Brain Tumor Analysis System

AI-Powered Segmentation and Clinical Interpretation

Version 1.0

Contents

1 Introduction

The Brain Tumor Analysis System is an advanced AI-driven framework that integrates **YOLOv11-segmentation** for precise tumor detection in brain MRI scans with **Retrieval-Augmented Generation (RAG)** powered by **Mistral-8x7B** (via Groq API). The system generates a detailed PDF report containing a comprehensive clinical analysis and tailored treatment recommendations for detected brain tumors, leveraging a vectorized medical knowledge base grounded in peer-reviewed literature. By combining computer vision and natural language processing, the system automates tumor segmentation, provides clinically relevant insights, and produces professional reports to assist radiologists, researchers, and medical educators. The output is designed for preliminary analysis and requires validation by qualified medical professionals to ensure clinical accuracy. This document details the system's development methodology, research design, limitations, and supporting literature, providing a robust foundation for its application in medical research and clinical support.

2 Problem Statement

Brain tumors represent a critical medical challenge due to their diverse histopathological profiles, complex imaging characteristics, and significant impact on patient outcomes. Manual segmentation of tumors in MRI scans is labor-intensive, subject to inter-observer variability, and requires extensive expertise, often taking hours per scan. Additionally, translating segmentation results into actionable clinical insights demands integration of vast medical knowledge, which is time-consuming and prone to inconsistencies. Existing automated systems, such as those based on U-Net or earlier YOLO models, achieve high segmentation accuracy but lack comprehensive clinical interpretation or treatment recommendations. The need for an integrated system that combines accurate tumor segmentation with detailed, evidence-based clinical analysis and treatment planning is evident. This project addresses these challenges by developing a pipeline that leverages YOLOv11 for segmentation and RAG-based LLM for generating detailed PDF reports, enhancing diagnostic efficiency and supporting clinical decision-making in neuro-oncology.

3 Methodology

3.1 Research Design

The research adopts a quantitative, experimental approach to design and evaluate an AI-driven pipeline for brain tumor analysis. The system combines **YOLOv11-segmentation** for tumor detection with **RAG** and **Mistral-8x7B** for clinical interpretation, producing a detailed PDF report. The design involves iterative development, including data preprocessing, model training, validation, and performance evaluation against medical benchmarks. The pipeline is modular, allowing for future enhancements such as 3D MRI support or hospital system integration. The research evaluates segmentation accuracy using metrics like Intersection over Union (IoU) and clinical relevance of LLM outputs against established neuro-oncology guidelines, ensuring alignment with clinical standards.

3.2 Data Collection

The dataset comprises brain MRI scans from publicly available repositories, including the Brain Tumor Segmentation (BraTS) dataset (Menze et al., 2015) and The Cancer Imaging Archive (TCIA) (Clark et al., 2013). These datasets include T1-weighted, T2-weighted, and FLAIR MRI images in `.jpg` or `.png` format, accompanied by ground-truth annotations for tumor boundaries. Approximately 10,000 images are collected, covering diverse tumor types (e.g., gliomas, meningiomas, metastases) across adult patients. Peer-reviewed medical literature, including papers on tumor classification and treatment protocols (Louis et al., 2021; Weller et al., 2021; Bakas et al., 2017), is gathered to populate the RAG knowledge base, ensuring evidence-based clinical analysis.

3.3 Data Preprocessing

MRI images are preprocessed to ensure compatibility with YOLOv11. Images are resized to 640x640 pixels using bilinear interpolation to preserve quality. Intensity normalization standardizes pixel values to a 0–1 range across imaging modalities. Ground-truth annotations are converted to polygon coordinates in YOLO format (normalized x, y coordinates). Data augmentation techniques, including rotation, flipping, and brightness adjustment, are applied to enhance dataset diversity and model robustness. Medical literature is extracted using `pdfplumber`, segmented into 500-word chunks, and embedded using `sentence-transformers/all-MiniLM-L6-v2` for storage in ChromaDB, enabling efficient retrieval for RAG.

3.4 Model Development

The system is developed as a modular pipeline with four core components:

- **Segmentation**: YOLOv11-segmentation (Ultralytics) detects and delineates tumors, outputting binary masks and polygon coordinates. OpenCV is used for post-processing, generating visualizations with a green mask and red outline.
- **Knowledge Base**: ChromaDB stores vectorized medical literature, enabling RAG to retrieve relevant context (e.g., tumor characteristics, treatment protocols).
- **LLM Interpretation**: Mistral-8x7B (via Groq API) integrates segmentation data and retrieved context to generate a detailed clinical analysis, including tumor type, symptoms, and treatment recommendations.
- **PDF Generator**: ReportLab compiles segmented images and LLM summaries into a professional PDF report, formatted for clinical and research use.

3.5 Model Training

YOLOv11 is fine-tuned on the preprocessed MRI dataset using a pre-trained model (`yolov11.pt`). Training parameters include a batch size of 16, learning rate of 0.001, momentum of 0.9, and 100 epochs, optimized with the Adam optimizer. Training is conducted on a GPU-enabled system (e.g., NVIDIA RTX 3080) to handle computational demands, with early stopping applied if validation loss plateaus for 10 epochs. The RAG pipeline leverages pre-trained embeddings from `sentence-transformers`, requiring no additional training, but the knowledge base is periodically updated with new literature to enhance LLM performance.

3.6 Model Evaluation

YOLOv11 performance is assessed using standard segmentation metrics:

- **Intersection over Union (IoU)**: Measures overlap between predicted and ground-truth masks (target: IoU > 0.85).
- **Mean Average Precision (mAP@0.5)**: Evaluates detection accuracy at IoU threshold of 0.5 (target: mAP > 0.90).
- **Dice Coefficient**: Assesses segmentation similarity (target: Dice > 0.87).

LLM outputs are evaluated for factual accuracy by comparing clinical analysis and treatment recommendations to peer-reviewed guidelines (e.g., Louis et al., 2021; Weller et al., 2021). A test set of 500 images is used for quantitative evaluation, with qualitative review by neuro-oncology experts to ensure clinical relevance.

3.7 Validation

The system is validated using a hold-out test set (20% of the dataset, approximately 2,000 images) to assess generalizability across tumor types, imaging modalities, and patient demographics. Five-fold cross-validation is performed to ensure robust segmentation performance. LLM-generated reports are validated against WHO tumor classifications (Louis et al., 2021) and EANO treatment guidelines (Weller et al., 2021), with expert feedback incorporated to refine report content and structure.

3.8 Ethical Considerations

The system adheres to ethical guidelines to ensure responsible development and deployment:

- **Non-Diagnostic Purpose**: The system is explicitly designed for research and preliminary analysis, with a disclaimer requiring radiologist validation to prevent misuse in clinical settings.
- **Data Privacy**: Only de-identified, publicly available datasets (e.g., BraTS, TCIA) are used, complying with ethical standards such as HIPAA.
- **Transparency**: Reports include a clear disclaimer emphasizing AI-assisted analysis and the need for professional review.
- **Bias Mitigation**: Training data is balanced across tumor types, ages, and genders, with regular audits to minimize algorithmic bias.
- **Equitable Access**: The system is designed to be open-source, promoting accessibility for research communities worldwide.

4 Limitation and Scope

4.1 Limitations

- **Clinical Use**: The system is not intended for direct clinical diagnosis and requires validation by a qualified radiologist.
- **Data Dependency**: Segmentation accuracy depends on the quality and diversity of the training dataset, which may not fully represent rare tumor subtypes.
- **2D Limitation**: The system processes 2D MRI slices, limiting its ability to analyze 3D tumor volumes.
- **LLM Accuracy**: The quality of clinical analysis relies on the completeness of the RAG knowledge base, which may lack data on emerging tumor types.
- **Computational Resources**: Training and inference require high-performance GPUs, potentially limiting accessibility in low-resource settings.

4.2 Scope of Research

The research focuses on developing an AI-driven system for 2D brain MRI tumor segmentation and clinical interpretation, targeting applications in medical research and clinical support. The system integrates YOLOv11-segmentation and Mistral-8x7B with RAG to produce detailed

PDF reports containing comprehensive clinical analysis and treatment recommendations. Future scope includes extending support to 3D MRI volumes, integrating with hospital Picture Archiving and Communication Systems (PACS), and incorporating real-time updates to the knowledge base for enhanced clinical relevance.

5 Literature Review

Brain tumor segmentation has seen significant advancements through deep learning. U-Net (Ronneberger et al., 2015) introduced a convolutional architecture for medical image segmentation, achieving IoU scores of 0.7–0.85 on brain MRI datasets. YOLO variants, including YOLOv8 and YOLOv11, have improved real-time segmentation performance, with mAP scores exceeding 0.9 for object detection tasks (Jocher et al., 2023). Retrieval-Augmented Generation (RAG) enhances LLM performance in knowledge-intensive tasks by retrieving relevant context, critical for medical applications (Lewis et al., 2020). The WHO classification of CNS tumors provides a standardized framework for tumor grading (Louis et al., 2021), while EANO guidelines outline multimodal treatment strategies for gliomas, including surgery, radiotherapy, and chemotherapy (Weller et al., 2021). The BraTS dataset (Menze et al., 2015; Bakas et al., 2017) serves as a benchmark for segmentation models, offering annotated MRI scans for gliomas and other tumors. Recent work by Havaei et al. (2017) demonstrates the efficacy of deep learning in multi-modal MRI analysis. This project advances prior work by integrating YOLOv11-segmentation with RAG-based LLM interpretation, addressing gaps in automated clinical analysis and treatment recommendation generation.

6 Timeline

Phase	**Description**
Month 1	Collect MRI datasets (BraTS, TCIA) and peer-reviewed literature
Month 2	Preprocess images and annotations; build and populate RAG knowledge base
Month 3–4	Train and validate YOLOv11 model for tumor segmentation
Month 5	Develop RAG pipeline and integrate Mistral-8x7B for clinical analysis
Month 6	Implement ReportLab for detailed PDF report generation
Month 7	Evaluate system performance (IoU, mAP, clinical accuracy) and refine components
Month 8	Conduct final validation with expert feedback; complete documentation

Table 1: Project Timeline

7 References

- Bakas, S., et al. (2017). Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data, 4*, 170117.
- Clark, K., et al. (2013). The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging, 26*(6), 1045–1057.

- Havaei, M., et al. (2017). Deep learning trends for focal brain pathology segmentation in MRI. *Machine Learning for Health Informatics*, 125–148.
- Jocher, G., et al. (2023). YOLO: A Decade of Object Detection and Segmentation. *arXiv preprint arXiv:2304.00501.*
- Lewis, P., et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *NeurIPS 2020.*
- Louis, D. N., et al. (2021). The 2021 WHO Classification of Tumors of the Central Nervous System. *Neuro-Oncology, 23*(8), 1231–1251.
- Menze, B. H., et al. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BraTS). *IEEE Transactions on Medical Imaging, 34*(10), 1993–2024.
- Ronneberger, O., et al. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *MICCAI 2015.*
- Weller, M., et al. (2021). EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nature Reviews Clinical Oncology, 18*(3), 170–186.