AI in Brain Tumour Detection: Comparative Analysis of YOLOv10 and PaliGemma2 with Public Perception Insights in Bosnia and Herzegovina

Fatima Hadžiahmetović[1*], Ali Abd Almisreb[1], Che Zawiyah Che Hasan[2], and Alessandro Cantelli-Forti[3]

[1]Department of Computer Science and Engineering,
International University of Sarajevo,
Ilidža 71210, Bosnia and Herzegovina.

[2]Department of Electrical Engineering,
Politeknik Ungku Omar, 31400 Ipoh, Perak, Malaysia.

[3]Radar and Surveillance Systems (RaSS) National Laboratory,
National Inter-University Consortium for Telecommunications (CNIT),
University of Pisa, 56124 Pisa, Italy.

*Corresponding Author's Email: fmuftic@student.ius.edu.ba

Article History: Received 10 July 2025; Revised 17 November 2025;
Accepted 18 November 2025; Published 30 November 2025

Abstract

Artificial Intelligence (AI) has emerged as a significant advancement in healthcare, particularly in medical imaging, where it assists in detecting abnormalities, segmenting lesions, and predicting conditions. However, comparative analyses between conventional object detection models and emerging vision-language architectures remain limited, especially for brain tumour detection. This study addresses this gap by comparing two algorithms: YOLOv10, a widely used object detection model, and PaliGemma2, a newer vision-language model that integrates image and text modalities. Despite the rapid development of AI tools, comparative studies evaluating their effectiveness and public acceptance in healthcare remain scarce. Therefore, this study aims to evaluate the technical performance of AI algorithms in brain tumour detection and to assess societal readiness for their adoption in Bosnia and Herzegovina. Performance was assessed using a labelled magnetic resonance imaging (MRI) dataset and evaluated through accuracy and precision metrics, while public perception was analysed through a survey involving 344 participants. The results indicate that YOLOv10 consistently outperformed PaliGemma2, likely due to its optimisation for object detection tasks, whereas PaliGemma2's multimodal design required greater computational resources. The findings from the survey revealed positive public acceptance of AI in healthcare, accompanied by calls for greater education, careful implementation, and appropriate professional training. Overall, results from this study provide empirical evidence supporting the practical applications of AI models in medical imaging and highlight the importance of integrating ethical and educational frameworks for AI adoption in developing healthcare systems, such as in Bosnia and Herzegovina.

Keywords: Artificial Intelligence, Brain Tumour Detection, Medical Imaging, PaliGemma2, Vision-Language Models, YOLOv10

1.0 Introduction

Brain tumours are abnormal cell growths that disrupt brain function, often

leading to severe neurological impairment or death [1]. Early and accurate detection is cardinal for effective treatment planning and improving patient outcomes. Magnetic resonance imaging (MRI) and computed tomography (CT) scans remain the standard imaging modalities; however, manual interpretation is often time-consuming, subjective, and susceptible to diagnostic variability. These underscore the important requirement for automated and reliable diagnostic support.

Artificial Intelligence (AI), particularly machine learning (ML) and deep learning (DL), has significantly advanced medical imaging by improving the speed and accuracy of tumour detection and classification [2]. Convolutional Neural Networks (CNNs) are widely used for imaging tasks such as classification and segmentation, supported further by transfer learning methods that improve performance on smaller datasets [3], [4], [5], [6]. Object Detection Models (ODMs) such as the YOLO series have shown strong performance in real-time tumour localisation, with YOLOv10 offering enhanced detection of small and overlapping abnormalities due to its optimised architecture [6], [7], [8], [9], [10].

In Bosnia and Herzegovina (BiH), the burden of brain tumours is increasing. In 2022 and 2023, malignant brain tumours ranked among the top ten causes of cancer-related deaths in the Federation of Bosnia and Herzegovina [11], [12]. However, fragmented healthcare data systems limit nationwide visibility. Access to advanced imaging is further constrained by long waiting times in public hospitals and high costs in private institutions, slowing the wider adoption of AI-assisted diagnostic tools. This technological gap creates a pressing need to evaluate AI solutions in both technical performance and societal acceptability within the country.

Although AI models such as CNNs and YOLO have achieved strong results, limited studies have compared traditional object detection models with emerging VLMs for brain tumour detection using MRI data, particularly in resource-constrained or developing healthcare systems. Furthermore, there is a lack of research exploring public perception and societal readiness for AI-based healthcare tools in Bosnia and Herzegovina, despite their increasing relevance to clinical workflows.

Given the rapid evolution of both object detection and vision–language models, this study compares YOLOv10 and PaliGemma2 for brain tumour detection using MRI images and examines public perception of AI's role in healthcare in Bosnia and Herzegovina. The novelty of this study lies in combining a technical model comparison with an assessment of societal acceptance, providing an integrated perspective that is rarely addressed in current literature. The main contributions include: (1) a performance benchmark between YOLOv10 and PaliGemma2 for tumour detection, (2) an analysis of the computational feasibility of VLMs in medical imaging, and (3) empirical insights into public attitudes towards AI in a developing healthcare system.

2.0 Methodology

2.1 Dataset Description

This study utilised the MRI for Brain Tumour with Bounding Boxes dataset from Kaggle, consisting of 5,249 MRI images annotated with bounding boxes in YOLO format [13]. The dataset was divided into a training set (90%) and a validation set (10%), covering four classes: Glioma, Meningioma, Pituitary, and No Tumour. Images were taken from sagittal, axial, and coronal planes, providing variability in anatomical visualisation.

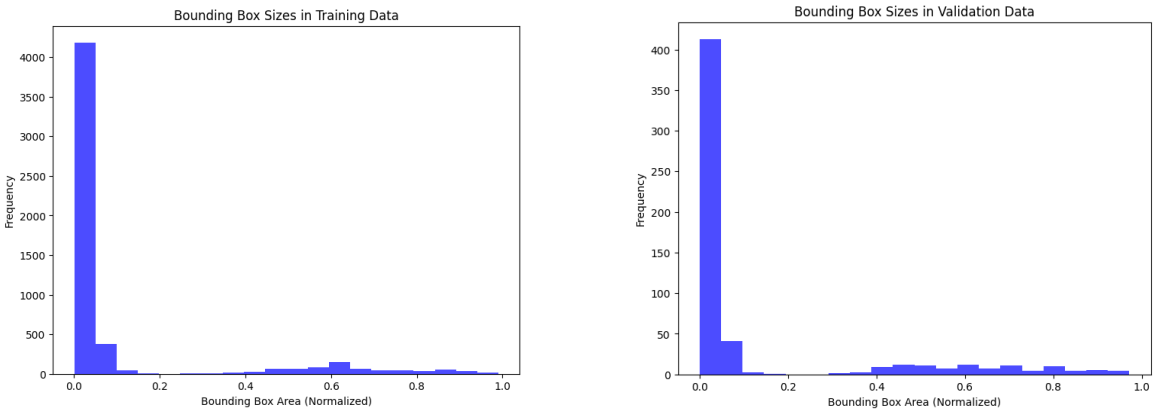
Table *1* summarises the dataset distribution across four tumour categories in both training and validation subsets.

Table 1: Dataset comparison

Class	Training Images	Validation Images	Total Images
Glioma	1,153	136	1,289
Meningioma	1,449	140	1,589
No Tumour	711	100	811
Pituitary	1,424	136	1,560
Total	4,737	512	5,249

Class distribution was imbalanced, with Meningioma and Pituitary tumours dominating, and the No Tumour class considerably underrepresented. Most bounding boxes were very small (normalised areas below 0.1), posing a challenge for tumour detection. Figure 1(a) and Figure 1(b) illustrate the distribution of bounding box sizes in the training and validation sets, showing similar patterns of small-object dominance across both subsets. Most tumours occupy a very small portion of the image, making localisation challenging.

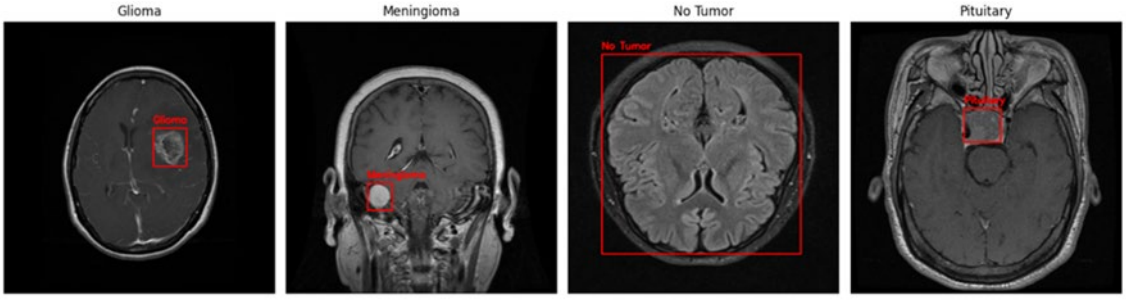
These findings underscore the inherent difficulty of detecting small tumours, which are not only more challenging to identify but also require the model to have a high degree of sensitivity and localisation precision.



(a) (b)

Figure 1: Distribution of normalised bounding box sizes in (a) the training set and (b) the validation set

Bounding box annotations and class labels were used for both detection and classification tasks. Figure 2 (a) – (d) presents examples of the MRI images for each tumour category, highlighting typical bounding box appearances and the diversity of scanning orientations. These samples illustrate the dataset variability essential for training optimisable models.



(a) (b) (c) (d)

Figure 2: Examples of MRI images with bounding box annotations for each tumour category: (a) Glioma, (b) Meningioma, (c) No Tumour, and (d) Pituitary

The sample images shown in Figure (a) – (d) illustrate the diversity in tumour types, sizes, and anatomical locations, underscoring the richness of the dataset. These variations are important for training models that need to generalise across different medical scenarios, including both tumour detection and the accurate identification of healthy scans. The annotations created with bounding boxes and tumour labels provide the necessary framework for object detection models like YOLOv10 and PaliGemma to learn effectively.

2.2 Selection of Models

Two models were selected to balance reliability and innovation. YOLOv10, a real-time object detection model, was chosen for its efficiency and strong performance in small-object detection [7]. PaliGemma2, a newer vision–language model, was selected due to its ability to integrate visual and textual information, offering a novel multimodal approach to brain tumour detection [14], [15]. Comparing these models enables a systematic evaluation of established object detection methods alongside emerging VLM-based pipelines.

2.3 Model Architectures

YOLOv10 introduces several innovations, including NMS-free training through a dual-assignment strategy and an efficiency-optimised architecture incorporating lightweight detection heads and rank-guided block allocation [16]. The architecture is designed to enhance detection accuracy for small tumours while maintaining low inference latency. Figure 3 provides the visualisation of YOLOv10's dual-assignment mechanism, which enables the removal of Non-Maximum Suppression during training.

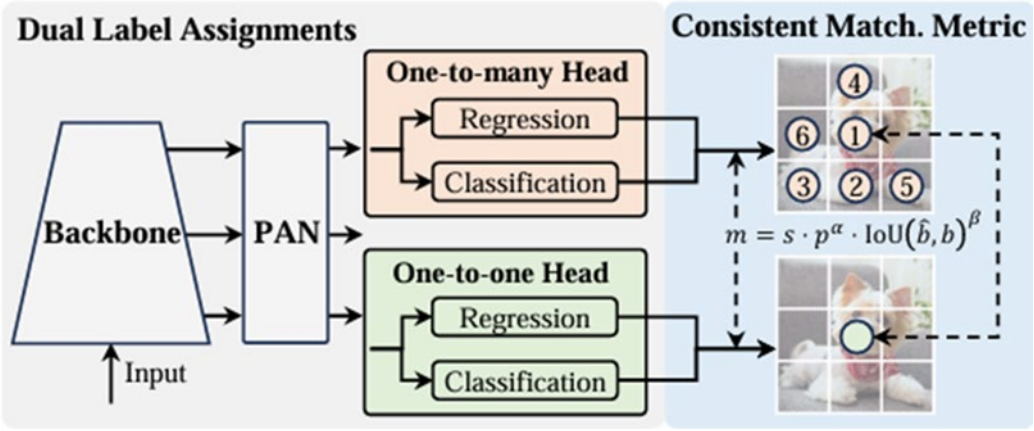


Figure 3: Visualisation of YOLOv10's consistent dual-assignment strategy for NMS-free training [16], showing how positive and negative sample assignments stabilise the optimisation process

PaliGemma2 combines a SigLIP vision encoder with a Gemma-2B language model, enabling cohesive multimodal processing of medical images and text prompts [15]. Figure 4 illustrates the staged architecture, including the vision encoder, cross-modal fusion, and transformer layers, which collectively support high-resolution visual understanding and contextual reasoning.

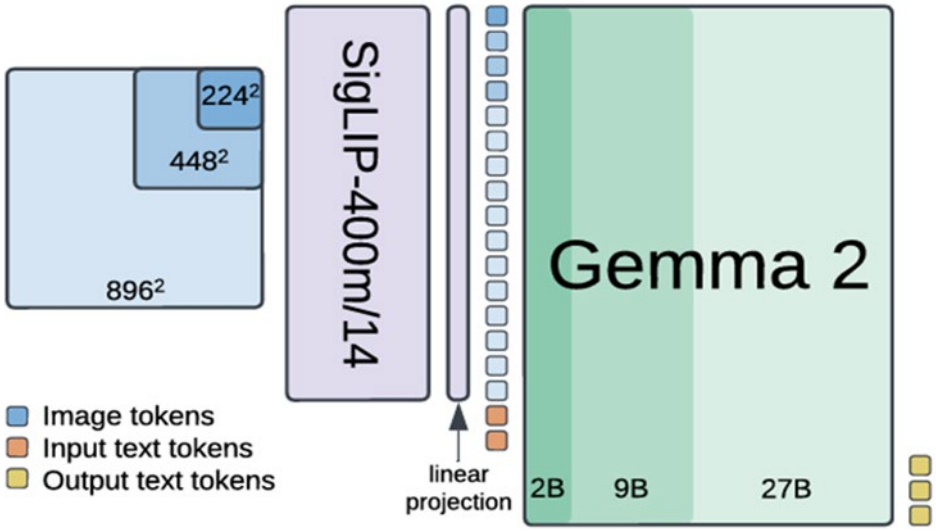


Figure 4: PaliGemma2 architecture [15], demonstrating its modular combination of a SigLIP visual encoder and a Gemma-based language model to support multimodal feature integration

2.4 Data Preprocessing

Custom preprocessing pipelines were developed for each model to meet its specific input requirements. For YOLOv10, images were resized to 640 × 640 pixels, grouped by class, and paired with YOLO-format bounding box labels. Missing annotations were manually added. For PaliGemma2, images were resized to 224 × 224 pixels, and YOLO annotations were converted into JSONL

format to incorporate both visual and textual descriptions. Visual validation checks ensured annotation consistency, and preprocessing aligned the dataset with the expected architecture of each model.

2.5 Model Training

YOLOv10 was initialised with COCO-pretrained weights and fine-tuned for 150 epochs with a batch size of 16 using a Tesla T4 GPU. Training adopted a multi-task loss function incorporating localisation, classification, and confidence objectives, supported by dynamic learning rate scheduling. Fine-tuning concentrated on higher network layers to adapt pretrained features to MRI-specific patterns.

PaliGemma2 was trained in a JAX environment with images resized to 256 × 256 pixels and normalised to the range [−1, 1]. Training combined text prediction and localisation losses, including cross-entropy and Intersection over Union (IoU)-based objectives. The visual encoder was frozen, while transformer layers were fine-tuned to support multimodal fusion. This enabled domain-specific adaptation for tumour detection and classification.

2.6 Evaluation Metrics

Model performance was assessed using both localisation and classification metrics. The IoU and Mean Average Precision (mAP) at IoU thresholds of 0.5 and 0.5–0.95 measured detection performance. Classification metrics such as precision, recall, F1-score, and confusion matrices were evaluated for tumour type prediction. Training and validation loss curves were monitored to assess convergence and detect overfitting.

2.7 Software and Tools

All experiments were conducted in Python 3.10 using Kaggle Notebooks equipped with Tesla T4 GPUs. YOLOv10 was implemented using the Ultralytics YOLO framework (v8.3.15), and PaliGemma2 fine-tuning was performed using JAX and the BigVision library. Supporting libraries included NumPy for numerical computation, Matplotlib for visualisation, Albumentations for data augmentation, TensorBoard for monitoring training progress, and Scikit-Learn for calculating classification metrics.

3.0 Results and Discussion

3.1 YOLOv10: Model Performance

YOLOv10 achieved strong results in brain tumour detection and classification, reaching a precision of 95.3%, a recall of 94.4%, and an F1-score of 94.8%. Detection performance was also high, with mAP@50 at 96.5% and mAP@50-95 at 81.2%. Figure 5 illustrates the confusion matrix, showing minimal misclassification and clear separation between tumour classes. Similar findings were reported by Priyadharshini et al. (2025) [17], who demonstrated that YOLO variants perform robustly in MRI-based tumour interpretation, where YOLOv11 demonstrated superior performance from

other algorithms, achieving 96.22% classification accuracy.

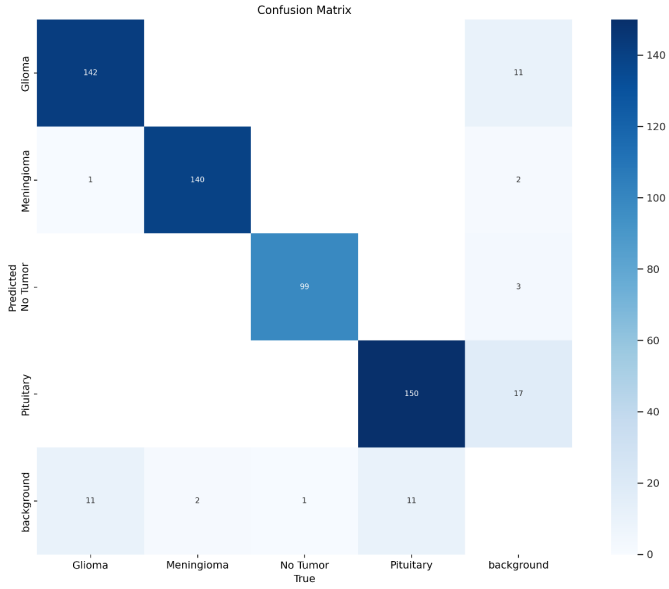


Figure 5: The confusion matrix shows YOLOv10's stability across all four tumour types, indicating strong feature extraction and localisation capability

Training and validation losses shown in Figure 6 decreased consistently over 150 epochs, reflecting effective learning. A slight increase in validation DFL loss toward later epochs suggests minor overfitting risks, but overall generalisation remained strong.

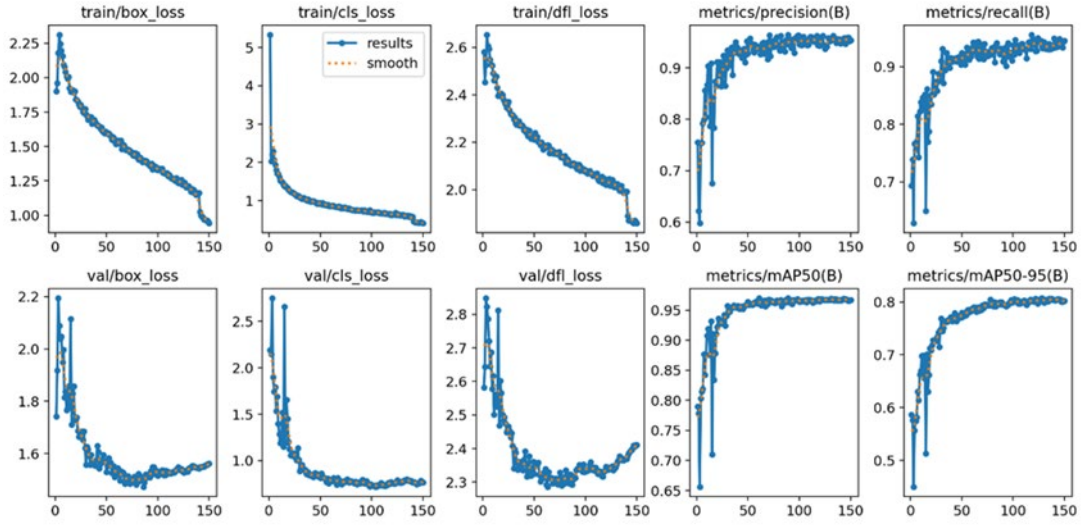


Figure 6: Training and validation loss curves for YOLOv10 indicate stable convergence and effective learning

Loss curves indicate well-converged optimisation, demonstrating YOLOv10's suitability for real-time, high-precision clinical imaging tasks. Similar convergence characteristics are reported by Sapkota et al. (2024)[7], who

highlighted YOLO models' robustness for medical image detection tasks.

3.2 PaliGemma2: Model Performance

PaliGemma2 achieved a precision of 92.5%, a recall of 84.7%, and an F1-score of 88.3%. In the detection task, the model reached a mAP@50 of 89.6% and mAP@50-95 of 62.9%. The confusion matrix in Figure 7 highlights strong classification performance for Meningioma, Pituitary, and No Tumour classes, although some confusion was noted for Glioma.

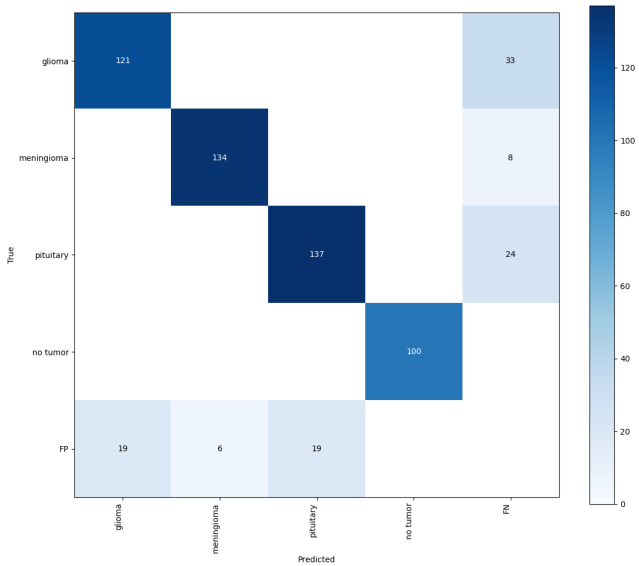


Figure 7: Confusion matrix for PaliGemma2, showing stronger performance on well-defined tumour types and more uncertainty in Glioma classification

Training and validation loss curves shown in Figure 8 demonstrated consistent downward trends, stabilising after initial fluctuations, indicating effective model learning. Although PaliGemma2's results were lower than YOLOv10's, they demonstrate strong potential for applications that integrate visual and textual information in medical imaging.

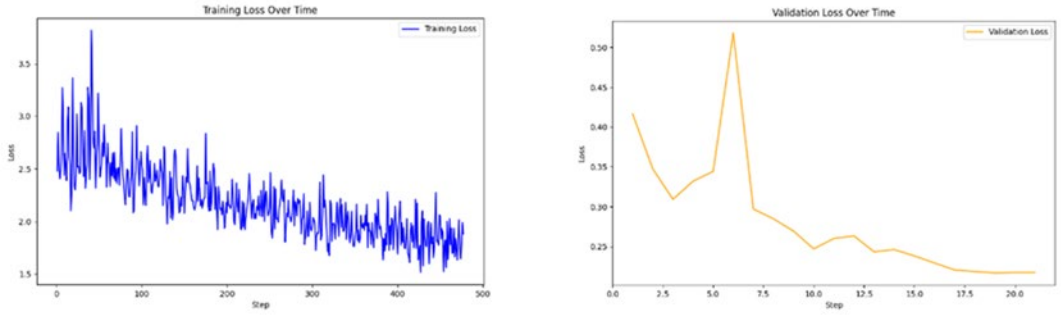


Figure 8: Training and validation losses for PaliGemma2, showing effective optimisation despite higher computational demands

3.3 Comparison Between YOLOv10 and PaliGemma2

As summarised in Table 2, YOLOv10 outperformed PaliGemma2 across all evaluated metrics. YOLOv10 achieved higher precision, recall, F1-score, and detection accuracy, while requiring less training time. This aligns with previous findings of Ragab et al. from 2024 [9] that the YOLO architecture remains highly competitive for small-object localisation in medical imaging. PaliGemma2, however, demonstrated strong potential despite being a general-purpose VLM. Its respectable performance suggests adaptability for medical image analysis, supporting findings by Steiner et al. from 2024 [15], who also reported effective domain transfer using PaliGemma-based models. In contrast, Klein et al. (2024) [18] in their study note limitations of VLMs in specialised tasks due to hallucination risk and computational burden, which matches the challenges observed in this paper.

Table 2: Result metrics comparison between YOLOv10 and PaliGemma2

Metric	YOLOv10	PaliGemma2
Precision	95.3 %	92.5 %
Recall	94.4 %	84.7 %
F1-Score	94.8 %	88.3 %
mAP50	96.7 %	89.6%
mAP50-95	81.2 %	62. 9 %
Training Epochs	150 epochs	477 epochs
Full Training Time	7.538 h	8.175 h

3.4 Interpretation of Results

This study evaluated the performance of two fundamentally different models, YOLOv10 and PaliGemma2, for brain tumour detection and classification. As expected, YOLOv10 achieved superior performance due to its architecture being specialised for object detection tasks, showing strong precision (95.3%), recall (94.4%), and F1-score (94.8%). Its real-time processing capabilities and efficiency with high-resolution images make it a strong candidate for clinical applications requiring rapid diagnostic support.

PaliGemma2, although not initially designed for object detection, demonstrated promising adaptability. Despite its higher memory demands and the need for downscaled inputs (244x244 pixels), it achieved respectable performance (Precision 92.5%, Recall 84.7%). Its ability to integrate visual and textual information suggests broader potential for future tasks such as automated radiology reporting. However, improvements in memory management and optimisation would be necessary to fully exploit its capabilities.

Recent studies provide both supporting and contrasting perspectives on the results obtained in this work. Priyadharshini et al. (2025) [17] reported that YOLO-based models achieve strong accuracy on MRI tumour datasets,

aligning with our finding that YOLOv10 excels in localisation tasks due to its detection-oriented architecture. In contrast, Elboardy et al. (2025) [19] benchmarked multiple vision–language models, including PaliGemma2, for radiology-report generation from multisequence MRI and found substantial performance variability across VLM families and model sizes. Their results showed that while large, specialised VLMs can match or exceed domain-specific baselines in radiological reasoning, smaller models such as PaliGemma2 lag behind more advanced VLMs and remain less reliable for clinical decision support. This partially contradicts our findings, as PaliGemma2 performed competitively in tumour detection despite being a general-purpose model. Together, these studies suggest that although YOLOv10 remains the stronger option for direct tumour localisation, VLM performance is highly dependent on scale, domain specialisation, and task design, indicating that more advanced versions of PaliGemma-based models may achieve stronger performance in future medical imaging applications.

4.0 Conclusion

This study successfully compared the performance of YOLOv10 and PaliGemma2 for brain tumour detection and classification using MRI data. The objectives were achieved by demonstrating that YOLOv10 consistently delivered superior localisation and classification performance, confirming its suitability for real-time clinical integration. PaliGemma2 showed promising adaptability despite being a general-purpose VLM, contributing novel evidence that such models can be fine-tuned for specialised medical imaging tasks. This research made three main contributions. From a technical perspective, it provided the first comparative evaluation of YOLOv10 and PaliGemma2 for MRI-based tumour detection, which clarified the trade-offs between specialised and multimodal AI systems. In terms of methodological, it successfully demonstrated that Vision–Language Models could be adapted for tumour classification, expanding the potential of VLMs within multimodal radiology workflows. Finally, from a societal viewpoint, it conducted a public perception survey in Bosnia and Herzegovina, which revealed cautious optimism towards AI in healthcare while highlighting concerns regarding oversight, data privacy, and the need for professional training. In addition to these contributions, the study also addressed ethical considerations, including privacy protection, interpretability, and the importance of human oversight, emphasising that AI should support rather than replace clinical decision-making. The survey findings further underscored the necessity for public education, infrastructure improvement, and training of healthcare professionals before the large-scale deployment of AI tools in Bosnia and Herzegovina.

Based on these findings, future research should explore advanced fine-tuning of VLMs like PaliGemma2 for radiology report generation, cross-modal reasoning, and complex diagnostic tasks. Apart from that, researchers should also pursue access to richer MRI datasets through collaborations with hospitals to improve generalisability. This would develop effective integration strategies for AI within Bosnia's healthcare institutions, including workflow

optimisation and policy development. Finally, future work should conduct several longitudinal studies to analyse the positive impact of AI tools on diagnostic accuracy, clinical efficiency, and patient outcomes. These findings suggest meaningful and potential opportunities for clinical adoption, particularly in Bosnia and Herzegovina, where radiology workflows are often resource-constrained. YOLOv10's strong accuracy and efficiency suggest near-term commercial viability for clinical decision-support systems, especially in resource-constrained environments. PaliGemma2, with further optimisation, shows long-term potential for multimodal radiology platforms capable of combining imaging, text, and patient metadata in a single diagnostic pipeline.

Acknowledgement

The authors would like to acknowledge the Faculty of Engineering and Natural Sciences, International University of Sarajevo, for providing the necessary resources and academic support for this research, as well as the participants whose valuable insights contributed to the study.

Author Contributions

Fatima Hadžiahmetović: Conceptualisation, Methodology, Software, Writing – Original Draft, Data Curation, Software, Formal Analysis, Investigation; **Ali Abd Almisreb**: Conceptualisation, Supervision, Visualisation, Validation, Resources, Writing – Review and Editing; **Che Zawiyah Che Hasan**: Writing – Review and Editing; **Alessandro Cantelli-Forti**: Writing – Review and Editing.

Conflicts of Interest

The manuscript has not been published elsewhere and is not under consideration by any other journal. All authors have reviewed and approved the manuscript, consent to its submission, and declare that there are no conflicts of interest.

References

[1] MedlinePlus, "Brain Tumor." Accessed: Dec. 24, 2024. [Online]. Available: https://medlineplus.gov/braintumors.html

[2] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018.

[3] L. O. Chua and T. Roska, "The CNN Paradigm," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 40, no. 3, pp. 147–156, 1993.

[4] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, Apr. 2020.

[5] L. Alzubaidi et al., "Towards a Better Understanding of Transfer Learning for Medical Imaging: A Case Study," *Applied Sciences*, vol. 10, no. 13, pp. 1–21, Jun. 2020.

[6] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: A literature review," *BMC Medical Imaging*, vol. 22, no. 1, pp. 1–13, Apr. 2022.

[7] R. Sapkota et al., "YOLOv10 to Its Genesis: A Decadal and Comprehensive Review of The You Only Look Once (YOLO) Series," Jun. 2024. Accessed: Dec. 25, 2024. [Online]. Available: https://arxiv.org/abs/2406.19407v4

[8] A. Baccouche, B. Garcia-Zapirain, C. C. Olea, and A. S. Elmaghraby, "Breast Lesions Detection and Classification via YOLO-Based Fusion Models," *Computers, Materials & Continua*, vol. 69, no. 1, pp. 1407–1425, 2021.

[9] M. G. Ragab et al., "A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023)," *IEEE Access*, vol. 12, pp. 57815–57836, 2024.

[10] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications," *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023.

[11] Institute for Statistics of FBiH, *Statistical Yearbook of the Federation of Bosnia and Herzegovina 2022*. Sarajevo, 2022.

[12] Institute for Statistics of FBiH, *Statistical Yearbook of the Federation of Bosnia and Herzegovina 2023*. Sarajevo, 2023.

[13] "MRI for Brain Tumor with Bounding Boxes." Accessed: Dec. 26, 2024. [Online].
Available: https://www.kaggle.com/datasets/ahmedsorour1/mri-for brain-tumor-with-bounding-boxes

[14] L. Beyer et al., "PaliGemma: A versatile 3B VLM for transfer," Jul. 2024. Accessed: Dec. 25, 2024. [Online]. Available: https://arxiv.org/abs/2407.07726v2

[15] A. Steiner et al., "PaliGemma 2: A Family of Versatile VLMs for Transfer," Dec. 2024. Accessed: Dec. 25, 2024. [Online]. Available: https://arxiv.org/abs/2412.03555v1

[16] A. Wang et al., "YOLOv10: Real-Time End-to-End Object Detection," May 2024. Accessed: Dec. 26, 2024. [Online]. Available: http://arxiv.org/abs/2405.14458

[17] S. Priyadharshini, R. Bhoopalan, D. Manikandan, and K. Ramaswamy, "A successive framework for brain tumor interpretation using Yolo variants," *Scientific Reports*, vol. 15, no. 1, Dec. 2025.

[18] L. Klein, K. Amara, C. T. Lüth, H. Strobelt, M. El-Assady, and P. F. Jaeger, "Interactive Semantic Interventions for VLMs: A Human-in-the-Loop Investigation of VLM Failure," Oct. 12, 2024. Accessed: Dec. 25, 2024. [Online].
Available: https://drive.google.com/file/d/1LGZBGOG1wK2-qqdSzqD-esvsEk8qVn_v/preview

[19] A. T. Elboardy, G. Khoriba, M. al-Shatouri, M. Mousa, and E. A. Rashed, "Benchmarking vision-language models for brain cancer diagnosis using multisequence MRI," *Informatics in Medicine Unlocked*, vol. 58, p. 101692, pp. 1–15, Jan. 2025.