Research on MRI Object Detection Using an Enhanced YOLOv8 Framework

Nan Gao, Fuguang Han, Yao Yao, Zhijie Gao, Yihan Zhang*

Tiangong University, Tianjin, China

Received: 03 Oct 2025
Revised: 09 Oct 2025
Accepted: 10 Oct 2025
Published: 11 Oct 2025

Abstract: Early diagnosis and accurate localization of brain tumors are crucial for improving patient survival rates. Among these, automated detection methods based on deep learning have become a research hotspot. However, tumor detection, especially for complex shapes and small targets, remains challenging. To address this, this study proposes an improved YOLO model—YOLO-CDF—aimed at enhancing the detection accuracy for complex and small targets in brain tumor MRI images. The model builds upon YOLO by incorporating the BRASPPF module (A combination of bi-level routing attention mechanism and spatial pyramid pooling), dilated convolution, and small object detection layers. Experimental results show that the YOLOv8-CDF model achieves a good balance between precision and recall, with an overall mAP@0.5 of 0.929 and an F1 score reaching 0.90, demonstrating excellent detection performance. When detecting tumors, the model's precision values are 0.974, 0.964, and 0.851, respectively. Validation results show that the model can provide accurate predictions at both high and low confidence levels, with strong detection capabilities and good generalization ability, making a significant contribution to the identification of brain tumors.

Keywords: YOLO; MRI; Deep learning; Comparative experiment; Image processing

1 INTRODUCTION

Brain tumours represent one of the significant diseases impacting global public health, where early diagnosis and accurate tumour localisation are crucial for improving patient survival rates. Traditional brain tumour detection methods rely on manual analysis by medical imaging specialists, a process that is both time-consuming and susceptible to human error. In recent years, with the rapid advancement of deep learning technologies, automated brain tumour detection systems based on computer vision have gradually emerged as a research focus. These approaches utilise convolutional neural networks to analyse features within medical images, enabling rapid and precise tumour detection and classification.

2 RELATED WORK AND HYPOTHESES

In this study, our work primarily encompasses the following aspects: we selected a brain tumour MRI dataset from Roboflow Universe and enriched the training data through data augmentation techniques. This ensured the model could adapt to diverse image variations, thereby enhancing its detection capabilities in complex scenarios. Subsequently, we designed a YOLO-CDF model, which represents an enhancement upon the YOLO framework. By incorporating the BRASPPF module (Combining a dual-layer routing attention mechanism with a spatial pyramid pooling structure), contextual information fusion during feature extraction was enhanced, thereby improving the model's detection accuracy for complex tumour morphologies. Simultaneously, dilated convolutions replaced conventional

convolutional layers to enhance the model's perceptual capabilities across different scales. A small object detection layer was introduced to improve performance when processing small lesions, thereby increasing accuracy in locating and identifying minute targets and bolstering the model's clinical efficacy. Finally, comparative experiments against conventional YOLO models (YOLOv5, YOLOv8s, and YOLOv9) validated the YOLO-CDF model's performance metrics, including accuracy, recall, and mAP.

3 DATASET PROCESSING

3.1 Dataset Introduction

The brain tumour MRI dataset utilised in this experiment was provided by Roboflow Universe, comprising 3,903 annotated MRI images categorised into three types: gliomas, meningiomas, and pituitary tumours. Each image in the dataset features bounding box annotations for precise tumour localisation, supporting object detection and classification tasks. The data is partitioned into a training set (70%), validation set (20%), and test set (10%), compatible with mainstream formats such as YOLO, making it suitable for real-time, efficient detection tasks [1].

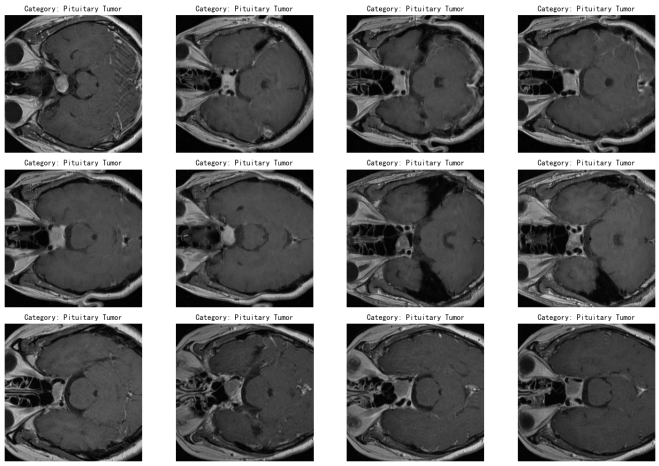


Fig. 1: Original Dataset.

Table. 1: Dataset Attribute Statistics.

Tumour type	Train	Test	Vald	All
glioma	1272	363	181	1816
meningioma	634	182	91	907
pituitary	826	236	118	1180
All	2732	781	390	3903

Table. 2: File Type Information.

Folder name	Data type
Images	PNG format
Labels	Txt message

3.2 Image processing

Insufficient dataset size may lead to overfitting or underfitting, thereby compromising deep learning performance. Consequently, this paper generates three augmented outputs for each training sample to enrich the training dataset and enhance the deep learning model's adaptability to image variations [2]. (1) Rotation: Images will undergo random rotation between -15° and +15°, simulating real-world variations caused by differing capture angles to enhance the model's robustness against image orientation shifts [3]. (2) Grayscaling: 100% of images will be converted to grayscale, enabling the model to focus on learning structural and textural information, thereby reducing sensitivity to lighting and colour variations. (3) Brightness Adjustment: Image brightness will be adjusted between 0% and +29% to simulate capture effects under varying lighting conditions, thereby strengthening the model's adaptability to changes in light intensity [4]. (4) Bounding Box Brightness Adjustment: The brightness within bounding boxes in images will be adjusted between 0% and 16%, aiming to improve the model's ability to recognise object boundaries under differing brightness conditions.

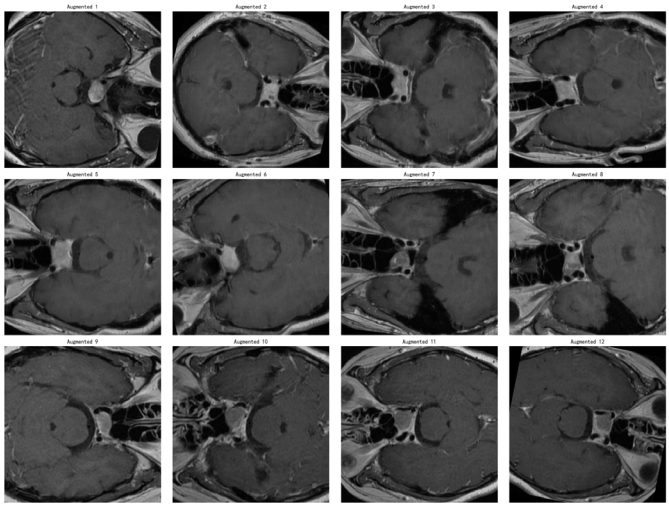


Fig. 2: Processed Data Image.

4 MODEL ESTABLISHMENT AND SOLUTION

4.1 model establishment

YOLO (You Only Look Once) is an object detection algorithm based on deep convolutional neural networks. It utilises convolutional neural networks to extract image features and employs multi-scale feature maps for comparative analysis, thereby achieving effective target detection. Due to issues such as varying morphological structures and sizes in MRI tumour data, this paper proposes an enhanced algorithm: YOLO-CDF (YOLO-based Compact Feature Detection). This algorithm builds upon the original YOLO model with optimisations and refinements aimed at improving detection accuracy and efficiency, particularly for complex object detection and localisation. The following outlines the design rationale and key structural optimisations of this model:

(1) Size Input Reduction

The image size is reset, and compared to higher resolution inputs, the original image size is $H \times W \times C_{\text{in}}$, where W_{out} and H_{out} represent the width and height of the original image, respectively. C_{out} is the number of channels in the image [5].

In this paper, the computational cost (FLOPs) of the convolution operation is proportional to the image size:

$$\text{FLOPs} \propto W_{\text{out}} = 640 \times H_{\text{out}} = 480 \times C_{\text{out}} = 3 \tag{1}$$

The input image size is reset to $W_{\text{out}} = 640$, $H_{\text{out}} = 480$, and the number of channels $C_{\text{out}} = 3$. This makes the model more efficient during inference while ensuring sufficient detection accuracy.

(2) Design of the BRASPPF Module

The YOLO convolutional network expands relevant features through the SPPF module. SPPF is an improved spatial pyramid pooling structure based on the spatial pyramid pooling (SPP) architecture, which integrates global and local information. However, there is some loss of feature information. Therefore, this paper integrates a bi-level routing attention (BRA) mechanism, resulting in a new module, BRASPPF [6].

The input feature map of size $H \times L$ can be divided into $S \times S$ non-overlapping regions and then linearly mapped to obtain the key tensor K, query tensor Q, and value tensor V:

$$K = X^r W^k, Q = X^r W^q, V = X^r W^v \tag{2}$$

Where X^r is the reconstructed feature map, and W^k, W^q, W^v are the weights corresponding to the key, query, and value tensors in the network.

Next, attention weights are calculated on the coarse-grained tokens, and only the top k largest elements are selected to represent the corresponding regions for fine-grained computation. The indices of these elements are returned to the matrix I^r:

$$A^r = Q^r (K^r)^T, I^r = \text{topkIndex}(A^r) \tag{3}$$

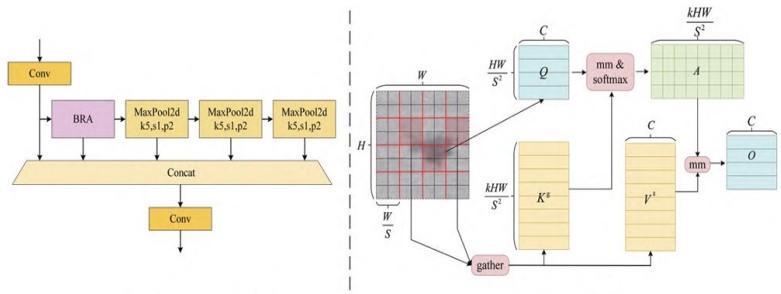
Finally, the Top-k coarse-grained regions in I^r are collected through the gather algorithm. These are used as keys and values in the final computation, and the matrix O is output. To enhance contextual information, a depthwise convolution is applied to the values.

$$K^g = \text{gather}(K, I^r) \tag{4}$$

$$V^g = \text{gather}(V, I^r) \tag{5}$$

$$O = \text{Attention}(Q, K^g, V^g) + \text{LCE}(V) \tag{6}$$

Where K^g and V^g are the aggregated key and value tensors.



(a) BRASPPF module (b) BRA module

Fig. 3: BRASPPF and BRA Module Structure.

(3) Hollow Convolution Module

In standard convolution, each input channel is convolved with all the convolution kernels of the output channels C_{out}, generating the corresponding feature map. The input tensor $X \in \mathrm{R}^{H\times W\times C_{\text{in}}}$, where H and W are the height and width of the input image, respectively, and C_{in} is the number of input channels. The kernel size is $k_h \times k_w$, and the output tensor has dimensions $Y \in \mathrm{R}^{H_{\text{out}}\times W_{\text{out}}\times C_{\text{out}}}$, where H_{out} and W_{out} are the height and width of the output feature map, and C_{out} is the number of output channels [7].

The Spatial Depthwise Convolution Block performs convolution operations individually on each input channel. It then combines the convolution results along the depth dimension to generate the output feature map. Each input channel c_{in} is convolved with its own independent kernel [9]. For each output pixel $y_{(h,w,c_{\text{in}})}$, the calculation formula is:

$$y_{(h,w,c_{\text{in}})} = \sum_{i=0}^{k_h-1}\sum_{j=0}^{k_w-1} x_{(h+i,w+j,c_{\text{in}})} \cdot k_{(i,j,c_{\text{in}})} \tag{7}$$

Where $x_{(h+i,w+j,c_{\text{in}})}$ is the pixel value in the $c_{\text{in}}-$th channel of the input tensor, and $k_{(i,j,c_{\text{in}})}$ is the value at position (i,j) in the convolution kernel $K_{(c_{\text{in}})}$.

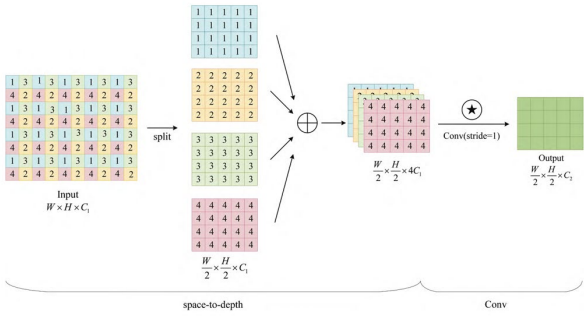


Fig. 4: SPDConv Module Structure.

(4) Small Object Detection Layer Design

The small object detection layer is designed to enhance the model's performance when processing small objects. In MRI images, numerous critical tumour lesions often manifest as minute targets within complex backgrounds. Consequently, the introduction of this layer enables the reinforcement of internal image details, thereby assisting the model in accurately localising and identifying small-scale lesions to support more precise medical diagnosis [8].

Its overall framework structure is as follows:

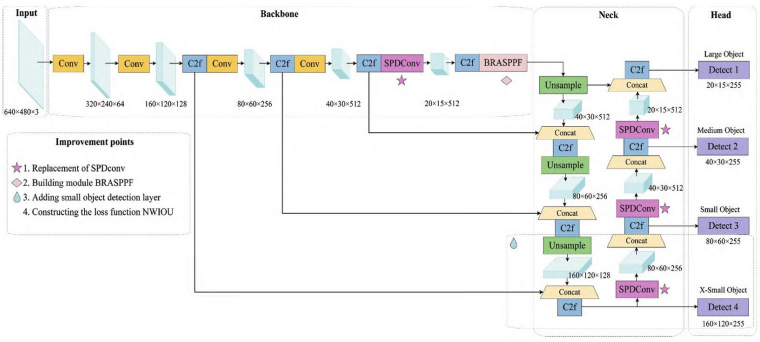


Fig. 5: YOLO-CDF Network Model Architecture.

4.2 model solution

This paper conducts comparative experiments between traditional YOLO models (YOLOv5, YOLOv8s, YOLOv9, YOLO10, YOLO11) and the designed YOLOv8s-CDF model. The results are as follows [11], [12]:

Table. 3: Comparative Evaluation of Multi-Model Performance.

Model	mAP50	mAP50-95	Precision
YOLOv5	0.92151	0.71058	0.90403
YOLOv8	0.92425	0.72112	0.90221
YOLOv11	0.92347	0.60071	0.91603
YOLOv9	0.9222	0.70664	0.91491
YOLOv10	0.90593	0.6825	0.89035
YOLOv8_CDF	0.92932	0.71775	0.89915

Table. 4: Continued.

Model	Recall	Train_Loss	Val_Loss
YOLOv5	0.87765	0.66518	0.95083
YOLOv8	0.86595	0.62414	0.94234
YOLOv11	0.877	0.89455	1.37169
YOLOv9	0.8771	0.62963	0.97373
YOLOv10	0.84204	1.34783	2.00352
YOLOv8_CDF	0.89767	0.67916	0.9408

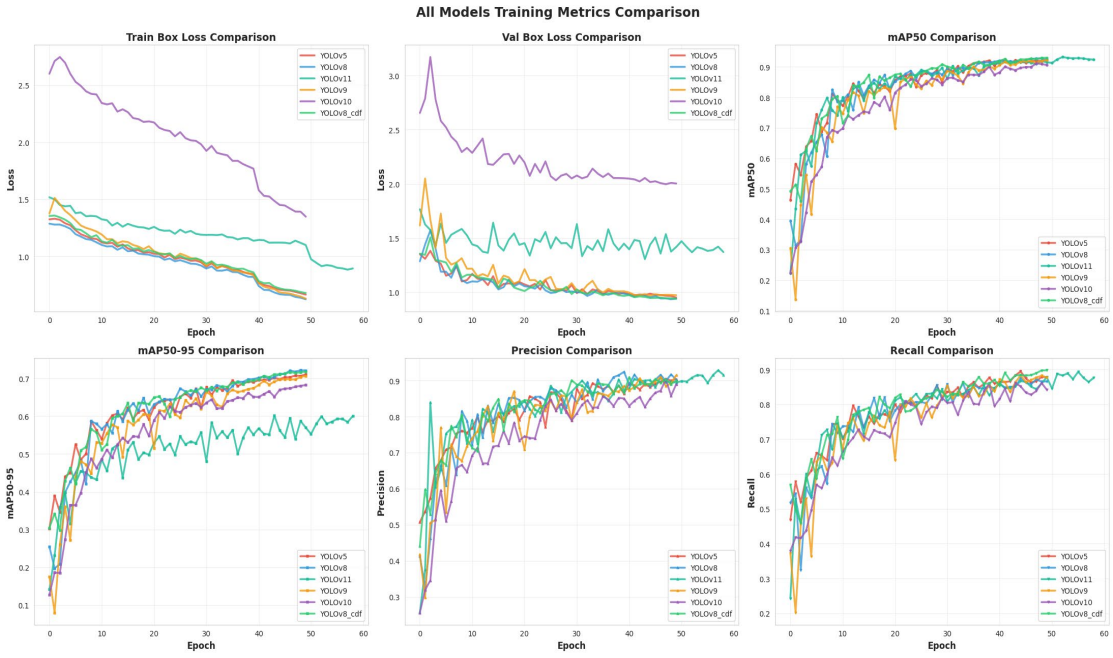


Fig. 6: Multi-model Visualisation Comparison.

Based on the performance metrics analysis of each model, YOLOv11 demonstrated the best accuracy (0.91603) and recall (0.89455), making it suitable for tasks requiring high detection

precision; YOLOv5 achieved the lowest training loss (0.66518) with stable overall performance, demonstrating high training efficiency; YOLOv8_CDF excelled in validation loss (0.9408) and mAP50 (0.92932), exhibiting strong generalisation capabilities; YOLOv10, however, demonstrated weaker performance, particularly in terms of precision (0.89035), recall (0.84204), and loss metrics (Training loss 1.34783, validation loss 2.00352). Overall, the YOLOv8_CDF model designed in this paper achieved the best comprehensive performance and is therefore identified as the optimal model.

The training process is as follows:

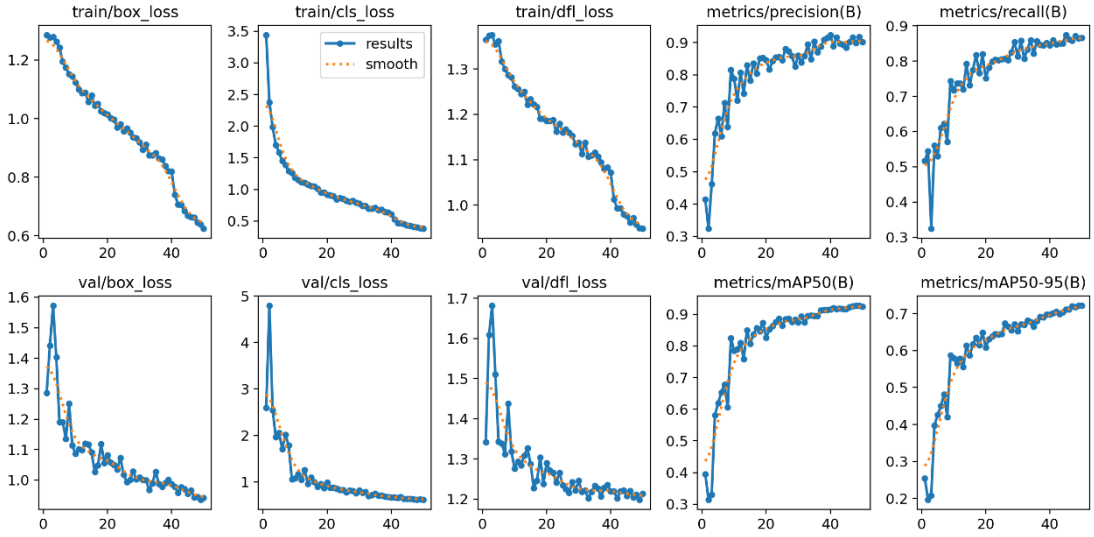


Fig. 7: Training Process.

The training process of the YOLOv8_CDF model demonstrated favourable convergence and performance enhancement.Regarding training loss metrics, train/box_loss, train/cls_loss, and train/dfl_loss progressively converged and optimised. For validation loss, the loss on the validation set similarly decreased, indicating improved generalisation capability. This validated the model's significant advancement in detection accuracy. YOLOv8_CDF demonstrated robust performance throughout both training and validation, exhibiting favourable convergence and strong generalisation capabilities.

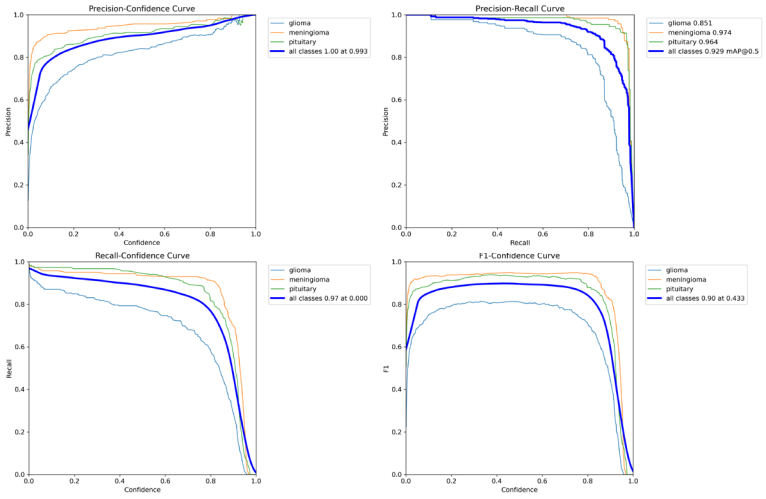


Fig. 8: Performance Convergence Curve.

The model performs excellently in brain tumor (Glioma, meningioma, pituitary) detection tasks. The overall mAP@0.5 reaches 0.929, indicating a good balance between precision and recall. As confidence increases, the precision approaches 1, and the model's performance remains stable, providing high-accuracy predictions at high confidence while maintaining a high recall rate at lower confidence, demonstrating strong detection capabilities and generalization ability.

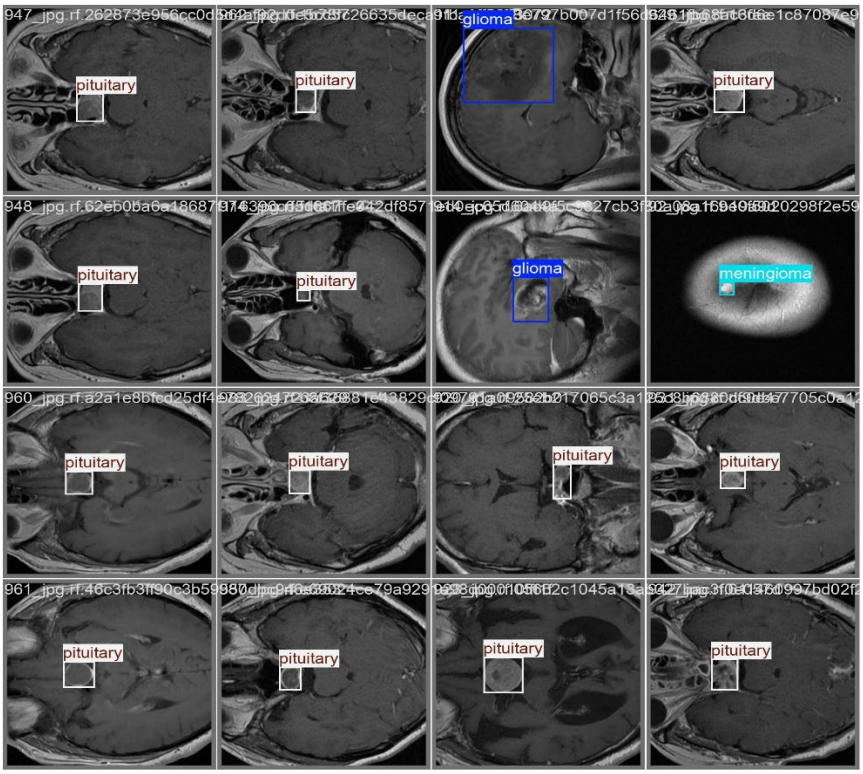


Fig. 9: Test Classification Results.

5 CONCLUSION

The experiment uses the Brain Tumor MRI dataset from Roboflow Universe, consisting of 3,903 annotated images categorized into glioma, meningioma, and pituitary tumors. The dataset is split into training (70%), validation (20%), and test (10%) sets. Data augmentation techniques (Rotation, grayscaling, brightness adjustment, and bounding box brightness adjustment) were applied to improve model adaptability and prevent overfitting. The proposed improved YOLO model—YOLO-CDF—combines the BRASPPF module, dilated convolution, and small object detection layers to enhance the detection accuracy of complex targets. Experimental results show that YOLOv8_CDF achieves a good balance between precision and recall, with an overall mAP@0.5 of 0.929. It performs excellently in detecting meningiomas and pituitary tumors, with precision values of 0.974 and 0.964, respectively, though glioma detection remains more challenging, with a precision of 0.851. Overall, YOLOv8_CDF demonstrates strong detection capabilities and excellent generalization ability.

REFERENCES

[1] Deng, X., Huang, T., Wang, W., & Feng, W. (2025). SE-YOLO: A sobel-enhanced frame work for high-accuracy, lightweight real-time tomato detection with edge deployment capa bility. *Computers and Electronics in Agriculture*, *239*, 110973. DOI: https://doi.org/10.10 16/J.COMPAG.2025.110973

[2] Duan, C., Guo, Y., Duan, X., Li, G., & Sheng, B. (2025). DD-YOLO: A dual-channel du al-path YOLO network for target detection of blurred vehicles. *Digital Signal Processing*, 105565. DOI: https://doi.org/10.1016/J.DSP.2025.105565

[3] Murat, A. A., & Kiran, M. S. (2025). A comprehensive review on YOLO versions for ob ject detection. *Engineering Science and Technology, an International Journal*, *70*, 102161. DOI: https://doi.org/10.1016/J.JESTCH.2025.102161

[4] Zheng, Y., Jing, Y., Zhao, J., & Cui, G. (2024). LAM-YOLO: Drones-based Small Object Detection on Lighting-Occlusion Attention Mechanism YOLO. *arXiv preprint arXiv:2411. 00485*. DOI: https://doi.org/10.1016/J.CVIU.2025.104489

[5] Gao, B., Tong, J., Fu, R., Zhang, Z., & Yuan, Y. Cbh-Yolo: A Steel Surface Defect Det ection Algorithm Based on Cross-Stage Mamba Enhancement and Hierarchical Semantic G raph Fusion. *Available at SSRN 5334461*. DOI: https://doi.org/10.1016/J.NEUCOM.2025. 131467

[6] Priya, S., & Amshakala, K. (2025). An adaptive fall detection system based on ensemble learning using variants of YOLO V8 retinanet and DETR. *Scientific Reports*, *15*(1), 33161. DOI: https://doi.org/10.1038/S41598-025-97634-8

[7] Mohamed, A., Nacera, Y., Ahcene, B., Teta, A., Belabbaci, E. O., Rabehi, A., ... & Beng hanem, M. (2025). Optimized YOLO based model for photovoltaic defect detection in ele ctroluminescence images. *Scientific Reports*, *15*(1), 32955. DOI: https://doi.org/10.1038/S4 1598-025-13956-7

[8] Zhou, W., Wang, J., Wang, J., Xi, M., Song, Y., & Liu, Z. (2024). Mp-Yolo: Multidime nsional Feature Fusion Based Layer Adaptive Pruning Yolo for Dense Vehicle Object Det ection Algorithm. *Available at SSRN 4952235*. DOI: https://doi.org/10.1016/J.JVCIR.2025. 104560

[9] Wang, Z., Zhou, W., & Li, Y. (2025). MSAF-YOLO: An Efficient Multi-Scale Attention Fusion Network for high-precision steel surface defect detection. *Measurement*, 118640. D OI: https://doi.org/10.1016/J.MEASUREMENT.2025.118640

[10] Zang, S., Hong, G., Liu, Y., Gan, H., Sun, F., & Lu, D. (2025). MCD-YOLO: An impro ved YOLOv8-based approach for surface defect detection of marine machinery component s. *Ocean Engineering*, *341*, 122624. DOI: https://doi.org/10.1016/J.OCEANENG.2025.122 624

[11] Madadum, H., Nasir, F. E., & Haruehansapong, K. (2025). Optimizing Watermelon Leaf Disease Detection Using SAM-based Augmentation with YOLO for Practical Agricultural Solutions. *Smart Agricultural Technology*, 101326. DOI: https://doi.org/10.1016/J.ATECH. 2025.101326

[12] Yao, J., Li, Y., Xia, Z., Nie, P., Li, X., & Li, Z. (2025). WTAD-YOLO: A Lightweight Tomato Leaf Disease Detection Model Based on YOLO11. *Smart Agricultural Technology*, 101349. DOI: https://doi.org/10.1016/J.ATECH.2025.101349