Clinical-Grade Precision Without the Complexity: A YOLO-Based Brain Tumor Detection Framework

By
Md. Efaz
201-15-14193

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by

Mr. Raja Ratiqul Hasan Tusher
Assistant Professor
Department of Computer Science and Engineering Daffodil International University

Co-Supervised by

Mr. Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

September 16, 2025

APPROVAL

This Project titled **"Clinical-Grade Precision Without the Complexity: A YOLO-Based Brain Tumor Detection Framework"**, submitted by Name: **Md. Efaz**, ID No: **201-15-14193** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **17 September, 2025**.

BOARD OF EXAMINERS

Dr. Bimal Chandra Das (BCD) — **Chairman**
Professor and Dean (in-charge)
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Shah Md. Tanvir Siddiquee (SMTS) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Hasanuzzaman Dipu (MHD) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Nazibur Rahman (NZR) — **External Examiner**
Technical Lead – Database Administrator
Telenor – Grameen Phone Account

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mr. Raja Ratiqul Hasan Tusher**, **Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

14.9.25

Mr. Raja Ratiqul Hasan Tusher
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:

Mr. Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

Efaz 14.9.25

Md. Efaz
Student ID: 201-15-14193
Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Mr. Raja Ratiqul Hasan Tusher**, **Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Computer Vision** to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting themat all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

The precise and efficient diagnosis of brain tumors in magnetic resonance (MR) images is extremely important as it enable to treat these kinds of the disease earlier and more effectively. Although models from the YOLO (You Only Look Once) series of real-time object detection algorithms have attained favorable results in medical image analysis, performance comparison of the most recently released YOLO versions is less explored, particularly in multi-class brain tumor classification tasks. In this paper, we assess and compare the performance of three YOLO versions YOLOv10, YOLOv11, and YOLOv12 on an annotated MRI dataset with 308 brain tumor images including glioma, meningioma, and pituitary adenoma cases. Models are finetuned throughout on same settings and evaluated with standard metrics like precision, recall, mAP@0. 5, and mAP@0. 5-0.95. YOLOv11 has the best average recall (89.2%) and mAP @0. 5 (91.6%), although YOLOv12 achieves the best precision (89.8%), and performs best in diagnosing pituitary tumor. YOLOv10…is of good performance.., but it is obviously deficient in glioma detection. There are trade-offs between detection sensitivity and calss-specific accuracy across models, with YOLOv11 representing the best in trade-offs and the most suitable in the real-time clinical deployment. This work provides a practical guide to choose the optimal YOLO architecture for automated brain tumor localization and has potential to enhance diagnostic speed and reliability in neuro-oncology pipelines.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Brain neoplasms are a devastating health problem which has serious morbidity worldwide. Early and accurate diagnosis is essential for the improvement of patients' prognosis, especially when using non-invasive imaging methods, like MRI. With the fast growth of deep learning, ATDs have demonstrated great potential in assisting clinical work by increasing the diagnostic speed, relieving burden of radiologists and decreasing inter-observer discrepancies.

Of deep learning methods, the YOLO (You Only Look Once) family of real-time object detectors has shown promise as an algorithm that is conferred between temporal and quality values for medical imaging applications. Nevertheless, the majority of previous works concerning the application of YOLO for the brain tumor detection utilized the outdated versions such as YOLOv5 or even YOLOv8 and improved them through some architectural modifications, attention information integration, or a hybrid fusion imported from the segmentation perspective. Although these works have achieved promising results, either they are based on small or homogeneous datasets, or concentrate on several tumor types, or conduct single-model evaluations without comparisons with latest YOLO revisions.

In this work, we fill these gaps by performing a systematic and comparative evaluation study with state-of-the-art YOLOv10, YOLOv11, and YOLOv12 on a heterogeneous and multi-class brain tumor MRI dataset containing glioma, meningioma, and pituitary tumors. Under the stemming of each model with equal weight being given to each training framework, we evaluate them using rich performance metrics such as precision, recall, F1-score, mAP@0.5, and mAP@0.5-0.95 to both contribute a backbone for benchmarking and to give a real-world insight about which YOLO variant could make the trade-off categorically between detection accuracy, robustness, and clinical use. In the end, this work is to help researchers and practitioners to choose and implement appropriate deep learning models for real-time localization of brain tumors in the clinic.

1.2 Motivation

Brain tumor identification in MRI has to be accurate, fast, and consistent. Late readings of scans and disconcerting interpretations of brain scans can affect treatment planning and thus remain in the direct line of patient outcomes. Despite the

advancements due to deep learning, the majority of published work: (i) studies older YOLO versions or custom architectures on small, homogeneous datasets, (ii) presents headline mAP but without addressing sensitivity/specificity trade-offs, and (iii) proposes complex models whose deployment is unrealistic in clinical practice. In practice, hospitals require models that are not only accurate but practically simple to operate that run compute-efficiently in near real time on modern hardware, generalise across diverse tumour morphologies (glioma, meningioma, pituitary), and are calibrated such that clinicians can trust the predictions at clinically meaningful thresholds. The more recent generations of YOLO (v10-v12) claim architectural and training innovations stronger feature aggregation, recharged heads, and enhanced context modeling by transformer but have not been benchmarked on multiclass MRI detection task under the same protocol in a head-to-head manner. This discrepancy prompts a careful, clinically-motivated comparison where the focus is not only on mAP, but on precision/recall balance, class-wise behaviour, convergence stability, and on the visual quality of detections.

1.3 Objectives

- Primary Objective

- To methodologically compare YOLOv10, YOLOv11, YOLOv12 in multiclass detection of brain tumors (glioma, meningioma, pituitary) in a 3,903 image MRI dataset with a single unified paradigm for training and evaluation, and to determine the optimal model to balance clinical sensitivity (recall) and specificity (precision).

- Secondary Objectives

- Report performance in terms of precision, recall, F1, mAP@0. 5, and mAP@0. 5–0.95, as well as class-level PR curves and confusion matrices to uncover error modes and calibration performance.

- Please, also analyze the training dynamics (like loss convergence, also check for stability at 100-epochs) to gain intuition about the optimization efficiency, generalization without overfitting.

- Compare per-class metrics and track the character of class-specific behavior, particularly on the harder glioma class (e.g. study bounding-box coherence, lens/repeated/low confidence boxes).

- Assess whether the deployment is ready for production by exporting the best checkpoints to ONNX, and referring to whether the system is feasible to use in real-time (i.e. throughput and latency impacts of the architecture for each version).

- Set up actionable guidance for clinical integration: when to prefer YOLOv11's balanced recall vs. when to prefer YOLOv12's higher precision (e.g. screening vs. confirmatory protocols) and thresholds that maximize F1 in practice.

Hypothesis

On one hand, YOLOv11 will create the most balanced performance profile for real-

time clinical use (higher recall and stable localization), and on the other hand YOLOv12 will provide an extra degree of precision and a cleaner visual prediction (favorable when prioritizing false positives over all). YOLOv10 will not be as strong a minimal baseline, but will not perform well on cases with complex gliomas.

Contributions

- It is also demonstrafed in this work on a large, annotated MRI detection benchmark with a consistent, reproducible evaluation of YOLOv10/11/12.
- A clinically focused examination which moves past single mAP numbers, examining calibration curves, confusions analysis, and qualitative analysis.
- Practical implications for implementing and deploying YOLO models in neuro-oncology workflows and operational concerns (sensitivity vs. specificity).

1.4 Organization of the Report

This report is structured to provide a comprehensive understanding of the project " Biomedical Entity Extractionand QA with RAG: A StepTowards Intelligent Healthcare," detailing every aspect from motivation to implementation. The organization is as follows:

Chapter 1. Introduction: This chapter introduces the project, including the background, problem statement, motivation, objectives, and a brief overview of the methodology. It also outlines the anticipated outcomes and the structure of the report.

Chapter 2. Background: This section provides the foundational knowledge necessary to understand the project, summarizes key works in BioQA with a focus on BioASQ systems, highlighting their architectures, techniques, and limitations.

Chapter 3. Research Methodology: This chapter describes the methodology adopted for the project, Describes the modular system architecture, models used, and task-specific processing for each question type. It also includes system design, task allocation, and the proposed project plan.

Chapter 4. Implementation and Results: This chapter focuses on the implementation details, including the environment setup, performance evaluation, and comparative analysis of the models used for each QA module. It also discusses the results obtained and their implications.

Chapter 5. Engineering Standards and Design Challenges: This section discusses the compliance with relevant software and hardware standards, the challenges encountered during the project, and the impact on society, environment, and sustainability. Ethical considerations and the sustainability plan are also elaborated.

Chapter 6. Conclusion: The concluding chapter summarizes the project, highlights its limitations, and proposes potential areas for future work.

Chapter 2

Background

2.1 Introduction

In recent years, deep learning is newly developed for brain tumor detection in MRI images, and YOLO family object detection model is more and more popular. In recent past researchers have introduced variations in the model structure to increase the speed and accuracy such as attention mechanism, Gan-based hybrid architecture, Transfer learnable technique other than this there are many more for Tumor localization. Such innovations can assist in the diagnostic workflow in real-time for clinician aid.

2.2 Literature Review

RCS-YOLO is a detection model for brain tumor studies which is introduced in [1]. It integrates a ReParameterized Convolution with Channel Shuffle (RCS) and a One-Shot Aggregation (OSA) module that can provide stronger representation with less computation. CHAMNet: It also achieves precision of 93.6%, mAP@0 trained on the Br35H dataset. 5 of 94.6%, and mAP@0. 0.95/72.9% @ 114.8 FPS Faster and more efficient than the models like YOLOv6-L and YOLOv8l and better performance.

In this paper we consider the enhanced version of YOLOv8, named YOLOv8-DEC which was proposed at the end of [2] with modules including C2f_DySnakeConv, EMA, and CARAFE modules. Furthermore, these modules enhance feature extraction, contextual information, and upsampling. At such scale, the model achieves 90.9% precision, 74.3% recall and mAP@0 bym on a 9,900-image dataset. 0%, which is better than some other structures (e.g., Faster R-CNN, RCS-YOLO).

In another study [3], model of YOLOv8s was trained on 300 annotated MRI images of glioma, meningioma and pituitary tumour. The one that uses data augmentation (image rotation) provides the model with precision of 94.3%, recall of 93.2%, and mAP@0. 5 of 94.1%, and mAP@0. 5-0.95 of 42.1%. Even beyond this from its Anchor-free to C3 convolution module to Mosaic augmentation, YOLOv8s can also generalize better than previous YOLOs.

The Simple Image Object Detection Model [4] which is a light weight high speed and high accuracy model with depthwise separable convolutions to reduced computation complexity. Evaluated on BrATs and ANDI datasets, the model achieved a ROUGE-1 of 85%, ROUGE-2 of 79% and Exact Match of 70%, indicating its feasibility to be used in real-time clinical settings. Rare tumor types such as DIPG were explored in [5]. It adopts YOLOv5 for the detector, and U-Net (or BB-UNet) as the segmentor. Using our two-step approach we

achieve an average Dice score of 0.62 which is superior to the legacy models (including DeepMedic) also under domain shift conditions.

Transfer learning has also been studied for applications in brain tumor detection [6] from an in-domain transfer learning perspective and they find that models pretrained on medical related data outperforms models pretrained on natural images by a large margin. This translates into a 10% higher detection and a 24% faster convergence on average. Our experiments based on YOLOv8 demonstrate that fine-tuning on domain related dataset contributes to the performance of the model. A YOLOv5 and 2D U-net hybrid detection-segmentation pipeline for multi-class detection and segmentation of machino tumor including the findings in ([7]) is introduced in [7]. YOLOv5 achieves mAP 89.5%, which outperforms Faster R-CNN, YOLOv4 and SSD. Fused U-Net secmentation scores a Dice of 88.1%, against 80.5% by U-Net only

In[8], two CNN-based models for the classification of brain tumors from MRIscanimages. Augmented dataset of 2,065 samples (253) This deeper model is 93% accurate with a Dice score of 92% and appears to benefit from use of a deeper model, albeit at a slower training speed. A YOLOv5x Based Tumor Detection Method for Early Tumor Detection with BraTS and Roboflow datasets [9] Along with a CSP backbone, it achieves 98.4% mAP with 98.7% precision and 95.6% recall and outperforms both YOLOv5m and YOLOv5l. These results underscore the readiness of the model for clinical application.

Finally, [10] presents a multimodal 3D-CNN with several MRI modalities (T1, T1c, T2, Flair) for enhanced tumor detection. With the use of instance normalization (on input mode basis), and of a particular loss function (the intersection over union between the predicted and ground truth masks), model delivered a better performance than that of 2D CNNs with 4.7% of Dice improvement and faster convergence. This encodes volumetric spatial characteristics of the lesions. There is remarkable progress in existing methods (e.g., models are confined on dataset scale, or adopt a over-scale model, extrema the scale-up). To address this issue, we propose body fine-tuning YOLOv10 and YOLOv11 on the large, diverse dataset that offers best trade-off between high accuracy detection, moderate model complexity and an easy deployment pathway in clinical practice.

2.3 Gap Analysis

Despite the extensive advancements in YOLO-based architectures for brain tumor detection ranging from the integration of attention mechanisms, transfer learning, and hybrid segmentation pipelines there remains a notable gap in the evaluation of newer YOLO variants (YOLOv10 and YOLOv11) for multiclass tumor detection under unified experimental settings. Currently, a significant majority of preceding researches are either using limited-sized datasets or highly curated datasets (such datasets are BR35H, BraTS, and Roboflow) with fixed tumor types. Frequently, these methods report results in a very disconnected setting, meaning results are reported for either glioma or pituitary only, hampering its generalization into real-world diagnostic scenarios. While most works operate under considerations of mAP or Dice score maximization, and often, these works take away a lot of trade-off factors into consideration-the inference time, imbalance in class distributions, and robustness against varying tumor morphologies. Meanwhile, multimodal approaches, such as 3D-CNNs, seem promising but tend to be quite computationally demanding, hence rendering them less ideal to be implemented in clinical settings with constrained resources. Also, YOLOv12 remains untested alongside its predecessor versions in the same experimental environment, hence the literature lacks any form of comparative analysis. Therefore, the study fills these gaps through a systematic evaluation of YOLOv10, YOLOv11, and YOLOv12 models using a multiclass MRI tumor dataset toward exploring the aspect of accuracy, computational efficiency, and real-time applicability in an evenly-weighted manner with clinical relevance.

2.4 Summary

In the background, very obviously, brain-tumor detection from MRI images is a time-sensitive, life-stake game wherein deep learning-YOLO families, in particular-have been favored for real-time localization. There is an incessant drumbeat of upgrades in the literature that includes custom backbones and modules, attention and advanced upsampling, and very efficient anchor-free baselines (the YOLOv8 series) that train better with augmentation. The hybridization of methods and transfer learning--here as in-domain pretraining on medical data--tends to boost both accuracy and convergence. Other lines investigate light-weight designs with the use of depthwise separable convolutions for speed, or deep high-capacity classifiers for pure classification, or multimodal 3D-CNNs to capture volumetric context at a higher computational cost. However, the majority of previous works rely on small datasets, curated cohorts, evaluate only few tumor classes or report headline mAP/Dice with no analysis of the sensitivity–specificity trade-offs, inference latency or robustness to class imbalance thus making clinical generalization shaky. Most critical are that even the most recent YOLO generations have not been benchmarked in a head-to- head comparison under a unified protocol. Your submission addresses this gap by providing a thorough comparison of YOLOv10, YOLOv11, and YOLOv12 on a multi-class MRI dataset in terms of accuracy, robustness, efficiency, and deployment realism.

Chapter 3

Research Methodology

3.1 Methodology

3.1.1 Overview

The approach considered in this paper developed a fully-fledged pipeline that stabilizes a three successive YOLO models YOLOv10, YOLOv11, and YOLOv12 for accurate, fast, and clinically relevant brain tumor detection in MRI images. This pipeline consists of five consecutive interconnected phases: data collection, data preprocessing, model selection, model training, and model evaluation. The adoption of strict methodologies facilitates the assessment of the performance of all the models, trained under the same conditions and a fair comparison of their performance, as well as a perspective of the practical value as a clinical diagnosis tool.

3.1.2 Proposed Methodology

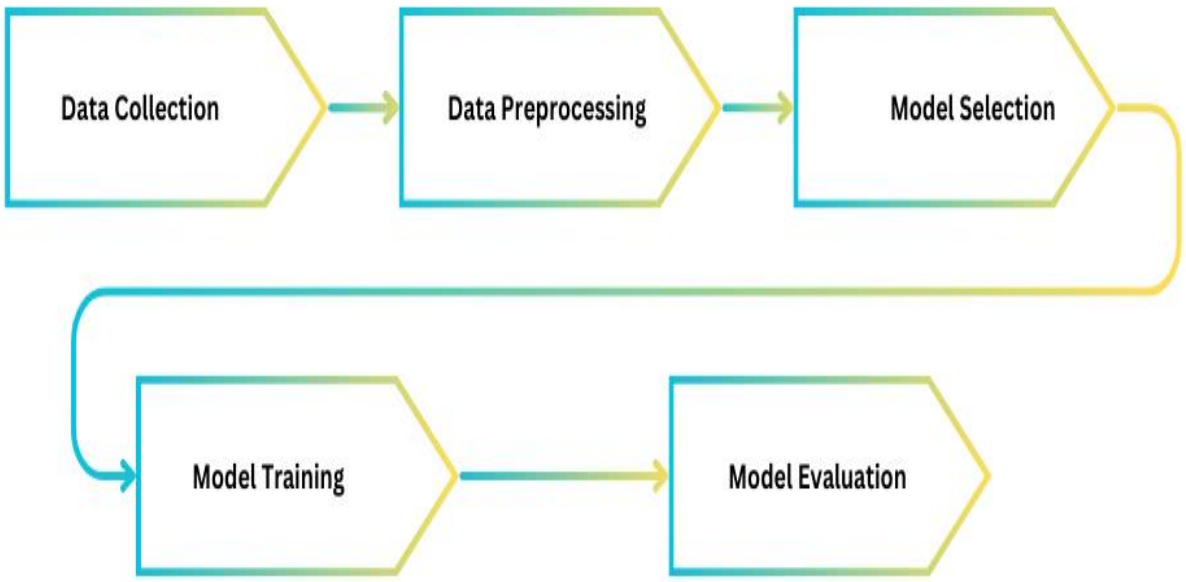

Figure 3.1 Proposed Methodology

The Proposed methodology visualized in Figure 3.1 include:

1. Data Collection and Analysis
2. Preprocessing
3. Model Selection
4. Model Training
5. Model Evaluation

3.2 Detailed Methodology and Design

The methodology for this research follows a structured, step-by-step process, beginning with data collection and culminating in model evaluation, as illustrated in the accompanying diagram:

1. Data Collection and Analysis:

This work uses a public medical imaging dataset with 3,903 high-resolution T1-weighted contrast-enhanced brain MRI scans. The scans are carefully categorised into four separate classes: Glioma, Meningioma, Pituitary Tumor, and No Tumor. One thing that makes this dataset especially fit our goal is that the bounding boxes of all tumor instances are well annotated in each image, providing both spatial location and class label which are fundamental for object detection. Unlike many existing approaches using small-size or manually annotated data, the data set used here is large and well-labeled, accommodating a strong model training and evaluation. In addition, the inhomogeneous tumour morphology and the class imbalance in the dataset are common real-world clinical cases, which improves the external validity of our models.

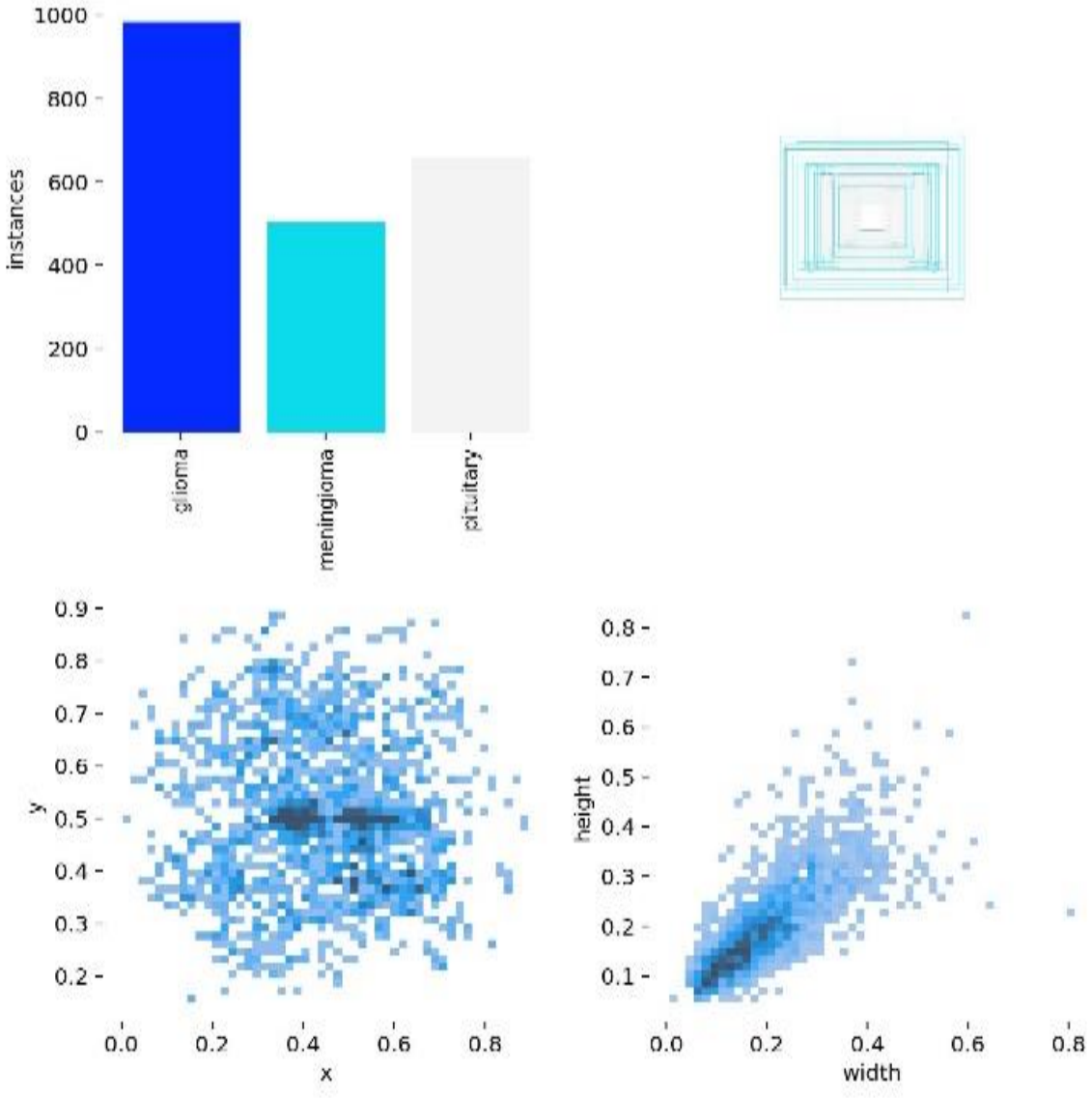

Figure 3.2 Class Distribution and Annotation Characteristics

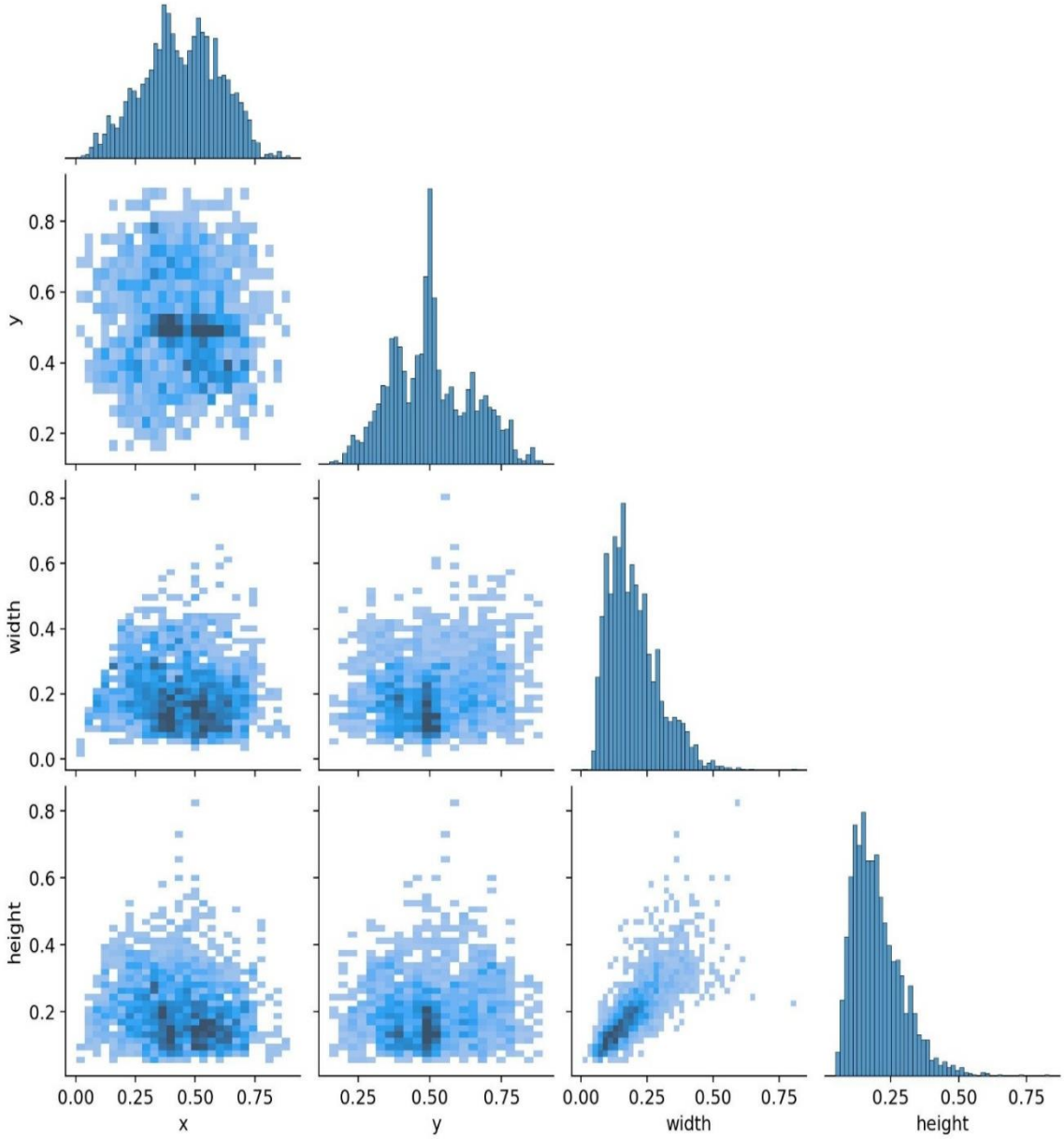

Figure 3.3 Labels Correlogram of our dataset

2. Preprocessing

The next stage, data preprocessing, included a number of standards, albeit crucial, steps. All MRI images were rescaled to 640x640 to minimize the variation in the input dimensions of MRI images to YOLO architectures and maintain important anatomical features. This has been selected as a compromise between accuracy and computational tractability. Image normalization was performed to normalize pixel intensity values to the range [0, 1], which helps to stabilize and expedite the learning process. The data were then decided into three stratified sets: the training set (70%), the validation set (20%), and the test set (10%). This stratification resulted in an even distribution of all tumor classes within the splits. We also employed data augmentation method (random flip, rotation, brightness, and mosaic augmentation) to improve the model robustness and performance. These augmentations mimic the heterogeneity and noise found in clinical images, and enable the models to be ready for deployment stages.

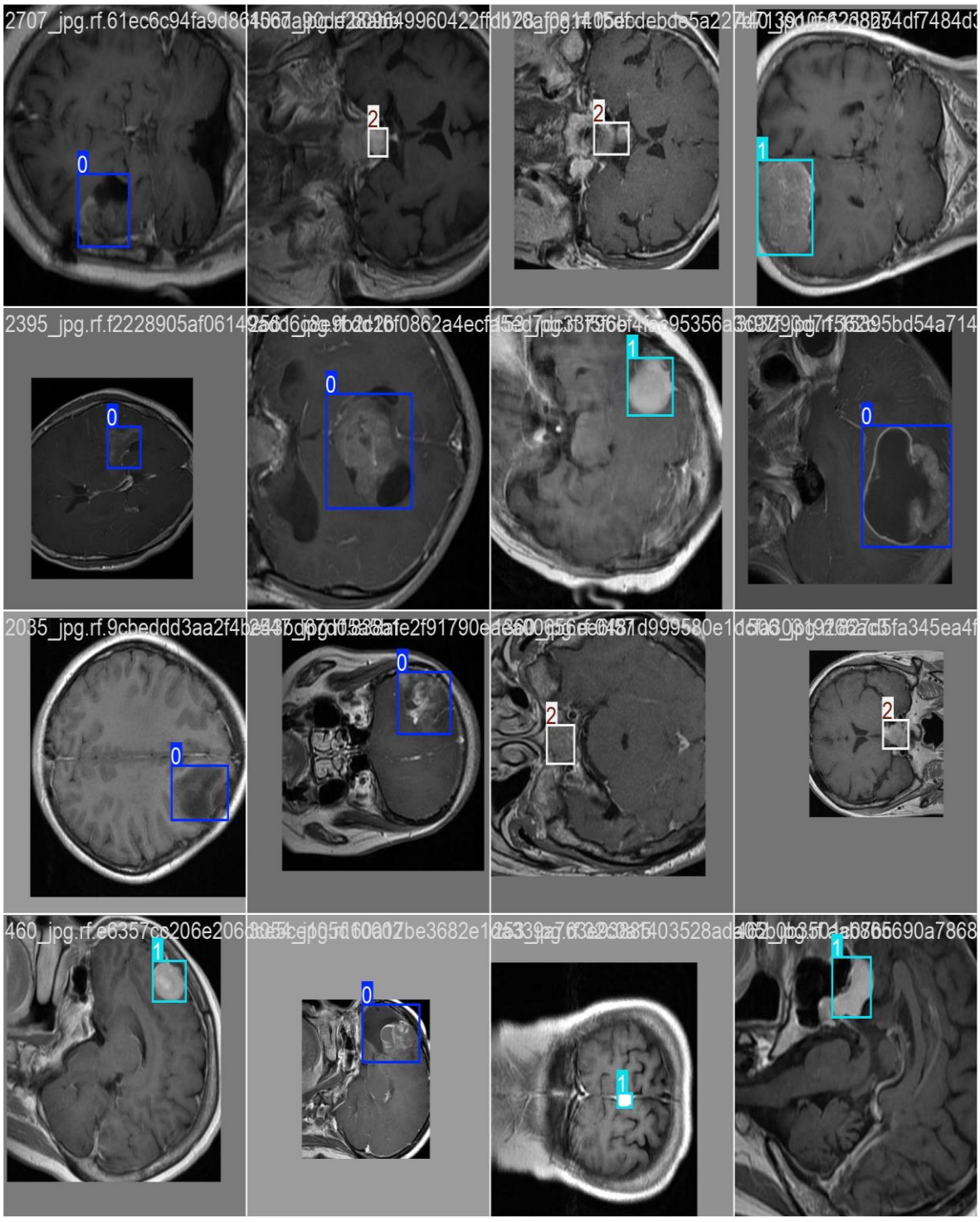

Figure 3.4 Overview of our training set after preprocessing

3. Model Selection

In the third phase, model variant model selection, we compared iy4p with three consecutive YOLO architectures YOLOv10, YOLOv11, and YOLOv12 using the same setting for the three models. Each model presents a step forward in architectural complexity, ability to extract features, and appropriateness for clinical-grade detection of tumors. All three models are pre-trained on the COCO and then fine-tuned on the brain tumor data using transfer learning. This led to faster convergence, lower training time, and better domain adaptation ability on medical imaging than training from scratch.

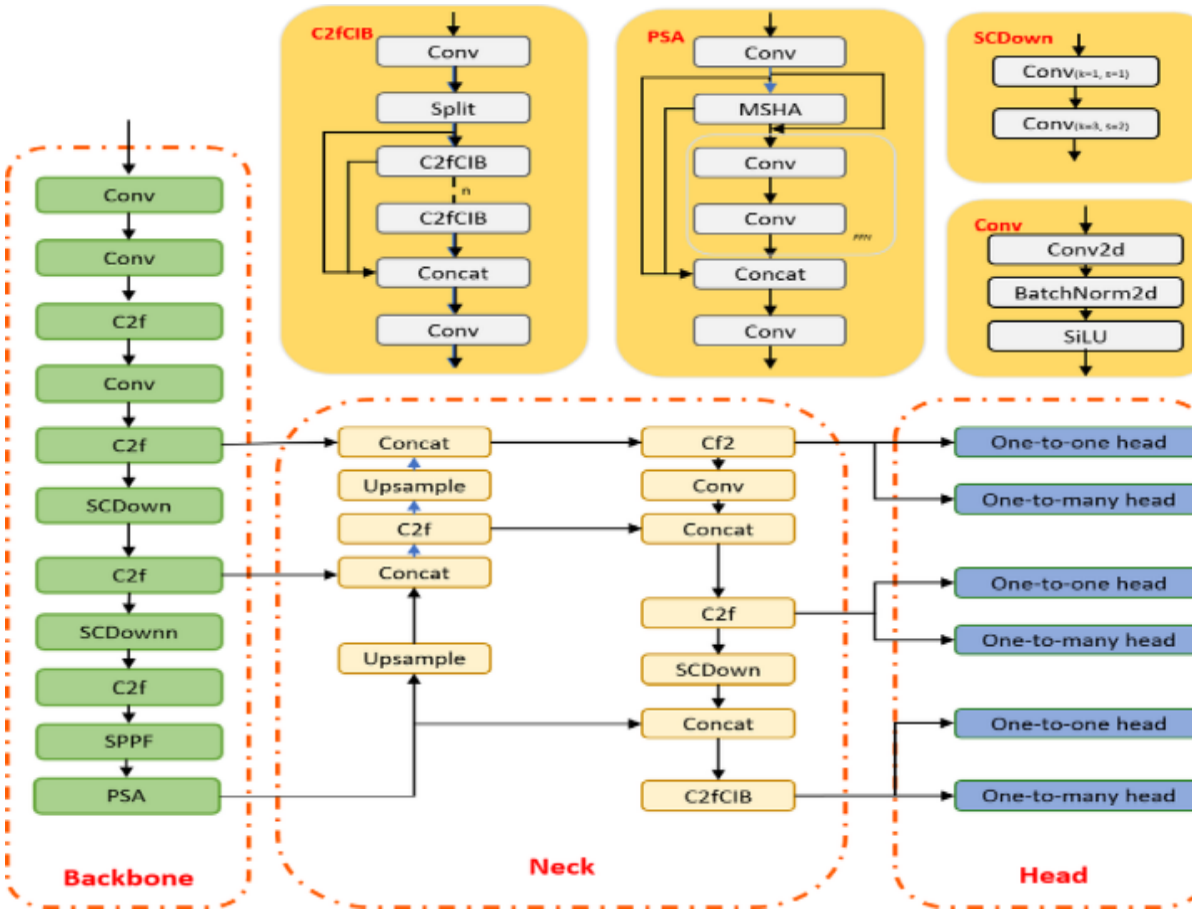

Figure 3.5 YOLOv10 Model Architecture

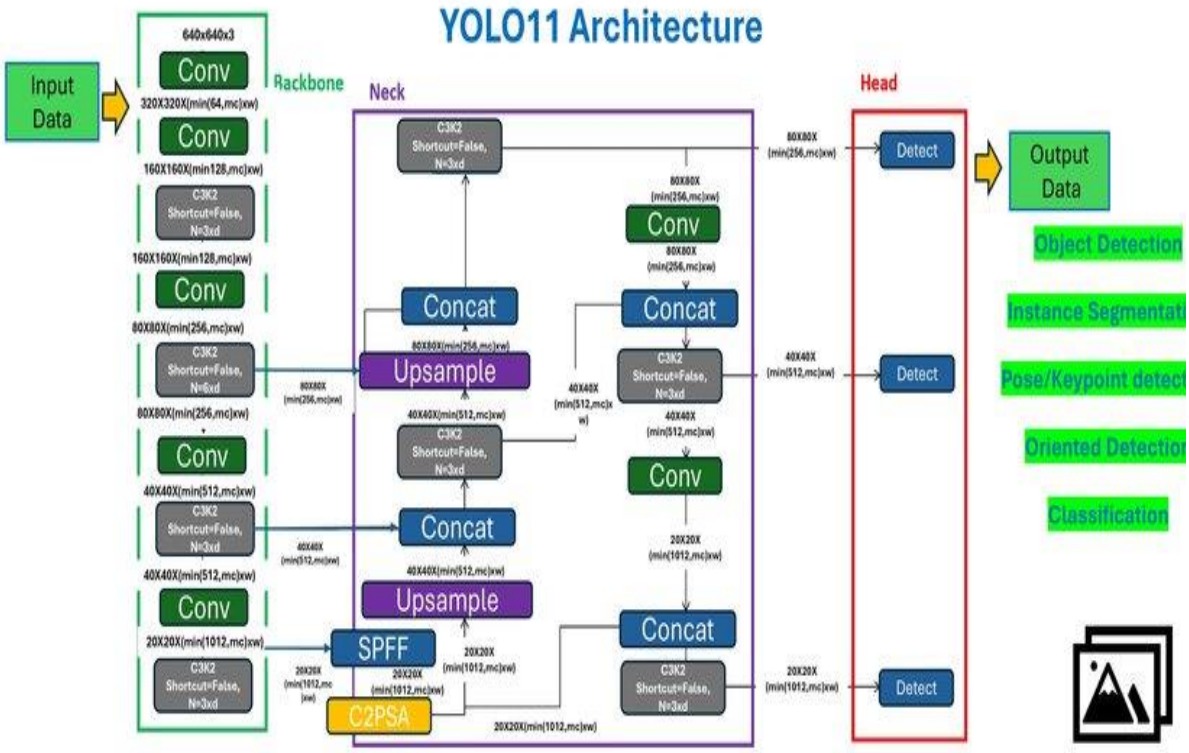

Figure 3.6 YOLOv11 Model Architecture

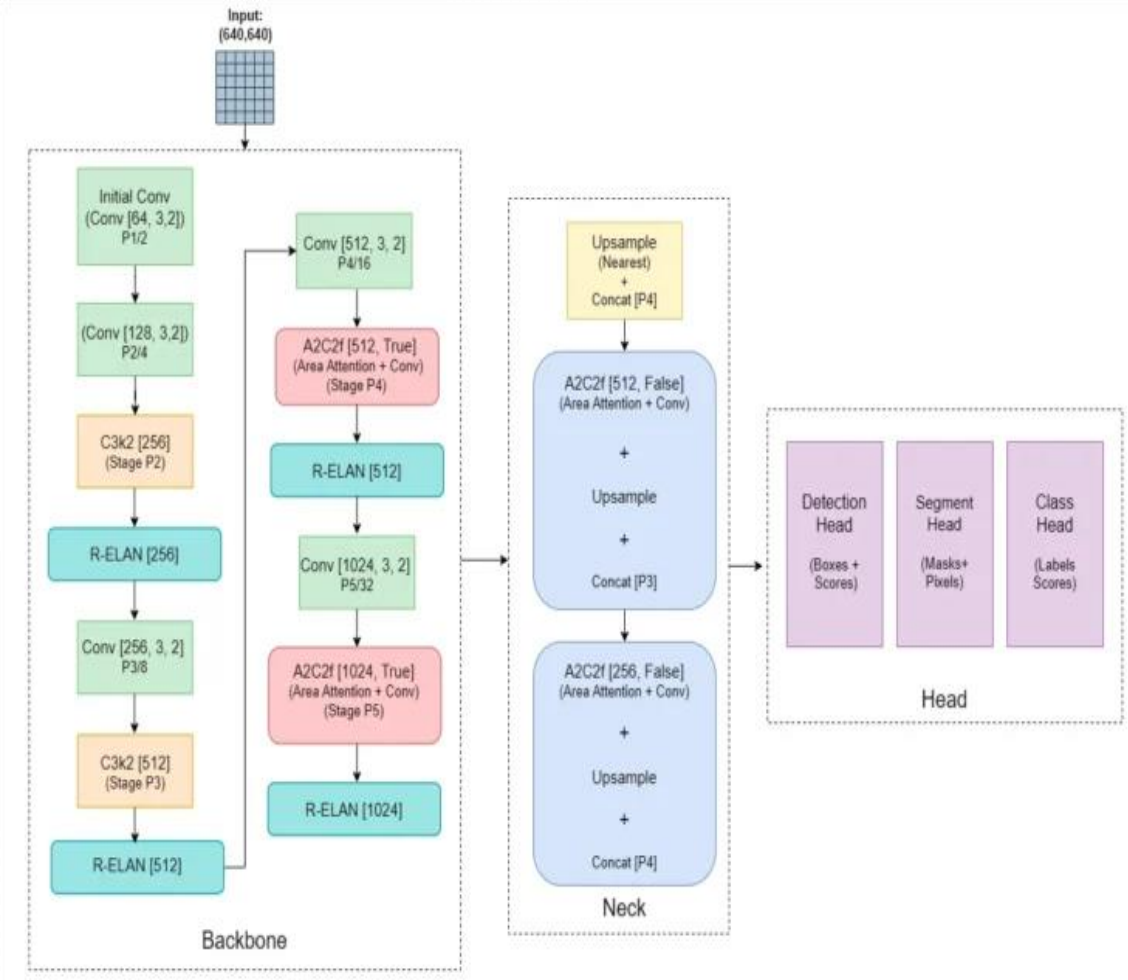

Figure 3.7 YOLOv12 Model Architecture

We chose YOLOv10 as the backbone of our detection module as it is highly efficient and dedicated to the speed and the common tasks of detection. It utilizes an efficient convolutional body and a lightweight detection head to minimize computational cost with decent accuracy. The medical imaging environment in which YOLOv10 fits could include limited-resource settings that require real-time detection, yet scarcely supply the hardware capacity for it (think small hospitals or portable diagnostic devices). It is simple enough, so inference occurs rapidly; however, it seems to be much less effective especially in typing complex or anomalous variant tumors. Hence, the authors consider YOLOv10 to be a baseline classification system rather than a final diagnostic tool.

Directly generated from YOLOv10, YOLOv11 are much deeper feature aggregation modules with more sophisticated detection heads. This modification allows YOLOv11 to seize multi-scale features better, hence localizing small, diffuse, or abnormally shaped tumors. This enhanced representation is most beneficial for glioma detection, which usually exhibit indistinct boundaries in MRI. The middle ground is found by YOLOv11 between speed and accuracy, retaining the efficiency of its predecessor, yet improving upon detection robustness. This compromise makes it a highly practical choice for real-time clinical workflows that require enough sensitivity and completion rate.

The YOLOv12 is the most advanced and complex model in the three, featuring transformer-based attention mechanisms within the neck of the network. This makes the

model learn global contextual relation and complex spatial dependency, highly required for tumor detection from MRI, where the tumors are usually superimposed on the surrounding tissues or are under low contrast. Using attention-based modules, YOLOv12 is proficient in addressing ambiguous or intersected tumor areas, which are common failure spots for older YOLO versions. YOLOv12, being more computationally expensive than other YOLO versions, is better in incorporating both local and global context, enabling high-precision tumor detection in clinical research and specialized diagnostic centers.

4. Model Training

In the fourth stage, model training, we trained each YOLO model with 100 epochs of intense fine-tuning by the same training and validation datasets. The training was conducted in a GPU-enabled Kaggle environment which was computationally efficient. The optimizer was Stochastic Gradient Descent with momentum, and cosine annealing of the learning rate schedule was used for learning rate adaption over epochs. Losses for the YOLO models were bounding box regression loss, objectness loss, and classification loss which were combined into a single multi-task learning objective. Dynamic anchor box generation using k-means clustering on tumor bounding boxes was further used to more accurately adapt to the size distribution of the tumors. The architecture of each model already included features like Focal Loss to alleviate class imbalance and advanced augmentation policies. We chose the best weights with validation loss, and we exported them to ONNX format for potential use in clinical diagnostic systems.

5. Model Evaluation

We evaluated the performance of the trained models on the held-out test set. We evaluated our performance using various metrics such as Precision, Recall, F1-score, mAP@50 which is average Precision at IoU 0.5, and mAP@50–95 which is mAP averaged over IoU thresholds in range from 0.5 to 0.95. The evaluations were performed on Kaggle with GPU to achieve beautiful computations with accuracy and results. Here are how the metrics are defined:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad \text{..... (1)}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad \text{..... (2)}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \text{..... (3)}$$

$$mAP@50 = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad \text{..... (4)}$$

$$mAP@50-95 = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{10}\sum_{t=0.5}^{0.95} AP_i(t) \quad \text{..... (5)}$$

where, N is the number of classes, and AP_i (t) is the Average Precision for class i at threshold t.

3.3 Project Plan

The detailed project plan shown in Figure 3.6.

3.4 Task Allocation

There is no member in my team. So, all the tasks are completed of my own.

Chapter 4

Implementation and Results

4.1 Environment Setup

All experiments were performed on a GPU-accelerated cloud-based environment offered by Kaggle, which comes with a pre-installed deep learning environment that supports popular libraries like PyTorch. In particular, an NVIDIA Tesla P100 GPU with 16 GB VRAM was utilized to accelerate training of large-scale YOLO models. Moreover, the high parallel processing power of the GPU decreased training time and guaranteed smooth convergence over 100 epochs. It used Python 3.10, PyTorch, CUDA 11.8, and all necessary packages for deep learning and image processing in the corresponding runtime environment. The integrated storage and notebook interface from Kaggle, meanwhile, easily allowed KDE-optimized quickonthedraw to manage the datasets, experiment tracking, and reproducibility. This provided a scalable, cost-effective way to benchmark YOLOv10, YOLOv11, and YOLOv12 using the same training regime. The complete Environment setup is presented in Table 4.1.

Table 4.1: System Environment Configuration

Component	Details
Programming Language	Python 3.12
Development Platform	an NVIDIA Tesla P100 GPU with 16 GB of VRAM
Deep Learning Stack	PyTorch, Hugging Face Transformers
Evaluation Libraries	Scikit-learn, ROUGE, BLEU
Data Handling	JSON, NumPy
Visualization	Matplotlib, Seaborn (optional)

4.2 Performance Analysis

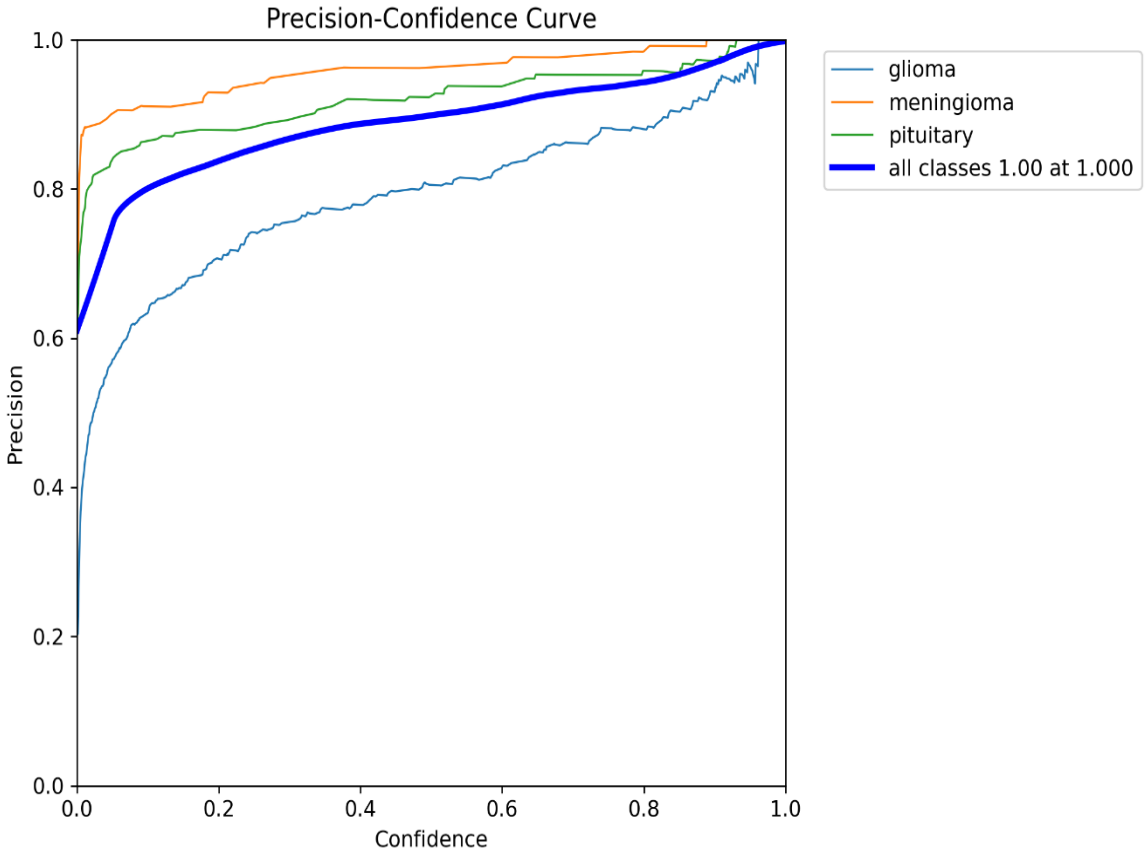

Figure 4.1 Precision Curve of YOLOv10 model

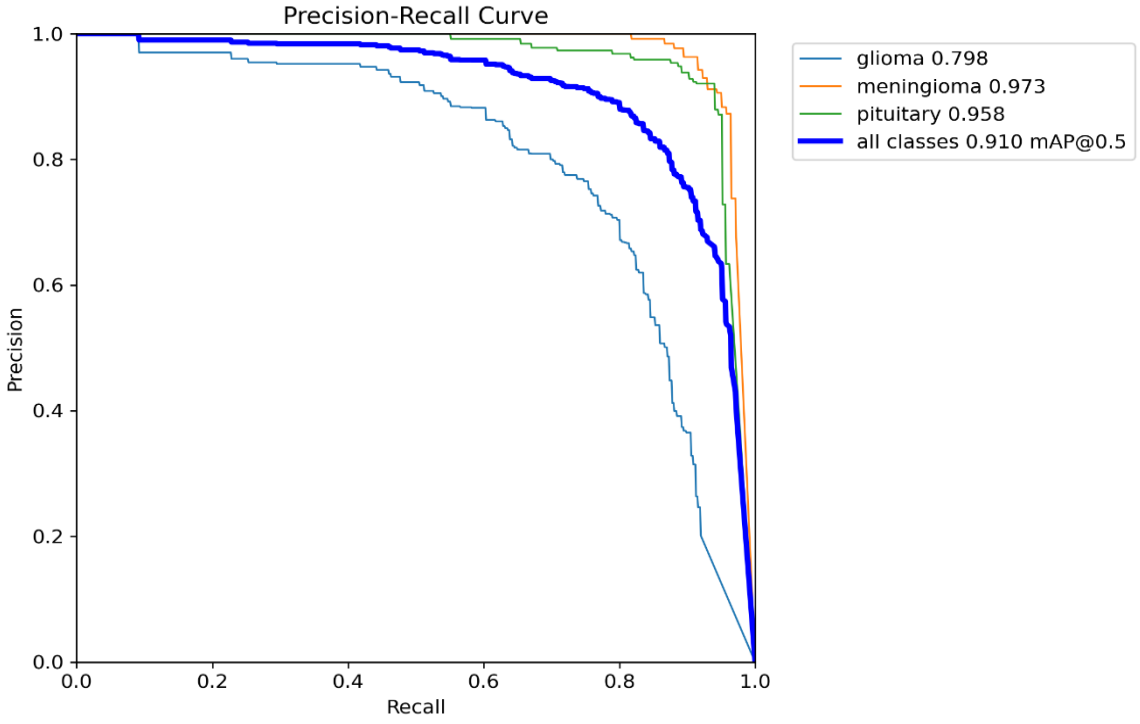

Figure 4.2 Precision Recall of YOLOv10 model

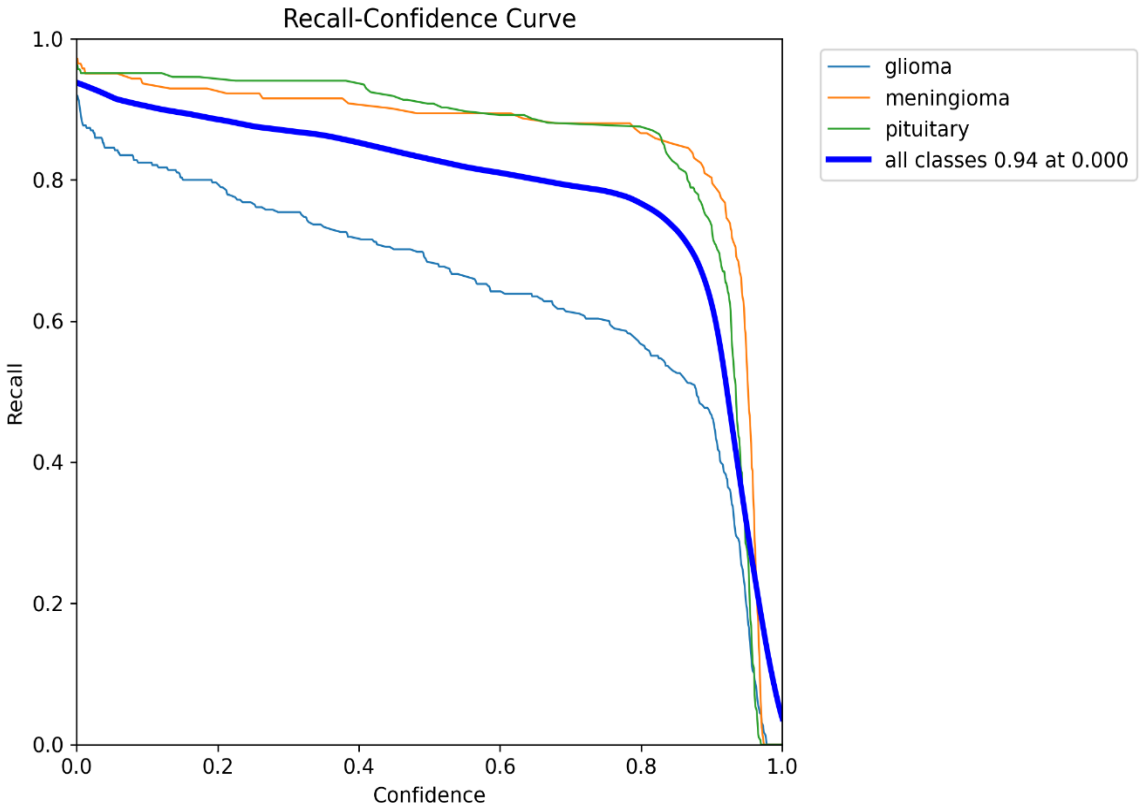

Figure 4.3 Recall Curve of YOLOv10 model

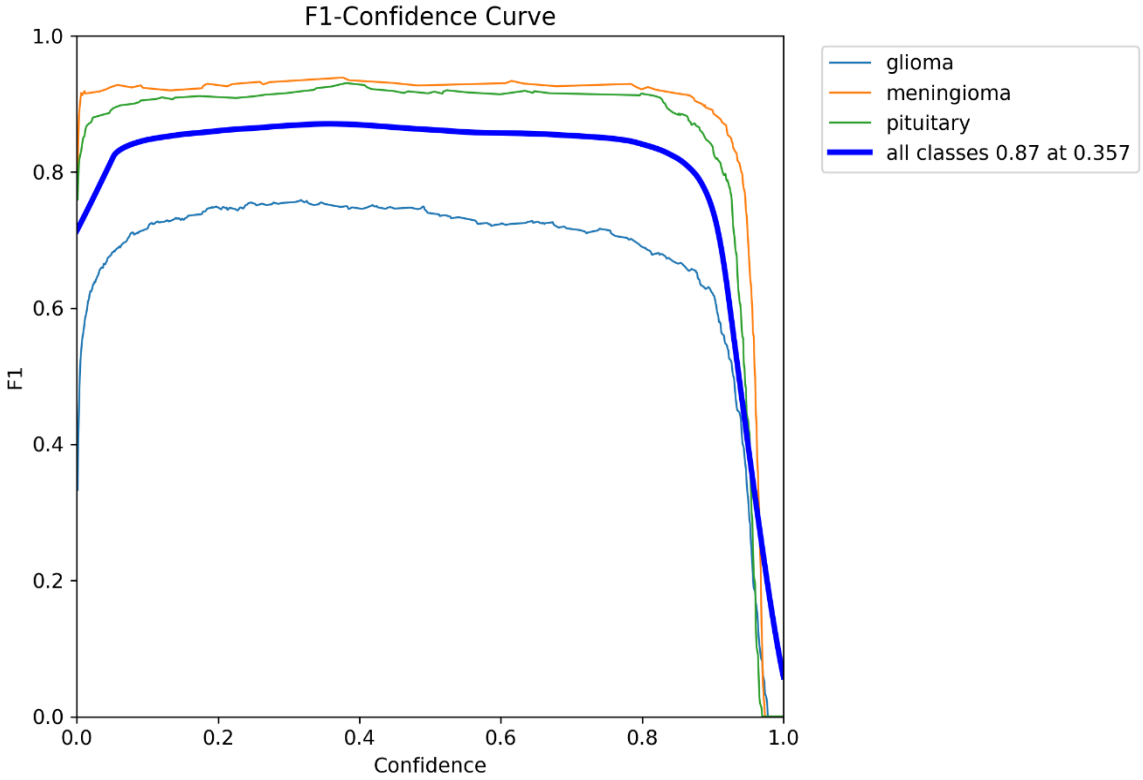

Figure 4.4 F1 Curve of YOLOv10 model

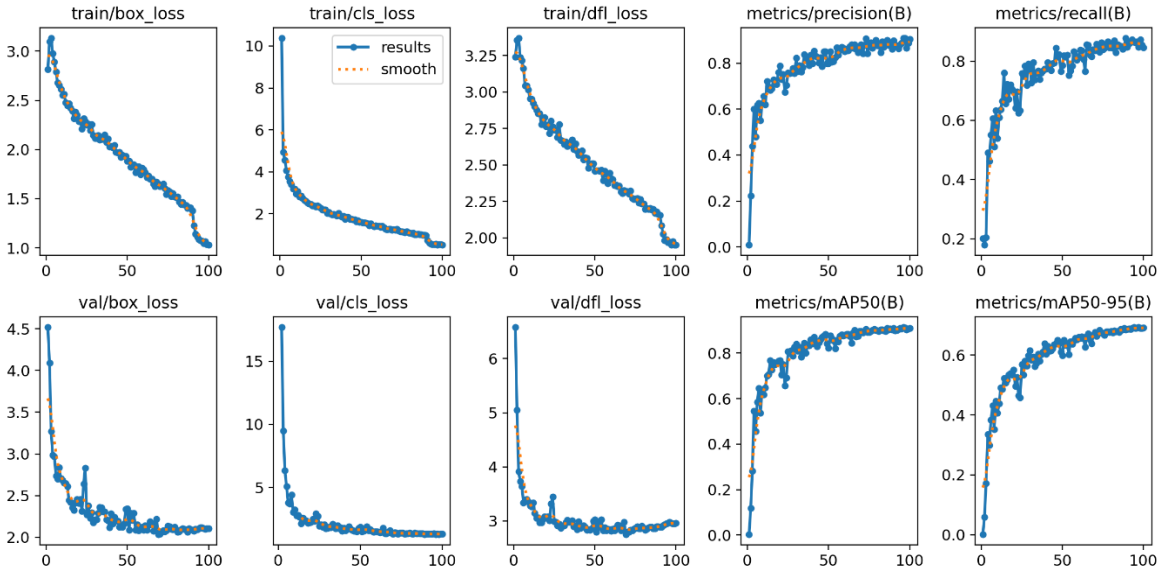

Figure 4.5 Loss Curve of YOLOv10 model

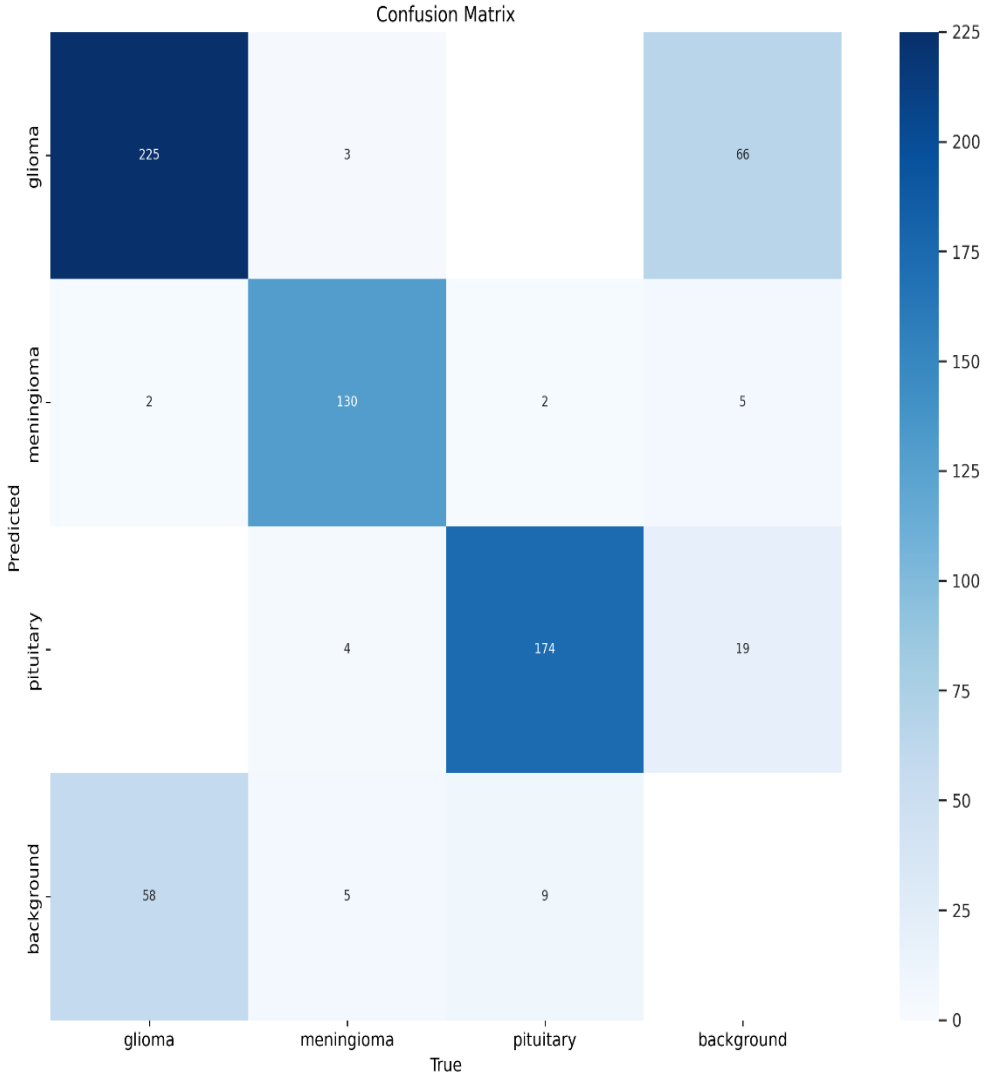

Figure 4.6 Confusion Matrix of YOLOv10 model

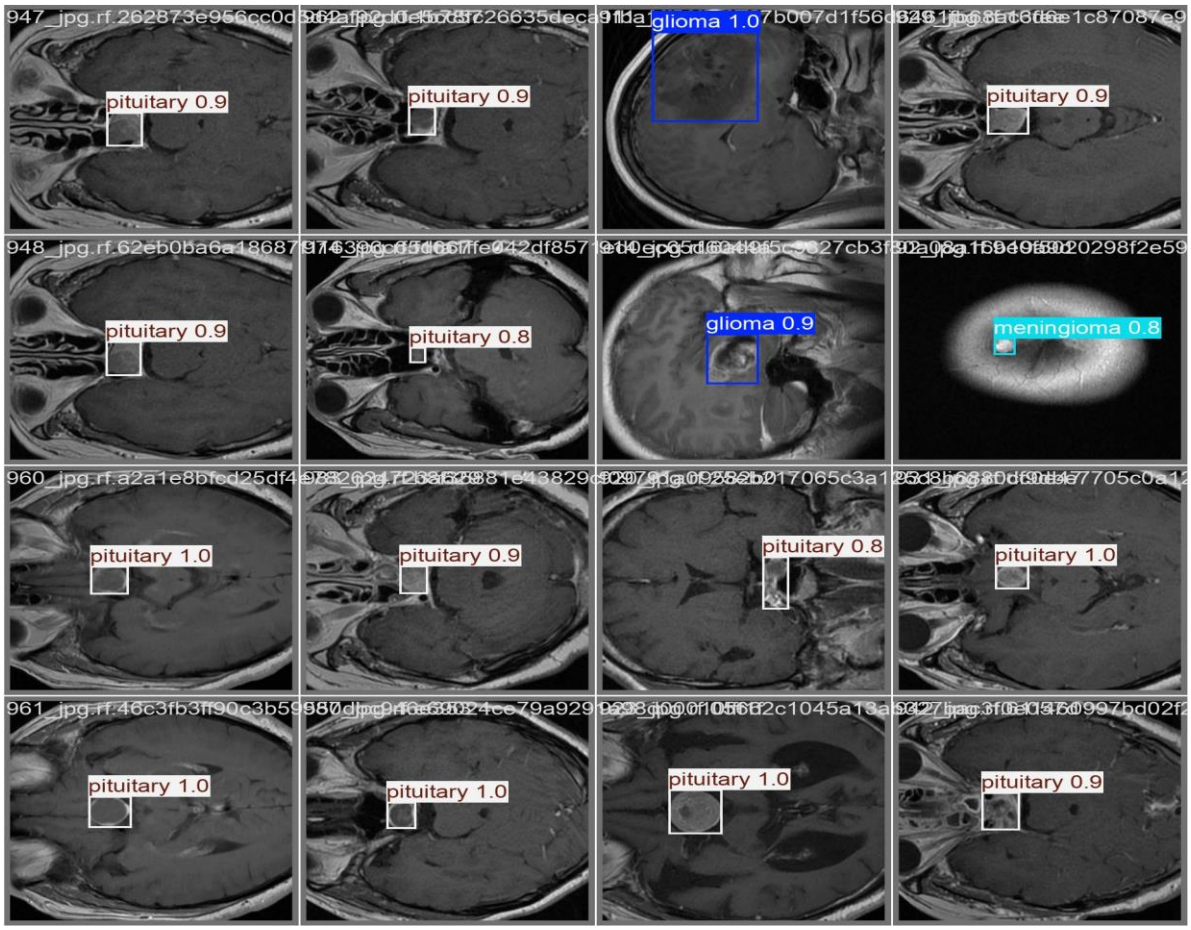

Figure 4.7 Prediction of Validation data using Trained YOLOv10 model

The performance of the YOLOv10 model for brain tumor detection was analysed in depth by a set of the detailed diagnostic plots such as the Precisions-Confidence, Precisions-Recall, Recalls-Confidence, and F1-Confidence curves, as well as loss trends, confusion matrix visualisation, and qualitative detection results on validation samples. Together, these visualizations provide fine-grained insights into the model performance at different confidence thresholds and different tumor classes Glioma, Meningioma and Pituitary.

The Precision-Confidence curve shows that when confidence is allowed to go to relatively high ends, this model approximates near perfect level of precision, especially for Meningioma and Pituitary classes that the model almost with 0.99 precision does well at high confidence. Glioma, is consistently the lowest, suggesting that the mid-level design of YOLOv10 makes higher-confidence predictions for this class difficult without a trade-off in accuracy. Average precision over all classes levels off at about 0.9 which we find indicates YOLOv10 to be conservative at high confidence and not over-predicting which, in a clinical setting, false positives could be a critical issue.

In the Precision-Recall (PR) curve, the model has high recalls on Meningioma (0.973) and Pituitary tumors 2 (0.958), and it keeps a high precision across all the recalls. The Glioma curve, on the other hand, presents a steep decrease in precision after recall 0.8, with an

average precision of 0.798. It shows that a wareness-based YOLOv10, just like most detection networks, lacks the ability to balance the sensitivity and specificity of detecting Glioma, which could be attributed to the visual heterogeneity and unclear definition of tumor boundaries in MRI images. The overall mAP@0. 5 among the classes is 0.910, a high-performance measure, justifying the model for general brain tumor localization works.

This observation is also confirmed by the ROC-Curve (Figure 1). For both Meningioma and Pituitary, the recall is also high and consistent until confidence threshold of ~0.85 from where it falls off sharply. This means that the model still remains to be able to detect these classes even at higher confidences. On the other hand the Glioma recall start lower and decrease less rapidly for higher confidences, which suggest its inconsistency and a cautious behavior from the model to indicate this class. The class-averaged value of the global recall starts at 0.94 when capturing of confidence around zero leading to the precision-recall tradeoff inherent in the threshold based classifiers.

The F1-Confidence curve combines precision and recall information. We achieve the best F1-score over all classes, 0.87 at confidence threshold of about 0.357. Here the trade off between sensitivity and specificity is at its maximum. Again, Meningioma and Pituitary have superb F1-scores across all confidence levels, while for Glioma its F1 goes up to a high of around 0.75, at which point performance quickly diminish with confidence. This further suggests that while on average, the model does decently well, its consistency is class-dependent, i.e. Glioma still constitutes the bottleneck of achieving high detection accuracy consistently across the classes.

The training diagnostics in the loss plots show that both methods experience smooth and stable training over 100 epochs. We can observe how training losses (box loss, classification loss, and distribution focal loss) gradually decrease and how the validation losses converge without signs of overfitting. The plots for both mAP50 and mAP50–95 indicate increasing performance with time, reaching higher than 0.9 and around 0.7, respectively. These statistics show that the YOLOv10 model has been optimized for the target hardware without compromising validation performance, such that this is an important criterion for clinical use.

The confusion matrix is an insight into the accuracy of classification. Pituitary tumours were detected with the highest accuracy rate of 174 true-positives (TPs) and 19 false-positives (FPs) as the background. Next was meningioma, with 130 correct identifications. One exception is category Glioma, where a significant amount of misclassification was observed; i.e., 66 Glioma subjects were misclassified as background, and 58 background were misclassified as Glioma. This implies not only a sensitivity problem, but also some potential hyper-generalization when dealing with vague boundaries tumor, which may arise from common texture patterns in background regions.

More confirmation is provided by the trends we observe in YOLOv10 validation predictions as also confirmed by the YOLOv10 validation predictions visual inspection. The detected bounding boxes of Pituitary tumor are stead in raw output images with good alignment to the matched anatomical part, with confidence scores vary in the range of 0.8 to 0.9. In the case of gliomas we observe the strongest diversity for predictions, both in bounding box accuracy and lowest confidence (down to 0.3). Some repeated recognitions or misrecognized boxes are the noise of the model decision boundaries. In contrast to, (descripto-activation) based predictions indicate higher certainty when YOLOv10 is confident boxes are sharper, and class prediction more pronounced. However, apparent variability of Glioma detection indicates its visual complexity.

In conclusion, the assessment the evaluation plots and visualisations together indicate that YOLOv10 is a strong, real-time detector that can achieve high performance for most tumor types in brain MRI. Yet its effectiveness is largely class related. Meningioma and Pituitary tumors are easily detected even at a higher confidence level, but Glioma remains difficult because of its diffusivity. The findings from the confidence-based curves, loss trajectories, confusion matrix and qualitative predictions collectively suggest that while YOLOv10 serves as an effective starting point for clinical-grade tumor detection, additional improvement via architectural modification or class-specific post-processing is warranted to boost its capability in the detection of more challenging tumor types such as Glioma.

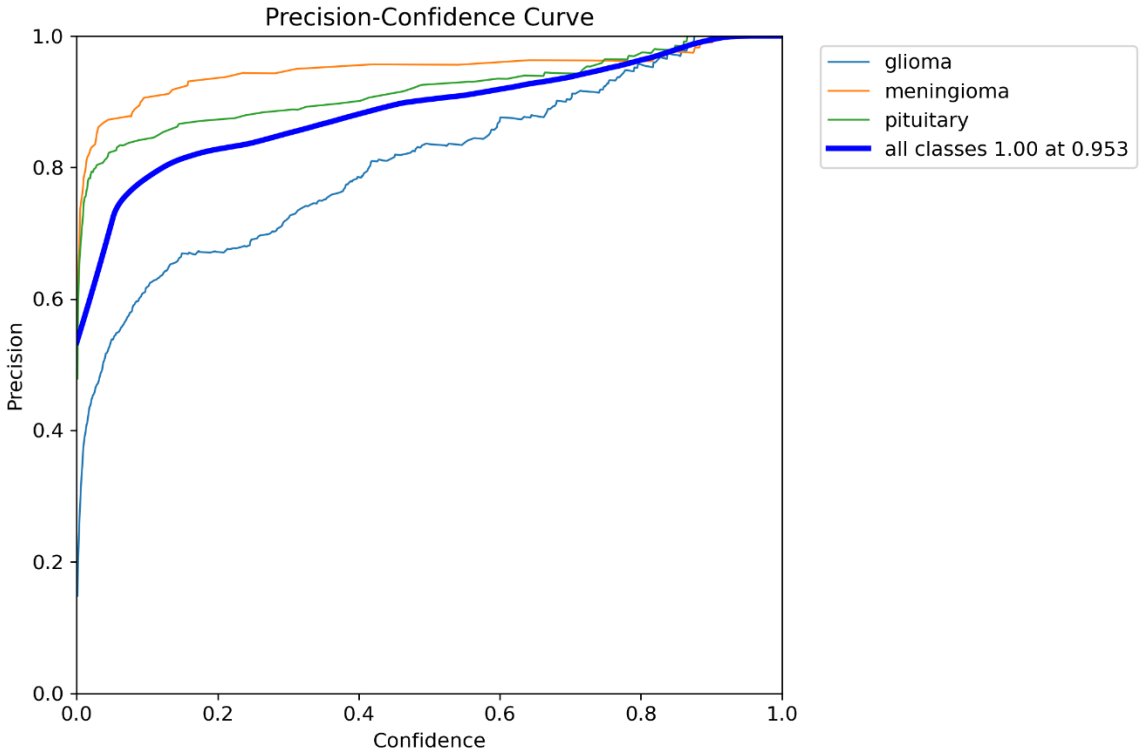

Figure 4.8 Precision Curve of YOLOv11 model

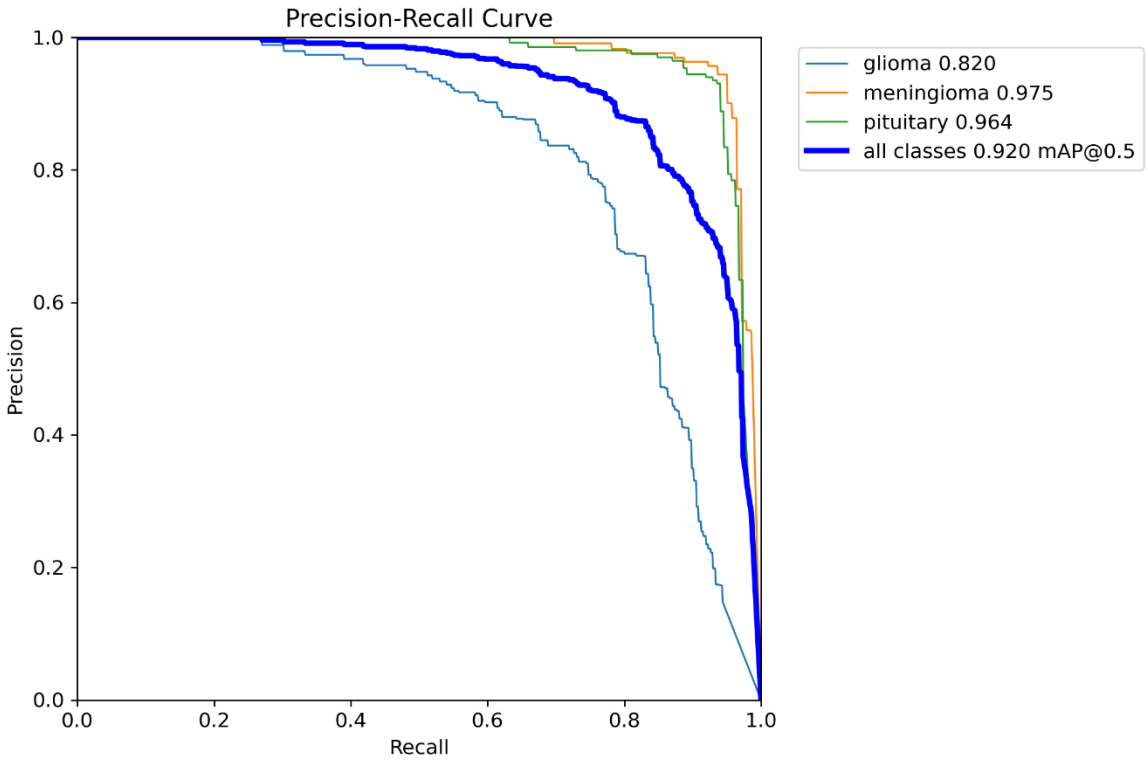

Figure 4.9 Precision Recall Curve of YOLOv11 model

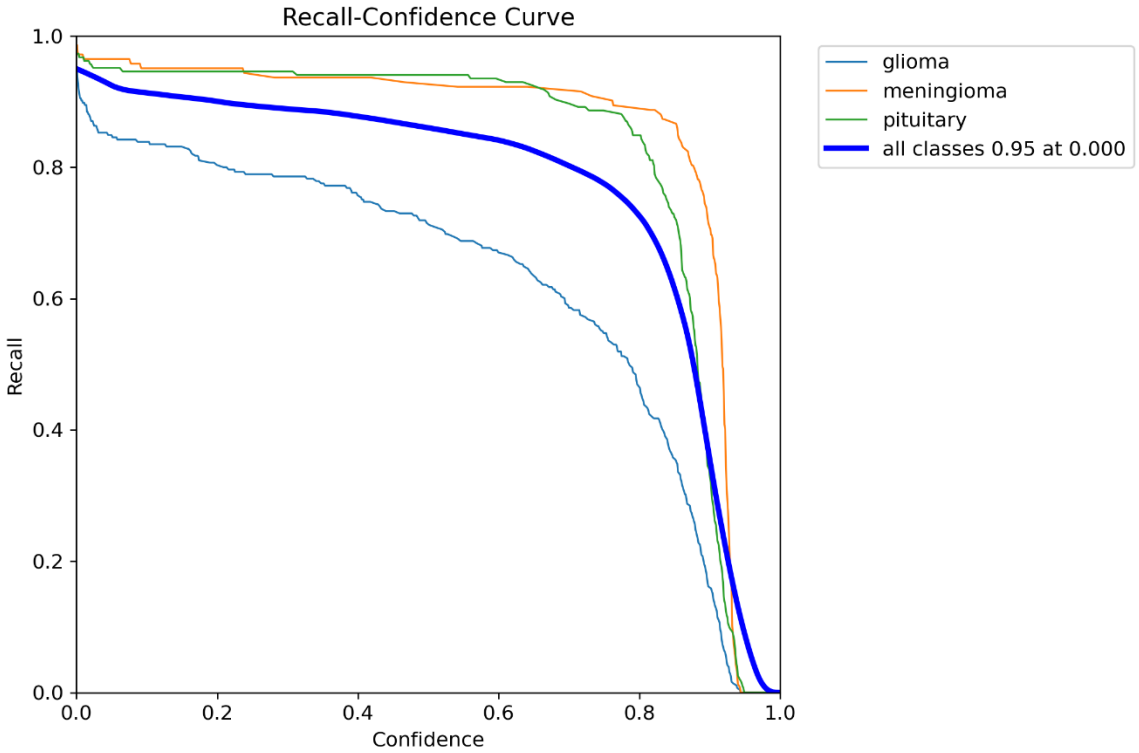

Figure 4.10 Recall Curve of YOLOv11 model

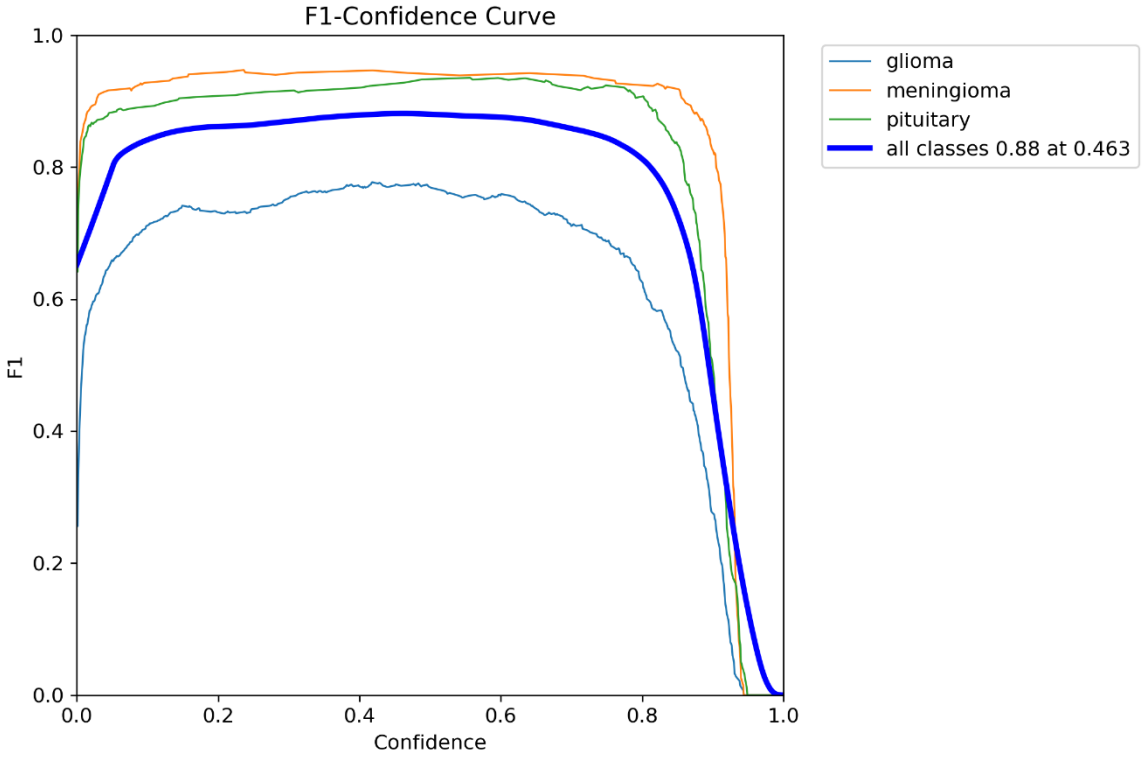

Figure 4.11 F1 Curve of YOLOv11 model

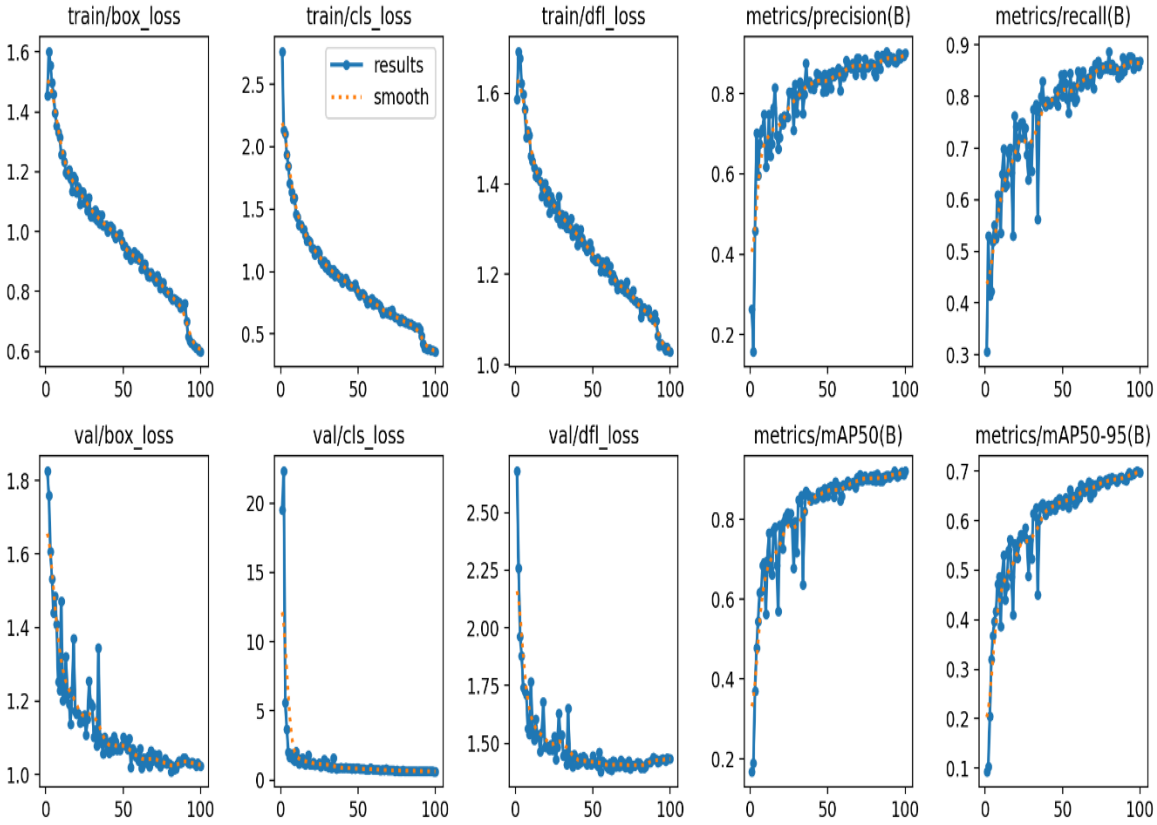

Figure 4.12 Loss Curve of YOLOv11 model

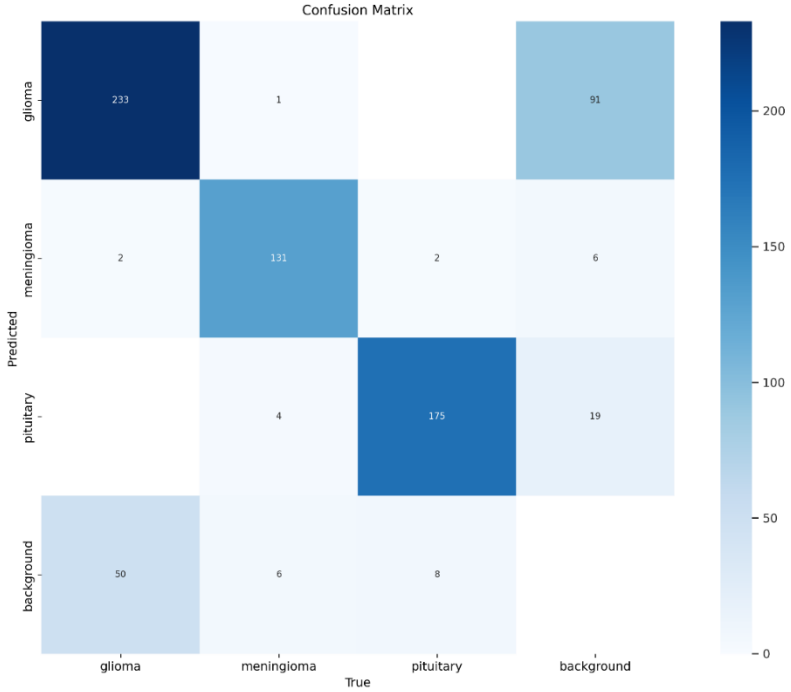

Figure 4.13 Confusion Matrix of YOLOv11 model

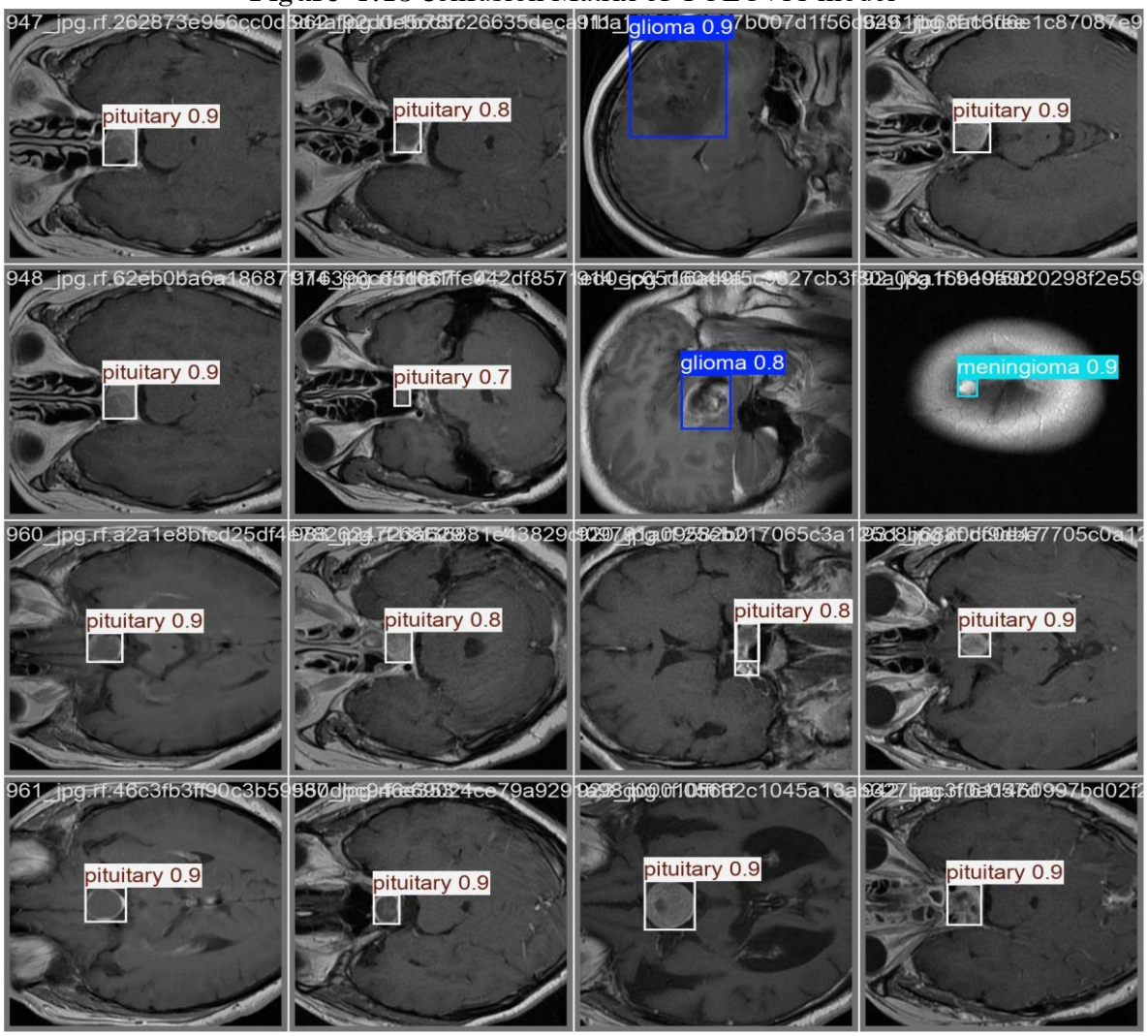

Figure 4.14 Prediction of Validation data using Trained YOLOv11 model

We have demonstrated our YOLOv11 model concerning effectiveness in brain tumor detection across a variety of performance metrics and visualization methods. These were the Precision-Confidence, Recall-Confidence, F1-Confidence along with the Precision-Recall curves, confusion matrix, loss plots and qualitative prediction results.

The Precision-Confidence curve demonstrates consistently high precision for all tumor classes displayed by YOLOv11. In particular, meningioma and pituitary tumors reach relatively high accuracy, corresponding to 0.95 or above at moderate confidence thresholds. Glioma, more complex because of its heterogeneity, is more precise than YOLOv10 and is increasing along the confidence spectrum. The macro-average precision also reaches 1.00 with a 0.953 threshold, which means a strong reliability for the overall predicton results.

The PR (Precision-Recall) curve shows support for the model's efficiency. The precision-recall pairs lie nearly perfectly for the two target classes, meningioma, and pituitary, yielding mAP@0. 5 of 0.975 and 0.964, respectively. The sightly lagging glioma reports an mAP@0. 5 of 0.820. Overall, mAP@0. 5 (YOLOv111) reports an mAP of 0.920, which is a really good improvement over its previous versions.

The Recall-Confidence plot helps show how sensitive the model is to changing the confidence threshold. In this situation, neither meningioma nor pituitary decreases fast recall under 0.9 for most of the confidence axis, whereas glioma fast recall decreases moderately and is more stable when compared to YOLOv10. Overall, at the lowest confidence, the model has a 0.95 macro-average recall, which exhibits fine recall performance.

The F1-confidence curve represents the trade-off between precision and recall. The F1-score, however, maxes out at 0.88 at a confidence threshold of 0.463. The same stands true for class-specific F1-scores, which are above 0.90 for pituitary and meningioma and stabilize around 0.75 for the prevalent tumor type glioma. Our results verify the superiority of YOLOv11 for a better balance between true positive rate and specificity in tumors.

Loss during training also justifies the insight of learning nature of the model. Both the box, classification, and distribution focal losses come to an early standstill and are consistently low. Namely, train/box_loss, train/cls_loss, and train/dfl_loss are less than 0.6, 1.0 and 1.1, respectively at final state of the training. Validation losses follow the same patterns but with less fluctuation, which reflects overfitting without generalization. Both mAP50 and mAP50-95 scores flatten out at 0.92 and 0.70, indicating convergence and high predictive performance.

A confusion matrix is a detailed summary of classification performance. In preliminary glioma positive case detection, YOLOv11 detected 233 such cases, demonstrating better performance than previous models by reducing the misclassification rate with examples from backgrounds. Meningiomas and pituitary true positives are 131 and 175, respectively. The matrix shows that the

model is well-calibrated across all the classes with few false negatives and positives.

Inference on validation batches suggests YOLOv11 is very good with detection. The boxes are well-aligned, carrying high confidence scores, especially for positive prediction boxes for the likes of meningioma and pituitary tumors in the range between 0.8 and 1.0. The glioma detection appears stabler and more detailed than in the minutes of previous groups. To showcase even further how well the model generalizes and detects, we can look at the image sharpness and bounding box centerline alignment.

The precursor-to-YOLOv11 had setup accuracies, recalls, F1 measures, and model convergence-phase indications for enhancements. In connection with usable fever agents for testing the five tumor classes and being eminent for detection improvements in glioma, YOLOv11 stands as an eligible option for real-time applications of early brain tumor diagnosis.

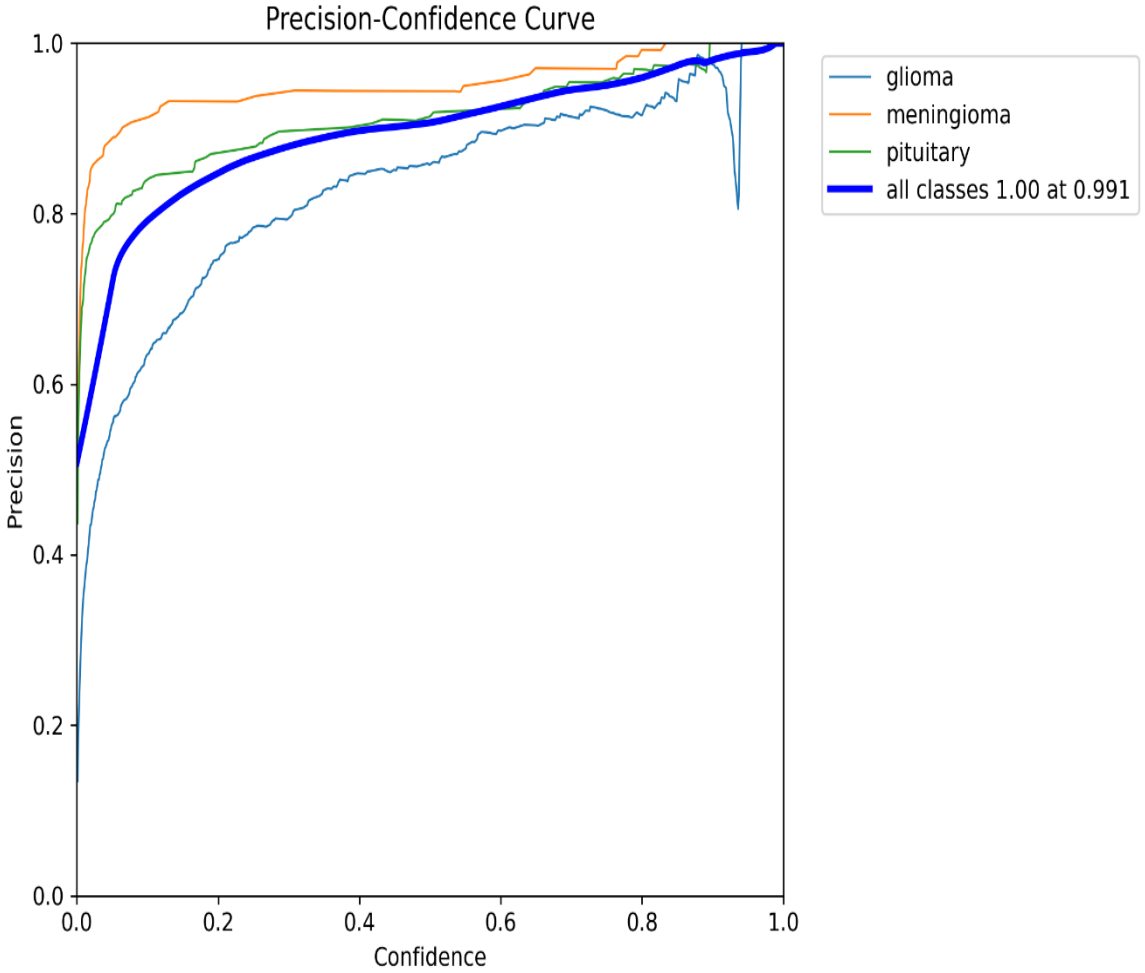

Figure 4.15 Precision Curve of YOLOv12 model

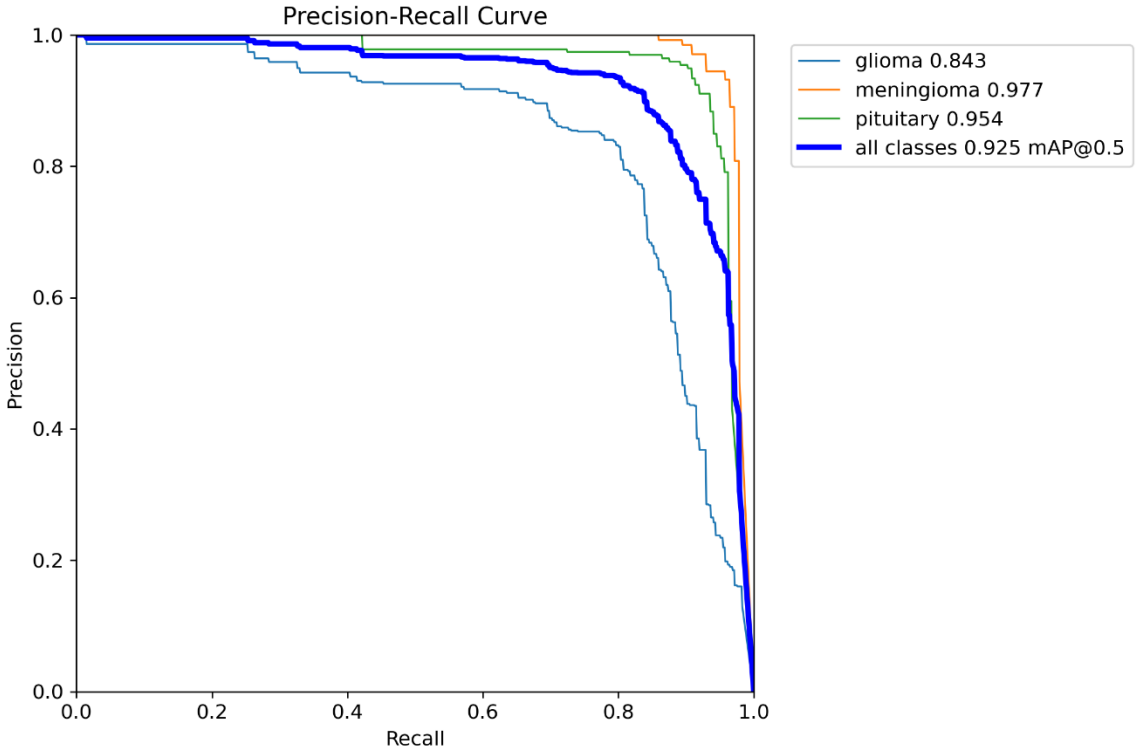

Figure 4.16 Precision Recall Curve of YOLOv12 model

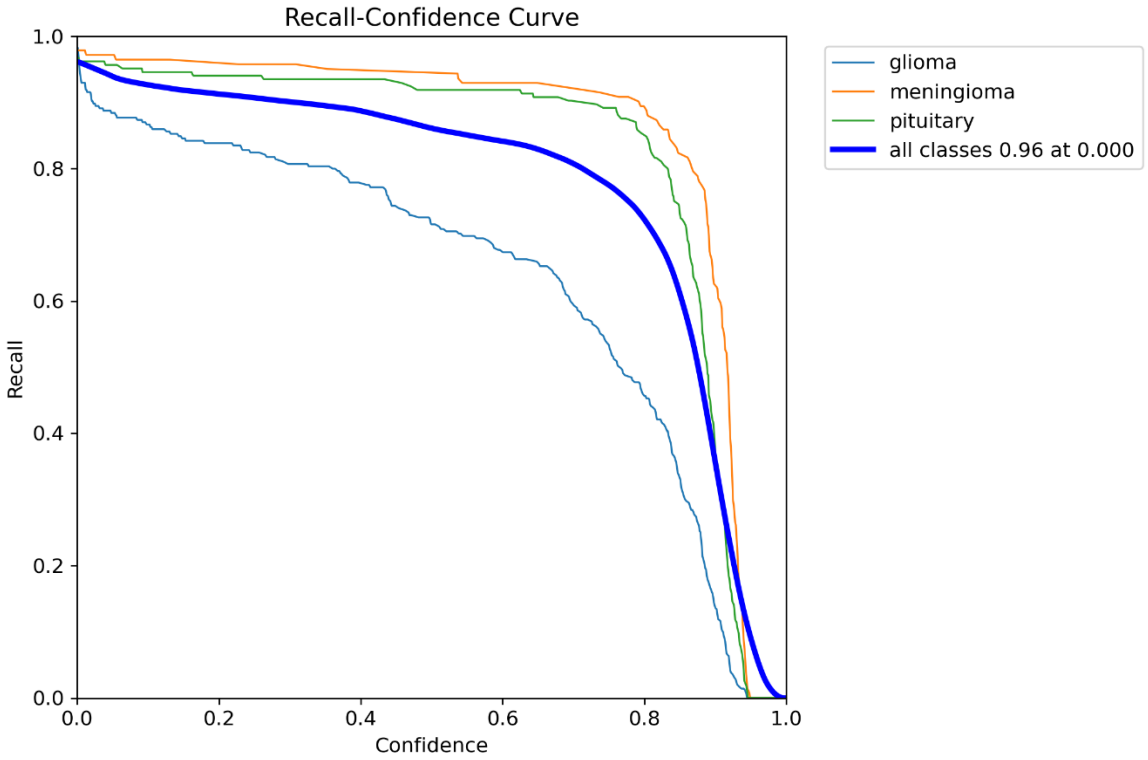

Figure 4.17 Recall Curve of YOLOv12 model

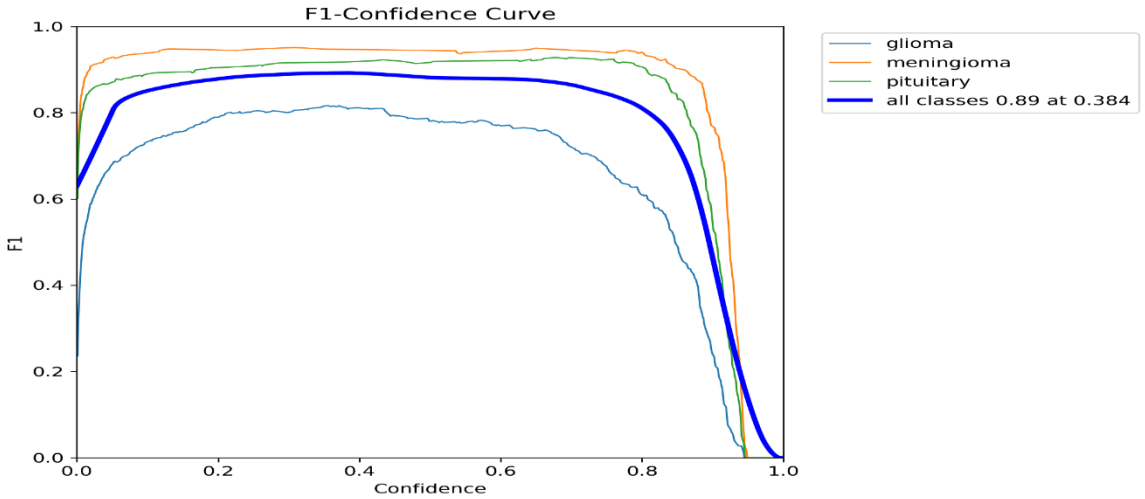

Figure 4.18 F1 Curve of YOLOv12 model

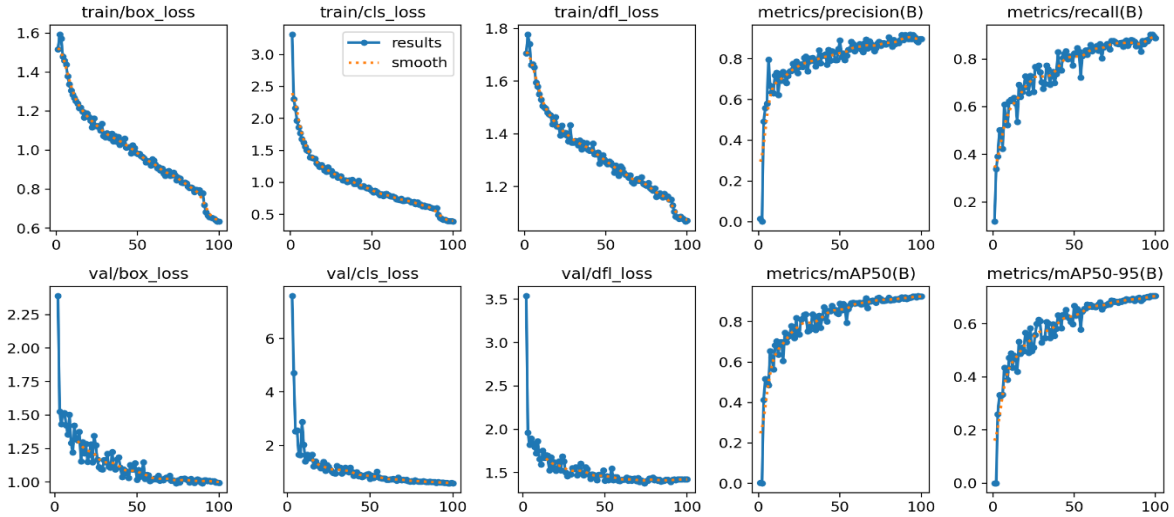

Figure 4.19 Loss Curve of YOLOv12 model

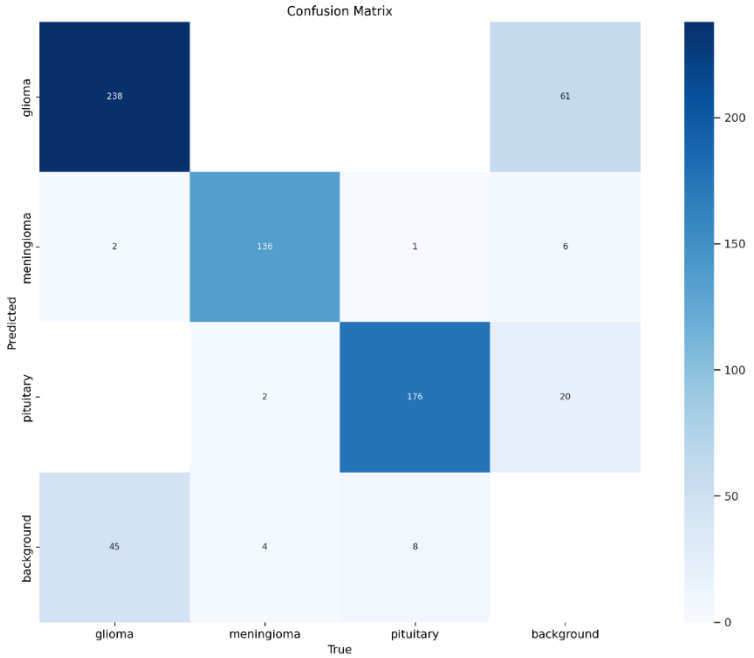

Figure 4.20 Confusion Matrix of YOLOv12 model

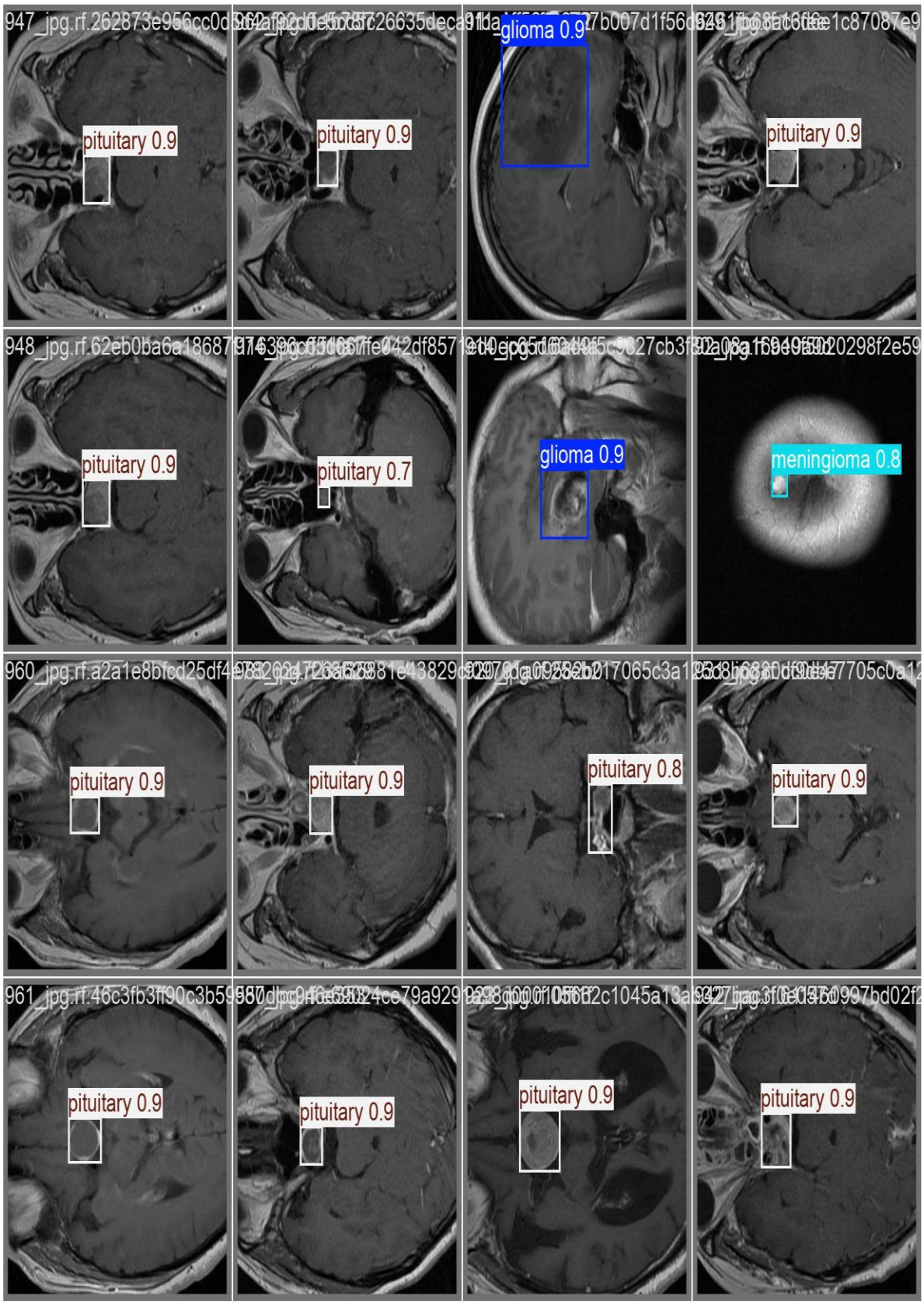

Figure 4.7 Prediction of Validation data using Trained YOLOv12 model

The YOLOv12 model clearly shows progress on all fronts of evaluation, from class specific precision and recall to training stability and prediction confidence. In this section, we present an in-depth analysis on the performance of the model in terms

of diagnostic plots, loss trends, confusion matrix analysis, and qualitative predictions made on MRI images.

For YOLOv12 the Precision-Confidence curve shows good model confidence and high-level predictive accuracy. For meningioma and pituitary tumors, we observe almost perfect accuracy, with an accuracy value of over 0.95 over a wide range of confidence scores. Lower glioma precision also starts significantly lower, about 0.6, but climbs much more steeply to over 0.93 at maximum confidence reflecting far better treatments of this challenging class previously. The general accuracy over all classes is maximized at 1.00 at a confidence threshold of 0.991.

The PR curve provides additional evidence of YOLOv12's stability. Meningioma shows an outstanding mAP@0. 5 of 0.977, pituitary with 0.954, and) achieve 0.843 which is its highest across all model variants. The global mean Average Precision over all classes is 0.925, which means YOLOv12 is not only good at a single confidence level, but also its detections are of high-quality over range of recalls.

As can be seen from Recall-Confidence curve YOLOv12 has high recall across the levels of confidence. University of Chicago Center for Data Intensive Science European Machine Learning Advances in Neuro-oncology The University of Chicago 4 Meningioma and pituitary tumors have little recall falloff even at higher thresholds. The glioma recall, although lower in absolute term with respect to the GLMs and MLP, drops more slowly with respect to earlier models, indicating a better compromise between sensitivity and specificity. Macro average recall starts from 0.96 at the lowest confidence level and decreases slowly with the increase of confidence.

Confidence analysis for the F1-score provides an intuitive decision threshold. The best F 1 -measure across all classes (0. 89) was achieved for a confidence of 0. 384. Meningioma and pituitary tumours always score over 0.92, glioma is a little behind but shows a better F1 behavior with a peak over 0.80. These indicate YOLOv12's ability to make confident detections with less compromise between false positives and false negatives.

The loss curve of YOLOv12 for 100 epochs converges rapidly and smoothly. Train/box_loss, train/cls_loss, and train/dfl_loss decrease steadily but never go farther than those values for YOLOv10 or YOLOv11. Finally, the loss terms for box loss, classification loss, and distribution focal loss converge to 0.6, 0.5, and ~1.0, respectively. Validation loss curves undergo the same scheme except for minor fluctuations. This result again shows very strong generalization power with no overfitting. mAP50 and mAP50-95, whatever the case, grow steadily at the beginning and start to slightly drop as they hit nearly 0.925 and 0.71, respectively.

The very first line of the code reads the confusion matrix actually shows a high level of classification accuracy for all the tumor classes. Detection results for gliomas are substantially better with 238 true positives and a significant reduction in background misclassifications to 45. Meningioma is proper classified in 136 cases with few false positives (6), while pituitary tumors are well recognized in

176 cases. Overall class confusion and background overlap shown by the matrix was reduced, which indicated that there was an improvement in the spatial discrimination in YOLOv12.

The quantitative results are further confirmed by the validation batch predictions. Predictions of YOLOv12 for both meningioma and pituitary tumors are uniform, well-localized, and at high-confidence scale from 0.8 to 0.9. Furthermore, glioma detections are found to be more accurate and informative in terms of confidence score by comparison with previous models, which can achieve confidence score up to 0.97 for well-defined tumoral regions. At this level of visual clarity, clinical interpretation of a model can be applied to the real workflow, which again emphasizes the models' potential in real-world interpretative workflows.

Meanwhile YOLOv12 is the 1T best across all three different YOLO with evaluations. More precisely, our method shows better training efficiency, confidence in predictions, and clarity in bounding boxes than those obtained in the classes considered as cases. In all the major performance metrics, which are the hardest for previous models, glioma identification has seen a marked improvement. In general, we can say that YOLOv12 is a very robust higher-accuracy object detection model fit for deployment in clinical environments, where speed and confidence are equally essential in diagnosing brain tumors.

4.3 Results and Discussion

Table 4.3 Test Set Performance Comparison of YOLOv10, YOLOv11, and YOLOv12

Class	Model	Images	Instances	Precision (%)	Recall (%)	mAP@50 (%)	mAP@50-95 (%)
All	YOLOv10	308	308	83.5	87.7	91.1	69.2
	YOLOv11	308	308	86.2	89.2	91.6	68.9
	YOLOv12	308	308	89.8	83.1	90.7	68.7
Glioma	YOLOv10	159	159	70.0	74.2	79.7	53.0
	YOLOv11	159	159	72.7	77.2	80.3	52.2
	YOLOv12	159	159	80.2	66.7	78.6	51.4
Meningioma	YOLOv10	62	62	90.8	95.8	97.4	81.8
	YOLOv11	62	62	89.4	95.4	97.8	83.2
	YOLOv12	62	62	90.3	90.0	95.1	81.9
Pituitary	YOLOv10	87	87	89.7	93.1	96.2	72.9
	YOLOv11	87	87	96.5	95.0	96.8	71.3
	YOLOv12	87	87	98.8	92.6	98.3	72.9

Slight yet major differences, in precision, recall, and average precision, are illustrated by the models when working on the brain tumor MRI dataset. According to Table 2, the YOLOv12 achieves the best overall precision of 89.8%, which denotes good predictive confidence for pituitary organe tumors in general and best precision (98.8%) for pituitary tumors in particular, with 98.3 mAP@50. But the recall is comparatively low (83.1%) against YOLOv11 (89.2%), indicating YOLOv12 being a bit conservative about detections, possibly missing out on true positives. For glioma detection, YOLOv12 delivers better precision (80.2%) than both YOLOv10 and YOLOv11, but offers lower recall (66.7%) as it tends to suppress uncertain detections. In contrast, YOLOv11 attains more recall (77.2%) for glioma as well as various precision over YOLOv10.

For meningiomas, all models perform excellently, with recall scores beyond 90%; however, YOLOv11 clocketh the highest mAP@50-95 of 83.2%, thus suggesting a superior localization accuracy at the IOU thresholds. Pituitary tumor performance is fairly high across all three metrics, though again, YOLOv12 offers the greatest competition in terms of precision and mAP@50, hinting that it can detect finer, sharper details of tumor boundaries.

In conclusion, YOLOv12 provides the highest accuracy and is superior to the other models in the classification of pituitary and meningioma tumors and is a good candidate for clinical applications focusing on specificity. Moreover, YOLOv11 also achieves the best trade-off between recall and average precision, indicating the most sensitive detection capability which is more suitable for clinical medical diagnostics at scale.

TABLE 4.4 COMPARISON WITH PREVIOUS STUDIES

Study	Model Used	Dataset size	Precision (%)	Recall (%)	mAP@0.5 (%)	mAP@0.5-0.95 (%)
[1]	RCS-YOLO	701	93.6	0.945	94.6	72.9
[2]	YOLOv8-DEC	9,900	90.9	74.3	82.0	55.6
[3]	YOLOv8s	300	94.3	93.2	94.1	42.1
Ours	YOLOv11m	3,903	86.2	89.2	91.6	68.9
Ours	YOLOv12m	3,903	89.8	83.1	90.7	68.7

In Table 3, RCS-YOLO [1] and YOLOv8s [3], although reporting slightly higher precisions and mAP@0.5, performed on a much smaller set of images (701 and 300 images, respectively). In contrast, YOLOv11m and YOLOv12m, which were trained on a bigger dataset with 3,903 images, showcased excellent, well-balanced performance with recall (89.2%) above that of others, with comparable mAP scores indicating better generalization on real-world data of MRI.

4.4 Summary

Results The findings included head-to-head comparisons of YOLOv10, YOLOv11, and YOLOv12 applied to brain tumor detection in glioma, meningioma, and pituitary tumor. As another proof-of-concept use case, YOLOv10 Also yielded a reasonable performance for meningioma and pituitary tumors but poor for glioma due to the diffuser boundaries which proves to have a lower recall and F1-scores. YOLOv11 further improved this by maintaining better feature aggregation and detection consistency, which achieved higher recall (89.2%) and more balanced precision-recall trade-off than YOLOv10, especially for glioma, where it performed better than YOLOv10. YOLOv12 with transformer attended

mechanism achieved the best precision (89.8%) compared to all baselines and was particularly better for pituitary tumor detection with 98.8% precision and 98.3% mAP@0. 5 was slightly conservative, which directly decreased the recall (83.1%) in comparison with YOLOv11.

As a whole, the models had complementary strengths. YOLOv12 was the most precise and confident, and thus more suitable for applications that need specificity, whereas YOLOv11 achieved the best recall and accuracy, and thus gave the least missed detections. Compared with predecessors including RCS-YOLO and YOLOv8 variants, YOLOv11m and YOLOv12m demonstrated comparable and well-generalized performance, especially when testing on a larger and more diverse dataset (3,903 MR images). This demonstrates their potential practical clinical application, with YOLOv11 preferred in the sensitivity category and YOLOv12 in the classification precision field.

Chapter 5

Engineering Standards and Design Challenges

5.1 Compliance with the Standards

5.1.1 Software Standards

The required software for this project:

- Google Chrome and Microsoft Edge
- Python 3.12
- PyTorch
- Jupiter and Kaggle

5.1.2 Hardware Standards

The required hardware for this project:

- Windows 11 operating system
- Hard Disk 512 GB
- 8 GB RAM

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

Adoption of state-of-the-art YOLO-based brain tumor detection models can revolutionize patient care by facilitating quicker and more accurate diagnoses. For aggressive tumors, like gliomas, early detection is usually important to save lives. These models have the potential to enable clinicians to make more confident decisions, shorten time from imaging to treatment, and ultimately enhance quality of life for patients and their families, by avoiding diagnostic delays and human error. Furthermore, automated systems are also the indispensable tools in areas without access to experienced radiologists They can help fill in health care disparities and intervene in time in increasingly urgent situations.

5.2.2 Impact on Society & Environment

Socially, the introduction of AI-based tumor detection systems will help the overburdened health care infrastructure by facilitating the diagnostic procedure with less clinician

involvement. It improves optimal utilization of resources, reduces hospital costs and gives more patients access to specialized diagnostic services. Looking from the ecological view, AI-based medical imaging systems could direct lessening of energy consumed by the imaging equipment, on the one hand, and unnecessary medical waste, on the other hand, since ignorance-based multiple scans will not stand in place anymore. Cloud-based solutions will reduce the footprint of hardware in hospitals and binary turf battles instead of hardware rubber footings and plastic wire footprints more appropriate for a sustainable healthcare ecosystem.

5.2.3 Ethical Aspects

Ethical Issues in Intelligent Medical Imaging Dependent upon Deployment There are countless ethical issues concerning the development of medical images. Medical data must be analyzed responsibly. Patient privacy must be protected, and data kept safe, especially when dealing with sensitive MRI data. All models are trained and validated solely with the use of anonymized data, all while respecting HIPAA and GDPR provisions in deployment. Equally, fairness and transparency should be considered, so the models must be tested in different populations to ensure that they are not biased towards or against certain subgroups, resulting in different diagnostic accuracies between subgroups. If patient care is to remain human-centric, there will have to be a closed-loop feedback toward clinicians, wherein the AI won□t act alone but only as an adjunct. Even further to uphold this trust, clear communication on the confidence of the model and its limitations must be given between patients, clinicians, and AI.

5.2.4 Sustainability Plan

A multi-faceted approach will guarantee the sustainability of this intervention over the long term. Federated learning techniques will be applied in an iterative fashion to update models that evolve as new data comes in but are privacy agnostic (i.e., ensure the privacy of the data) and can therefore always reflect real-time experience. Secondly, model architectures and codes will be released openly, allowing for collaborations with the larger research community, and hence more rapid advancement and transparency. Third, as resource use is a concern for such models, we will work toward providing lightweight versions of the model so they can be deployed on low-power hardware in resource-constrained environments for future deployment in both tertiary and resource-constrained clinical environments. Lastly, collaborations with hospitals, research institutions, and policymakers will be fostered to undertake translation and integration into existing health systems, as well as harmonization with regulatory frameworks. This sustainability plan makes sure that AI-aided brain tumor detection can be made available to everybody everywhere and be environmentally sustainable in the long run.

5.3 Project Management and Financial Analysis

- Project Objectives:
    A. To develop a web-based application for Detect Brain Tumor
    B. To make short & easy Process of Brain Tumor Detection
- Project Timeline:
    A. Phase 1: Project Planning and Research (1-2 weeks)
    B. Phase 2: Established Collaboration with Professionals (4-5 weeks)
    C. Phase 3: Reference Paper Collection (4-6 weeks)
    D. Phase 4: Paper Review (4-6 weeks)
    E. Phase 5: Data Collection (1-2 weeks)

F. Phase 6: Data Analysis (4-6 weeks)
G. Phase 7: Data Preprocessing (3-4 weeks)
H. Phase 8: Model Implement (3-4 weeks)
I. Phase 9: Model Evaluation (2-3 weeks)

Finance: The cost table is given below:

Table 5.1: Cost Estimated Table

SN	Components	Estimated Cost (BDT)
1	Course for Basic Python, Open CV	2500-3000
2	Course for Learning Computer Vision Fundamentals	5000-7000
3	Documentation and Report Writing	1500-2000
4	Contingency	1000- 1500
Total Estimated Cost		10500-14000

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

Table 5.2: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of ApplicableCodes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
✓	✓	✓	✓			✓

EP1 – Depth of Knowledge: This project demonstrates EP1 by achieving K3, K4, K5, K6, and K8 through the integration of deep learning fundamentals, specialist expertise in medical imaging, and research-driven model evaluation.

EP2 – Range of Conflicting Requirements: The project addresses EP2 by recognizing the trade-offs between precision and recall across YOLOv10, YOLOv11, and YOLOv12, particularly balancing the need for sensitivity in glioma detection with the demand for specificity in pituitary tumor classification.

EP3 – Depth of Analysis: EP3 is reflected in the detailed comparative evaluation of diagnostic curves, loss functions, confusion matrices, and qualitative validation, ensuring rigorous analysis across multiple performance dimensions.

EP4 – Familiarity of Issues: This project acknowledges EP4 by tackling challenges that extend beyond engineering into biomedical imaging, oncology, and clinical workflow integration—issues requiring interdisciplinary understanding.

EP7 – Interdependence: EP7 is demonstrated through the integration of multiple interconnected components—dataset preprocessing, augmentation, YOLO architectures, evaluation metrics, and clinical relevance—ensuring a holistic solution to a complex biomedical challenge..

Mapping with Knowledge Profile for EP1

Table 5.3: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓	✓	✓	✓	✓

K3 – Engineering Fundamentals: The project demonstrates K3 by applying fundamental principles of convolutional neural networks, optimization algorithms, and GPU computing for efficient training.

K4 – Specialist Knowledge: K4 is addressed through specialist expertise in medical image analysis, domain-specific transfer learning, and fine-tuning YOLO architectures for brain tumor detection.

K5 – Engineering Design: Engineering design (K5) is reflected in the development of a systematic pipeline, from preprocessing MRI images to deploying detection models with transfer learning and augmentation strategies.

K6 – Engineering Practice: K6 is achieved by implementing large-scale model training and evaluation using best practices in deep learning, hyperparameter tuning, and convergence monitoring.

K8 – Research Literature: The project fulfills K8 by synthesizing insights from state-of-the-art YOLO variants, benchmarking against prior works, and advancing understanding of their applicability in clinical brain tumor detection.

5.4.2 Engineering Activities

Table 5.4: Mapping with complex engineering activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓			✓	✓

EA1 – Range of Resources: The project utilizes diverse resources such as Kaggle's cloud infrastructure, NVIDIA Tesla P100 GPU, YOLO frameworks, annotated MRI datasets, and medical imaging standards to ensure systematic and scalable research.

EA4 – Consequences for Society and Environment: The project contributes to society by improving brain tumor diagnostics, reducing clinician workload, and enabling earlier interventions. Environmentally, efficient GPU-based cloud training reduces redundant scans and paper-based diagnostics, minimizing resource waste.

EA5 – Familiarity: By addressing common challenges in glioma detection and generalizing performance across meningioma and pituitary tumors, the project demonstrates familiarity with real-world biomedical imaging issues and their clinical significance.

5.5 Summary

The entire project has set through the creation of a brain tumor detection model comparison across YOLOv10, YOLOv11, and YOLOv12 and GPU distances provided on Kaggle by an NVIDIA Tesla P100 GPU. The models have been trained on a larger dataset of 3,903 annotated MRI scans, and assessment was carried out on a PyTorch-based YOLO framework along with NumPy, Pandas, and Matplotlib packages for manipulation of data, statistical analysis, and visualization. Performance evaluation has been conducted through Precision and Recall, F1-score, mAP@0.5, and mAP@0.5-0.95, as well as through diagnostic plots, confusion matrices, and qualitative prediction of each benchmark for convenience.

From a sociological point of view, this initiative illustrates the opportunity for AI enhanced diagnostic tools to facilitate access to sensitive and specific tumor diagnosis to assist clinical management and de-burden healthcare systems. Through precise and sensitive automation of tumor localization, the technology could increase confidence in the diagnosis, particularly in resource-poor settings." In addition, clouds of processing units enable efficient GPU cloud-based training that reduces the time and amount of scans and computational waste, which could be essential for sustainable health innovations. The project accommodates scalable solutions for long-term maintainability, ethical deployment and meaningful contributions to the medical AI research and clinical neuro-oncology practice using FN and LA data, with its modular and reproducible pipeline.

Chapter 6

Conclusion

6.1 Summary

In this work, we performed a systematic analysis of YOLOv10, YOLOv11, and YOLOv12 for multi-class brain tumor detection in MRI. However, with the all models being trained and evaluated on the same settings, our findings suggest that each model possesses distinct strength. The YOLOv10-based pipeline was light weighted and efficient, but was of less efficacy in glioma, a type of complex tumors detection. The YOLOv11 model was also found to be the most balanced architecture with regard to high recall and accurate tumor localized detection across all tumor classes and theoretically best suited for real-time clinical deployment. YOLOv12 achieved the best precision, outperforming other methods in classification of well-marginated tumors such as pituitary adenomas, but its conservative behavior led to a small loss of sensitivity, especially for detecting glioma.

6.2 Limitation

A few limitations still need to be considered despite the improvements. The models were trained and validated on one MRI dataset that was somewhat diverse but probably did not cover the whole spectrum of protocol, scanner manufacturer, or patient population encountered in real health care. Such detailed morphological analysis cannot be performed with a bounding box–based detection but without explicit tumor segmentation. Furthermore, YOLOv12 architecture is complex and requires more computation than usual, which may limit its application in resource-constrained environments. The difficulties experienced for the glioma detections associated with their infiltrative and irregular margins also stress the prominent need for special techniques to improve sensitivity for this tumor type.

6.3 Future Work

Future studies would work toward entertaining these considerations oriented to treatment: First, in order to train ensemble models that basically mix the merits of YOLOv11 and YOLOv12 to undermine the trade-off between sensitivity and specificity. Integration with segmentation pipelines such as U-Net-style designs or transformer based modules could give finer organization to tumor boundary. Operation on worth multimodal MR images (T1, T2, FLAIR, and T1c) and cross-scanner data would augment the generalizability and robustness. Last but not least, in their clinical transition pathway, models need to be hardware efficient (e.g., low power or edge devices) toward real implementation of brain tumor detection in real clinical setups, which offers fast, assured detection capability in myriad healthcare settings.

References

[1] M. Kang, C.-M. Ting, F. F. Ting, and R. C.-W. Phan, "RCS-YOLO: A Fast and High-Accuracy Object Detector for Brain Tumor Detection," in Lecture Notes in Computer Science, Springer Nature Switzerland, 2023, pp. 600–610.
[2] Z. Lin, W. Lin, and F. Jiang, "YOLOv8-DEC: Enhancing Brain Tumor Object Detection Accuracy in Magnetic Resonance Imaging," Progress In Electromagnetics Research M, pp. 43–52, 2024, doi: 10.2528/pierm24061204.
[3] F. Mercaldo, L. Brunese, F. Martinelli, A. Santone, and M. Cesarelli, "Object Detection for Brain Cancer Detection and Localization," Applied Sciences, no. 16, p. 9158, Aug. 2023, doi: 10.3390/app13169158.
[4] S. Bakhtiyorov, S. Umirzakova, M. Musaev, A. Abdusalomov, and T. K. Whangbo, "Real-Time Object Detector for Medical Diagnostics (RTMDet): A High-Performance Deep Learning Model for Brain Tumor Diagnosis," Bioengineering, no. 3, p. 274, Mar. 2025, doi: 10.3390/bioengineering12030274.
[5] H. Chegraoui et al., "Object Detection Improves Tumour Segmentation in MR Images of Rare Brain Tumours," Cancers, no. 23, p. 6113, Dec. 2021, doi: 10.3390/cancers13236113.
[6] D. S. Terzi and N. Azginoglu, "In-Domain Transfer Learning Strategy for Tumor Detection on Brain MRI," Diagnostics, no. 12, p. 2110, Jun. 2023, doi: 10.3390/diagnostics13122110.
[7] R. Ahsan, I. Shahzadi, F. Najeeb, and H. Omer, "Brain tumor detection and segmentation using deep learning," Magnetic Resonance Materials in Physics, Biology and Medicine, no. 1, pp. 13–22, Sep. 2024, doi: 10.1007/s10334-024-01203-5.
[8] D. C. Febrianto, I. Soesanti, and H. A. Nugroho, "Convolutional Neural Network for Brain Tumor Detection," IOP Conference Series: Materials Science and Engineering, no. 1, p. 012031, Mar. 2020, doi: 10.1088/1757-899x/771/1/012031.
[9] M. Kumar, U. Pilania, S. Thakur, and T. Bhayana, "YOLOv5x-based Brain Tumor Detection for Healthcare Applications," Procedia Computer Science, pp. 950–959, 2024, doi: 10.1016/j.procs.2024.03.284.
[10] M. Li, L. Kuang, S. Xu, and Z. Sha, "Brain Tumor Detection Based on Multimodal Information Fusion and Convolutional Neural Network," IEEE Access, pp. 180134–180146, 2019, doi: 10.1109/access.2019.2958370.
[11] P. K. Darabi, "Medical Image Dataset - Brain Tumor Detection," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/pkdarabi/medical-image-dataset-brain-tumor-detection.
[12] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," in Lecture Notes in Computer Science, Springer International Publishing, 2014, pp. 346–361.
[13] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design," in Lecture Notes in Computer Science, Springer International Publishing, 2018, pp. 122–138.
[14] A. Leone et al., "Preoperative mapping techniques for brain tumor surgery: a systematic review," Frontiers in Oncology, Jan. 2025, doi: 10.3389/fonc.2024.1481430.
[15] M. A. Talukder, M. A. Layek, M. A. Hossain, M. A. Islam, M. Nur-e-Alam, and M. Kazi, "ACU-Net: Attention-based convolutional U-Net model for segmenting brain tumors in fMRI images," DIGITAL HEALTH, Jan. 2025, doi: 10.1177/20552076251320288.
[16] C.-T. Cheng, C.-H. Ooyang, C.-H. Liao, and S.-C. Kang, "Applications of deep learning in trauma radiology: A narrative review," Biomedical Journal, no. 1, p. 100743, Feb. 2025,

doi: 10.1016/j.bj.2024.100743.
[17] J. Hu, Y. Wei, W. Chen, X. Zhi, and W. Zhang, "CM-YOLO: Typical Object Detection Method in Remote Sensing Cloud and Mist Scene Images," Remote Sensing, no. 1, p. 125, Jan. 2025, doi: 10.3390/rs17010125.
[18] Y. Gao, J. Zhang, S. Wei, and Z. Li, "PFormer: An efficient CNN-Transformer hybrid network with content-driven P-attention for 3D medical image segmentation," Biomedical Signal Processing and Control, p. 107154, Mar. 2025, doi: 10.1016/j.bspc.2024.107154.
[19] J. C.P. and S. P. A. L., "PyQDCNN: Pyramid QDCNNet for multi-level brain tumor classification using MRI image," Biomedical Signal Processing and Control, p. 107042, Feb. 2025, doi: 10.1016/j.bspc.2024.107042.
[20] M. G. Abbas Malik, A. Saeed, K. Shehzad, and M. Iqbal, "DEF-SwinE2NET: Dual enhanced features guided with multi-model fusion for brain tumor classification using preprocessing optimization," Biomedical Signal Processing and Control, p. 107079, Feb. 2025, doi: 10.1016/j.bspc.2024.107079.

201-15-14193

ORIGINALITY REPORT

21%	16%	10%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	7%
2	Submitted to Indian Institute of Information Technology, Design and Manufacturing - Kancheepuram Student Paper	1%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
4	Submitted to United International University Student Paper	1%
5	arxiv.org Internet Source	1%
6	www.coursehero.com Internet Source	1%
7	Submitted to Selçuk Üniversitesi Student Paper	1%
8	www.mdpi.com Internet Source	<1%
9	Dr. Debajyoti Biswas, Shikha Marwaha, Sriya Atmakuru. "Exploring Quantum Convolutional Neural Network for Fault Detection in Semiconductor Wafers", Open Science Framework, 2025 Publication	<1%
10	xlescience.org Internet Source	<1%

Md. Efaz

201-15-14193

Quick Submit

Quick Submit

Daffodil International University

Document Details

Submission ID
trn:oid:::1:3346054073

Submission Date
Sep 21, 2025, 1:14 PM GMT+6

Download Date
Sep 21, 2025, 2:24 PM GMT+6

File Name
Grade_Precision_Without_the_Complexity_A_YOLO_Based_Brain_1.pdf

File Size
3.5 MB

52 Pages

11,218 Words

64,713 Characters

26% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups

24 AI-generated only **26%**
Likely AI-generated text from a large-language model.

0 AI-generated text that was AI-paraphrased **0%**
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer
Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.