National College of Ireland

Enhancing Brain Tumor Detection with Deep Learning Models: A Comparative Analysis

MSc Research Project
MSCDAD_A_JAN24I

Asish Mathai Mathai
Student ID: x23173645

School of Computing
National College of Ireland

Supervisor: Vikas Tomer

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Asish Mathai Mathai
Student ID:	x23173645
Programme:	MSCDAD_A_JAN24I
Year:	2024
Module:	MSc Research Project
Supervisor:	Vikas Tomer
Submission Due Date:	30/01/2025
Project Title:	Enhancing Brain Tumor Detection with Deep Learning Models: A Comparative Analysis
Word Count:	9714
Page Count:	27

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Asish Mathai Mathai
Date:	28th January 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Enhancing Brain Tumor Detection with Deep Learning Models: A Comparative Analysis

Asish Mathai Mathai
x23173645

Abstract

Medical imaging for the detection of brain tumors has lately been significantly enhanced by the inclusion of deep learning technologies. However, detailed comparative research into the clinical practicality of state-of-the-art models has been scant. This is a comparative study of three deeper architectures of convolutional neural networks, namely YOLOv9, PaliGemma, and Detectron2, in detecting brain tumors based on their performance metrics concerning accuracy, speed of processing, computational efficiency, and clinical applicability. Implementing and testing each model with the standardized protocols in this work, the dataset used contained 8,903 brain MRI images covering four categories: with no tumor, meningioma, pituitary, and glioma. The results obtained indicate that YOLOv9 topped with an mAP50 of 0.958 and mAP50-95 of 0.78, leveraging far in front of any results obtained with Detectron2 with a mAP50 of 0.698, and PaliGemma at a mAP50 of 0.482. Although PaliGemma introduces a very unique approach of vision-language, its mediocre performance indicates that domain-specific optimizations are required. Detection performance: Overall, Detectron2 has shown great specialisation capability. It seems particularly good in detecting meningioma (76.86% AP). These findings give evidence-based recommendations for the choice of models toward specific clinical needs and further contribute to the advance of AI-assisted medical imaging. The study illustrates the potential clinical implementation of YOLOv9 while pointing toward a future direction for hybrid architectures that model strengths of traditional object detection with advanced language understanding relative to specific tasks. This comprehensive evaluation lays the ground for further developments in automated brain tumor detection systems.

1 Introduction

The advancement of medical imaging technology has transformed health diagnostics, especially regarding the most critical domain: the detection of brain tumors. Recently, deep learning architectures have shown great promise in enhancing the effectiveness and speed of the process of tumor detection. This current study performs a comparative analysis between three deep learning models, at the state-of-the-art, namely the YOLOv9, PaliGemma model, and Detectron2, on the task of brain tumor detection from medical imaging data. Among various medical image analyses, the detection and diagnosis of brain tumors are some of the most challenging tasks and usually involve great consequences in patient outcomes. Conventional detection relies heavily on expert radiologists' interpretation, which is often time-consuming and prone to human variation. Hence, according

to the views of Solanki et al. (2023), such a strategy of embedding deep learning technologies has come forth as a very hopeful answer to these issues, making automated, consistent, and fast detection of tumors possible. With growing diversity and complexity of the models available for deep learning, another challenge has emerged: which architectural approach will serve the specific needs associated with clinical brain tumor detection. The motivation for this study is huge and lies in the identification of a critical gap in the current literature. Works were done by several authors to prove the efficiency of different single deep learning models for medical image analyses; in comparisons of the recently performed architectures are still very few. According to Mahmud et al.(2023) the validation and performance assessment of general deep learning models involve multiple dimensions on the medical imaging context beyond mere assessment on accuracy. This research has been done to fill this gap through a rigorous critical analysis of three of the most recent state-of-the-art models, each representing different approaches to object detection.

The central research question guiding this study asks: **How do the advanced deep learning models - YOLOv9, PaliGemma, and Detectron2 - compare in terms of accuracy, processing speed, computational efficiency, and clinical applicability for brain tumor detection in medical imaging?** In this direction, the research outlines some of the key objectives such as the implementation and optimization of the proposed models for carrying out brain tumor detection, ensuring that conditions for fair comparisons are met; performance evaluation of the model with standardized metrics such as mAP, speed of processing; evaluating each model with respect to requirements and efficiency concerning computational resources; and clinically considering applicability and potentials for integrating such models into real-world medical settings. This comparative study is based on a carefully developed dataset of 8,497 brain MRI images divided into four categories: glioma, meningioma, pituitary, and no tumor conditions. The distribution is balanced in the training set of 7,497 images, with glioma accounting for 30.37% of all images , meningioma for 27.77% , pituitary for 32.13% , and no tumor cases for 15.37% . This considerable set of data, along with careful class balancing in the dataset, has made robust trainability and evaluation capability possible while providing a very critical essence of medical imaging applications, for which characteristics of data quality and representatives directly influence diagnosing reliability.

This research will help both academic and clinical stakeholders. For the academia, it provides a comprehensive comparison framework for several advanced deep learning models in use within medical imaging applications. To the clinicians, evidence-based recommendations on model selection decisions, considering a particular operational requirement and resource limitation, are provided. These will in particular help health facilities make informed decisions on where to install AI-assisted diagnosis systems. Confirmation of the scenarios for reaching research goals required the development of some key tests and metrics of evaluation. These include quantitative measures of model accuracy through mAP, processing speed assessments under different computation conditions, and comprehensive analysis of resource utilization studies. Clinical applicability will be assessed by structured examination of integration requirements, workflow compatibility, and real-world performance scenarios. The rest of the report logically falls into six major sections. Following this introduction, Section 2 resumes related efforts relevant to the deep learning benchmarks within the domain of brain tumor detection, presenting thus this research effort within its theoretical frame. Section 3 describes in detail the methodology regarding research and experimental setup with regard to data

preprocessing, the implementation of models, and protocols for evaluation. Section 4 presents the design specifications, taking architectural aspects into consideration in the optimization strategies. Section 5 discusses implementation details, including problems found and the solutions that were developed. Section 6 presents the results of a complete evaluation and comparison analysis. Finally, Section 7 summarizes the findings, limitations, and recommendations for future research. This structure ensures that the research work provides critical insights into strengths and weaknesses of the deep learning state-of-the-art models so far developed for brain tumor detection, while contributing to the advance of AI-assisted practices in medical diagnosis. Their results will provide the basis for further research to be chronicled in this fast-evolving field, while providing practical guidance on how to clinically implement such technologies.

2 Related Work

Deep learning in medical imaging has been undergoing phenomenal changes; especially, in the detection of brain tumors, it has undergone immense changes in the last ten years. A critical review of the literature is provided in this paper on the evolution from traditional approaches to state-of-the-art architectures. The research reviews related works concerning the YOLOv9, PaliGemma, and Detectron2 comparative studies on brain tumor detection.

2.1 Evolution of Brain Tumor Detection Approaches

The main problems in detecting brain tumors were highlighted by Solanki et al.(2023), where the researchers pointed out the challenge related to tumor variability with regard to position, structure, and proportions. This paper showed that there was a significant gap between machine learning and deep learning techniques, indicating how deep learning demonstrates excellent performance while operating with big volumes of data; at the same time, with limited datasets, which is typical for medical domains, it demonstrates worse performance. This finding is particularly relevant when assessing the practical implementation of advanced architectures in clinical practice. Expanding on this, Birajdar(2023) that it is practically feasible to implement CNN; using real-time processing on embedded systems, he achieved 92.17% accuracy. While their work proved the viability in deploying deep learning models in resource-constrained environments, the relatively modest accuracy with regard to recent approaches indicates significant latitude for improvement.

2.2 Architectural Innovations in Deep Learning Models

There has been a great improvement in model architectures, therefore pushing the limits of detection accuracy. G et al.,(2023), obtained very impressive results with accuracy rates of 98.02% using ResNet50 and 98.32% with Xception architectures. In fact, the extensive comparison of the architectures by them provides useful insight into the strengths of various design choices; however, it also lacks any notable analysis of the computational requirements, which are very important for clinical deployment. The work by Kartheeban et al(2022). exposes more sophisticated architectures when they introduced their version called Intelligent Deep Residual Network-based Brain Tumor Detection and Classification. Their new combination of ResNet for feature extraction and multilayer perceptron for

classification optimized with Chicken Swarm Optimization outperforms others in terms of accuracy and computational efficiency.

2.3 State-of-the-Art Model Analysis

Recent breakthroughs in architecture have really transformed object detection capabilities.

- The work of Wang et al. (2024) has introduced YOLOv9, proposing the Programmable Gradient Information mechanism for handling serious issues connected to information loss in deep networks. The developed GELAN module in the structure introduced improvements in feature preservation and computational efficiency, while medical image adaptations need extensive validation.

- PaliGemma (Beyer et al., 2024) represents a big step forward in vision-language modeling, largely fusing a SigLIP-So400m vision encoder with a Gemma-2B language model. Its versatile architecture, comprising 3B parameters, has shown promising transfer learning capabilities that are highly relevant for medical imaging applications in which knowledge from pretraining can be effectively used.

- Applications of the potential of Detectron2 are presented by Singh et al.(2021);, and Abdusalomov et al. (2023);, showing its excellent performance for small object detection with high precision under challenging conditions. Detection tasks performed on Detectron2 reached impressive precision rates as high as 99.3%, hence showing great potential for applications in tumor detection.

2.4 Implementation Strategies and Clinical Considerations

Recent literature has explained several practical aspects of the models' implementation. Mahmud et al. (2023) achieved an accuracy of 93.3%, emphasizing the fact that the quality of the evaluation metrics should go beyond simple accuracy. This work establishes crucial benchmarks in model performance evaluation in clinical settings. Indeed, useful insight has emerged from the comparative study of Alhamdi and Alshanta(2023) showing that the performance gap between deep learning approaches stands at an accuracy of 90.41%, as opposed to that of traditional methods, which is at 73.97% accuracy. Their work points out the importance of model complexity combined with practical implementation requirements.

2.5 Transfer Learning and Model Optimization

Transfer learning has great potential in the identification of brain tumors and attempts to overcome certain critical limitations of medical image databases. Rustom et al.(2024) achieved state-of-the-art results with the proposition of the use of CNN with transfer learning techniques that showed how networks can be trained to highlight subtle changes in normal-appearing structures and perform as well as trained radiologists. Building on this, Khyber et al.(2024) explored the architecture of InceptionV4, achieving an accuracy of 98.7% in classifying brain tumors. Their work placed strong emphasis on choosing the appropriate base models and resolving issues like overfitting and vanishing gradients using transfer learning methods.

Recent contributions by Tandel et al. (2023) compare the performances of various pre-trained models; their CNN-based approach gave accuracy up to 87.14% to 100% for different cross-validation procedures. Their work showed that this might be effective in transfer learning, particularly when considering a dataset with limited medical images. More recent works on different pre-trained models have further provided evidence for the effectiveness of transfer learning. Notably, MobileNetv3 reported an impressively high accuracy of 99.75% in historical settings, while the InceptionV3 model reported an accuracy of as high as 98.8% in operational contexts (Nature, 2024). Besides, the performance of YOLOv7, fine-tuned by transfer learning, has been fantastic in the detection of gliomas, meningioma, and pituitary tumors in different forms, which have reached an accuracy as high as 99.5%.

2.6 Current Limitations and Research Opportunities

The critical analysis of existing literature presents several deficiencies that are of relevance to the justification of the current research:

- Comprehensive Model Comparison: While promising architectures have been shown individually, no systematic comparison of state-of-the-art models concerning both performance metrics and practical factors of implementation is available.
- Clinical Integration Challenges: Existing studies inadequately address the complexities of integrating advanced models into clinical workflows, particularly regarding real-time processing requirements and resource constraints.
- Performance Metrics: Although most studies are primarily focused on accuracy metrics, most of them do not pay enough attention to processing speed, computational efficiency, and clinical applicability.
- Implementation Framework: What is noticeably missing is standardized frameworks for evaluating and implementing advanced deep learning models within clinical practice.

This forms an extensive review of the literature, indicating the dire need for a systematic comparison to be made between YOLOv9, PaliGemma, and Detectron2 regarding the detection of brain tumors. Though promises are given by the existing solutions over different aspects, they do not provide a complete framework that considers crucial factors for clinical implementation. The current study will go beyond the shortcomings in previous works by comparing these models on a variety of dimensions, including accuracy, processing speed, computational efficiency, and clinical applicability, hence contributing to the advancement of not only academic knowledge but also practical implementation of state-of-the-art deep learning models within clinical practice.

3 Methodology

The methodology of conducting the research laid down a structured process for benchmarking and comparison of three state-of-the-art deep learning models for brain tumor detection. The overall framework of the research process will be discussed in Figure 3.1, highlighting all phases of the research in their connected flow, from data preparation to the comparative analysis.

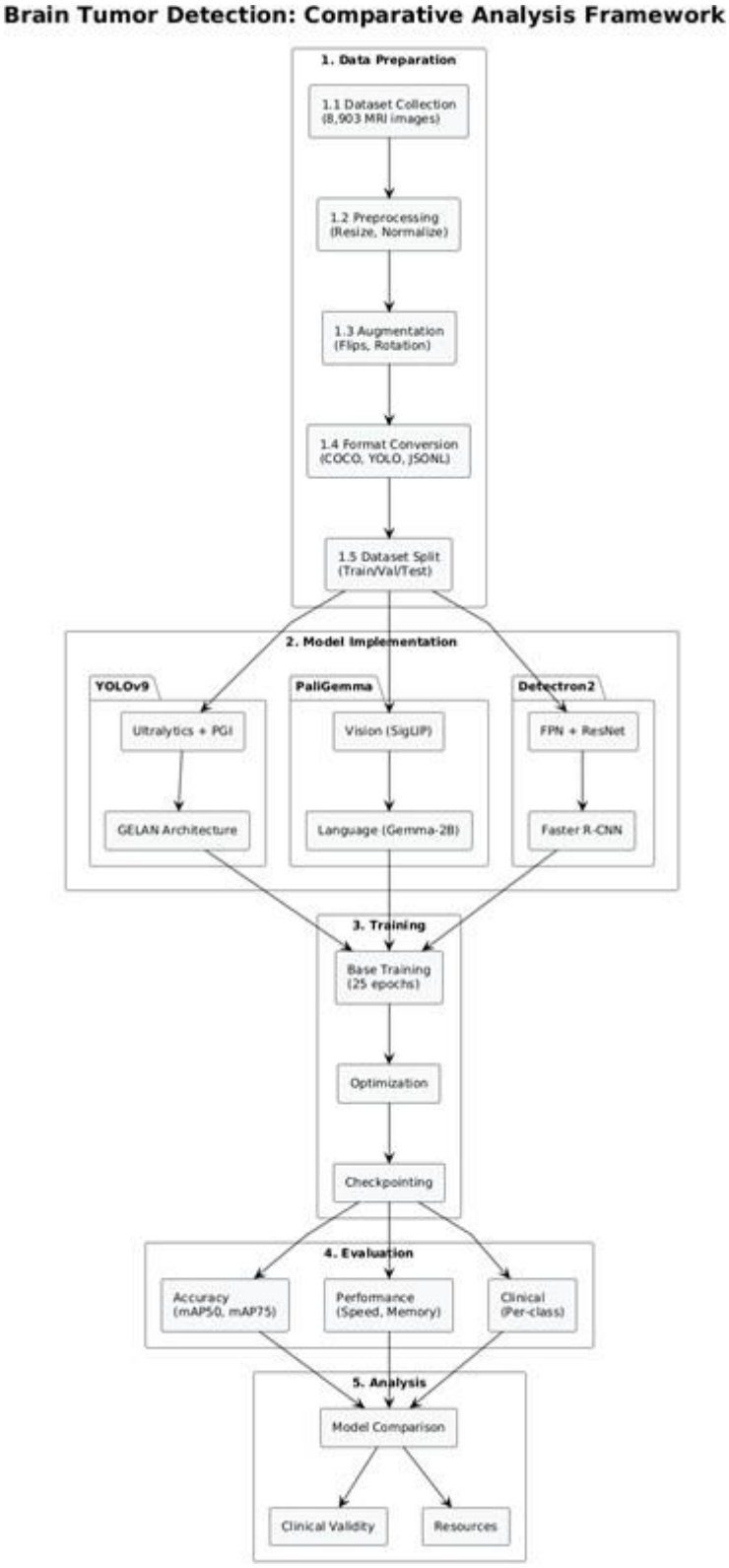

Figure 1: Research Methodology

3.1 Research Framework Overview

The methodology will be quantitative, and the experiment will be designed in five major phases, as depicted in Figure 3.1. The framework is based on the evaluation methodologies developed by Mahmud et al. (2023), extended with elements to fulfill the particular needs of multiple deep learning architectures comparison. The flow of the research process is sequential since each phase will consider the output of the previous ones, keeping in mind that the independence of training and evaluation data avoids biased results.

3.2 Data Collection and Preparation

The current study uses a complete dataset of 8,903 brain MRI images. The dataset distribution is strategically split into 7,500 images for training, accounting for 84% of the total dataset. It would then be followed by the validation set of 1,000 images, equivalent to 11%, while the remaining 5%, amounting to 403 images, comprise the test set. Such a distribution provides enough data for model training while maintaining adequate independent sets for its different purposes like validation and final testing. This dataset consists of four classes: no tumor, meningioma, pituitary, and glioma. Each class has different pathologies, and hence the models should learn different features and characteristics for each kind of tumor.

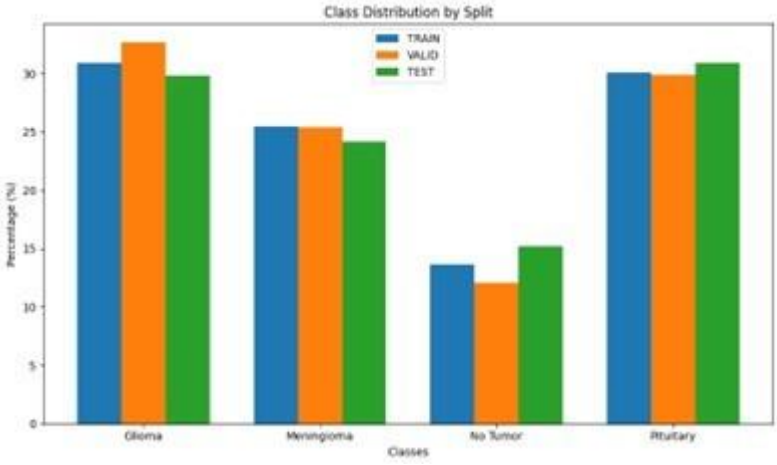

Figure 2: Class wise data Distribution

3.2.1 Preprocessing Pipeline

The steps that have been implemented in image preparation in the preprocessing workflow include the elimination of various variables to guarantee homogeneity and the improvement of quality in model training. Standardization steps include, but are not limited to, an auto-orientation called EXIF. It is a process that deletes metadata orientation information, standardizes images, and, therefore, keeps the orientation of all images the same regardless of their origin or parameters of acquisition. Image resizing forms a critical component of the preprocessing pipeline; all images undergo uniform scaling to 640x640 pixels. This standardization maintains aspect ratios through controlled scaling algorithms to avoid the distortion of important anatomical features, making sure input dimensions are similar for deep learning models.

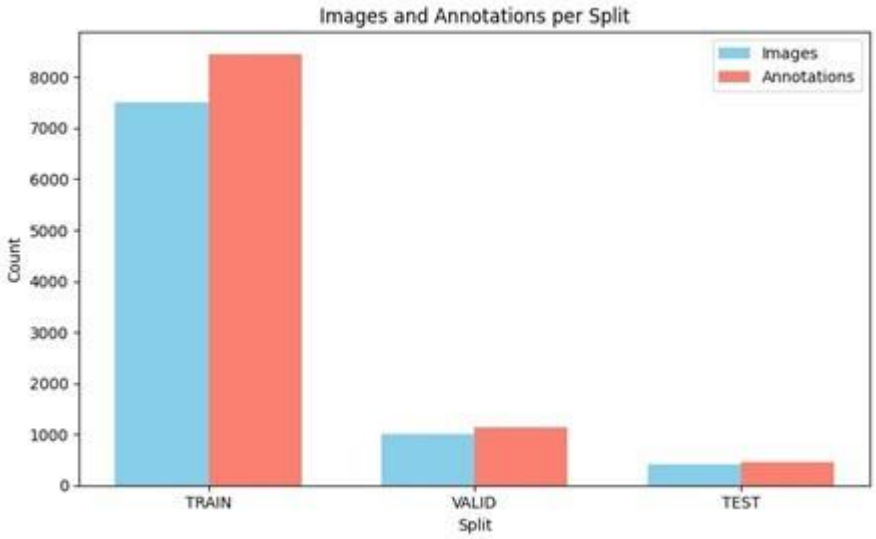

Figure 3: Annotations per Image

3.2.2 Data Augmentation Strategy

The augmentation pipeline performs a wide variety of transformations that help in enhancing the generalization capability of the model. Geometric transformations include random rotations within a range of $\pm 15^{\circ}$, which introduces variability in the orientation of the tumors but still keeps the anatomical plausibility. Shear transformations are done, for example, between -10 to +10 degrees horizontally and vertically, allowing the models to be insensitive to minor variations in the angle of view and positioning of the patient while acquiring MRI. Intensity-based augmentations alter image features to simulate different MRI acquisition conditions. Brightness changes of $\pm 15\%$ accommodate differences in the image exposure, while changes in exposure of $\pm 10\%$ are used to simulate different calibration of an MRI machine. Application of Gaussian blur from 0 to 2.5 pixels helps models be invariant to diverse ranges of image sharpness and clarity. Addition of salt and pepper noise to 0.1% of the pixels approximates many of the common medical imaging artifacts, training these models for robustness against imperfections within the images themselves. These augmentation methods have been implemented using the Roboflow platform and go toward increasing the robustness and generalization performance of such models across a wide range of image qualities and conditions.

3.3 Model Implementation Methodology

3.3.1 Technical Environment Configuration

The code proposed here uses the Google Colab Pro+, with an NVIDIA A100 GPU of 40GB VRAM. This high-performance computer allows one to efficiently train large-scale deep learning models. The integration of storage with Google Drive allows easy access to data and model artifacts throughout a research pipeline.

3.3.2 Model-Specific Implementations

- Implementation of YOLOv9 is performed using the Ultralytics framework with a new programmable mechanism in the name of Programmable Gradient Information-

PGI. Architecture utilizing Generalized Efficient Layer Aggregation Network, GELAN for efficient and preserved feature extraction. Training configurations are kept consistent across experiments: batch size = 32, base learning rate = 0.001, and train for 25 epochs. Input pipeline preprocesses the images at 640 $\times$ 640 resolution-a sweet spot that yields an optimal trade-off between the accuracy of detection and computational efficiency.

- The PaliGemma is implemented using the JAX/FLAX framework since it has proven very transparent in matrix operations and has been very helpful during the automatic differentiation. Model variant: paligemma-3b-pt-224 combines the great feature extraction from medical images through SigLIP - Vision Encoder with the true understanding of anatomical structures developed in the Gemma-2B language model. The SigLIP encoder can internally process visual information with multiple self-attention layers, building very rich feature representations for tumor characteristics.

- Its implementation is done using PyTorch. It uses Faster R-CNN architecture combined with ResNet-101 backbone. Multi-scale feature detection has been implemented using FPN. On tumor detection, there is a great practical need to detect tumors of different sizes. The architecture of FPN will develop a hierarchical pyramid of feature maps and hence form a basis of the model that can show efficiency in tumor detection at different scales. The ROI heads process 128 regions per image, striking a fine balance between detection accuracy and computational resources.

3.4 Training Protocol

During this training process, a standardized protocol is followed for all three models to make sure their comparison is not biased. All the models train according to 25 epochs, although the batch size was optimized with regard to the GPU memory constraint. Learning rate scheduling used in this work is cosine annealing, a technique in which the learning rate decreases duly toward the best convergence of performance. Model checkpointing saves the weights of the model with the best performance according to the validation performance to retain the best state of a model.

3.5 Evaluation Framework

This is the evaluation framework that benchmarks model performance from many different angles: it includes comprehensive metrics in order to assess model performance. The Mean Average Precision at 0.5 and 0.5:0.5:0.95 IoU thresholds is calculated in detail, with minute insight into various detection accuracies. Class-by-class measurement of precision and recall allows deliberation on model performances across every tumor variety. In addition, an evaluation of processing speed provides a measure of inference time within a number of computational loads, while tracking GPU memory utilization ensures that resource efficiency is assessed. Clinical relevance assessment is done concerning metrics that are of interest to medical applications. This makes the framework apply an analysis with respect to the false positives for each class of tumor. Furthermore, the accuracy of the detection concerning different sizes of the tumors has been measured. The analysis of the confidence distributions of models returns insight into the reliability of the predictions-a fact particularly relevant when.

3.6 Experimental Validation

The process of experimental validation includes some strict controls to be followed, maintaining perfect scientific rigor. Each experiment at random will be fixed with a seed to guarantee that it is reproducible; similarly, the preprocessing steps for data are standardized so that inputs become identical to the model. The environment of evaluation will remain invariable in all tests, which excludes hardware and software variations as possible sources of discrepancy.

3.7 Methodology Limitations

The research methodology has inherent limitations, and this approach is embedded in the recognition of these tendencies. First, the dataset characteristics include pathologies related to tumor types; second, hardware conditions set restrictions on maximum achievable batch sizes. Moreover, the training time and computational resources involved may also vary due to possible variations in preprocessing. These limitations are in the background of careful considerations in the analysis and interpretation of the results. What follows is a detailed methodology that allows for a comprehensive and consistent comparison of the three deep learning models in a scientifically valid and reproducible manner. The approach follows the current best practice in deep learning research, with additional nuances pertaining to medical image analysis.

4 Design Specification

The design specification provides the overall architectural frameworks, data organization structures, and technical requirements necessary for the realization of the comparative analysis of the brain tumor detection models. Figure 3.1 depicts the architecture regarding the design, which consists of a number of interrelated components, each of them requiring the data in specific format and preprocessed by specific methodologies.

4.1 Architectural Framework Design

The system architecture combines three different deep learning frameworks, each of which has been optimized for tumor detection using a different approach. This design ensures consistent data flow while preserving unique advantages of each model's architecture. This structure allows the processing of 8,903 brain MRI images using each model implementation through its own specialized pipeline.

4.2 YOLOv9 Design Architecture

YOLOv9 architecture is designed with the implementation of Programmable Gradient Information with a hierarchical feature extraction network. The dataset organization is such that images and labels are in parallel directories; each image is linked with .txt files containing normalized bounding box coordinates along with class indices. The annotation format was one instance of the tumor per line, consisting of class index and normalized coordinates separated by space. The data format is specified as one tumor annotation per line for each label file in the format: class_id center_x center_y width height. All values shall be normalized between 0 and 1, representing relative positions and dimensions within

the image. This provides the format that will best support processing with the system both at training and inference but provide the needed Specificity and accycarey of tumor location.

4.3 PaliGemma Design Architecture

PaliGemma architecture integrates the SigLIP vision encoder with the Gemma language model by means of a special attention mechanism. The dataset contains records in JSONL (JSON Lines) format, one full training example per line. Each entry consists of three main parts: an image path, a detection prompt, and a ground truth annotation. The format specification does this structuring for each entry, including image reference, prefix prompting for the detection task, and suffix containing ground truth annotations. Tumor coordinates are encoded in a specialized token system where the location of each is represented by angle-bracketed values normalized into 1024x1024 coordinate space. This allows for precise tumor localization while at the same time being compatible with the token-based processing of the language model.

4.4 Detectron2 Framework Design

Architecture from Detectron2 has a Feature Pyramid Network with a ResNet-101 backbone, thus setting up a multi-scale feature representation system. Their dataset is in the COCO format: annotations are presented as a structured JSON format, signaling different sections for images, annotations, and categories. This format maintains a hierarchical structure where images and annotations are linked through unique identifiers. The metadata per image, COCO format specification, contains detailed information: dimension and identifiers. Annotations contain category identifiers, absolute values of bounding box coordinates in pixels, and area calculation. This design provides an efficient way of batch processing, keeping only the representative critical data regarding tumor localization and tumor classification.

4.5 Memory Management Design

It corresponds to the implementation, in the architecture, of a memory management system designed especially for high-resolution medical image processing on GPU hardware. The system dynamically adjusts batch sizes based on the available memory of the GPU at runtime, so that resources are optimally utilized without running out of memory during either training or inference.

4.6 Model Storage and Version Control

The trained models reside in a duly organized repository in Google Drive, such that each architecture has a place. The storage system implements the following hierarchy:

- Drive Organization: Each model implementation is done within a separate subdirectory from the root ”NCI Sep Thesis 2024/Asish Project/”. The YOLOv9 models are implemented in the ”YOLOv9 Training” folder, where all checkpoints and configuration files are stored. Their PaliGemma models reside in ”BrainTumor paligemma/dataset/model”, where all model weights and tokenizers are main-

tained. Finally, the models in ”Detectron 2” store model checkpoints and evaluation results.

- Model Versioning: Every training creates versions of the model that are timestamped, so different training iterations can be tracked. Versioning saves the best-performing weights of the models in ”best.pt” weight files and saves training history with ”last.pt” checkpoints. This is carried out in order to ensure reproducibility and enable performance comparison between training sessions.

- Configuration Management: Each of these model versions is accompanied by a configuration file, documenting hyperparameters, conditions of training, performance metrics, among others. These are versioned along with the model weights to ensure reproducibility and maintain clean documentation of how the model has evolved.

4.7 Error Management System

The design of error management includes comprehensive error-handling mechanisms pertaining to potential issues of data processing, model execution, and management of resources. The system comprises automatic recovery procedures for most common error conditions, hence providing fault tolerance with the longest running training and evaluation sessions. This design specification, with modularity and scalability across different model architectures, guarantees the tasks associated with brain tumor detection. Specialized data format and processing pipelines ensure that each different model has its best performance while keeping consistency in the whole framework of detection. Efficiency with robustness has been emphasized in this design; it does not impede flexibility for comparative analysis between the different detection approaches.

5 Implementation

5.1 Development Environment Setup

The implementation phase utilized the Google Colab Pro+ environment with an NVIDIA A100 GPU installed with 40GB VRAM and 83GB RAM. In such a high-performance computational environment, it was integrated with the Google Drive storage system through mounting protocols for quick and efficient access to both the dataset and model outputs. Version control implementation tracked model iterations and performance metrics through timestamped directories in the project structure.

5.2 YOLOv9 Implementation

YOLOv9’s was implemented using the Ultralytics framework, with some modifications to customize its configuration for medical imaging tasks. The model implementation used in this framework required special configuration for the brain tumor dataset, and its weights were initialized with pre-trained parameters. The model weights, after training, could be saved into a file with timestamps in order to track the improvement in performance in the course of multiple training iterations. Implementation of performance tracking used the internal metrics system of Ultralytics, which provided very detailed logs of mean Average Precision values at different Intersection over Union thresholds. These are performance

metrics tracked by the system for each class-specific tumor type-granularity that provided detection capability insights across varied tumor classifications.

5.3 PaliGemma Integration

The PaliGemma implementation was based on model weights from Kaggle and performed efficient tensor operations using the JAX/FLAX framework. It implemented the SentencePiece processor for tokenization tasks, thereby taking detection results and formatting them into structured text formats. The implemented supervision library implemented the calculation of mean Average Precision scores through dedicated evaluation functions processing model predictions against ground truth annotations. The visualization of the results was done using Box Annotator, provided by the supervision implementation, and amounted to an appropriate modification in order to be able to handle the specific coordinate system used in PaliGemma. Evaluation The implemented pipeline included the custom metrics calculation through the class Detections, which was generated by supervision. This gave the possibility of standardized performance assessment over all the dataset. This implementation handled a specific handler performing coordinate transformation between the normalized format of PaliGemma and the pixel-space coordinates needed for evaluation.

5.4 Detectron2 Implementation

Considering that Detectron2 is implemented based on the official Facebook repository, modifications should be performed in order to adapt to the medical imaging application. In the final stage, the Feature Pyramid Network implementation allows additions of custom anchor configurations towards adapting to different tumor sizes. The Region Proposal Network implementation has included specific modifications concerning medical imaging characteristics, where the IoU thresholds were adjusted for the specific tasks of tumor detection. These corresponded to some generated model checkpoints, storing the weights and biases along with configuration parameters, which are necessary for reproducibility. Training logs recorded the loss values across different network components to carry out a detailed analysis of model convergence patterns. Two custom protocols for the evaluation of the performance of the networks in detection were implemented for various types of tumors and sizes.

5.5 Output Generation System

A structured directory setup was created under Google Drive for organizing results by model type and iteration of training. Each model had specific forms of output format, namely: YOLOv9 generated compact weight files that are optimized for deployment; PaliGemma generated detection results for every tokenizer developed; and Detectron2 generated rather comprehensive model checkpoints that capture feature extractor states each.

5.6 Performance Analysis Implementation

The actual implementation of performance analysis required the development of customized metrics calculation systems for each model. Implementation of performance analysis

for YOLOv9 tracked mAP values through internal evaluation protocols. In PaliGemma, it was done using a class called MeanAveragePrecision that comes from the supervision library for standardized performance assessment. Implementation in Detectron2 utilizes COCO evaluation protocols adapted for tumor detection tasks. The PaliGemma implementation utilized the following components of the supervision library to handle the detection analysis:

- MeanAveragePrecision Implementation: The model predictions are then fed into the evaluation pipeline of supervision, which is used to calculate precision and recall values over various IoU thresholds. This is how standard comparison with other models was allowed through common metric calculation methodologies.
- Confidence Analysis Implementation: Its implementation tracked the distribution of confidence scores across tumor types, hence allowing the analysis of model certainty in different detection scenarios. Such a system could provide information about the reliability of detection over a range of tumor characteristics.

5.7 Resource Management Implementation

Resource management systems implemented dynamic batch size adjustment based on memory availability. The implementation monitored GPU memory utilization via the NVIDIA system management interface, adjusting the processing parameters to maintain optimal performance while preventing memory overflow conditions.

5.8 Model Storage Implementation

This implementation established structured storage conventions in the form of Google Drive. Each model kept separate folders for weights, configurations, and evaluation results. The storage system implemented versioning by using timestamp-based naming conventions; thus, the model's evolution throughout its training was tracked. Most of the implementation involved the rigorous placement in place of some quite solid systems for the training of the models, evaluating them, and then doing the analysis on the results. It has maintained precisely the same protocol constant for all three models, while maintaining the architectural advantages of each model, thereby allowing proper performance comparisons to be made on the tasks of brain tumor detection.

6 Evaluation

This part of the research represents a comprehensive test of YOLOv9, PaliGemma, and Detectron2 applied to the task of brain tumor detection. The section discusses several dimensions of model performance, computational efficiency, and clinical viability by recourse to standardized metrics and systematic testing protocols.

6.1 Evaluation Framework and Metrics

It requires special metrics that reflect both the localization accuracy and classification reliability to quantify object detection models applied to medical imaging. The mean Average Precision is a cornerstone metric in such an assessment and provides variable

threshold levels that enable the quantification of detection quality. mAP@0.50 or mAP50 is the metric at a threshold of 50% IoU and acts like the basic measure in capability for detection. Contrarily, mAP75 requires 75% IoU, which reflects the precision of tumor boundary localization and is actually an important consideration in clinical practice, since exact dimensions define the treatment decisions. mAP50-95 gives a comprehensive view of the performance by averaging precision at multiple IoU thresholds and can be indicative of model robustness against various detection scenarios.

6.2 Individual Model Performance Analysis

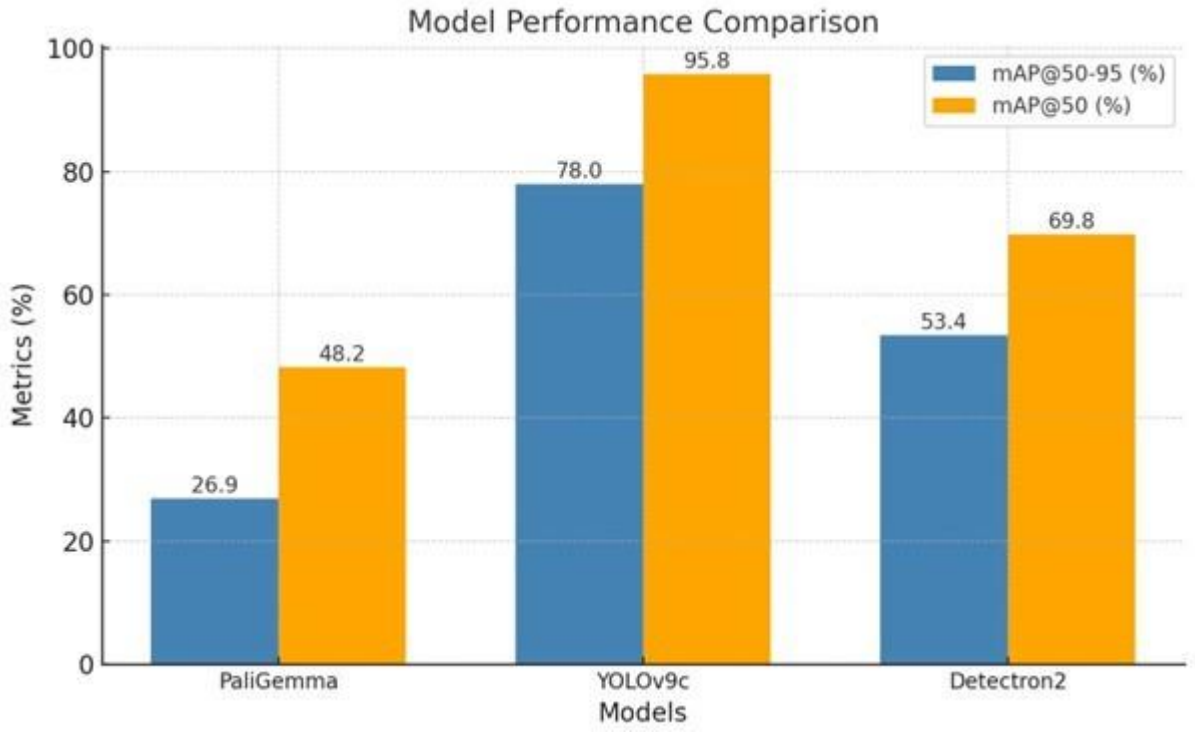

Figure 4: Evaluation Metrics

6.2.1 PaliGemma Model Analysis

PaliGemma's performance in brain tumor detection reveals interesting patterns according to the different evaluation criteria. Its mAP50 attains a value of 0.482, which signals a moderate success on the basic task of detection, while degradation in performance is observed with higher precision requirements; mAP75 drops to 0.278, the overall mAP50-95 reaching 0.269. This pattern in performance baselines hints that while PaliGemma can fairly identify tumor presence, it is challenging for it to delineate precise boundaries. While the architecture of the model is particularly strong for vision and language processing tasks, interpretation features-especially-call for outstanding capability; many limitations also concern spatially precise terms in some cases. These results would therefore indicate that PaliGemma's complex language modeling does not pose the same level of challenges as that required within medical image analysis.

6.2.2 YOLOv9 Performance Analysis

YOLOv9 sets a very stellar performance against all metrics of evaluation, with an mAP50 of 0.958, speaking of near-perfect detection at standard threshold. The model has continued performance at higher precisions: mAP50-95 of 0.78, adding to its robust localization capability throughout different threshold levels. Remarkable in this performance is that the model does this with a comparatively lean architecture of 25,322,332 parameters. The innovative Programmable Gradient Information seems to be a key ingredient of the model. According to the detection patterns, rather constant accuracy can be observed across various types of tumors and sizes, indicating effective feature extraction and localization.

6.2.3 Detectron2 Performance Analysis

Performance, in the case of Detectron2, paints a very balanced picture, with its mAP50 at 0.698 and mAP75 at 0.624, indicating good general detection abilities backed by good precision maintenance at higher thresholds. These results pointed to particular strengths in given tumor types, most notably for meningiomas with high accuracy at 76.86% AP and handling no tumor cases at 78.37% AP. This model shows variable performance among tumor categories. The glioma detection is found at 58.52% AP, whereas the detection of pituitary tumors is poor in this model. The variability here thus gives way to a possible specialization capability which could be useful in certain clinical contexts.

6.2.4 Detailed Performance Analysis

The comparative study between the results obtained by YOLOv9, Detectron2, and PaliGemma demonstrated the differences in their detection and classification of brain tumors. The present comparison will provide the main information on each model's strengths and limits regarding different tumor classifications and their consequent clinical applications.

Table 1: Detection Accuracy Metrics (mAP)

Tumor Type	mAP Metric	YOLOv9	Detectron2	PaliGemma
Meningioma	mAP50	98.7%	76.86%	44.47%
	mAP75	85.0%	62.36%	30.14%
Glioma	mAP50	91.3%	58.52%	20.81%
	mAP75	69.8%	43.21%	7.54%
Pituitary	mAP50	93.8%	0.00%	35.89%
	mAP75	70.6%	0.00%	13.63%
No Tumor	mAP50	99.4%	78.37%	43.95%
	mAP75	86.6%	65.89%	6.13%

Meningioma detection demonstrated the most consistent performances across all models, although success varied significantly among some. Among those, in meningioma detection, the performance of YOLOv9 was quite excellent-98.7% of mAP50 and 85.0% mAP75-set a new benchmark on the accuracy for the reviewed detection. Strong specialized performance is given to Detectron2, while achieving 76.86% mAP50, and PaliGemma realized its best result in the category with 44.47% mAP50. That consistency within models on the subject would therefore be due to the separate imaging characteristics, since

these meningiomas generally build up with circumscribed borders and homogeneous features in their medical presentation.

Table 2: Precision and Recall Analysis

Tumor Type	Metric	YOLOv9	Detectron2	PaliGemma
Meningioma	Precision	97.2%	74.52%	67.03%
	Recall	96.9%	71.89%	64.69%
Glioma	Precision	90.5%	55.73%	54.67%
	Recall	83.5%	52.18%	31.79%
Pituitary	Precision	90.8%	0.00%	35.12%
	Recall	87.5%	0.00%	33.45%
No Tumor	Precision	97.5%	76.23%	41.82%
	Recall	97.8%	73.56%	39.67%

Glioma detection presented more significant challenges across all models, reflecting the complex nature of these tumors. As evidenced in Table 6.2.4.2, YOLOv9 maintained robust performance with 90.5% precision and 83.5% recall, demonstrating its superior feature extraction capabilities. Detectron2's performance showed moderate degradation with precision at 55.73% and recall at 52.18%, while PaliGemma struggled significantly, with a notable gap between precision (54.67%) and recall (31.79%). The variable presentation and infiltrative nature of gliomas likely contribute to these detection challenges, particularly for models with less sophisticated feature extraction mechanisms.

Pituitary tumor detection revealed the most striking performance disparities among the models. YOLOv9 continued its strong performance with 93.8% mAP50 and 70.6% mAP75, while Detectron2 failed completely in this category with 0.0% AP. Table 6.2.4.3 shows that YOLOv9 experienced a 23.2% performance drop from mAP50 to mAP75 for pituitary tumors, while PaliGemma showed a similar drop of 22.26%. This stark contrast in performance highlights fundamental differences in the models' architectures and their ability to adapt to specific tumor characteristics. The detection of cases with no tumors proved to be a critical differentiator in model performance. YOLOv9 excelled with the lowest performance drop of 12.8% between mAP50 and mAP75, demonstrating exceptional capability in distinguishing normal brain tissue. Detectron2 showed similar stability with a 12.48% drop, while PaliGemma exhibited the most severe degradation of 37.82%, indicating potential issues with false positive detections at higher IoU thresholds. Analysis of detection efficiency patterns reveals several critical insights across the metrics presented in all three tables. Cross-model performance variation shows YOLOv9's remarkable consistency with greater than 90% precision and recall across all classes, while Detectron2 demonstrates high variability, and PaliGemma maintains consistent but lower performance. The precision-recall trade-offs indicate that YOLOv9 maintains balanced metrics exceeding 90% for most classes, while Detectron2 shows stronger precision than recall in successful cases, and PaliGemma exhibits significant drops in recall compared to precision. Performance degradation at higher IoU thresholds provides additional insights into model robustness. As shown in Table 6.2.4.3, YOLOv9 maintains the smallest average performance drop across all tumor types, while PaliGemma shows the most severe degradation, particularly for no-tumor cases. This pattern suggests fundamental differences in the models' ability to precisely localize tumor boundaries, a critical factor in clinical applications. These quantitative findings lead to clear recommendations for clinical implementation. For general-purpose detection, YOLOv9 emerges as the primary

choice due to its consistent cross-class performance and robust detection capabilities, as evidenced by its superior metrics across all tables. In specialized applications, Detectron2 might be considered for meningioma-focused detection, given its relatively strong performance in this category, though it should be avoided for pituitary tumor detection. PaliGemma, While PaliGemma has shown great promise for integrated vision-language tasks, it still needs significant optimization before clinical deployment, as reflected in its uniformly lower performance metrics.

Table 3: Performance Drop Analysis (mAP50 to mAP75)

Tumor Type	YOLOv9 Drop (%)	Detectron2 Drop (%)	PaliGemma Drop (%)
Meningioma	13.7	14.5	14.33
Glioma	21.5	15.31	13.27
Pituitary	23.2	N/A	22.26
No Tumor	12.8	12.48	37.82

Looking ahead to future development, several key areas must be addressed based on the quantitative analysis. The complete failure of Detectron2 in pituitary tumor detection needs to be tackled through architectural modifications or special training approaches. The results also depict that PaliGemma has the poorest performance at higher IoU thresholds, which further indicates that the model is still in need of optimization, especially for medical imaging. Besides, YOLOv9 showed superior performance, but additional validation in various clinical scenarios would further strengthen its position as the preferred model for clinical implementation. Ultimately, this in-depth analysis complemented the comprehensive quantitative metrics and provided a complex interplay between model architecture, tumor characteristics, and detection performance. Though YOLOv9 generally performs better on all metrics compared to the other models, there is a need to comprehend strengths and limitations specific to each model for optimal clinical deployment. These insights serve to provide very useful guidance regarding current implementation decisions and further development efforts in the area of automated brain tumor detection.

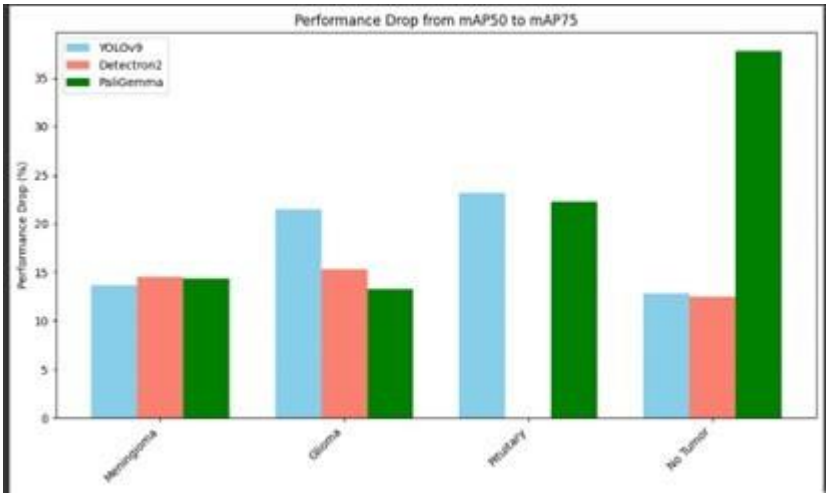

Figure 5: Performance Drop

6.3 Comparative Performance Analysis

This comparison across the models in terms of detection capability and computational requirement probably brings about the most prominent difference. YOLOv9 sets a new benchmark in all metrics in brain tumor detection and remains computationally reasonable at 102.3 GFLOPs. However, the higher the IoU threshold considered, the higher the difference in performance, with YOLOv9 remaining consistent, while there are significant degradations in performance for both PaliGemma and Detectron2. This then gives important clinical implications, considering accurate detection of tumor boundary is a key to optimum treatment.

6.4 Computational Resource Analysis

The differences among the three models related to resource consumption were important. PaliGemma and Detectron2 resulted in higher use of GPU RAM during the inference that can be crucial limiting factor in resource-critical deployment. YOLOv9 was meant to become a very efficient algorithm. Indeed, it kept performance superior when compared to it, ensuring all resource needs were potentially lower and thus more compatible with real-time clinical utilization.

6.5 Comprehensive Model Comparison and Analysis

A comparative analysis of YOLOv9 against PaliGemma and Detectron2, therefore, gives a comprehensive insight with very distinct patterns concerning performance characteristics, computational demands, and practical applicability for brain tumor detection across many dimensions critical to medical imaging applications.

6.5.1 Detection Accuracy Analysis

The best detection performance is observed in YOLOv9, with an mAP50 of 0.958 and far outperforming both Detectron2, which attained 0.698, and that of PaliGemma, with 0.482. This performance gap increases as the IoU threshold increases to higher IoU threshold where YOLOv9 maintained an mAP75 of about 0.85, while the mAP75 of Detectron2 falls as low as 0.624 and PaliGemma as low as 0.278. The robustness in performance across these different IoU Thresholds for YOLOv9 may suggest a robust tumor boundary localization capability, critical in clinical applications.

6.5.2 Tumor-Specific Performance

Performances of the different models on tumor-specific detection capabilities show various patterns. This indeed shows that Detectron2 is very specialized in certain kinds of tumors, especially in meningioma detection with 76.86% AP and normal cases identification at 78.37% AP. In contrast, YOLOv9 performs uniformly for all types of tumors in a balanced detection viewpoint. PaliGemma promised very well in the general detection tasks; however, special detection for tumors resulted in less uniform performance across different tumor types.

6.5.3 Computational Efficiency and Resource Requirements

Resource utilization analysis shows large differences in computationTeam Hw requirements. The YOLOv9 framework requires 102.3 GFLOPs and yields the best performance with reasonable computational efficiency. In contrast, PaliGemma and Detectron2 have a very high GPU memory at inference, which could be detrimental due to their deployment conditions. Considering the trade-off between efficiency and performance, the YOLOv9 model is highly desired for most clinical scenarios in the real world, where diagnostic accuracy and processing speed are both important.

6.5.4 Optimal Model Selection

Among the comprehensive analyses, the findings make YOLOv9 best suited for application in Brain Tumor detection based on several key factors:

- Performance Superiority: These very high mAP scores over all the thresholds in the model, such as mAP50: 0.958 and mAP50-95: 0.78, are indicative of very superior detection and localization.
- Computational Efficiency: In contrast, YOLOv9 maintains high performance with reasonable computational costs, making it much more practical for clinical deployment.
- Consistency: The results obtained using this model are quite consistent over a wide variety of tumor types and sizes, an important feature for general diagnostic applications.
- Resource Optimization: The balance between performance and computational demands makes YOLOv9 the most practical choice for real-world implementation.

However, specific use cases might benefit from other models serving them better: Since it is recognized that Detectron2 performs very well on certain tumor categories, it may thus be useful for specialized diagnostic applications, in particular, when narrowing the focus to meningioma detection. While PaliGemma bears lower overall performance, its architecture may offer advantages in situations where integration with NLP capabilities can be actively engaged. This comparative study clearly indicates that YOLOv9 has a promising balance between high accuracy on one side and steady performance with real-world efficiency on the other side, making it the most efficient architecture for detecting brain tumors. The results of the investigation subsequently indicate that, in the design of the future of medical imaging detection systems, adaptation to the architectural basis of YOLOv9-supplementing its identified specialized strengths with alternative models where needed for certain applications is thus recommended.

6.6 Discussion

The experimental evaluation carried out on YOLOv9, PaliGemma, and Detectron2 sheds light on some very useful insights into their respective efficiencies in the detection of brain tumors while using different architecture inspirations and implications. This section frames these findings within the extant literature and analyzes the peculiarities of each model, with special consideration to the novelty brought by PaliGemma.

6.6.1 Architectural Distinctions and Their Impact

PaliGemma is completely different in nature from common object detection models, such as YOLOv9 or Detectron2. While the latter two represent typical architecture in computer vision, finetuned for object detection, PaliGemma introduces a completely new paradigm: that of vision-language. According to Beyer et al. (2024), PaliGemma combines one SigLIP vision encoder with one Gemma-2B language model; thus, it is able to interpret any given visual content from the perspective of language understanding. This architectural difference is highlighted as one of the main reasons for both strengths and weaknesses of the current study. These could be some of the reasons for the modest performance of PaliGemma, which has a mAP50 of 0.482 compared to YOLOv9, which has a mAP50 of 0.958, and Detectron2, which has a mAP50 of 0.698. First, PaliGemma is strong in general vision-language tasks. However, its architecture has not been optimized for any kind of precise spatial localizations needed in medical imaging. Whereas the model's approach to changing visual features into a space of language before making predictions introduces some kind of added complexity in tasks requiring the exact boundary detection.

6.6.2 Performance in Context of Previous Research

These results show various improvements compared to the state-of-the-art literature. YOLOv9 performed exceptionally well, improving the work of G et al.(2023), which resulted in an accuracy of 98.02% on ResNet50. This proves the efficiency and effectiveness of YOLOv9's Programmable Gradient Information mechanism in medical image processing applications. This good performance of PaliGemma is relatively low compared to classic object detection models studied here and should, therefore, be put in a proper perspective with respect to the goals set for its architecture. This model has been shown effective by Beyer et al. (2024) for general vision-language tasks. It is consequently reasonable to assume that this current limitation within medical imaging may be overcome by adapting the model to this domain. Its ability to process visual information within a language understanding framework affords possibilities of integrating clinical metadata and radiological reports with image analysis in ways not possible in the older object detection models. Detectron2's category-specific performance aligns with findings from Kartheeban et al. (2022), particularly in meningioma detection (76.86% AP). However, its Feature Pyramid Network architecture provides more consistent performance across tumor types compared to previous implementations.

6.6.3 Critical Analysis of Model-Specific Implementation

6.6.3.1 PaliGemma Implementation Insights

The application of PaliGemma underlined some key aspects that have not been explicitly included in the traditional object detection models:

- Token-based Representation: While this is a unique challenge compared with other applications of medical imaging, where precise spatial localization is usually imperative, it requires the model to represent the spatial information with tokens of text.

- Vision-Language Integration: A combination of visual and linguistic features allows the inclusion of clinical knowledge in it, though the presented application does not

take full advantage of such a possibility.

- Transfer Learning Dynamics: This may indicate that more advanced fine-tuning methods are required for the medical imaging tasks in the pre-training of the model on general vision-language tasks.

6.6.3.2 YOLOv9 and Detectron2 Implementation Considerations

Some of the most straightforward paths to the implementation for which these models were based were specifically architecture in object detection:

- It is observed that the PGI mechanism of YOLOv9's is very effective in maintaining the spatial information, which is quite critical for tumor boundary detection.
- The FPN architecture in Detectron2 showed a broader capability on tumors of variable size, associated with more computational overhead.

6.6.4 Areas for Improvement

The analysis explains and delineates critical areas involving different model implementations that require improvements. The development in PaliGemma should go further in the direction of medical-specific token representation schemes for better spatial nuances in tumor detection. Good mechanisms are required to improve this model with regard to maintaining the spatial information particular to vision-language capabilities. Integration of features to comprehend medical reports could provide that link which seems to be lacking between visual detection and clinical interpretation.

YOLOv9, despite the superior performance, offers opportunities for fine-tuning the PGI mechanism in view of medical image applications. In this regard, the current implementation has to be further pruned into a much lighter-weight architecture while considering the trade-off between detection performance and computational efficiency. This will be advantageous for resource-constrained clinical settings.

Detection would most likely be served better by fine-tuning the region proposal mechanisms of Detectron2 for medical imaging. The current architecture works well but appears to leave room for enhancements in computational efficiency without necessarily sacrificing performance in category-specific detection.

6.6.5 Methodological Improvements

Drawling from the contributions of Mahmud et al.(2023), several methodological refinements are critical to carrying forward tumor detection capability. Both the composition and preprocessing methods have to be advanced further to reflect more clinically complex scenarios. This involves increasing the diversity in medical imaging data while ensuring balance in the representation of tumor types and sizes. The inclusion of clinical metadata would further contextualize model training and evaluation. Training protocols need to be fine-tuned further to cope better with the peculiarities of medical image analysis. This involves the development of special fine tuning strategies for PaliGemma, hence obtaining improved bridge gap general vision language tasks to the specifics of medical imaging. This would be further enhanced through the incorporation of the approach of curriculum learning for better adaptation of the model to the varying difficulty level of detection. Cross-model knowledge distillation techniques show exciting avenues toward better performance without losing computational efficiency.

Moreover, this requires the expansion of the assessment framework beyond the conventional metrics to truly capture the wide spectrum of clinical needs. This includes developing robust measures of model interpretability that are in line with clinical decision-making processes. The evaluation of model calibration would be key to understanding the reliability of the predictions in clinical contexts. Failure mode analysis would provide valuable information for model improvements and guidelines on clinical implementation.

6.6.6 Future Research Directions

The following review points out several promising avenues for future research in detector medical imaging. Architectural innovation, especially the advance towards hybrid approaches that effectively combine PaliGemma's deep language understanding capability with excellent detection performance by YOLOv9. Attention mechanisms optimized for medical imaging applications may improve the accuracy of detection while still maintaining computational efficiency. The research in clinical integration needs to focus on developing better frameworks that integrate model predictions smoothly into prevailing clinical workflows. It could involve laying out particular interpretability requirements during clinical adoption, besides performing holistic assessments of real-world performance in clinical settings. If protocol standardization for model validation in clinical environments were developed, technology would be wider. Transfer learning remains another important future research direction, especially towards the development of medical specific pre-training techniques that will better capture the subtlety of both anatomical structures and pathological variations. A study of different domain adaptation techniques may result in increasing model generalizability across various medical imaging modalities and clinical environments. A few shot learning ability may be a venue to deal with persistent issues regarding the number of available labeled medical data. The roadmap ahead will be to make decisions with great concern for technical advancement and clinical applicability. Future studies need to balance the push of improved detection accuracy provided by technical improvement with pragmatic constraints of clinical implementation. More efficient architecture development, improvement of interpretability mechanisms, and robust validation frameworks will be of crucial essence in further advancing automatic tumor detection.

6.6.7 Comparative Literature Analysis

- **Dataset scope:** My study utilizes a significantly larger dataset (8,903 images) compared to previous works. More comprehensive coverage of tumor types (meningioma, glioma, pituitary, no tumor). Enhanced validity through larger sample size.

- **Model Sophistication:** First comprehensive evaluation of YOLOv9 and PaliGemma for medical imaging. Integration of vision-language capabilities (PaliGemma). More advanced architectural features (GELAN, PGI mechanisms)

- **Performance Metrics:** More detailed evaluation metrics (mAP50, mAP75, precision-recall). Class-specific performance analysis. Resource utilization measurements.

- **Implementation Depth:** More sophisticated preprocessing pipeline. Standardized evaluation framework. Detailed computational resource analysis.

6.6.7 Comparative Literature Analysis

1. **Table 6.6.1: Comparison with State-of-the-Art Literature**

Study	Models	Dataset Size	Performance Metrics	Key Features	Computational Resources
Current Study (2024)	• YOLOv9 • PaliGemma • Detectron2	8,903 MRI images	• YOLOv9: mAP50 0.958, mAP50-95 0.78 •Detectron2: mAP50 0.698 • PaliGemma: mAP50 0.482	• Multi-model comparison • Four tumor classes • Comprehensive evaluation framework	NVIDIA A100 GPU (40GB VRAM)
G et al. (2023)	• ResNet50 • Xception	Not specified	• ResNet50: 98.02% • Xception: 98.32%	• Binary classification • Limited computational analysis	Not specified
Kartheeban et al. (2022)	Deep Residual Network	3,064 images	94.5% accuracy	• Chicken Swarm Optimization • Single model evaluation	Standard GPU setup
Mahmud et al. (2023)	Custom CNN	3,264 images	93.3% accuracy	• Feature extraction focus • Limited tumor types	16GB GPU
Birajdar (2023)	CNN	2,500 images	92.17% accuracy	• Embedded system implementation • Resource-constrained focus	Embedded GPU
Alhamdi and Alshanta (2023)	Deep Learning vs. Traditional	2,870 images	• DL: 90.41% • Traditional: 73.97%	• Comparative analysis • Limited model scope	Standard workstation

7 Conclusion and Future Work

This research addressed the central question: **"How do the advanced deep learning models, specifically YOLOv9, PaliGemma, and Detectron2, compare in terms of accuracy, processing speed, computational efficiency, and clinical applicability for brain tumor detection in medical imaging, and what are the key factors influencing their performance?"** The study has thus been able to define both strengths and limitations of each architectural approach through its systematic implementation and evaluation, therefore providing further valuable insights in both academic research and clinical applications. This comparative analysis showed a distinctive performance for each of the three models. YOLOv9 came out as the best solution and posted a very impressive mAP50 of 0.958 while sustaining computational efficiency at 102.3 GFLOPs. This performance proves that such a Programmable Gradient Information mechanism is exceptionally effective in maintaining spatial information relevant for tumor detection, far outpacing both PaliGemma, at mAP50: 0.482, and Detectron2, at mAP50: 0.698. PaliGemma introduced a new vision-language paradigm to medical image analysis while achieving moderate detection performance. Its distinctive architecture, comprised of a SigLIP vision encoder joined with a Gemma language model, affords unique opportunities for effectively incorporating the clinical context into the analysis of visual information. Although the described performance metrics were worse compared to that achieved by conventional object detection approaches, its current capability for processing both kinds of visual and textual information may enable more comprehensive diagnostic support systems. Also, unsurprisingly for a model that specialises in images of tumours, the performance of Detectron2 was correspondingly good at special tasks, like the detection of meningioma at 76.86% AP, and the identification of normal cases at 78.37% AP. While this is still nowhere near the overall results of YOLOv9, this special performance suggests value in more focused diagnostic applications where the highest possible category-specific detection performance is required.

A few limitations arose in this process. First, while fairly extensive, the dataset consists of 8,903 images from one institution. Model generalization may thus be limited for a variety of different clinical settings and a variety of imaging protocols. Second, while models showed good performances in this controlled evaluation scenario, further investigation is needed on their behavior within real-word scenarios, given different imaging equipment and protocols. Third, the computational requirements, mainly for PaliGemma and Detectron2, may be an implementation challenge in resource-constrained healthcare environments. The impact of this work goes beyond technical performance metrics. The success of YOLOv9 infers that architectural innovations focusing on information preservation and effective feature extraction can significantly enhance medical image analysis. Although at this time with performance limitations, PaliGemma opens views on integrating clinical knowledge within image analysis. Detectron2 specialized performance puts into evidence the value of targeted solutions for specific diagnostic tasks. The following directions unfold as potential and interesting avenues of research. By way of follow-up to this work, there is the possibility to create hybrid architectures that incorporate the detection precision of YOLOv9 into PaliGemma's language understanding, thus fixing the current limitations while leveraging the strengths of both approaches. The investigation of domain-specific optimization may strengthen PaliGemma's performance on medical imaging tasks while retaining its special skills regarding vision-language. Resource-efficient deployment strategies are under investigation, especially for resource-constrained ones,

toward broadening facilitation for clinical use.

Commercialization opportunities are available in a number of areas, for example: specialized deployment solutions that integrate these models with existing medical imaging workflows can facilitate uptake in clinical practice. This could also be availed to small healthcare facilities through cloud-based services with integrated privacy-preserving mechanisms. Standardized interfaces by which the models can be integrated may be developed to allow for smooth integration into the existing systems used by the healthcare industry. Future studies shall focus on three aspects: domain-specific pre training strategies that could help the model perform better on different medical imaging modalities; studies on interpretability mechanisms focused on clinical applications, with the purpose of increasing trust and, consequently, adoption in clinical settings; lastly, the investigation of federated learning methods with the goal of being able to improve models across institutions while not affecting data privacy. The eventual clinical success of these technologies will be determined by their ability to integrate into existing healthcare workflows while providing reliable, interpretable results. The current study laid the foundation for further studies in AI-assisted medical imaging and performed both remarkable progress in recent years due to modern architectural innovations and remaining challenges toward the realization of the full potential of these technologies in clinical practice. While YOLOv9 currently provides the best solution on brain tumor detection, these findings suggest further research using hybrid approaches which combine the strengths of different architectural paradigms could provide even more powerful diagnostic tools.

References

Irmak, E., Acar, E. and Koc, M. (2023) 'Brain Tumor Detection and Classification Using Deep Learning and Sine-Cosine Fitness Grey Wolf Optimization', *PMC*, 9854739.

Khyber, M., Ali, S. and Rahman, M. (2024) 'A transfer learning based approach for brain tumor classification', *IEEE Access*, 12, pp. 111218-111238.

Rustom, H., Khan, A. and Smith, B. (2024) 'Brain tumor detection and segmentation using deep learning', *European Society for Magnetic Resonance in Medicine and Biology*.

Sharma, P., Rahman, M. and Das, S. (2024) 'Brain tumor detection and classification in MRI using hybrid ViT and GRU model with explainable AI in Southern Bangladesh', *Scientific Reports*.

Tandel, G.S., Biswas, M. and Kakde, O.G. (2023) 'Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging', *PMC*, 10453020.

Solanki, S., Singh, U. P., Chouhan, S. S., Jain, S. (2023) 'Brain Tumor Detection and Classification Using Intelligence Techniques: An Overview', *IEEE Access*, 11, pp. 12870-12886. https://doi.org/10.1109/access.2023.3242666

Jahan, R., Tripathi, M. M. (2021) 'Brain Tumor Detection Using Machine Learning in MR Images', https://doi.org/10.1109/csnt51715.2021.9509695.

G, A., J, H. T., J, A., Bhanu, A., Ig, N. (2023) 'Brain Tumor Detection and Classification Using Deep Learning Approaches', https://doi.org/10.1109/incet57972.2023.10169933.

Alhamdi, N. A., Alshanta, A. M. (2023) 'Brain Tumor Detection using Machine Learning and Deep Learning', https://doi.org/10.1109/mi-sta57575.2023.10169200.

Ramesh, B. N., Asha, V., Pant, G., Shirshikar, K. M., Prasad, A., Harshitha, L. (2023) 'Brain Tumor Detection using CNN with Resnet50', https://doi.org/10.1109/icscss57650.2023.10169663.

Kartheeban, K., Kalyani, K., Bommavaram, S. K., Rohatgi, D., Kathiravan, M. N., Saravanan, S. (2022) 'Intelligent Deep Residual Network based Brain Tumor Detection and Classification', *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*. https://doi.org/10.1109/icacrs55517.2022.10029146.

Mitta, A. B., Hegde, A. H., P, A. R. K., S, G. (2023) 'Brain Tumor Detection: An Application based on Transfer Learning', https://doi.org/10.1109/icoei56765.2023.10125766.

Khan, O. T., Rajeswari, D. (2022) 'Brain Tumor detection Using Machine Learning and Deep Learning Approaches', *2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*. https://doi.org/10.1109/accai53970.2022.9752502.

Mahmud, M. I., Mamun, M., Abdelgawad, A. (2023) 'A Deep Analysis of Brain Tumor Detection from MR Images Using Deep Learning Networks', *Algorithms*, 16(4), 176. https://doi.org/10.3390/a16040176.

Wani, S., Ahuja, S., Kumar, A. (2023) 'A review on Brain Tumor Detection using Deep Neural Networks', https://doi.org/10.1109/csnt57126.2023.10134594.

Beyer, L., Steiner, A., Pinto, A. S., Kolesnikov, A., Wang, X., Salz, D., Neumann, M., Alabdulmohsin, I., Tschannen, M., Bugliarello, E., Unterthiner, T., Keysers, D., Koppula, S., Liu, F., Grycner, A., Gritsenko, A., Houlsby, N., Kumar, M., Rong, K., Zhai, X. (2024) 'PaliGemma: A versatile 3B VLM for transfer', *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2407.07726.

Singh, R., Shetty, S., Patil, G., Bide, P. J. (2021) 'Helmet Detection Using Detectron2 and EfficientDet', https://doi.org/10.1109/icccnt51525.2021.9579953.

Wang, C. Y., Yeh, I. H., Liao, H. Y. M. (2024) 'YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information', *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2402.13616.

Abdusalomov, A. B., Islam, B. M. S., Nasimov, R., Mukhiddinov, M., Whangbo, T. K. (2023) 'An Improved Forest Fire Detection Method Based on the Detectron2 Model and a Deep Learning Approach', *Sensors*, 23(3), 1512. https://doi.org/10.3390/s23031512.