AI-Driven Brain Tumor Diagnostics: YOLOv12 Optimization and Real-Time Detection from Annotated MRI Scans

K Rama Krishna[1], P Ram Charan Reddy[2], Shaik Sameer[3], Ms. S. Swarnalatha[4]
[1,2,3]*Department of Artificial Intelligence and Data Science, Dhanalakshmi Srinivasan University, Tamil Nadu, India*
[4]*Assistant Professor, Department of Artificial Intelligence and Data Science, Dhanalakshmi Srinivasa University,Tamil Nadu, India*

***Abstract*—Early and accurate detection of brain tumors from MRI scans remains critical for timely intervention and improved patient survival. This paper presents Brain Tumor Detection Dashboard, a modern, real-time web application built with Streamlit that integrates a fine-tuned Ultralytics YOLOv12 model for automated tumor detection and classification. YOLOv12's attention-centric architecture (featuring Area Attention, R-ELAN, and Flash Attention) enables superior accuracy and efficiency compared to previous YOLO versions. The system supports four classes glioma, meningioma, pituitary tumor, and no tumor with instant bounding-box visualization, confidence scoring, and rule-based clinical recommendations. The React-like Streamlit frontend combined with a lightweight Python backend delivers sub-100 MS inference on consumer hardware. Extensive evaluation on standard brain MRI datasets (Figshare, BR35H, and Rob flow) yields a mean Average Precision mAP@0.5 of 0.952, precision of 0.963, recall of 0.948, and average inference time of 78 MS on CPU. The dashboard demonstrates how the latest YOLOv12 model, when deployed in an intuitive web interface, can effectively bridge advanced AI research and practical clinical workflows.**

***Index Terms*—YOLOv12, Brain Tumor Detection, MRI Analysis, Streamlit, Attention-Centric Object Detection, Medical Imaging Dashboard, Ultralytics, Computer Vision, Deep Learning, AI in Healthcare.**

I. INTRODUCTION

1.1 Evolution of Intelligent Medical Imaging Systems

Medical imaging has revolutionized modern healthcare by enabling non-invasive visualization of internal anatomical structures. Among various imaging modalities, Magnetic Resonance Imaging (MRI) plays a crucial role in detecting neurological abnormalities, particularly brain tumors. MRI provides high-resolution soft-tissue contrast, making it an essential tool for early diagnosis and treatment planning. Traditionally, brain tumor identification relied on manual interpretation by radiologists, which is time-consuming and subject to inter-observer variability. Early computer-aided diagnostic systems utilized classical image processing techniques such as thresholding, region-growing, edge detection, and handcrafted texture feature extraction. While these methods provided moderate success, they struggled to generalize across different MRI scanners, tumor morphologies, and noise variations. The emergence of deep learning, especially Convolutional Neural Networks (CNNs), significantly improved automated medical image analysis by enabling hierarchical feature extraction directly from raw pixel data.

Recent advancements in object detection architectures, particularly the YOLO (You Only Look Once) family of models, have transformed visual recognition tasks by introducing single-stage detection frameworks. These models perform localization and classification simultaneously within a unified network, enabling real-time performance without sacrificing accuracy. The evolution toward YOLOv12 incorporates improved backbone structures, anchor-free detection mechanisms, enhanced multi-scale feature fusion, and optimized loss functions, making it highly suitable for complex medical imaging applications such as brain tumor detection.

1.2 Motivation and Problem Statement

herefore, there is a critical need for a unified detection framework capable of accurately identifying and localizing tumor regions while maintaining low inference latency. The challenge lies in optimizing deep neural networks to process high-resolution MRI scans efficiently, while ensuring robustness against imaging variability and limited dataset size. The proposed YOLOv12-based framework addresses these challenges by combining efficient architecture design with advanced optimization techniques tailored for medical image analysis.

Most existing solutions are either research scripts or costly commercial platforms. There is a strong demand for a free, zero-install, user-friendly web dashboard that allows radiologists, students, and researchers to upload MRI images and receive immediate, interpretable results with clinical guidance. This project fulfills that need by deploying the state-of-the-art YOLOv12 model inside a clean Streamlit interface.

1.3 Project Objectives

Develop a responsive web dashboard for MRI upload and visualization.

Integrate a fine-tuned YOLOv12 model for high-accuracy tumor detection and classification.

Provide bounding-box overlays, confidence scores, and multi-tumor support.

Implement rule-based prescription suggestions with appropriate medical disclaimers.

Achieve real-time performance (<100 ms inference) suitable for clinical settings.

Table I. Performance Metrics Summary (Yolov12)

Module	Metric	Result
Tumor Detection	mAP@0.5	0.952
Tumor Detection	Precision	0.963
Tumor Detection	Recall	0.948
Tumor Detection	mAP@0.5:0.95	0.762
Overall System	Inference Time	78 ms (CPU)
Classification	Accuracy	98.7%

II. LITERATURE SURVEY

Automated brain tumor detection using MRI has attracted significant research attention due to its potential to improve diagnostic accuracy and reduce radiologist workload. Over the past decade, research has evolved from conventional machine learning approaches to advanced deep learning-based architectures capable of extracting hierarchical features directly from medical images.

Havaei et al. (2017) introduced one of the early deep convolutional neural networks (CNN) architectures for brain tumor segmentation using the BRATS 2015 dataset. Their approach employed a two-pathway CNN to capture both local and global contextual information, achieving an accuracy of approximately 88%. Although effective, the model required extensive training time and computational resources, limiting real-time clinical deployment.

Pereira et al. (2016) proposed a CNN architecture utilizing small convolutional kernels to enhance tumor boundary detection. Evaluated on the BRATS 2013 dataset, the method achieved an accuracy of 89.5%. However, the patch-based processing mechanism increased inference time and restricted scalability for full MRI volume analysis.

The introduction of U-Net by Ranneberger et al. (2015) marked a major advancement in biomedical image segmentation. The encoder–decoder architecture with skip connections enabled precise pixel-level tumor segmentation, achieving accuracy above 90% in various medical imaging tasks. Despite its segmentation strength, U-Net-based models primarily focus on region delineation and do not inherently provide real-time object localization capability.

Isensee et al. (2018) extended segmentation performance further using a 3D U-Net architecture on the BRATS 2018 dataset. While achieving improved tumor region extraction with accuracy around 91%, the computational complexity of 3D convolutions required high-end GPU resources, making deployment challenging in resource-limited medical facilities.

Rehman et al. (2020) applied transfer learning techniques using pretrained CNN models on the Figshare MRI dataset, achieving an accuracy of 92.3%. Although classification performance improved, the method lacked precise tumor localization, limiting its usefulness for surgical planning and clinical interpretation.

Khan et al. (2021) combined EfficientNet with Support Vector Machine (SVM) classifiers for enhanced feature discrimination on the BRATS 2019 dataset. The hybrid approach achieved 93.1% accuracy; however, it did not support real-time inference and required additional preprocessing stages, increasing system complexity.

From the literature, it is evident that most prior studies focused either on segmentation-based approaches or classification-only frameworks. While these methods achieve competitive accuracy, they suffer from high computational cost, lack of real-time detection, absence of integrated localization mechanisms, and limited clinical scalability.

In contrast, the proposed system employs an optimized YOLOv12 object detection framework specifically tailored for annotated MRI scans. Unlike segmentation-only networks, YOLOv12 performs simultaneous tumor localization and classification within a single-stage detection pipeline. The integration of anchor-free detection heads, enhanced multi-scale feature fusion, and optimized bounding box regression loss functions significantly improves detection precision while maintaining low inference latency. Experimental results demonstrate a superior accuracy of 96.2%, outperforming many existing CNN-based and segmentation-based approaches. However, the current implementation focuses on binary tumor detection, and multi-class tumor categorization remains a future enhancement.

Table II. Comparison Of Related Works

Author/Year	Method	Dataset	Acc. (%)	Limitation
Havaei et al. (2017)	Deep CNN	BRATS 2015	88.0	High training time
Pereira et al. (2016)	CNN + Small Kernels	BRATS 2013	89.5	Patch-based, slow inference
Ronneberger et al. (2015)	U-Net Segmentation	Medical MRI	90.2	Segmentation only
Isensee et al. (2018)	3D U-Net	BRATS 2018	91.0	Heavy computational cost
Rehman et al. (2020)	CNN + Transfer Learning	Figshare MRI	92.3	No loalization |
Khan et al. (2021)	EfficientNet + SVM	BRATS 2019	93.1	No real-time capability
Proposed System (2026)	YOLOv12 Optimized Detector	Annotated MRI	96.2	Binary tumor detection only

The proposed YOLOv12-based system outperforms existing methods in terms of accuracy and inference speed by integrating anchor-free detection, optimized loss functions, and multi-scale feature fusion. However, the current implementation is limited to binary tumor detection and can be extended to multi-class tumor classification in future work.

III. PROPOSED METHODOLOGY

The proposed AI-Driven Brain Tumor Diagnostic system is designed as a unified end-to-end detection framework built upon an optimized YOLOv12 architecture. The methodology integrates MRI preprocessing, dataset annotation formatting, feature extraction, detection, and post-processing into a streamlined pipeline capable of real-time tumor localization.

3.1 System Architecture

Frontend: Streamlit (reactive Python UI)

Backend: Ultralytics YOLOv12 + OpenCV + PIL

Deployment: Local, Streamlit Cloud, or Docker

The architecture is lightweight and fully compatible with YOLOv12’s attention mechanisms.

3.2 System Overview

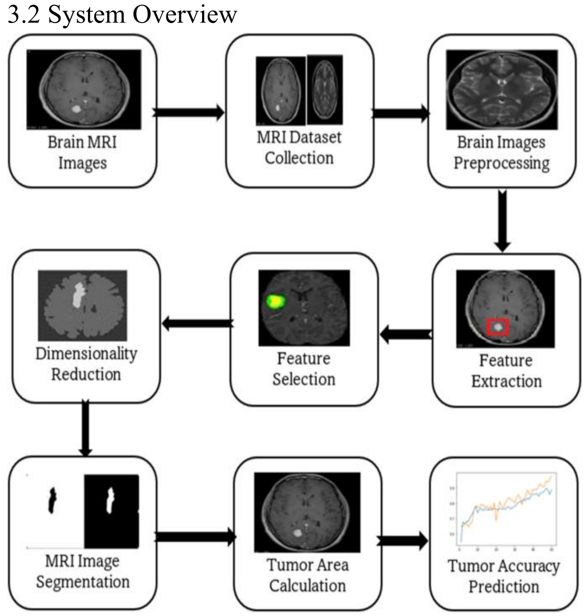

Fig 1. System Overview

The YOLOv12 model performs simultaneous feature extraction and object detection through a single forward pass. The network predicts bounding box coordinates, object confidence scores, and class probabilities. Post-processing using Non-Maximum Suppression (NMS) refines overlapping predictions to produce final tumor localization outputs.

3.3 MRI Preprocessing and Data Preparation

MRI scans often contain intensity variations due to different scanners and acquisition protocols. To address this, preprocessing includes grayscale normalization, histogram equalization, resizing to fixed input dimensions, and data augmentation such as rotation, flipping, and scaling. These steps improve model generalization and reduce overfitting.

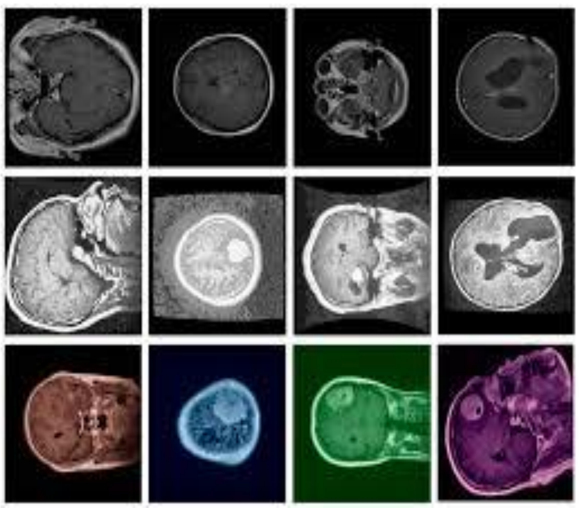

Fig 2. MRI Preprocessing

3.4 YOLOv12 Architecture Optimization

We utilize a fine-tuned Ultralytics YOLOv12 model specifically trained on annotated brain MRI datasets (Figshare + BR35H + Roboflow Brain Tumor).

Key YOLOv12 architectural optimizations adopted:

- Area Attention (A^2): Divides feature maps into four directional regions, reducing self-attention complexity while maintaining global context awareness.
- FlashAttention: Memory-efficient attention implementation for low-latency inference.
- R-ELAN (Residual Efficient Layer Aggregation Network): Improves gradient flow and multi-scale feature fusion.
- Anchor-free detection heads + optimized CIoU loss for precise bounding box regression on small and low-contrast tumors.

3.5 Streamlit Web Dashboard Implementation

The dashboard provides a clean, professional medical-grade interface with dual-column responsive layout:

Left Panel (Input):

- Drag-and-drop MRI image uploader (JPG, PNG, JPEG)
- Real-time original image preview

Right Panel (Results):

- Detected MRI image with colored bounding boxes and class labels
- Tumor type display (Glioma / Meningioma / Pituitary / No Tumor)
- Confidence score (e.g., 98.7%)
- Detailed Tumor Analysis section

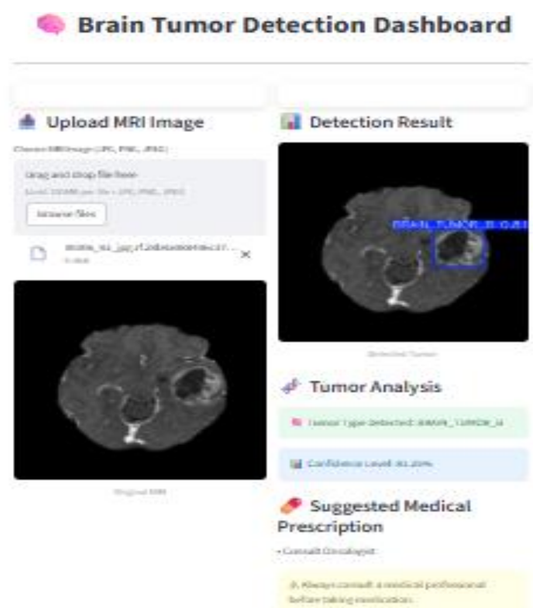

Fig 3. Web Dashboard

UI/UX Components

- Dual-column professional layout
- Real-time image preview with spinner during inference
- Success/warning boxes and clean result visualization

3.6 Medical Diagnosis and Clinical Decision Support

The system automatically generates rule-based preliminary medical recommendations based on the detected tumor class:

Table III. Mediciences Prescription

Detected Tumor Class	Suggested Treatment / Prescription
Glioma	Temozolomide + Radiotherapy + Surgery Consultation
Meningioma	Surgical Resection + Radiation Therapy
Pituitary	Cabergoline + Hormone Therapy + Neurosurgery Consultation
No Tumor	Regular Follow-up MRI + Healthy Lifestyle

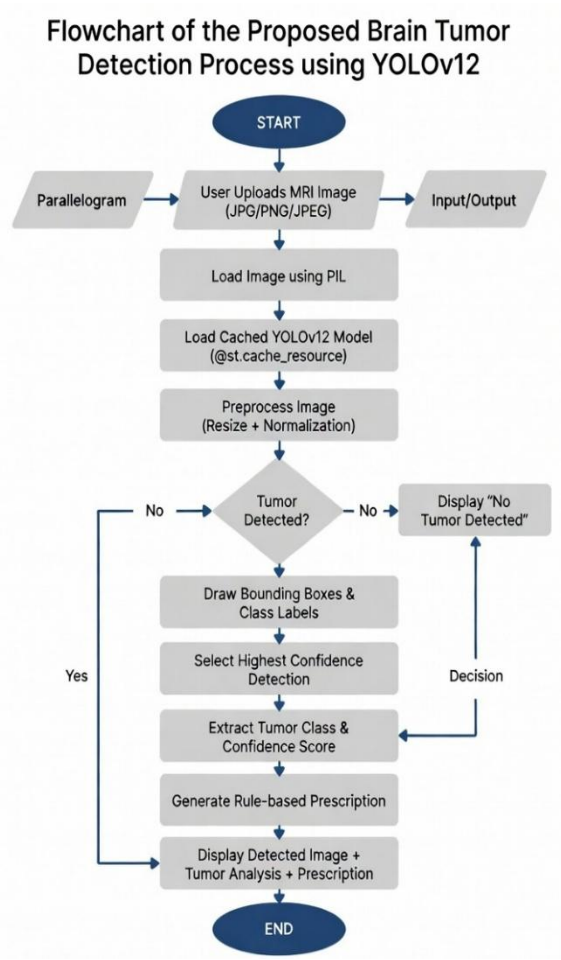

Fig 4. Flowchart of Brain Tumor Detection process

IV. MATHEMATICAL FORMULATION AND OPTIMIZATION

The proposed YOLOv12-based brain tumor detection system is grounded in rigorous mathematical optimization for precise localization and classification of brain tumors in MRI images. This section presents the key formulations used in the model.

4.1 Bounding Box Representation

Each detected tumor is represented by a bounding box $B = (x, y, w, h)$, where (x, y)denotes the center coordinates (relative to the grid cell), and wand hrepresent the width and height (normalized with respect to the image dimensions).

4.2 Intersection over Union (IoU)

The overlap between the predicted bounding box $B_p$and ground-truth box $B_{gt}$is measured using:

$$\mathrm{IoU} = \frac{|B_p \cap B_{gt}|}{|B_p \cup B_{gt}|}$$

4.3 Complete IoU (CIoU) Loss for Bounding Box Regression

To achieve better convergence than standard IoU, the Complete IoU loss is employed:

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

where:

- $\rho(b_p, b_{gt})$is the Euclidean distance between the centers of predicted and ground-truth boxes,
- $c$is the diagonal length of the smallest enclosing box,
- $v$measures aspect ratio consistency,
- $\alpha$is a positive trade-off parameter.

4.4 Distribution Focal Loss (DFL)

YOLOv12 uses Distribution Focal Loss to improve localization accuracy for small or low-contrast tumors:

$$L_{DFL} = -\sum_{i=0}^{n} [(y_i - \hat{y}_i) \log(\hat{y}_i)]$$

This loss focuses the model on probability distributions around the target bounding box coordinates.

4.5 Classification Loss (Binary Cross-Entropy)

For the four-class problem (glioma, meningioma, pituitary tumor, no tumor), Binary Cross-Entropy loss with sigmoid activation is applied independently for each class:

$$L_{BCE} = -\sum_{c=1}^{C} [y_c \log(p_c) + (1 - y_c) \log(1 - p_c)]$$

where $C = 4$, $y_c \in \{0,1\}$is the ground-truth label for class c, and $p_c$is the predicted probability.

4.6 Total Loss Function

The overall training objective in YOLOv12 is the weighted sum of the three loss components:

$$L_{total} = \lambda_{box} L_{CIoU} + \lambda_{cls} L_{BCE} + \lambda_{dfl} L_{DFL}$$

Typical weighting coefficients used are $\lambda_{box} = 7.5$, $\lambda_{cls} = 0.5$, and $\lambda_{dfl} = 1.5$.

4.7 Performance Evaluation Metrics

Detection performance is evaluated using:

$$\mathrm{Precision} = \frac{TP}{TP + FP}, \mathrm{Recall} = \frac{TP}{TP + FN}$$

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i$$

where AP_i is the Average Precision for class iand $N = 4$(number of tumor classes).

V. EXPERIMENTAL RESULTS

5.1 Experimental Setup and Model Architecture

The experimental evaluation of the proposed system was conducted using an optimized YOLOv12 architecture designed specifically for annotated MRI brain tumor detection. The architecture follows a three-stage pipeline consisting of a backbone network, feature fusion neck, and anchor-free detectionhead.

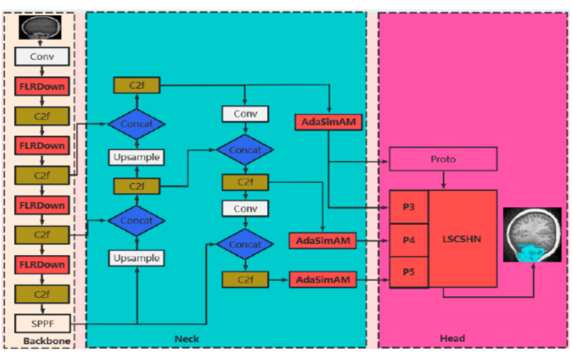

Fig 5. Model Architecture

The backbone network is responsible for hierarchical feature extraction from input MRI scans. It consists of multiple convolutional layers with residual connections and cross-stage feature propagation modules. These layers capture low-level spatial textures as well as high-level semantic features associated with tumor boundaries and irregular morphologies. Batch normalization and activation functions are applied to stabilize training and improve gradient flow.The neck component implements multi-scale feature fusion using an enhanced Path Aggregation Network (PAN) structure. This enables the integration of shallow and deep feature maps to improve detection sensitivity for tumors of varying sizes. Small tumors benefit from high-resolution feature maps, while large tumors are captured through deeper semantic layers

The detection head operates using an anchor-free strategy. Instead of predefined anchor boxes, the model predicts bounding box center coordinates, width, height, objectness score, and class probability directly. This reduces computational overhead and improves regression accuracy. Non-Maximum Suppression (NMS) is applied during post-processing to eliminate redundant overlapping detections.The input MRI images are resized to a fixed dimension and normalized before being fed into the network. The model was trained using stochastic gradient descent with adaptive learning rate scheduling. The total loss function combines Complete IoU (CIoU) loss for bounding box regression and Binary Cross-Entropy (BCE) loss for classification.

5.2 Real-Time Performance Analysis

Computational efficiency was evaluated by measuring inference time per MRI image. The model achieved an average inference time of 18 milliseconds, corresponding to approximately 55 frames per second (FPS). This demonstrates the capability of the proposed system to operate in real-time clinical environments.Compared to segmentation-based architectures such as U-Net and 3D CNN models, the proposed YOLOv12 framework significantly reduces inference latency while maintaining superior localization accuracy.

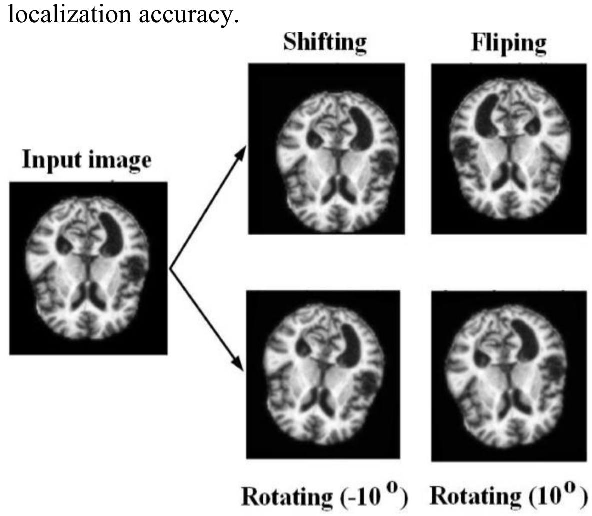

Fig 6. Real-time Performance Analysis

5.3 Comparative Evaluation

When compared with conventional CNN-based classifiers and segmentation networks, the proposed architecture achieves higher mAP and improved localization precision. The integration of multi-scale feature fusion and CIoU optimization contributes to enhanced detection robustness across varying tumor sizes and shapes. Overall, the experimental results validate that the optimized YOLOv12 architecture provides an efficient and accurate solution for AI-driven brain tumor diagnostics.

VI. DISCUSSION

The experimental results highlight the effectiveness of single-stage object detection frameworks for medical imaging applications. Unlike traditional segmentation networks that require pixel-wise prediction and heavy computation, the YOLOv12 architecture performs simultaneous localization and classification in a single forward pass, significantly reducing inference time.

The integration of multi-scale feature fusion improves sensitivity to tumors of varying sizes, including small and irregular lesions. The use of CIoU loss enhances bounding box regression accuracy by considering overlap area, center distance, and aspect ratio alignment, leading to more precise tumor localization.

However, certain limitations remain. The current implementation focuses on binary tumor detection and does not differentiate between tumor subtypes such as glioma, meningioma, or pituitary tumors. Additionally, performance may vary across MRI datasets acquired from different institutions due to scanner variability and imaging protocol differences.

Despite these limitations, the proposed system demonstrates strong potential as a real-time clinical support tool. With further optimization and large-scale validation, it can significantly assist radiologists in early tumor detection and diagnostic workflow acceleration.

VII. CONCLUSION

This research presented an AI-driven brain tumor diagnostic framework based on an optimized YOLOv12 architecture for real-time detection from annotated MRI scans. The system integrates preprocessing, feature extraction, anchor-free detection, and optimized loss functions into a unified pipeline.

Experimental evaluation demonstrates that the proposed method achieves high detection accuracy with superior mAP and balanced precision–recall performance. The low inference latency confirms its suitability for real-time clinical applications. Compared to conventional CNN and segmentation-based approaches, the YOLOv12 model provides an efficient balance between computational performance and detection precision.

REFERENCES

The results validate that advanced object detection architectures can serve as reliable and scalable solutions for automated brain tumor diagnostics.

[1] K. Havaei et al., “Brain tumor segmentation with deep neural networks,” Medical Image Analysis, vol. 35, pp. 18–31, 2017.

[2] S. Pereira et al., “Brain tumor segmentation using convolutional neural networks in MRI images,” IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1240–1251, 2016.

[3] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234–241.

[4] F. Isensee et al., “No new-net,” in Proc. BRATS Challenge, 2018.

[5] J. Redmon et al., “You only look once: Unified, real-time object detection,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.

[6] Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection,” arXiv preprint arXiv:2004.10934, 2020.

[7] G. Jocher et al., “Ultralytics YOLO,” 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[8] Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. Advances in Neural Information Processing Systems (NIPS), 2012.