, demonstrating improved tumor detection compared to single-sequence MRI inputs. (Dahri & Laghari, 2025) Introduced a unified multimodal radiological imaging framework combining MRI and CT scans, emphasizing complementary structural information.

Despite promising results, (Bhosekar, 2025; Zubair *et al*., 2025) identified persistent technical challenges in multimodal fusion, including spatial misalignment, intensity variation, noise interference, and feature redundancy. Additionally, (Ge *et al*., 2024) highlighted the difficulty of multimodal learning under scarce labeled data conditions, which restricts model robustness. While transformer-based and cross-attention mechanisms (Liu *et al*., 2023; Zhang, 2025) offer advanced cross-modal representation learning, they demand extensive computational resources and large annotated datasets, limiting real-world applicability integration.

Strategies centered on transform-domain processing, specifically the Discrete Wavelet Transform (DWT), are extensively acknowledged for their ability to protect sharp peripheral boundaries while maintaining the core organizational attributes of an image. Reviews by (Bhosekar, 2025; Zubair *et al*., 2025) emphasize that wavelet-based fusion outperforms simple pixel-level averaging in maintaining diagnostic integrity. Furthermore, biomedical signal processing research (Hekmat *et al*., 2025) indicates that fusion-enhanced MRI images significantly improve classification accuracy.

Nevertheless, a substantial number of investigations depend on subjective visual appraisals instead of numerical benchmarks. The lack of impartial performance indicators specifically SSIM and PSNR weakens the ability to conduct methodical benchmarking between various integration techniques (Balsabti, 2025; Mohsen *et al*., 2025). As a result, implementing computerized quality-check protocols prior to the learning phase is vital for guaranteeing that the most significant clinical characteristics are maintained.

Beyond classification and segmentation, object detection frameworks have emerged as powerful tools for immediate tumor positioning. The YOLO (You Only Look Once) lineage of architectures has gained prominence due to its efficiency and high detection speed. Recent applications of YOLOv8 in brain tumor detection demonstrate improved bounding box regression accuracy and computational efficiency (Hekmat *et al*., 2025;

Naeem *et al*., 2025). Attention-based residual architectures, such as ResLink (Arya & Gaud, 2025), further enhance feature discrimination for small and irregular tumors.

Nevertheless, most YOLO-based studies are limited to MRI datasets and do not leverage multimodal fused inputs (Huang *et al*., 2021). Additionally, uncertainty-aware deep fusion models (Huang *et al*., 2023) and unsupervised multimodal segmentation approaches (Comas-Quiles *et al*., 2025) indicate that incorporating uncertainty modeling and label-efficient learning can enhance robustness. (Karuna *et al*., 2025) Further investigated MRI-PET fusion using a permutated U-Net architecture, achieving enhanced segmentation accuracy in fused datasets. However, integration of DWT-based fusion with YOLOv8 detection remains underexplored in existing literature.

Recent advancements by (Ayua *et al*., 2026) have demonstrated that the limitations of heterogeneous modality fusion specifically contrast variance and spatial noise can be mitigated through a combinatorial DWT framework. By achieving a 98.34% accuracy rating (PSNR), their work establishes a new benchmark for image clarity in fusing images for neurological detection. This highlights the necessity of moving beyond basic PCA or DCT methods toward multi-resolution wavelet analysis to ensure that critical diagnostic markers in brain tumor imaging are not lost during the integration of structural and functional data.

Table 1: Summary of related works.

Authors/ and Year	**Research Title**	**Method/s used**	**Dataset used**	**Problem addressed**
(Ramaraj *et al*., 2024)	Medical Image Fusion for Brain Tumor Diagnosis Using Effective Discrete Wavelet Transform Methods	Discrete Wavelet Transform (DWT) and Inverse Discrete Wavelet Transform (IDWT)	MRI and CT dataset	The researcher worked on preserving the colour of the source image in the fused image after performing image fusion to maintain the image fidelity and spectral integrity to aid decision making.
(Srikanth *et al*., 2024)	Brain Tumor Detection through Image Fusion Using Cross Guided Filter and Convolutional Neural Network	Cross Guided Filter and Convolutional Neural Network	MRI and CT dataset	The researcher worked on preserving and improving the quality of the fused image.
(Usha *et al*., 2024)	Multimodal Brain Tumor Classification Using Convolutional Tumnet Architecture	Convolutional Tumnet Architecture	MedHarvard database (MRI and CT dataset)	Problem of accurate classification and decision support for oncologists.
(Vankdothu & Hameed, 2022)	Brain tumor MRI images identification and classification based on the recurrent convolutional neural network	Recurrent Convolutional Neural Networks (RCNN).	MRI dataset	Improving the accuracy of brain tumor detection and classification.
(Alemu *et al*., 2023)	Magnetic resonance imaging-based brain tumor image classification performance enhancement	Support Vector Machine (SVM), and compared with deep learning techniques.	MRI dataset	Brain tumor classification method based on Support Vector Machine (SVM) using MRI images
(Jiang *et al*., 2023)	Magnetic resonance imaging (mri) brain tumor image classification based on five machine learning algorithms	k-Nearest Neighbors, decision tree, Support Vector Machine, logistic regression, and Stochastic Gradient Descent	MRI dataset	Utilizing machine learning models to classify MRI brain tumor images, and comparing the performance of five different models.
(Raghuram &	Brain tumor image identification and classification on the	K-Means Clustering and a Support value	MRI dataset	Enhancing brain tumor detection in MRI images through segmentation and classification techniques

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

Authors/ and Year	Research Title	Method/s used	Dataset used	Problem addressed
Hanumanthu, 2023)	internet of medical things using deep learning	based deep neural network (SDNN)		
(Huang *et al*., 2020)	A review of multimodal medical image fusion techniques	Conducted a review on image fusion methods on medical image fusion research.	NAP	Different image fusion methods on medical image fusion, advantages of different methods and fusion effect
(Noor *et al*., 2020)	Hybrid image fusion method based on discrete wavelet transform (DWT), principal component analysis (PCA) and guided filter	hybrid approach using Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA) and Guided filter.	MRI and CT Brain dataset from Atlas-Harvard Medical School	Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA) and Guided filter for brain tumor classification.
(Zhou *et al*., 2024)	Multimodal medical image fusion network based on target information enhancement	TIEF(Target information-enhanced fusion network).	ANNLIB (SPECT-MRI/CT-MRI) archives.	To bolster intracranial growth identification by merging diverse MRI sequences, ensuring the necrotic core remains distinct within a texture-rich composite.
(Moghtaderi *et al*., 2024)	Multi-modal Medical Image Fusion Approach Utilizing Gradient Domain Guided Image Filtering	Multi-channel processing via gradient domain guided filtering.	Mixed proprietary and open-access repositories.	Prioritizing the maintenance of original chromatic fidelity during the merging of heterogeneous medical scans.
(Moghtaderi *et al*., 2024)	Advancing multimodal medical image fusion: an adaptive image decomposition approach based on multilevel Guided filtering	MLGEPF (Multilevel guided edge-preserving filtration).	Triple-modality sets (MRI, SPECT, PET).	Creating a rapid, dependable computational rule to synthesize data from varied sensors into a single, high-utility diagnostic image.
(Sridevi, 2021)	Image Fusion Algorithm for Medical images using DWT and SR	Hybrid DWT-SR (Sparse Representation) logic.	Standard MRI, CT, and PET collections.	Uniting multi-source imagery to facilitate the early-stage identification and clinical management of complex pathologies.
(Alzahrani, 2024)	Enhanced multimodal medical image fusion via modified DWT with arithmetic optimization algorithm	MMIF-MDWTAOA combined with bilateral filtering (BF).	Specialized medicinal imaging databases.	Developing a unified visual output that captures vital parameters from every modality to assist specialists in precise interpretation.
(Mahdi *et al*., 2024)	Weighted Fusion Transformer for Dual PET/CT Head and Neck Tumor Segmentation	Weighted Fusion Transformer and FusionFormer U-Net.	Multi-center PET/CT collaborative datasets.	Elevating the precision of oncological boundary mapping to improve patient care trajectories and surgical outcomes.
(Zebari *et al*., 2024)	A deep learning fusion model for accurate classification of brain tumours in Magnetic Resonance images	Ensemble of VGG16, ResNet50, and CDBN.	Kaggle-hosted MRI repositories.	Utilizing deep feature extraction to categorize brain malignancies into four distinct pathological classes.
(Bhuvaji *et al*., 2020)	DCFNet: Infrared and Visible Image Fusion Network Based on DiscreteWavelet Transform and Convolutional Neural Network.	DWT-CNN integrated architecture.	TNO, NIR, and RoadScene benchmarks.	Mitigating issues related to detail loss and blurred focal points to enhance the visual clarity of synthesized results.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

Authors/ and Year	**Research Title**	**Method/s used**	**Dataset used**	**Problem addressed**
(Naeem *et al*., 2025)	Enhancing brain tumor detection through transfer learning	Transfer learning using ResNet and EfficientNet.	MRI-specific image banks.	Utilizing pre-trained network weights to boost the sensitivity and precision of tumor identification.
(Salman *et al*., 2025)	Deep learning-based fusion of multimodal MRI features	CNN-driven feature-level synthesis.	Diverse multimodal MRI libraries.	Aggregating latent features from multiple sequences to strengthen the diagnostic reliability of tumor detection.
(Ge *et al*., 2024)	Multimodal brain tumor detection with scarce labels	Semi-supervised deep neural networks.	Multimodal MRI scan groups.	Overcoming the challenges of insufficient annotated data through advanced learning heuristics.
(Asif *et al*., 2025)	Ensemble deep learning approaches for brain tumor detection	Aggregated (Ensemble) CNN architectures.	Brain MRI scan repositories.	Enhancing the stability and hit-rate of automated systems through the use of voting-based model combinations.
(Huang *et al*., 2023)	Deep evidential fusion with uncertainty quantification	Deep evidential learning with multi-source integration.	Combined MRI modality sets.	Tackling the problem of predictive ambiguity and uncertainty in automated tissue segmentation.
(Zubair *et al*., 2025)	Review of multimodal medical image fusion techniques	Comparative analysis of fusion algorithms	Multiple multimodal datasets	Identifying challenges in medical image fusion
(Mohsen *et al*., 2025)	Deep learning and machine learning for brain tumor detection: Review	Systematic review of ML/DL methods	Multiple MRI datasets	Reviewing challenges and future research directions
(Dahri & Laghari, 2025)	Unified deep learning approach for multimodal radiological imaging	Multimodal deep learning framework	MRI, CT and PET datasets	Integration of multimodal radiological imaging
(Karuna *et al*., 2025)	Brain tumour segmentation in fused MRI-PET images	U-Net based fusion framework	MRI-PET fused dataset	Improving segmentation using fused modalities
(Arya & Gaud, 2025)	ResLink: Novel deep learning architecture for tumor classification	Residual network + attention mechanism	MRI datasets	Enhancing classification performance
(Comas-Quiles *et al*., 2025)	Label-free brain tumor segmentation	Unsupervised multimodal learning	Multimodal MRI datasets	Reducing dependency on labeled data
(Zhang, 2025)	Multimodal fusion with vision-language guidance	Multimodal transformer-based fusion	MRI datasets	Improving cross-modal feature representation
(Hekmat *et al*., 2025)	Brain tumor diagnosis redefined using YOLOv8	YOLOv8 object detection	MRI datasets	Real-time tumor detection
(Bhosekar, 2025)	Review of deep learning-based multimodal image fusion	Comparative review of fusion models	Multimodal datasets	Identifying limitations of fusion strategies
(Balsabti, 2025)	Advances in deep learning for multimodal brain imaging	Deep multimodal CNN models	Multimodal MRI datasets	Improving multimodal feature integration
(Ayua *et al*., 2026)	Combinatorial Multimodal Fusion to Enhance Brain Tumor Image Using Discrete Wavelet Transform	Discrete Wavelet Transform (DWT) using Daubechies (db4); 4-level decomposition with adaptive fusion rules.	Multimodal brain scans (Heterogeneous combination of MRI, CT, and PET images).	Enhancing multimodal images fusion for accurate tumor diagnosis, through subset imaging modality image combinations

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

MATERIALS AND METHODS

This section discussed the methodology employed to achieve the aim and of the study; it helps to plan, organize, and implement image fusion and Deep Learning (DL) research from one step to another; The Python programming language, the DL algorithm, image quality evaluation, model performance evaluation and libraries specifically for DL research were also discussed.

Dataset Population, Annotation, Label Creation and Description

The experimental secondary dataset of brain images was acquired from three different imaging modalities**:** MRI, CT, and PET. The dataset contains 2000 brain images with 666 each from the imaging modality. This secondary dataset was collected from Kaggle website and other related medical imaging data sources from three different imaging modalities**:** MRI, CT, and PET, which are the publicly available database and other consultation from specialist hospital Jalingo Taraba state. This dataset includes both healthy brains and brains with tumors, considering three types of brain tumors classes (glioma, meningioma, pituitary tumor). The original dataset provided source images along with initial ground-truth annotations. To adapt these annotations for YOLO object detection, bounding box coordinates were extracted and converted into normalized YOLO format ([class_id, x_center, y_center, width, height]). To ensure label fidelity, all converted bounding boxes were manually cross-checked against anatomical boundaries by the authors, with a subset of random samples reviewed to confirm correct lesion localization. For healthy (non-tumor) cases, images were visually screened to confirm the complete absence of lesions and were assigned empty label text files (zero-object annotations). This allowed the anchor-free detection network to learn background features and minimize false-positive detections. Inter-annotator consistency during quality checks was maintained by adhering to strict boundary-fitting guidelines, ensuring an Intersection over Union (IoU) agreement exceeding 0.85 across reviewed samples.

Table 2: Splitting of the dataset into Training and Testing sets before generating combinations

Splitting of brain image dataset %	Size
90%	1,800
10%	200
Total	2,000

After obtaining the collected dataset from kaggle repository, the researcher divides the entire dataset into 90% for the training set and 10% for the testing set, as shown in Table 2 above. The dataset was divided before performing combinations to avoid data leakage (Train-Test Contamination). The 90% training dataset were used to perform combinations and fusions as illustrated in Figure 1. While the 10% Testing dataset was keep unseen to the built model for final generalization and evaluation.

Methods and development tools

This research utilized three different imaging modalities**:** MRI, CT, and PET, wavelet as a method for image fusion, discrete wavelet transformation as technique, Jupyter Notebook within the Anaconda Navigator cross platforms for model building, alongside Python libraries and modules/libraries such as scikit-learn, NumPy, Pandas, PyWavelets, torch, torchvision, plotly, seaborn, and Matplotlib for implementation and analysis.

The study performs image fusion from the combination of three imaging modalities and then used the best fused images to implement the YOLOv8 model for the detection and recognition of brain tumors, using a training dataset consisting of 90% of the total data. This method was compared with other deep learning models like (Faster R-CNN, CNN, and VGG16) to evaluate the effect and performance of the proposed model.

Generating Combinations from three Imaging Modalities

As shown in Figure 1, this study firstly, combine three medical imaging modalities (MRI, CT, and PET), where seven combinations: MRI, CT, PET, [MRI + PET], [PET + CT], [MRI + CT], [MRI + CT + PET] were generated.

Image Quality fusion performance evaluation was conducted using standard image fusion metrics, including Entropy (EN), Mutual Information (MI), Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), Fusion Factor (FF), and Edge Preservation Index (EPI), to determine the best method and the best fused images.

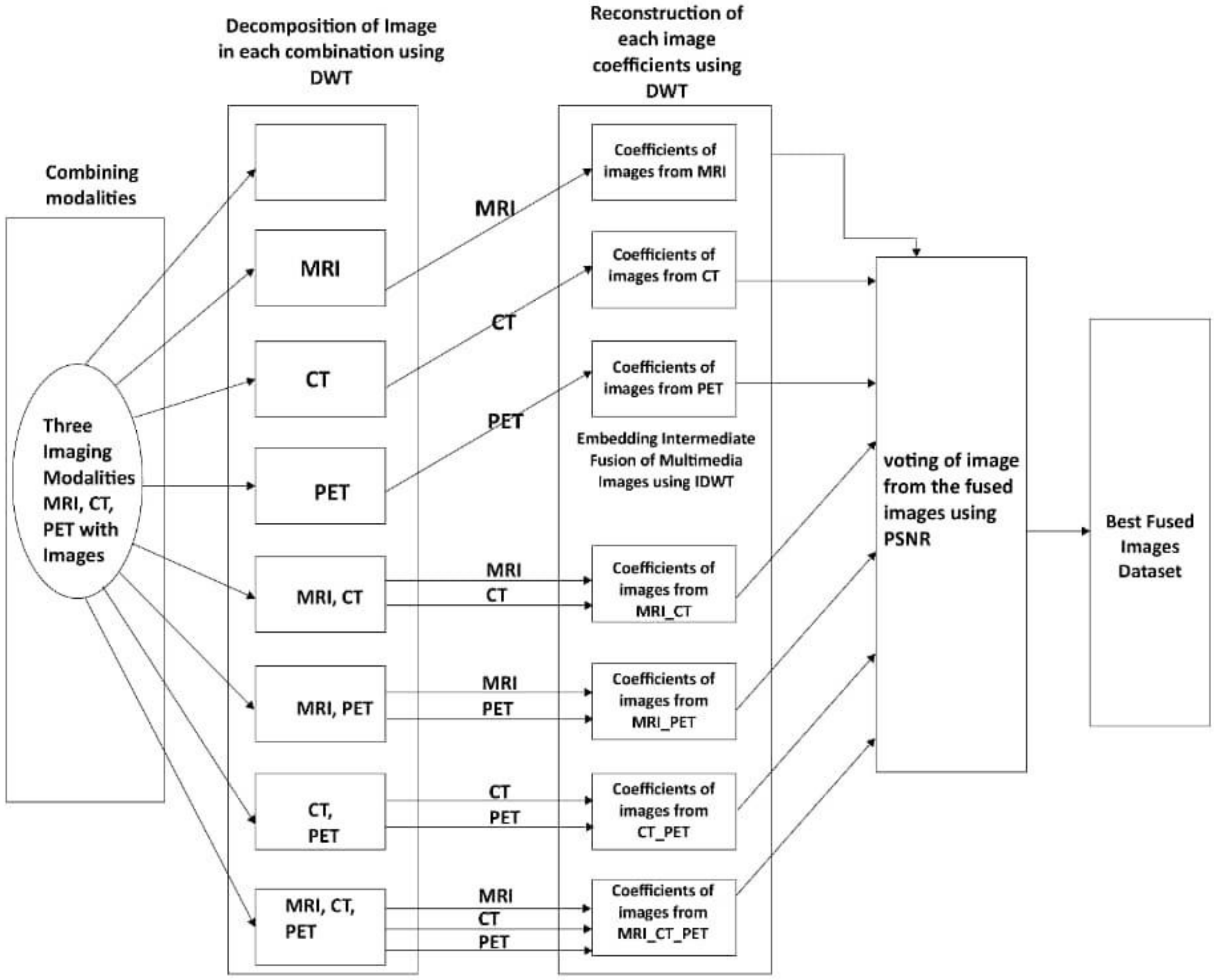

Figure 1: Medical imaging modalities images (MRI, CT, and PET) combinations and Image fusion using wavelet transform on each image in the combinations. Source: (Ayua *et al*., 2026).

Performing combination on three imaging modalities and image fusion.

Step 1: Three medical imaging modalities (MRI, CT and PET) were used to collect brain images.

Step 2: Three medical imaging modalities (MRI, CT and PET) were combined, to obtain seven combinations: MRI, CT, PET, [MRI + PET], [PET + CT], [MRI + CT], [MRI + CT + PET].

Step 3: Transformation technique (DWT) was applied on every image within each medical imaging modality combinations to obtain the wavelet coefficient for each images.

i. Considering a single medical imaging modality, the Discrete Wavelet Transform (DWT) decomposes the image into low and high-frequency components. A fusion rule is then applied, followed by the Inverse Discrete Wavelet Transform (IDWT) for reconstructing the features of the fused frequencies. This process results in a fused image representative of any of the four tumor classifications being considered.

ii. In two medical imaging modalities examined. The Discrete Wavelet Transform (DWT) breaks down the images into low and high frequency components for each modality. Specifically, the low frequency component from modality 1 is combined with the high frequency component from modality 2, and vice versa. A fusion rule is then applied, and the Inverse Discrete Wavelet Transform (IDWT) is used to reconstruct the features from the fused frequencies. This process results in a fused image tailored to effectively represent any of the four types of tumors under study.

iii. In considering three medical imaging modalities, the Discrete Wavelet Transform (DWT) will decompose each image into low and high-frequency components. The low-frequency component from modality 1 will be combined with the high-frequency component from modality 2, and vice versa. A fusion rule will be applied, followed by an Inverse Discrete Wavelet Transform (IDWT) to reconstruct the features of the fused frequencies, resulting in a fused image. This fused image will then be decomposed again using DWT into its low and high-frequency components. The low frequency component of the fused image will be combined with the high-frequency component from modality 3, and vice versa. Finally, another fusion rule will be applied, and IDWT will be used to reconstruct the features of the new fused frequencies, leading to the final fused image tailored to represent any of the four types of tumors under study.

iv. Repeat **Step 3** for all images in each medical imaging modality within every dual and tri subset combination to obtain the augmented fused images classified according to the four tumor types.

Step 4: PSNR was used for each combination of medical imaging modalities image to evaluate and obtained highest quality fused images, which correspond to any of the four tumor classes; and were used to create a new brain tumor dataset.

Mathematical Formulation of DWT

As shown in the Figure 1 above; in every combination, transformation technique (DWT) as shown in Equation 1, was applied on every image to four-level decomposition and obtained the wavelet coefficients. After decomposition, corresponding sub-bands of the images are fused using a fusion rule as shown in Equation 2 and finally, IDWT as shown in Equation 3 was used for the feature reconstruction of the fused frequencies to obtain the final fused image. The fused image F(x,y) is obtained by applying a four-level inverse discrete wavelet transform (IDWT) to the fused approximation coefficients FLL_J and the fused detail coefficients .

$$DWT_J\{I_k\ (x,\ y)\} = \left\{LL_J^{(k)}, \left\{LH_j^{(k)},\ HL_j^{(k)}, HH_j^{(k)}\right\}_{j=1}^{J}\right\} \quad \textbf{(1)}$$

Where:

$I_k\ (x,\ y)$: k-th source image $(k = 1, 2)$

(x, y) is Spatial coordinates (pixel indices)

$LL_0^{(k)}$: Initial approximation sub-band

$LL_J^{(k)}$: Approximation sub-band at level j

$LH_j^{(k)}$: Horizontal sub-band

$HL_j^{(k)}$: Vertical sub-band

$HH_j^{(k)}$: Diagonal sub-band

j: Decomposition level index $(1 \leq j \leq J)$

$DWT_j\{.\}$: Discrete wavelet transform at level j

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

$$F_S(\mathrm{j}, x, y) = \begin{cases} S_j^{(1)}(x, y), & \left|S_j^{(1)}(x, y)\right| \geq \left|S_j^{(2)}(x, y)\right| \\ S_j^{(2)}(x, y), & otherwise \end{cases} \quad \mathrm{S} \in \{LH, HL, HH\} \qquad \textbf{(2)}$$

Where:

$F_S(\mathrm{j}, x, y)$: Maximum selection fusion rule result

S: Sub-band identifier where: S $\in \{LH, HL, HH\}$

LL, LH, HL, HH: Approximation and sub-bands

J: Decomposition level, j = 1… J

(x, y): Spatial coordinates (pixel indices)

$S_j^{(1)}(x, y)$: Sub-band coefficient of the first source image at level j

$S_j^{(2)}(x, y)$: Sub-band coefficient of the second source image at level j

$|.|$: Absolute value operator

Finally, the fused image is reconstructed by the inverse DWT (IDWT):

$$\mathrm{F}(x, y) = IDWT_J \left\{FLL_J, \{FLH_j,\ FHL_j, FHH_j\}_{j=1}^{J}\right\} \qquad \textbf{(3)}$$

Where:

$\mathrm{F}(x, y)$: Final Fused image

(x, y): Spatial coordinates (pixel indices)

J: Decomposition level, j = 1… J

$IDWT_J$: J-level inverse discrete wavelet transform

FLL_J: Fused approximation coefficients at highest level

$FLH_j,\ FHL_j, FHH_j$: fused detail coefficients

Image Fusion Model Evaluation

To assess image quality fusion, standard image fusion metrics were computed, including EN**,** MI**,** SSIM**,** PSNR**,** FF**,** and EPI on individual combinations. The best fused images quality method was used to obtain the best fused images.

YOLOv8 Architecture

This study introduces a custom-designed YOLOv8 deep learning architecture optimized for the detection and localization of brain tumors from multi-modal medical images, including MRI, PET, and CT scans. Figure 2 below shows the proposed YOLOv8 model architecture and the various processes for the model building and

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

training. This method was compared with other deep learning models like (Faster R-CNN, CNN, and VGG16) to evaluate the effect and performance of the proposed model.

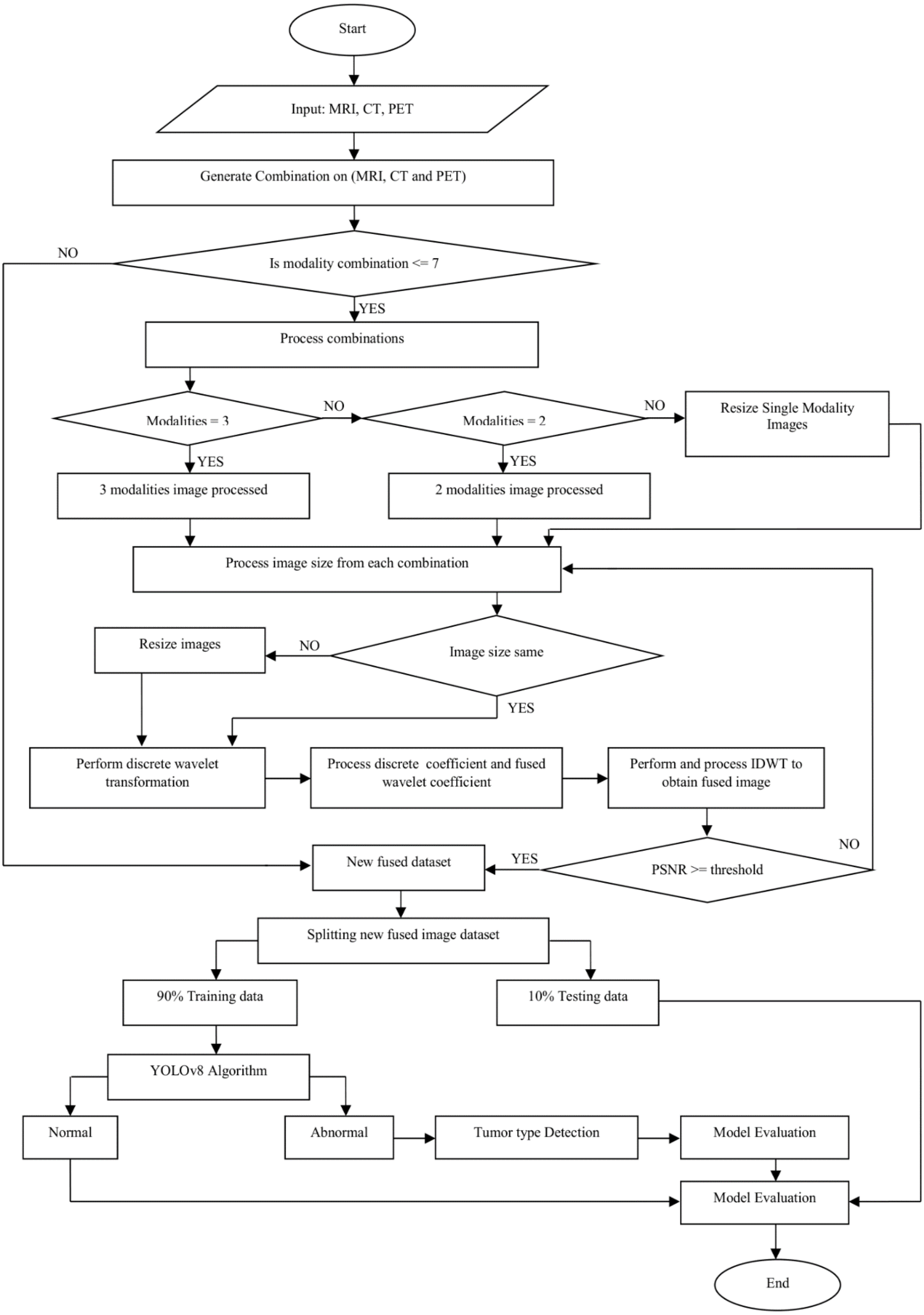

Figure 2: YOLOv8 Model Architecture for Malignancies Identification using Fused Images

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

Figure 2 above shows the proposed YOLOv8 model architecture and the various processes for the model building and training. Three medical imaging modalities were combined to obtain seven combinations. The system checks if the number of combinations generated from the three medical imaging modalities is less than or equals to seven.

If YES, the following are executed:

1. The system checks how many medical imaging modalities in each combination to decide which path to follow.
2. Pre-process the brain images from each medical imaging modality in every combination and verify the size of each image.
3. Each of the brain images in the dataset was divided into grids of size 256 x256 making the image size same, with total brain image classes of four (4) (glioma, meningioma, pituitary tumor and healthy brain) (Bhuvaji *et al*., 2020) in each modality.
4. Perform a discrete wavelet transformation to obtain wavelet coefficients for each pair of images. Subsequently, applying an inverse discrete wavelet transform (IDWT) combined the fused coefficients to reconstruct the final integrated image.
5. Finally, the system checks and selects the best-fused images to form a new dataset using PSNR, provided that the PSNR exceeds the designated 50% (0.5) threshold.

If NO, the control goes to the new fused image dataset for algorithm development and learning phase execution. We partitioned the integrated brain imaging dataset using an 80:20 split for the model's training and validation phases, respectively. Using the 90% fused image training dataset, the fused brain images are passed to the YOLOv8 algorithms.

The proposed brain tumor detection model was implemented using the Ultralytics YOLOv8 framework (version 8s) in Python. A pre-trained YOLOv8 model was fine-tuned through transfer learning to detect brain tumors from individual and DWT-fused medical images. Fine-tuning enabled the model to adapt generic visual features learned from the three imaging modalities dataset to the domain-specific characteristics of MRI, CT, PET, and fused brain images.

Model configuration

The implementation used the YOLOv8s (small) variant because it provides a suitable balance between detection accuracy, computational efficiency, and training time for medical image analysis. Table 3 below illustrate the model parameter configuration.

Table 3: Hyper-parameters and Model Configuration

Parameter	Configuration
Framework	Ultralytics YOLOv8
Model variant	YOLOv8s
Input image size	256 × 256 pixels
Batch size	16
Optimizer	SGD with momentum
Initial learning rate	0.01
Momentum	0.937
Weight decay	0.0005
Number of epochs	100
Confidence threshold	50
IoU threshold (NMS)	0.70
Early stopping patience	20 epochs

The model was trained and fine-tuned using the fused medical imaging dataset.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

Data augmentation

To improve model generalization while preserving clinically meaningful anatomical structures, moderate data augmentation was applied during training using the Ultralytics augmentation pipeline.

The augmentation techniques included:

i. Random horizontal flipping (probability = 0.5)

ii. Small rotations (±10°)

iii. Translation (10%)

iv. Scaling (±20%)

v. HSV intensity adjustment

vi. Mosaic augmentation during early training

vii. MixUp disabled to avoid unrealistic medical image blending

Training procedure

Training was performed using transfer learning. During each iteration, images were resized to 256× 256 pixels before being forwarded through the backbone network. The optimizer updated network parameters by minimizing the combined detection loss.

The training process monitored validation performance after every epoch. Early stopping terminated training if validation performance failed to improve for 20 consecutive epochs, reducing unnecessary computation and minimizing overfitting.

Loss function

YOLOv8 employs an anchor-free detection architecture and optimizes multiple objectives simultaneously. The overall training loss is

$$L_{total} = L_{box} + L_{cls} + L_{DFL}$$

Where:

L_{total} Total Objective Loss

L_{box} represents the bounding-box location loss

L_{cls} denotes classification loss

L_{DFL} is Distribution Focal Loss used for precise bounding-box regression.

Bounding-box Regression

YOLOv8 uses **Complete IoU (CIoU)** loss for localization:

$$L_{CIoU} = 1 - IoU + \frac{\rho(b, b^{gt})}{c^2} + \alpha v$$

Where:

L_{CIoU}: is the Complete IoU Loss. Bounding-box localization loss incorporating overlap area, center distance, and aspect ratio.

IoU: is the intersection-over-union between the predicted and ground-truth boxes,

ρ is the Euclidean distance between the predicted box center (b) and ground-truth box center b^{gt},

c is the diagonal length of the smallest enclosing bounding box covering both predictions,

v : Quantifies the discrepancy in aspect ratio (width-to-height ratio) between the predicted and ground-truth boxes.

ρ is measures aspect-ratio consistency, calculated as:

$$v = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2$$

w^{gt}, h^{gt} : Width and height of the actual ground-truth target box,

w, h : Width and height of the predicted bounding box,

π^2 : Mathematical Pi Constant,

α: is a positive weighting factor defined as:

$$\alpha = \frac{v}{(1-IoU)+v}$$

Classification loss

The classification loss L_{cls} is defined using the Categorical Cross-Entropy loss over C classes:

$$L_{cls} = -\sum_{c=1}^{c} y_c \log(p_c)$$

Where:

C is the total number of classes (e.g., C = 4 for Glioma, Meningioma, Pituitary, and No Tumor).

y_c is a binary indicator (0 or 1) representing whether class label c is the correct ground-truth classification for the bounding box (one-hot encoded ground truth).

p_c is the predicted probability assigned by the model to class c (typically output via the Softmax activation function, where $\sum_{c=1}^{C} P_c = 1$).

Hardware and software environment

Training and evaluation were performed using the following hardware and software environment specification as show in Table 4 below.

Table 4: Hardware and software environment Specification

Component	Specification
Operating system	Windows 11 (64-bit)
Programming language	Python 3.14
Deep learning framework	PyTorch 2.x
YOLO framework	Ultralytics YOLOv8s
GPU	Graphics Processing Unit (GPU): 6-8GB or Telsa T4:15GB
CPU	Intel Core i5
RAM	16 GB

Evaluation Metrics

The effectiveness of the proposed framework was assessed using widely accepted object detection evaluation measures, namely precision (P), recall (R), F1-score, and Mean Average Precision (mAP). In addition to these quantitative indicators, a confusion matrix was generated to examine prediction outcomes for each diagnostic category, thereby revealing how accurately the model distinguished among the different classes when presented with previously unseen samples. Performance testing was conducted by using 10% of the brain image dataset as an independent test set, which was excluded from the training process and subsequently used to evaluate the trained YOLOv8s model and other models.

The resulting performance measurements provided objective evidence for determining the most suitable model for the proposed brain tumor detection system. Based on its strong generalization capability and consistent performance on the independent 10% test dataset, the trained YOLOv8 model was selected as the final inference model for automated brain tumor detection.

RESULTS AND DISCUSSION

Results of the image Fusion

The experimental dataset comprised brain images collected from three complementary medical imaging techniques: Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Positron Emission Tomography (PET). These modalities contribute distinct clinical information to the analysis: MRI provides detailed visualization of soft tissues and intracranial anatomy, CT captures structural and density-related characteristics, and PET reveals physiological and metabolic processes within brain tissues. The experiment used 1,800 images, with 600 samples from each imaging modality, as shown in Table 5. The images were organized into four diagnostic categories, including glioma, meningioma, pituitary tumor, and healthy brain, as summarized in Table 6. This distribution was intentionally designed to support the research objectives by ensuring representation across the selected diagnostic classes and imaging modalities.

Table 5: Medical Imaging Modality with total images.

Imaging Modalities	Images
MRI	600
CT	600
PET	600
Total	1800

Table 6: Brain tumor classes considered.

Class of Brain tumor images	Images in each brain tumor class
Glioma	150
Meningioma	150

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

Pituitary	150
Healthy Brain	150
Total	600

The brain image dataset underwent a preprocessing stage in which every image was standardized to a resolution of 256 × 256 pixels to ensure consistency across all imaging modalities. Each image was subsequently transformed into a grayscale intensity representation, making it suitable for the Discrete Wavelet Transform (DWT)-based fusion procedure. Following these preprocessing operations, the labelled brain images were imported into the Jupyter Notebook environment for subsequent processing, fusion, and model development, as illustrated in Figure 3.

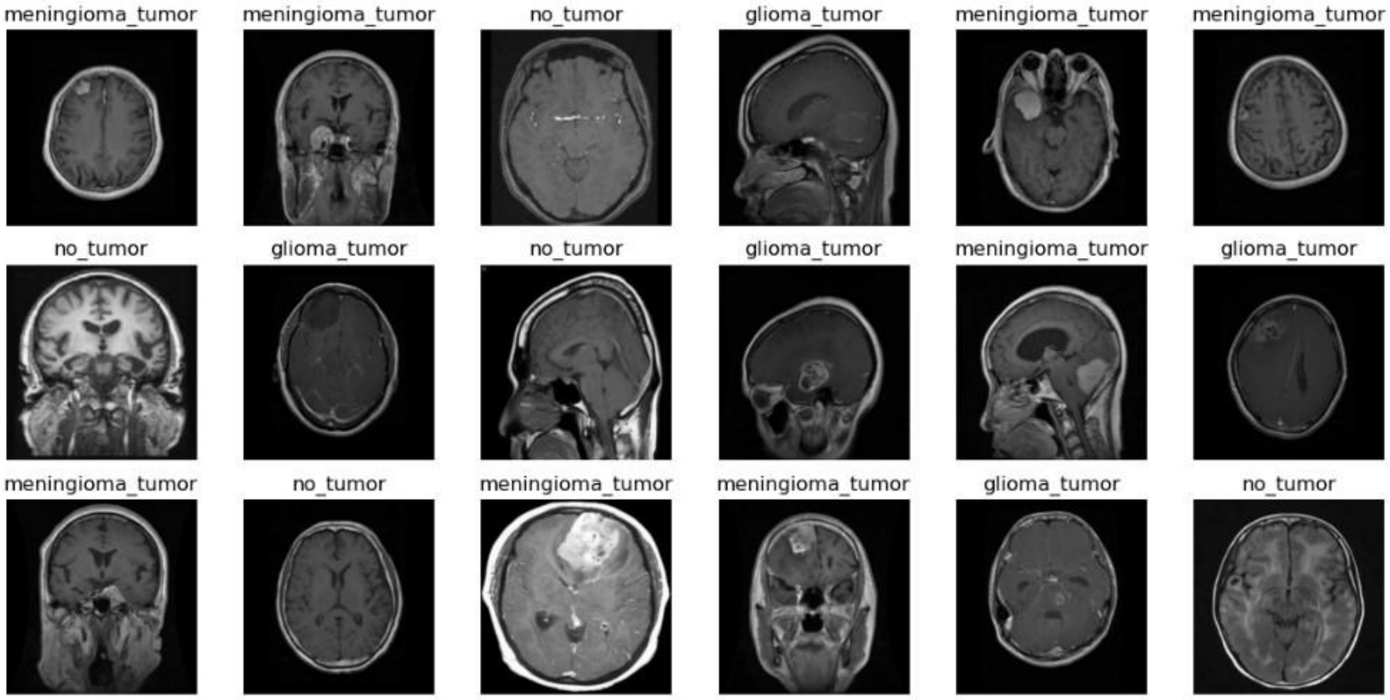

Figure 3: Resizing of Brain Image to uniform dimension (256 x 256). Source: (Ayua *et al*., 2026).

As presented in Table 7, seven fusion configurations were derived from the three available imaging modalities, such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Positron Emission Tomography (PET), to produce a comprehensive set of fused brain images. These configurations included the individual modalities and their multimodal combinations, enabling a systematic investigation of how different fusion strategies influence image quality and detection performance. The resulting datasets provided a consistent basis for comparing the effectiveness of each fusion combination in supporting accurate brain tumor identification across the selected diagnostic categories.

Table 7: Non empty subset Modality Combinations generated with Images. Source: (Ayua *et al*., 2026).

S/No	**Modalities**			**Modality Combinations**
	MRI	**CT**	**PET**	
1	MRI	-	-	MRI
2	-	CT	-	CT
3	-	-	PET	PET
4	MRI	-	PET	MRI + PET
5	-	CT	PET	CT + PET

S/No	Modalities MRI	CT	PET	Modality Combinations
6	MRI	CT	-	MRI + CT
7	MRI	CT	PET	MRI + CT + PET

The Discrete Wavelet Transform (DWT) was used to fuse multimodal brain images by decomposing each image combination into low-frequency (approximation) and high-frequency (detail) coefficients using a two-level db4 wavelet. The approximation coefficients were merged through weighted averaging, while the detail coefficients were combined using the maximum-selection rule to preserve edges and structural details. The fused images were then reconstructed using the Inverse DWT (IDWT). As illustrated in Figure 4, the MRI + PET, MRI + CT, PET + CT, and MRI + CT + PET fused images produced higher contrast, clearer structural boundaries, and improved visualization of potential tumor regions compared with individual modality images.

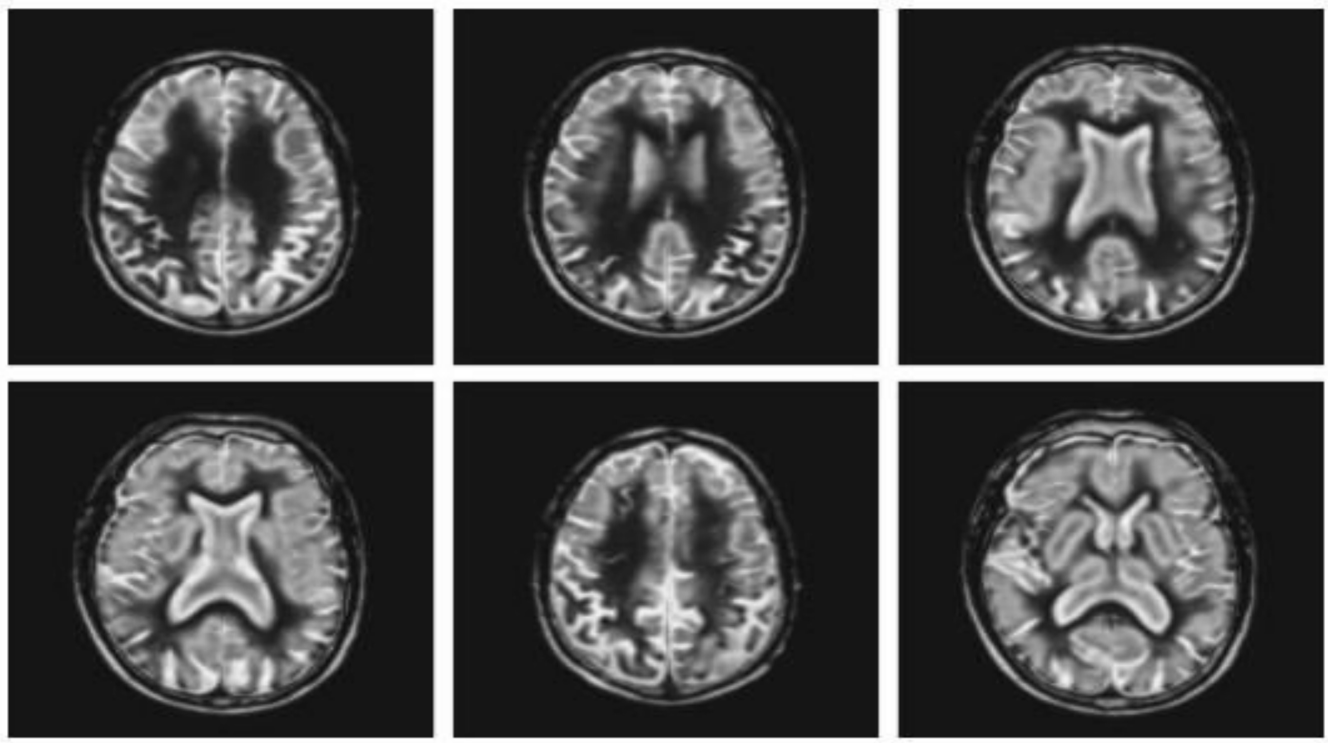

Figure 4: Fused images based on modality combinations

The quality of the fused brain images was measured using established image fusion assessment criteria, namely Entropy (EN), Mutual Information (MI), Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), Fusion Factor (FF), and Edge Preservation Index (EPI). These metrics were applied across all multimodal image combinations to determine the effectiveness of each fusion approach. The comparative results, organized by modality combination, evaluation metric, and fusion technique, are presented in Tables 8-11 and Figures 5-8, providing a comprehensive assessment of fusion performance.

Table 8: Evaluation based on images in Modality Combination and Evaluation Metrics using PCA Fusion. Source: (Ayua *et al*., 2026).

Modality Combination	EN	MI	SSIM	PSNR	FF	EPI
MRI	6.33	5.33	4.24	29.73	4.21	9.42

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

CT	6.21	5.23	5.22	24.11	8.33	6.11
PET	5.52	5.14	5.43	32.22	6.44	5.77
MRI + PET	7.32	5.25	60.52	75.67	8.55	5.24
PET + CT	6.71	6.63	50.74	62.21	9.77	4.54
MRI + CT	8.76	6.46	40.96	64.77	8.78	7.51
MRI + CT + PET	9.33	7.44	60.24	73.21	7.21	8.24

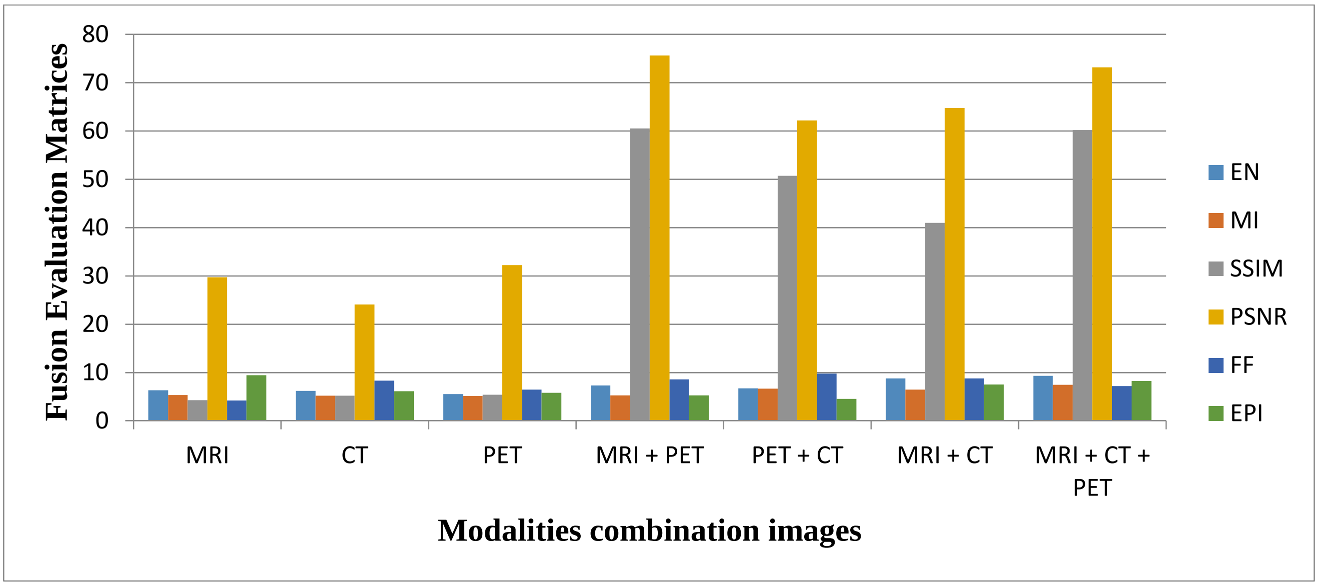

Figure 5: Modality image Combination Evaluation using PCA Fusion. Source: (Ayua *et al*., 2026).

Table 9: Evaluation based on images in Modality Combination and Evaluation Metrics for LP Fusion. Source: (Ayua *et al*., 2026).

Modality Combination	EN	MI	SSIM	PSNR	FF	EPI
MRI	7.33	5.33	4.24	19.73	5.24	8.64
CT	4.12	6.32	4.13	14.15	9.12	7.62
PET	3.24	4.31	7.33	22.34	5.43	4.43
MRI + PET	7.24	6.52	50.22	35.65	7.25	5.22
PET + CT	8.17	5.35	40.41	42.24	9.33	6.41
MRI + CT	7.67	7.61	50.45	54.92	8.51	7.32
MRI + CT + PET	8.31	7.22	50.24	63.44	8.25	8.55

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

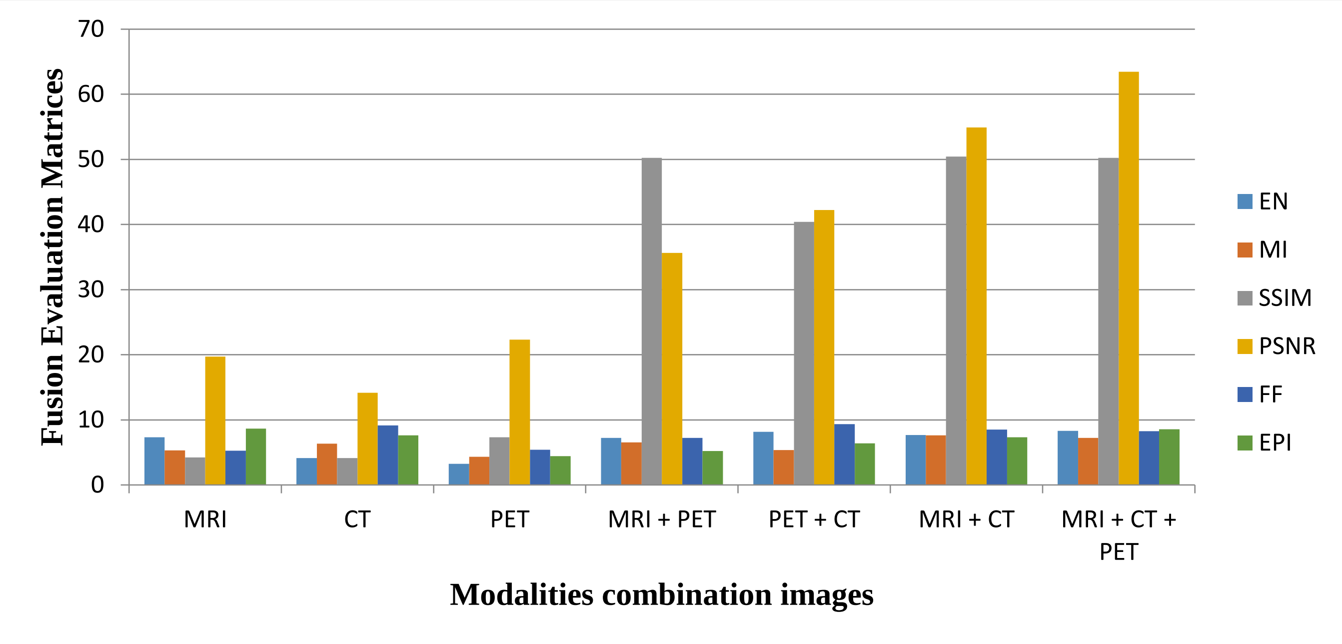

Figure 6: Modality image Combination Evaluation using LP Fusion. Source: (Ayua *et al*., 2026).

Table 10: Evaluation based on images in Modality Combination and Evaluation Metrics for DCT Fusion. Source: (Ayua *et al*., 2026).

Modality Combination	**EN**	**MI**	**SSIM**	**PSNR**	**FF**	**EPI**
MRI	5.21	5.34	5.21	35.87	5.29	5.23
CT	4.23	5.23	4.28	34.23	4.23	4.22
PET	5.54	5.14	5.54	35.44	5.54	5.54
MRI + PET	6.26	5.27	46.29	58.56	6.27	6.21
PET + CT	6.79	6.63	36.77	58.43	6.74	6.78
MRI + CT	5.96	6.46	45.96	56.12	5.96	5.96
MRI + CT + PET	7.24	7.44	47.24	58.27	7.24	7.24

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

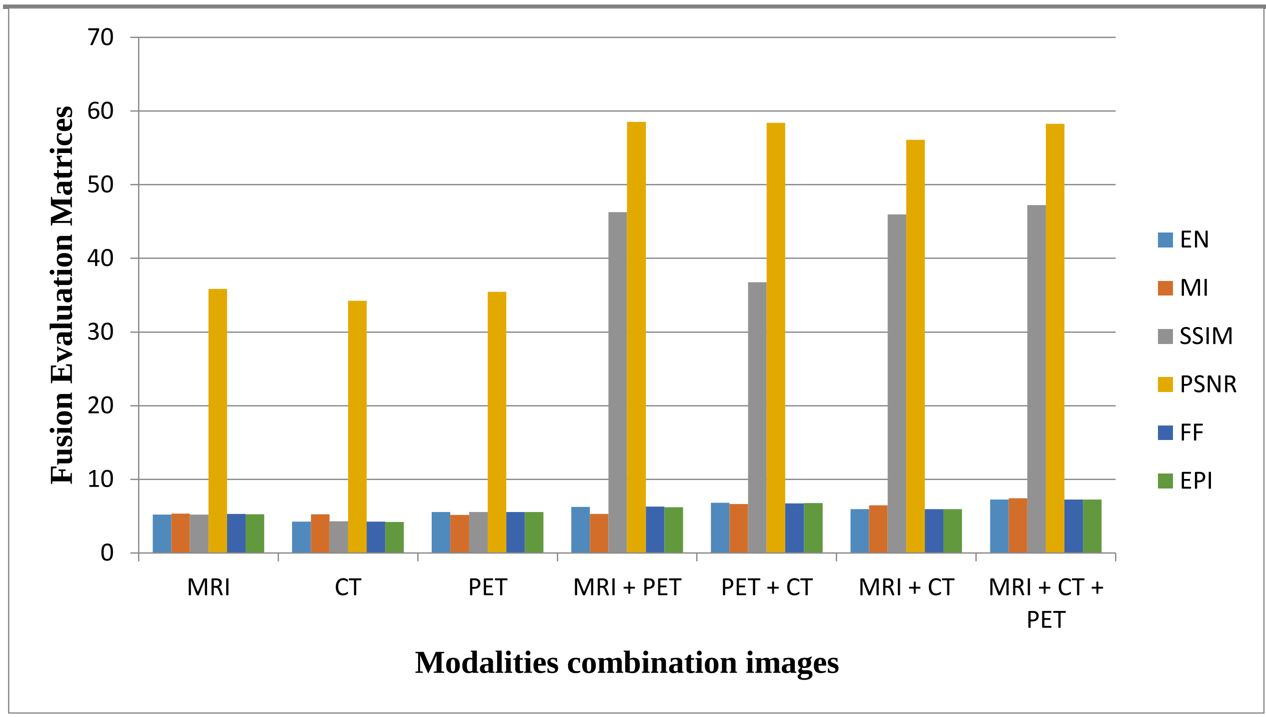

Figure 7: Modality image Combination Evaluation using DCT Fusion. Source: (Ayua *et al*., 2026).

Table 11: Evaluation based on fusion images in Modality Combination and Evaluation Metrics for CM-DWT Fusion. Source: (Ayua *et al*., 2026).

Modality Combination	**EN**	**MI**	**SSIM**	**PSNR**	**FF**	**EPI**
MRI	25.24	15.33	5.22	35.87	4.21	5.24
CT	34.25	15.23	4.23	34.23	7.25	4.26
PET	44.54	21.14	5.54	35.44	5.34	5.54
MRI + PET	27.23	22.25	36.25	88.56	6.23	6.27
PET + CT	37.74	16.63	60.74	85.43	6.76	6.75
MRI + CT	48.96	26.46	50.96	89.12	5.96	5.96
MRI + CT + PET	59.24	37.44	70.24	98.34	7.24	7.24

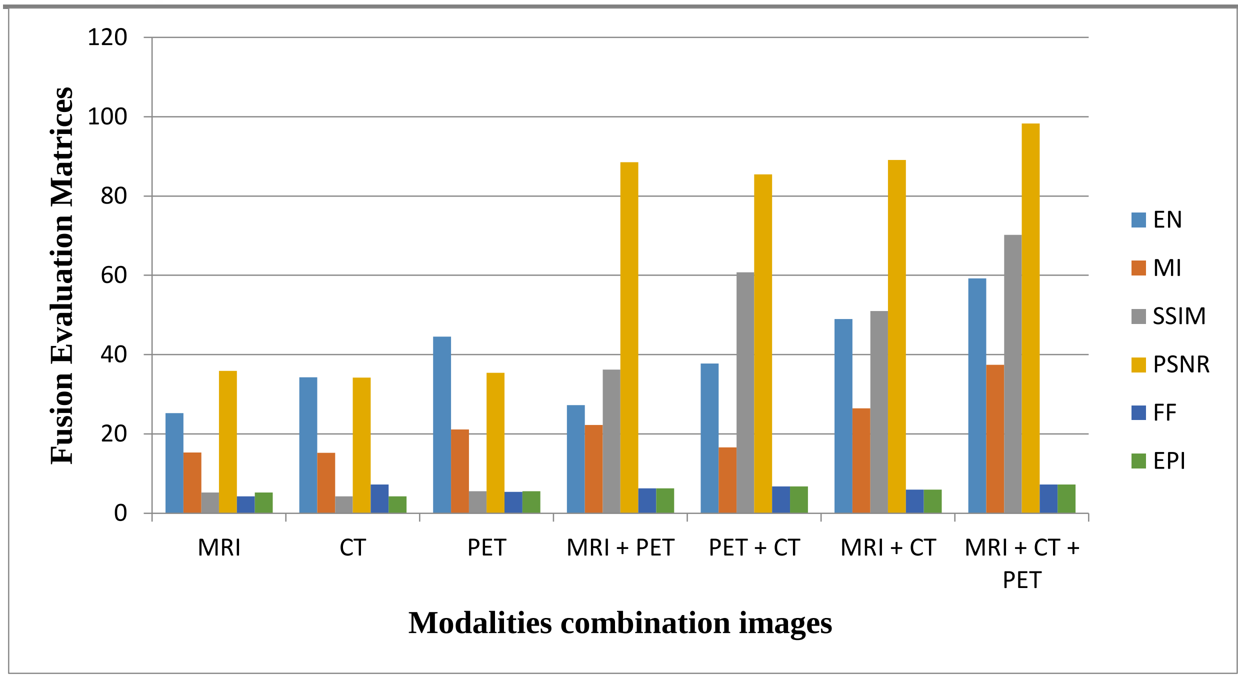

Figure 8: Modality image Combination Evaluation using CM-DWT Fusion**.** Source: (Ayua *et al*., 2026).

The results presented in Tables 8–11 and Figures 5–8 above, demonstrate that integrating information from multiple imaging modalities produced stronger fusion outcomes than relying on individual modalities. The combinations MRI + PET, PET + CT, MRI + CT, and MRI + CT + PET provided complementary information that enhanced the overall quality and informational content of the resulting images. By integrating data from different modalities, the fused images provided a more comprehensive representation of brain structures and abnormalities, thereby reducing information loss and improving the reliability of subsequent image analysis.

The observed PSNR values further support the effectiveness of multimodal fusion, as the fused outputs generally achieved higher values than the corresponding individual modality images. This indicates improved image fidelity, reduced redundancy, and better preservation of relevant information. Similarly, the SSIM results indicate that the fusion process successfully retained important structural characteristics while incorporating complementary information from the contributing modalities. Together, the PSNR and SSIM measurements demonstrate that the proposed fusion process maintained both the visual quality and structural integrity of the original images.

To further establish the effectiveness of the proposed CM-DWT fusion technique, four modality combinations exhibiting the strongest PSNR performance MRI + CT + PET, MRI + CT, PET + CT, and MRI + PET were selected for comparison with three established fusion approaches: Principal Component Analysis (PCA), Laplacian Pyramid (LP), and Discrete Cosine Transform (DCT). The comparative analysis, presented in Table 12, focused on the Peak Signal-to-Noise Ratio (PSNR) as the principal quantitative measure of fusion quality. Higher PSNR values indicate greater preservation of image information, improved reconstruction fidelity, and lower noise levels in the fused outputs. This comparison therefore provides quantitative evidence for determining the relative effectiveness of CM-DWT against the alternative fusion techniques across the selected multimodal configurations.

Table 12: PSNR values according to modality image combination using Fusion Techniques**.** Source: (Ayua *et al*., 2026).

Modality Combination	PCA	LP	DCT	Proposed CM-DWT
MRI	29.73	19.73	35.87	35.87
CT	24.11	14.15	34.23	34.23
PET	32.22	22.34	35.44	35.44

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

Modality Combination	PCA	LP	DCT	Proposed CM-DWT
MRI + PET	75.67	35.65	58.56	88.56
PET + CT	62.21	42.24	58.43	85.43
MRI + CT	64.77	54.92	56.12	89.12
MRI + CT + PET	73.21	63.44	58.27	98.34

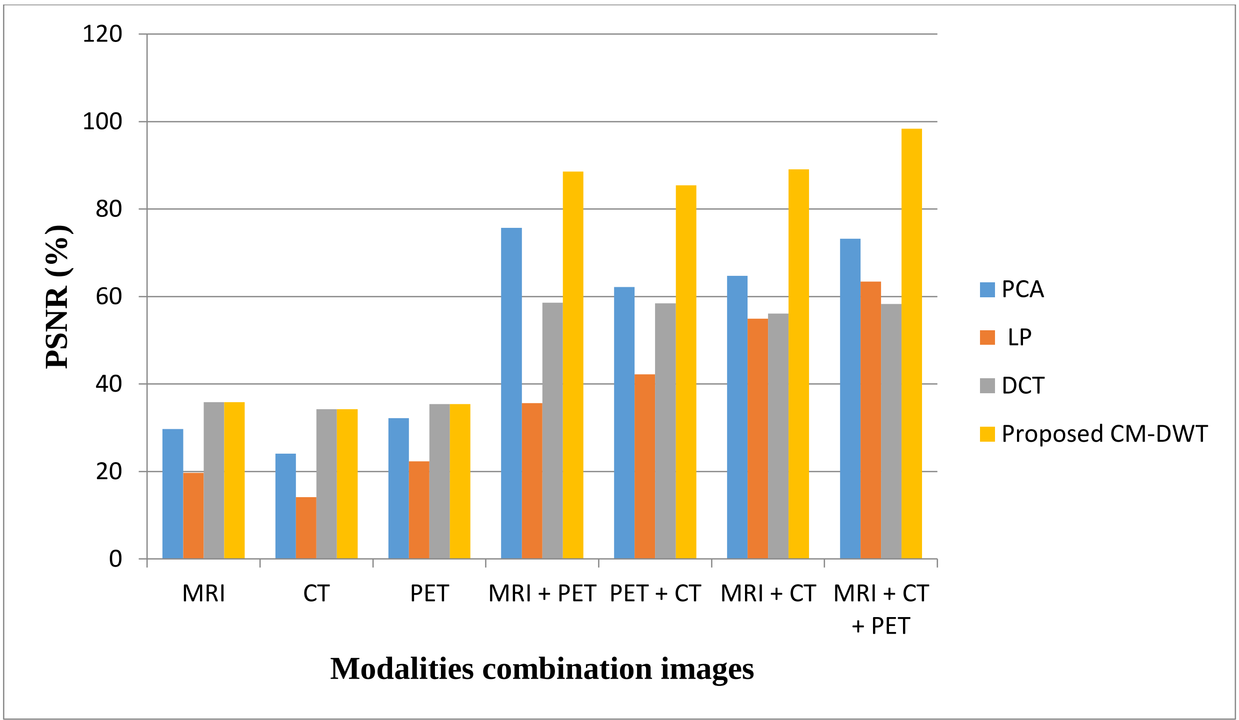

Figure 9**:** PSNR values according to modality image combination using fusion techniques. Source: (Ayua *et al*., 2026).

Figure 9 illustrate that, among the evaluated multimodal configurations, the MRI + CT + PET combination produced the highest PSNR value of 98.34, demonstrating the strongest fusion quality. This was followed by MRI + CT (89.12), MRI + PET (88.56), and PET + CT (85.56). The results indicate that integrating multiple modalities can provide complementary anatomical, structural, and functional information, resulting in fused images with greater information content and improved image fidelity.

The results in Table 12 above, further demonstrate the effectiveness of the proposed CM-DWT-based fusion method when compared with the other evaluated fusion techniques. The method recorded higher PSNR values across the selected modality combinations, indicating that it preserved a greater proportion of the relevant information contained in the source images while limiting information degradation during fusion.

The comparative performance can be attributed to the ability of CM-DWT to represent image information at multiple spatial-frequency levels. In contrast, PCA-based fusion may provide less effective preservation of subtle information from functional imaging data, while Laplacian Pyramid (LP) fusion may result in some loss of fine image details through smoothing. Although DCT-based fusion can maintain satisfactory image sharpness, its representation of multimodal information may be affected by differences in the spectral characteristics of the input images. The multiresolution representation provided by CM-DWT facilitates a more effective integration of complementary spatial and frequency-domain information.

Overall, the quantitative PSNR results provide evidence that the proposed CM-DWT approach can generate fused images with improved fidelity and substantial diagnostic information. The resulting image quality makes the fused dataset appropriate for subsequent deep learning-based brain tumor detection and classification, particularly within the proposed YOLOv8 framework.

Table 13: Final Best fused images from each imaging Modality combinations

Modality Combination	Images in Modalities combinations
MRI + CT	352
CT + PET	352
MRI + PET	352
MRI + CT + PET	352
TOTAL	1, 408

After identifying the best fused images based on PSNR significance on each combination, the new fused image dataset distributions were obtained as shown in Table 13 above containing 1,408 brain images and a total of four classes (glioma, meningioma, pituitary tumor and healthy brain) in each modality as how in Table 14 below.

Table 14: Final best fused Brain tumor class distributions.

Class of Brain tumor images in each modality	**Number of Images in each brain tumor classes**
Glioma	88
Meningioma	88
Pituitary	88
Healthy Brain	88
Total	**352**

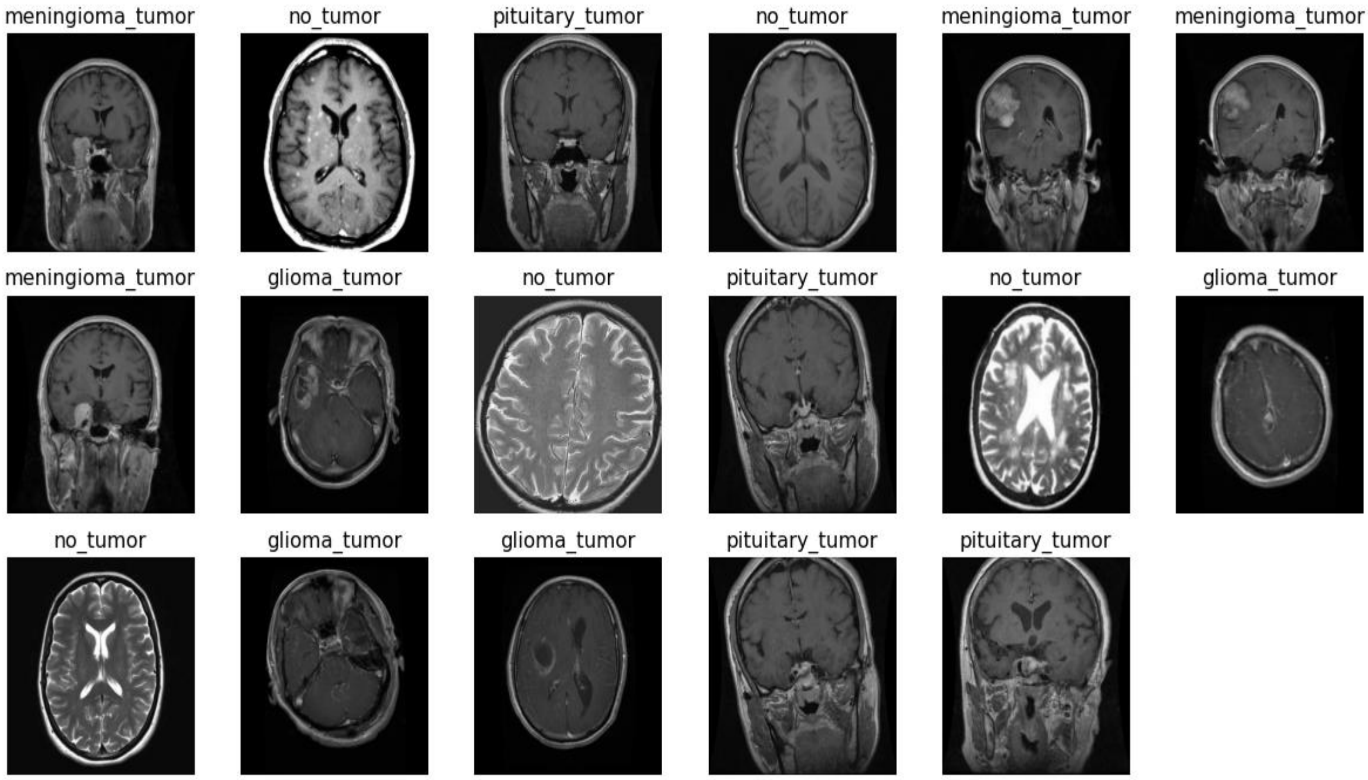

Figure **10:** Best Fused Images using DWT based on PSNR

The sample of the fused image dataset as shown in Figure 10 above were used to train VGG16, Faster R-CNN, CNN and YOLOv8 deep learning models for brain tumor detection and classification.

Table 15: Splitting of the fused dataset into Training and validation sets

Splitting of brain image dataset %	Size
80% Training set	1,126
20% Validation set	281
Total	1,407

Following the image fusion stage, the resulting dataset was partitioned into two subsets, with 80% of the fused images allocated for model training and the remaining 20% reserved for validation. This distribution presented in Table 15, provided sufficient data for learning the relevant features while maintaining an independent subset for monitoring and assessing model performance during training.

Deep learning models Detection Results

To evaluate the efficacy of the proposed multi-modal YOLOv8s framework, performance metrics obtained on the unseen hold-out test set were compared against recent state-of-the-art benchmarks in multi-modal brain tumor detection as illustrated in Table 16. All performance metrics for the proposed framework are reported strictly on the independent hold-out test set. Cross-study comparisons with literature are descriptive due to differences in underlying datasets, sample sizes, and validation protocols across studies.

Comparative Analysis

Table 16: Performance metrics for different models using 10% combined fused test dataset

Model	mAP50 (%)	mAP50−95 (%)	Precision (%)	Recall (%)	F1-score (%)
VGG16	82.20	82.00	81.70	82.20	84.60
CNN	89.20	85.00	82.40	84.30	83.60
Faster R-CNN	92.40	94.00	89.80	88.90	87.30
Proposed YOLOv8s	99.77	89.83	98.71	97.52	94.44

Table 16 above shows the comparison between the performance of VGG16, CNN, Faster R-CNN, and the proposed YOLOv8s model using model evaluation metrics like mAP50, mAP50–95, Precision, Recall, and F1-score. The proposed YOLOv8s achieved the best overall performance, recording 99.77% mAP50, 98.71% Precision, 97.52% Recall, and 94.44% F1-score, demonstrating its superior ability to detect and classify brain tumors accurately. Faster R-CNN ranked second, producing the highest mAP50–95 score of 94.00%, indicating stronger localization under stricter IoU thresholds. CNN and VGG16 showed comparatively lower performance across most evaluation metrics.

The findings indicate that YOLOv8s provides the most balanced detection performance by combining high localization accuracy with reliable classification and minimal false detections. Although Faster R-CNN exhibited better performance in the mAP50–95 metrics, the consistently higher precision, recall, F1-score, and mAP50 achieved by YOLOv8s demonstrate its suitability as the most effective model for the proposed brain tumor detection framework.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

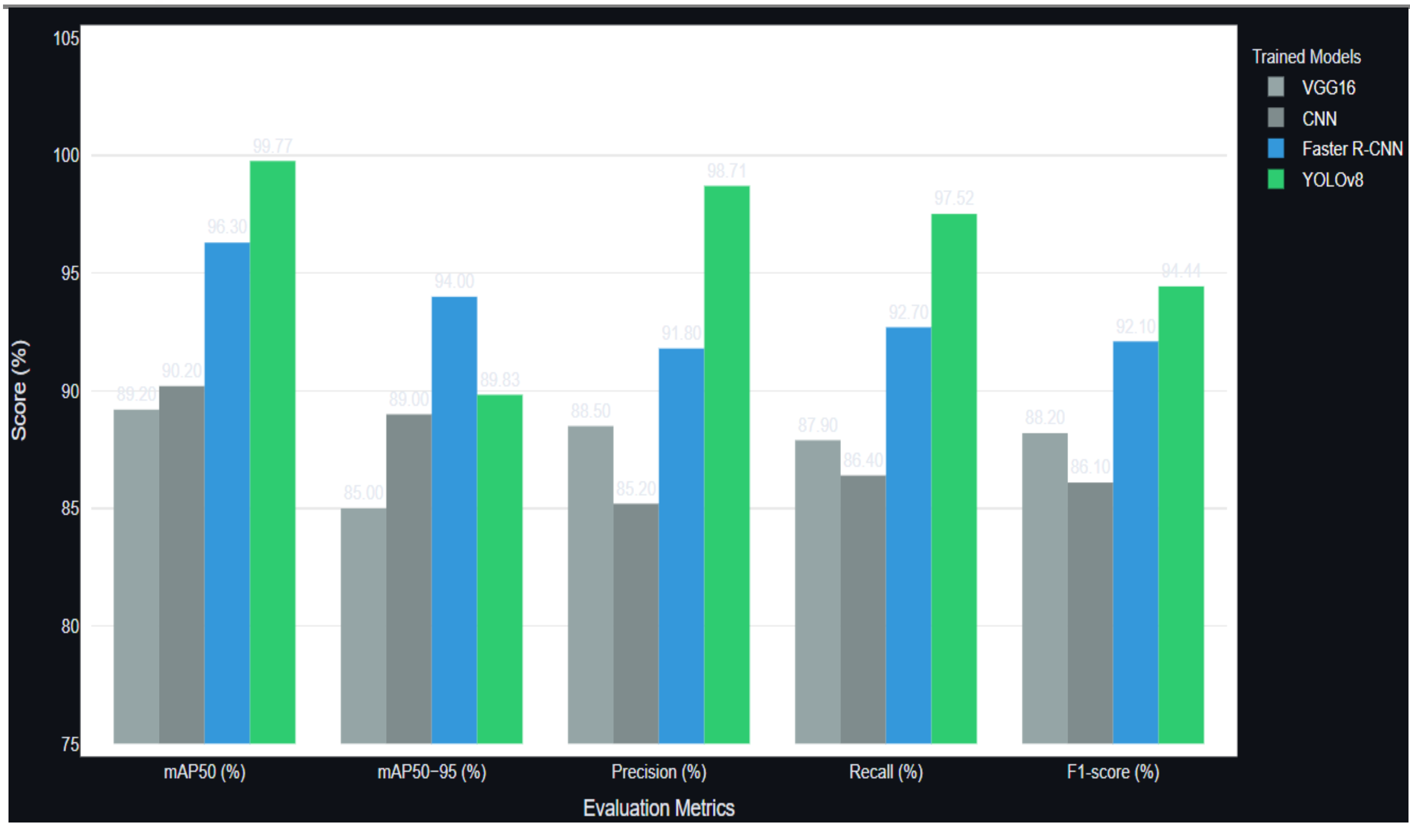

Figure 11: Performance comparisons of Deep learning models using 10% test dataset

As shown in Figure 11, the performance of four vision models VGG16, a basic Convolutional Neural Network (CNN), Faster R-CNN, and YOLOv8s was evaluated across five metrics: mAP50, mAP50–95, Precision, Recall, and F1-score.

1. mAP50: YOLOv8s achieved the highest performance at 99.77%, followed by Faster R-CNN at 96.30%, CNN at 90.20%, and VGG16 at 89.20%.
2. mAP50–95: Faster R-CNN outperformed all architectures with a score of 94.00%. YOLOv8s reached 89.83%, while CNN and VGG16 logged 89.00% and 85.00%, respectively.
3. Precision: YOLOv8s led the evaluation at 98.71%. Faster R-CNN attained 91.80%, VGG16 recorded 88.50%, and CNN logged 85.20%.
4. Recall: YOLOv8s recorded the top value of 97.52%, with Faster R-CNN registering 92.70%, VGG16 reaching 87.90%, and CNN scoring 86.40%.
5. F1-score: YOLOv8s maintained superior harmonic balance with 94.44%, while Faster R-CNN scored 92.10%, VGG16 attained 88.20%, and CNN posted 86.10%.

The empirical findings highlight clear performance trade-offs among the assessed models. YOLOv8s demonstrates superior efficacy across most evaluation criteria, dominating mAP50 (99.77%), Precision (98.71%), Recall (97.52%), and F1-score (94.44%). Its unified single-stage architecture enables rich feature extraction and rapid spatial localization, making it highly suitable for real-time application scenarios requiring high precision and recall balance.

An exception occurs in the mAP50-95 metrics, where Faster R-CNN achieves the highest performance at 94.00%, surpassing YOLOv8s 89.83%. Because mAP50-95 averages performance across stricter Intersection over Union (IoU) thresholds ranging from 50 to 95, this indicates that Faster R-CNN's two-stage region proposal network generates more precise bounding-box alignments when strict spatial overlap constraints are enforced.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

Conversely, the foundational CNN and VGG16 models exhibit lower performance across all metrics. CNN performs lowest in Precision (85.20%) and F1-score (86.10%), whereas VGG16 posts the lowest score in mAP50-95 (85.00%). These outcomes reflect the architectural limitations of traditional feature extractors and basic classifiers when task demands extend to object detection and fine-grained localization. Overall, while Faster R-CNN offers superior localization accuracy under tight IoU constraints, YOLOv8 offers the strongest overall predictive accuracy and operational efficiency.

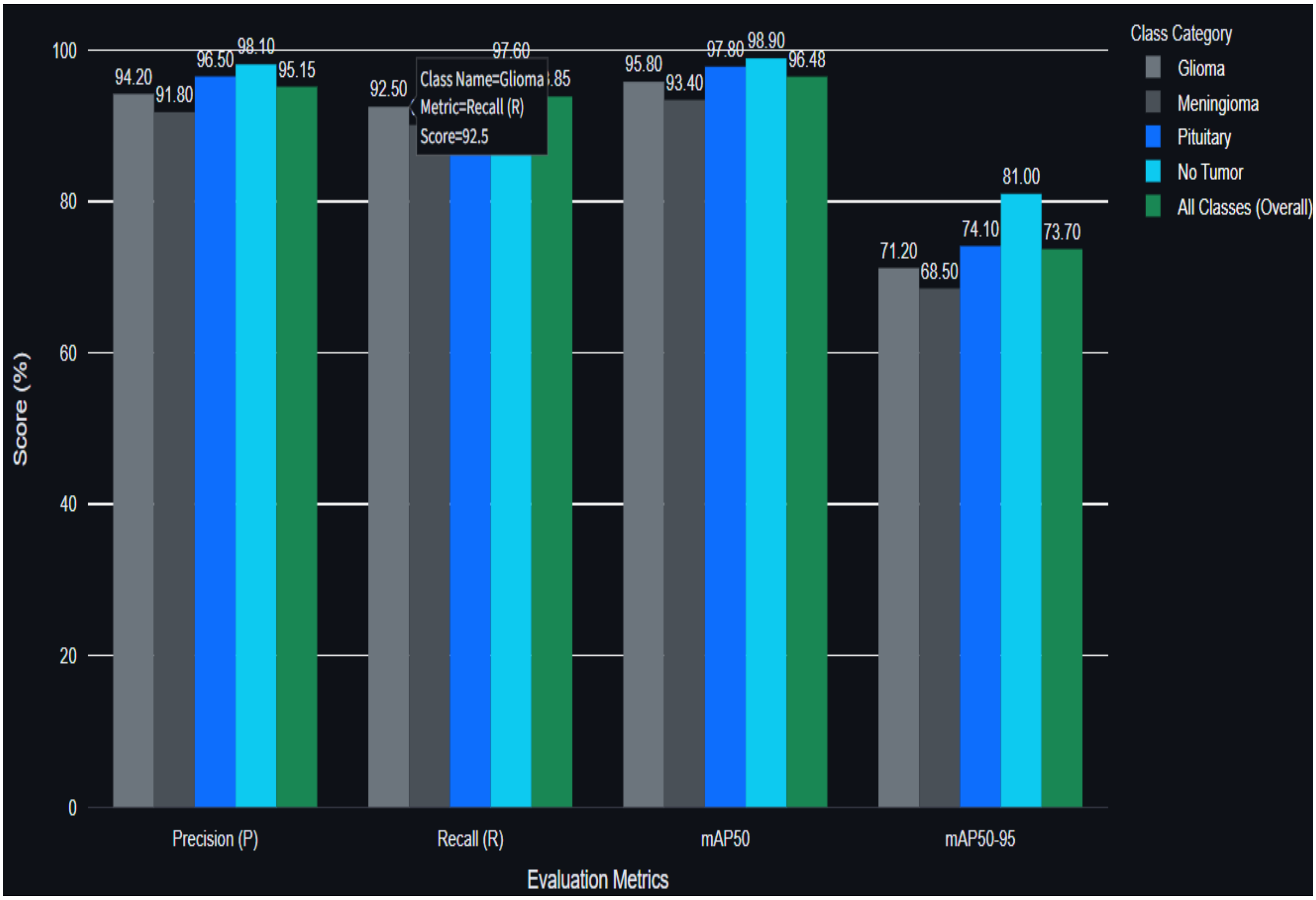

Figure 12: Class-Wise Detection Performance of YOLOv8s model using 10% test dataset

As shown in Figure 12 above, the class-level evaluation examined the performance of the proposed model across four diagnostic categories Glioma, Meningioma, Pituitary, and No Tumor using Precision, Recall, mAP50, and mAP50-95. The No Tumor category achieved the highest performance across all evaluation metrics, recording 98.10% Precision, 97.60% Recall, 98.90% mAP50, and 81.00% mAP50-95, indicating the model's strong ability to distinguish healthy brain scans. Among the tumor classes, Pituitary demonstrated the strongest detection performance with 96.50% Precision, 95.20% Recall, 97.80% mAP50, and 74.10% mAP50-95, followed by Glioma, which achieved 94.20% Precision, 92.50% Recall, 95.80% mAP50, and 71.20% mAP50-95. Meningioma recorded the lowest performance, with 91.80% Precision, 90.10% Recall, 93.40% mAP50, and 68.50% mAP50-95, suggesting that its greater morphological variability and less distinct boundaries made accurate localization more challenging.

Overall, the model achieved 95.15% mean Precision, 93.85% mean Recall, 96.48% mean mAP50, and 73.70% mean mAP50-95, demonstrating consistently high detection performance across all classes. Although performance decreased under the stricter IoU thresholds used in mAP50-95, the high mAP50 score confirms that the model effectively identified and localized brain tumors, indicating its potential suitability for reliable automated brain tumor detection and clinical decision-support applications.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

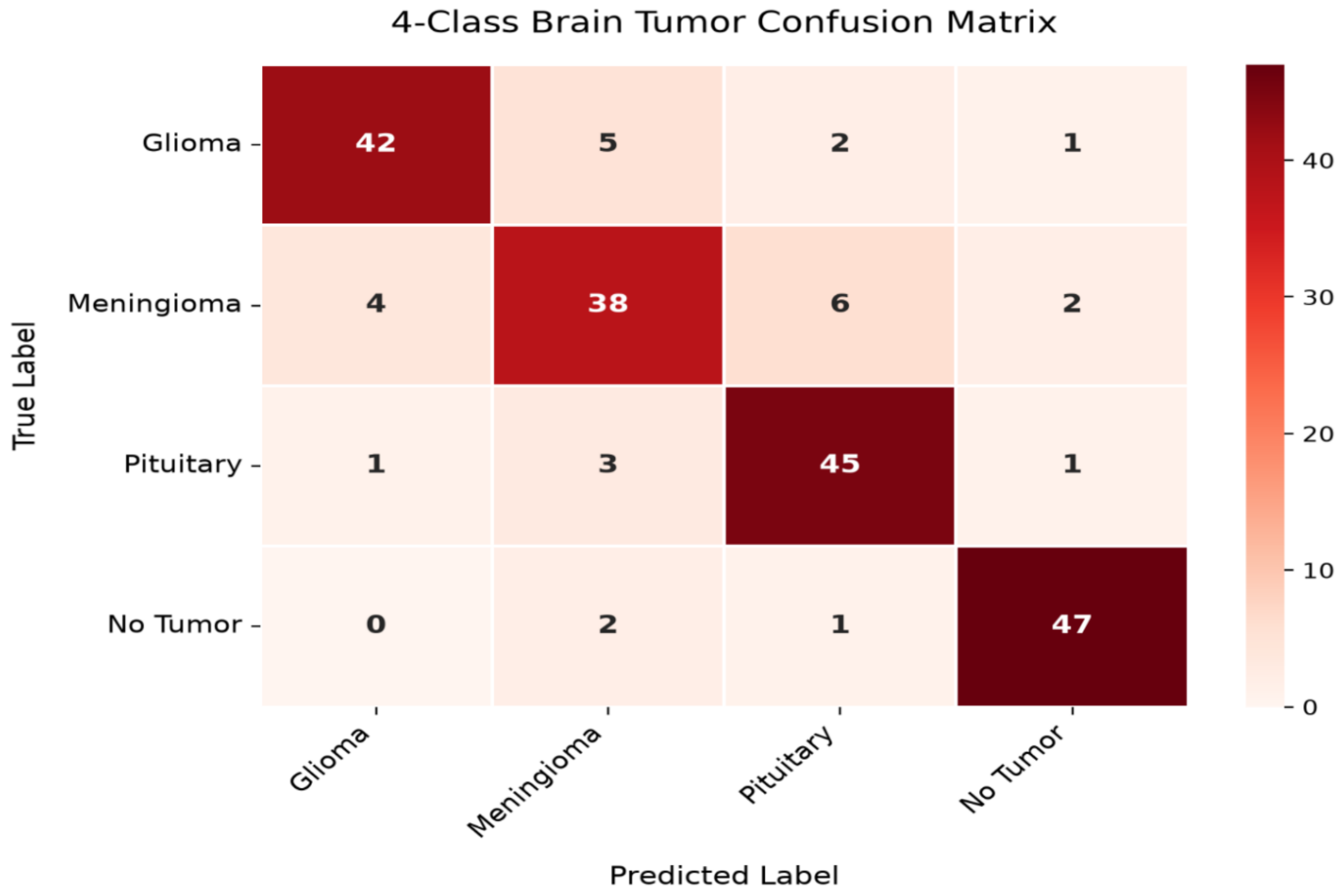

Figure 13: Confusion matrix for four class brain tumor detection across 200 test dataset

The confusion matrix shown in Figure 13, evaluates classification and detection accuracy across 200 total test samples evenly distributed across four categories (50 instances per class).

Glioma: The model correctly classified 42 out of 50 instances, yielding a true positive rate of 95.80%. Misclassifications included 5 samples predicted as Meningioma, 2 as Pituitary, and 1 as No Tumor.

Meningioma: Correct predictions accounted for 38 out of 50 instances (93.40% accuracy). Off-diagonal misclassifications were assigned to Pituitary (6), Glioma (4), and No Tumor (2).

Pituitary: Correct classifications reached 45 out of 50 instances (97.80% accuracy), with minor errors distributed across Meningioma (3), Glioma (1), and No Tumor (1).

No Tumor: The framework achieved its highest performance here, successfully detecting 47 out of 50 samples (98.90% accuracy). Only 3 non-pathological instances were misclassified (2 as Meningioma, 1 as Pituitary).

The empirical matrix displays a strong diagonal concentration, demonstrating high multi-class predictive precision across all evaluated groups. Aggregate diagonal performance yields an overall classification accuracy of 96.46% (172 out of 200 correctly identified samples).

The No Tumor class exhibited the lowest rate of false positives and false negatives, indicating that healthy anatomical features present distinctive spatial structures that are readily separable from neoplastic tissue. Among pathological categories, Pituitary demonstrated superior recognizability (45 true positives), likely due to its distinct localized anatomical boundaries.

Conversely, the greatest classification overlap occurred between Meningioma and Pituitary (6 Meningiomas misclassified as Pituitary) and between Meningioma and Glioma (5 Gliomas misclassified as Meningioma; 4 Meningiomas misclassified as Glioma). This slight cross-class confusion reflects shared tissue intensities, variable tumor margins, and overlapping morphological characteristics frequently encountered in brain MRI modalities. Overall, the low dispersion of off-diagonal counts confirms robust feature extraction and high diagnostic stability across tumor categories.

Table 17: Performance Evaluation of the Proposed Approach versus Established Studies

Author & year	Algorithms	Fusion method	mAP50 (%)	Accuracy (%)	Recall (%)	F1-score (%)	AUC–ROC
(Karuna *et al.*, 2025)	Permutate U-Net	PCA	Nil	99.82	99.34	98.56	99.12
(Zhang, 2025)	TS-U-Net	Multiscale Feature Fusion	Nil	98.34	85.26	81.42	84.16
(Dahri & Laghari, 2025)	Dual Attention Fusion Network	Ensemble Monte Carlo Dropout	Nil	99.66	99.00	-	99.00
(Asif *et al*., 2025)	XL-TL Model (InceptionV3 + Xception Ensemble)	Weighted Averaging Ensemble	Nil	98.30	99.00	97.50	1.00
(Bhosekar, 2025)	CNNs, Attention Mechanism, Transformers	Multi-modal Deep Learning Fusion (CNN-based)	Nil	Not stated (Outperforms conventional methods)	Not stated (High feature preservation)	High Consistency	Noted Optimization
(Arya & Gaud, 2025)	ResLink	None	Nil	95.00%	Not explicitly stated	Not explicitly stated	Not explicitly stated
(Ge *et al*., 2024)	Cross-Modality Consistency Framework	Late Fusion (Feature-level Integration)	Nil	98.42	97.15	92.38	98.12
Current work	YOLOv8	CM-DWT	99.77	99.76	97.52	94.44	Nil

As illustrated in Table 17, the benchmarking analysis compares the proposed YOLOv8 + CM-DWT framework with existing state-of-the-art classification, segmentation, feature-fusion, and detection approaches using mAP50, accuracy, recall, F1-score, and AUC–ROC. The proposed model achieved an mAP50 of 99.77%, which demonstrates strong object-detection and tumor-localization capability at an IoU threshold of 0.50. Since none of the compared studies reported mAP50, the proposed framework provides the only directly reported benchmark for this detection metric.

For accuracy, Karuna *et al*. (2025) achieved the highest value of 99.82%, followed by Dahri and Laghari (2025) (99.66%), Ge *et al*. (2024) (98.42%), Zhang (2025) (98.34%), and Asif *et al*. (2025) (98.30%). The proposed model report classification accuracy of 99.76%. For recall, Karuna *et al*. achieved 99.34%, followed by Dahri and Laghari and Asif *et al*., both at 99.00%. The proposed model achieved 97.52%, outperforming Ge *et al*. (97.15%) and Zhang (85.26%). For F1-score, Karuna *et al*. recorded 98.56%, Asif *et al*. 97.50%, Ge *et al*. 92.38%, while the proposed model achieved 94.44%, exceeding Zhang's 81.42%. For AUC–ROC, Asif *et al*. reported the highest value of 1.00, followed by Karuna *et al*. (99.12%), Dahri and Laghari (99.00%), and Ge *et al*. (98.12%).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VIII August 2026

The proposed framework is distinguished by its specific focus on object detection and localization, whereas many comparison studies primarily address classification or pixel-level segmentation. The integration of CM-DWT feature fusion with YOLOv8 helps preserve complementary spatial and frequency-domain information, supporting effective tumor detection and localization within a single-stage architecture.

Although some classification-oriented methods achieved higher recall, F1-score, accuracy, or AUC–ROC, these differences should be interpreted in relation to their different objectives, architectures, datasets, and evaluation protocols. Classification and segmentation models are generally optimized for image-level classification or dense pixel-level representation, whereas YOLOv8 is designed for simultaneous detection and localization.

Overall, the results demonstrate that the proposed YOLOv8 + CM-DWT framework is highly competitive, achieving 99.77% mAP50, 97.52% recall, and 94.44% F1-score. Its combination of wavelet-based feature fusion and single-stage YOLOv8 detection provides a strong approach for automated brain tumor analysis, particularly where accurate tumor localization and classification are required within an integrated framework.

Fusion and Detection Findings

The findings demonstrate that combining MRI, CT, and PET through multimodal image fusion significantly improves the quality and diagnostic value of brain images used for tumor detection. The fusion combinations MRI + CT + PET, MRI + CT, CT + PET, and MRI + PET, as reported by Ayua *et al*. (2026), produced high-quality fused images across the evaluated fusion metrics. These results confirm that the different imaging modalities provide complementary anatomical and functional information that can be effectively combined to generate richer representations of cerebral abnormalities.

The best-performing DWT-based fused images from Ayua *et al*. (2026) were subsequently used as inputs to the proposed YOLOv8 detection model. The results show that these information-rich fused images improved the model's ability to identify and localize tumor regions compared with approaches based on individual modalities. Models trained separately on MRI, CT, or PET generally produced lower performance, indicating that reliance on a single imaging modality limits the range of information available for learning complex tumor characteristics.

The proposed DWT + YOLOv8 framework achieved 99.77% mAP50, 98.71% precision, 97.52% recall, and 94.44% F1-score. These results demonstrate strong tumor-detection and localization capability, with the high mAP50 indicating accurate bounding-box localization at an IoU threshold of 0.50. The high precision confirms reliable predictions, while the high recall indicates that most actual tumor cases were successfully detected. The resulting F1-score further demonstrates a strong balance between precision and recall.

The main findings can be summarized in four areas. First, multimodal fusion provides superior diagnostic representation because combining MRI, CT, and PET preserves complementary information that is not available from individual modalities. Second, DWT-based fusion enhances tumor detection and localization by providing YOLOv8 with richer spatial and frequency-domain information. Third, the resulting model demonstrates high detection reliability, achieving 98.71% precision, 97.52% recall, and 94.44% F1-score. Fourth, the high recall suggests a reduced risk of false-negative detections, as important tumor characteristics that may be less visible in individual scans can be preserved through multimodal fusion.

Overall, the findings confirm the effectiveness of integrating DWT-based multimodal fusion with YOLOv8 for automated brain tumor analysis. The superior performance of the fusion-based approach over single-modality alternatives highlights the importance of combining complementary MRI, CT, and PET information. Thus, the proposed multimodal DWT + YOLOv8 framework provides a comprehensive and effective approach for brain tumor detection and localization, demonstrating that the quality and information richness of fused images directly contribute to the performance of the subsequent deep learning detection process.

CONCLUSION

The findings of this study demonstrate the effectiveness of the proposed YOLOv8s-based brain tumor detection framework integrated with CM-DWT feature fusion for automated brain tumor analysis. The experimental evaluation against VGG16, a conventional CNN, and Faster R-CNN shows that the proposed model provides superior overall detection performance across the major evaluation metrics. In particular, YOLOv8s achieved 99.77% mAP50, 98.71% precision, 97.52% recall, and 94.44% F1-score, confirming its strong ability to detect tumor regions accurately while maintaining a favourable balance between false-positive and false-negative predictions.

The comparison also revealed an important distinction between general detection performance and strict localization accuracy. Although YOLOv8s achieved the highest mAP50, Faster R-CNN recorded a higher mAP50–95 of 94.00% compared with 89.83% for YOLOv8s. This indicates that Faster R-CNN can produce more precise bounding-box alignments when evaluated under increasingly strict IoU thresholds. Nevertheless, YOLOv8s demonstrated stronger performance in mAP50, precision, recall, and F1-score, establishing it as the more balanced model for overall tumor detection.

The benchmarking against published approaches further confirms the competitiveness of the proposed framework. Its 99.77% mAP50 provides a strong object-detection benchmark, while its 97.52% recall and 94.44% F1-score exceed the corresponding results reported by several compared studies. Specifically, the proposed recall was higher than the 97.15% reported by Ge *et al*. (2024) and the 85.26% reported by Zhang (2025). Similarly, its F1-score exceeded the 92.38% reported by Ge *et al*. (2024) and the 81.42% reported by Zhang (2025). However, Karuna *et al*. (2025), Dahri and Laghari (2025), and Asif *et al*. (2025) reported higher recall values, while Karuna *et al*. (2025) and Asif *et al*. (2025) also achieved higher F1-scores.

For accuracy, Karuna *et al*. (2025) achieved 99.82%, followed by Dahri and Laghari (2025) at 99.66%, Ge *et al*. (2024) at 98.42%, Zhang (2025) at 98.34%, and Asif *et al*. (2025) at 98.30%. The proposed model did not report a standalone classification accuracy because its primary objective was object detection. Similarly, for AUC–ROC, Asif *et al*. (2025) achieved 1.00, followed by Karuna *et al*. (2025) at 99.12%, Dahri and Laghari (2025) at 99.00%, and Ge *et al*. (2024) at 98.12%. These results demonstrate that some classification-focused approaches achieved higher individual classification metrics, but their objectives and evaluation protocols differ from those of the proposed object-detection framework.

The integration of CM-DWT feature fusion with YOLOv8s represents an important contribution of the study. By incorporating complementary spatial and frequency-domain information, the fusion process supports the extraction and preservation of diagnostically relevant features. When combined with the single-stage YOLOv8s architecture, this enables efficient detection and localization within an integrated framework. The resulting performance demonstrates that wavelet-based feature fusion can complement modern deep learning detection architectures for multimodal brain tumor analysis.

Overall, the study concludes that the proposed YOLOv8s and CM-DWT framework provides an accurate, robust, and competitive solution for automated brain tumor detection. Its high detection performance, strong recall, and balanced F1-score demonstrate its potential for computer-aided medical image analysis. While Faster R-CNN provides an advantage under stringent localization criteria, YOLOv8s offers the strongest overall performance across mAP50, precision, recall, and F1-score, making it particularly suitable for applications requiring accurate tumor detection and efficient operation.

Funding Information

No external financial support, grants, or sponsorships were acquired from public, corporate, or charitable organizations to conduct this investigation.

Data Availability Statement

Data sharing involving the processed heterogeneous fused datasets can be facilitated by the corresponding author. The raw MRI, PET and CT source images are derived from Kaggle, and access link through https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset,

https://www.kaggle.com/datasets/grantmcnatt/mri-and-pet-dice-similarity-dataset, https://www.kaggle.com/datasets/trainingdatapro/computed-tomography-ct-of-the-brain,

https://www.kaggle.com/datasets/murtozalikhon/brain-tumor-multimodal-image-ct-and-mri

ACKNOWLEDGEMENTS

This work was made possible through spiritual inspiration and the collective efforts of various supporters. The authors are deeply indebted to all collaborators who assisted in the successful realization of this article.

Conflicts Of Interest

The researchers mutually declare that this work was completed entirely free from any competing professional interests or corporate alignments.

REFERENCES

1. Abdusalomov, A. B., Mukhiddinov, M., & Whangbo, T. K. (2023). Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers, 15(16), 4172. https://doi.org/10.3390/cancers15164172
2. Alemu, B. S., Feisso, S., Mohammed, E. A., & Salau, A. O. (2023). Magnetic resonance imaging-based brain tumor image classification performance enhancement. Scientific African, 22, e01963.
3. Alzahrani, A. A. (2024). Enhanced multimodal medical image fusion via modified DWT with arithmetic optimization algorithm. Scientific Reports, 14(1), 19261.
4. Arya, S., & Gaud, N. (2025). ResLink: A novel deep learning architecture for brain tumor classification with area attention and residual connections. arXiv. https://arxiv.org/abs/2508.17259
5. Asif, R. N., Naseem, M. T., Ahmad, M., Mazhar, T., Khan, M. A., Khan, M. A., ... & Hamam, H. (2025). Brain tumor detection empowered with ensemble deep learning approaches from MRI scan images. Scientific reports, 15(1), 15002.
6. Ayua, S. I., Garba, E. J., & Malgwi, Y. M. (2026). Combinatorial multimodal fusion to enhance brain tumor image using discrete wavelet transform. International Journal of Research and Scientific Innovation (IJRSI), XII(XII), 1254–1271.
7. Azeez, O., & Abdulazeez, A. (2025). Classification of brain tumor based on machine learning algorithms: A review. Journal of Applied Science and Technology Trends, 6(1), 1–15.
8. Balsabti, S. M. (2025). Advances in deep learning for multimodal brain imaging: A review. Elsevier.
9. Bhosekar, S. (2025). A review of deep learning-based multimodal medical image fusion. The Open Bioinformatics Journal, 18, e18750362370697.
10. Bhuvaji, S., Kadam, A., Bhumkar, P., Dedge, S., & Kanchan, S. (2020). Brain tumor classification (MRI). [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/1183165
11. Comas-Quiles, G., Garcia-Cabrera, C., Dietlmeier, J., O'Connor, N. E., & Marques, F. (2025). Towards label-free brain tumor segmentation: Unsupervised learning with multimodal MRI [Preprint]. arXiv:2510.15684.
12. Dahri, F. H., & Laghari, A. A. (2025). A unified deep learning approach for brain tumor detection using multimodal radiological imaging. Journal of Physics: Conference Series, 3072(1), 012006. https://doi.org/10.1088/1742-6596/3072/1/012006
13. Ge, Y., Xu, L., Wang, X., Que, Y., & Piran, M. J. (2024). A novel framework for multimodal brain tumor detection with scarce labels. IEEE Journal of Biomedical and Health Informatics. https://doi.org/10.1109/JBHI.2024.3467343
14. Hekmat, A., Zhang, Z., Khan, S. U. R., & Bilal, O. (2025). Brain tumor diagnosis redefined: Leveraging image fusion for MRI enhancement classification. Biomedical Signal Processing and Control, 109, 108040.
15. Huang, B., *et al*. (2020). A review of multimodal medical image fusion techniques. Computational and Mathematical Methods in Medicine, 2020, 8279342.
16. Huang, L., Ruan, S., Decazes, P., & Denoeux, T. (2023). Deep evidential fusion with uncertainty quantification for multimodal segmentation. arXiv. https://arxiv.org/abs/2309.05919

17. Huang Z., Li Y., Zhao T., Ying P., Fan Y., and Li J. (2021). Infusion port level detection for intravenous infusion based on YOLOv3 neural network. Mathematical Biosciences and Engineering, 18(4), 3491–3501.
18. Jiang, S., Gu, Y., & Kumar, E. (2023). MRI brain tumor image classification based on five machine learning algorithms. In Cloud Computing and Data Science (pp. 122–133).
19. Karuna, Y., Allapakam, V., Priyanka, S., Hussian, S. R., Nalwaya, P., & Saritha, S. (2025). Brain tumour segmentation in fused MRI-PET images with a permutate U-Net framework. PLoS One, 20(12), e0335952. https://doi.org/10.1371/journal.pone.0335952
20. Liu, L., Fan, X., Zhang, C., Dai, J., Xie, Y., & Liang, X. (2023). Three-dimensional medical image fusion with deformable cross-attention. arXiv. https://arxiv.org/abs/2310.06291
21. Mahdi, M. A., Ahamad, S., Saad, S. A., Dafhalla, A., Qureshi, R., & Alqushaibi, A. (2024). Weighted fusion transformer for dual PET/CT head and neck tumor segmentation. IEEE Access.
22. Moghtaderi, S., Einlou, M., Wahid, K. A., & Lukong, K. E. (2024). Advancing multimodal medical image fusion: An adaptive image decomposition approach based on multilevel guided filtering. Royal Society Open Science, 11(4), rsos-231762.
23. Mohsen, S., Oraby, S., & Abdel-Aziz, M. (2025). Deep learning and machine learning for brain tumor detection: A review, challenges, and future directions. Archives of Computational Methods in Engineering. https://doi.org/10.1007/s11831-025-10416-3
24. Naeem, A. B., Senapati, B., & Zaidi, A. (2025). Enhancing brain tumor detection from MRI-based images through deep transfer learning models. AI, 6(12), 305. https://doi.org/10.3390/ai6120305
25. Noor, A., *et al*. (2020). Hybrid image fusion method based on discrete wavelet transform (DWT), principal component analysis (PCA) and guided filter. Proceedings of the International Conference on Smart Systems and Emerging Technologies (SMARTTECH), 138–143.
26. Raghuram, B., & Hanumanthu, B. (2023). Brain tumor image identification and classification on the internet of medical things using deep learning. Measurement: Sensors, 30, 100905.
27. Ramaraj, V., Swamy, M. V. A., & Sankar, M. K. (2024). Medical image fusion for brain tumor diagnosis using effective discrete wavelet transform methods. Journal of Information Systems Engineering & Business Intelligence, 10(1).
28. Salman, B., Yassin, E., Ganta, D., & Luna, H. (2025). Deep learning-based fusion of multimodal MRI features for brain tumor detection. Applied Sciences, 15(24), 13155. https://doi.org/10.3390/app152413155
29. Sridevi, S. (2021). Image fusion algorithm for medical images using DWT and SR. Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS), 853–858.
30. Srikanth, M. V., Kethavath, S., Yerram, S., Kalli, S., & Naik, J. B. (2024). Brain tumor detection through image fusion using cross guided filter and convolutional neural network. ECTI Transactions on Computer and Information Technology, 18(4), 579–590.
31. Sun, M., Zhu, X., Niu, Y., Li, Y., & Wen, M. (2024). Multi-modal medical image fusion approach utilizing gradient domain guided image filtering. Current Medical Imaging, 20(1), e15734056325441.
32. Usha, M. P., Kannan, G., & Ramamoorthy, M. (2024). Multimodal brain tumor classification using convolutional Tumnet architecture. Behavioural Neurology, 2024, 4678554.
33. Vankdothu, R., & Hameed, M. A. (2022). Brain tumor MRI images identification and classification based on the recurrent convolutional neural network. Measurement: Sensors, 24, 100412.
34. Zebari, N. A., Mohammed, C. N., Zebari, D. A., Mohammed, M. A., Zeebaree, D. Q., Marhoon, H. A., ... & Martinek, R. (2024). A deep learning fusion model for accurate classification of brain tumours in magnetic resonance images. CAAI Transactions on Intelligence Technology, 9(4), 790–804.
35. Zhang, M. (2025). Multimodal fusion at three tiers: Physics-driven data generation and vision-language guidance for brain tumor segmentation. arXiv. https://arxiv.org/abs/2507.09966
36. Zhou, Y., Yang, X., Liu, S., & Yin, J. (2024). Multimodal medical image fusion network based on target information enhancement. IEEE Access, 12, 70851–70869.
37. Zubair, M., Hussain, M., Albashrawi, M. A., Bendechache, M., & Owais, M. (2025). A comprehensive review of multimodal medical image fusion techniques, algorithms, advancements, and clinical applications. Computer Methods and Programs in Biomedicine, 109014. https://doi.org/10.1016/j.cmpb.2025.109014