An Hybrid Lightweight Model for Brain Tumor Detection

Notsa Jeff Rakotozafy*[1], Andriamasinoro Rahajaniaina[1], Adolphe Andriamanga Ratiarison[2]

[1]Department of Mathematics, Computer Science and Applications, University of Toamasina, Toamasina, Madagascar.

[2]Department of Physics and Applications, University of Antananarivo, Antananarivo, Madagascar.

***Corresponding Author**

DOI: https://doi.org/10.51584/IJRIAS.2025.100900087

Received: 16 Sep 2025; Accepted: 22 Sep 2025; Published: 23 October 2025

ABSTRACT

In the last decade, deep transfer learning (TL) approaches are most widely used to detect and classify brain tumours imagines. However, current models are either complex and require significant computer resources, or they are lightweight but use a small dataset. To overcome these problems, in this paper we suggested a hybrid lightweight model for brain tumor detection in MRI images dataset efficiently and accurately. Our model used MobileNetV3Small as backbone followed by a single conv layer (the neck) to adjust the channel count and YOLO11 as detection component. So, YOLO11's inference time remains slower than that of MobileNetV3Small. The main difficulty lies in YOLO11's feature extractor, which, while performant, requires significant resources, limiting its use on mobile devices. To reduce complexity and improve efficiency on mobile devices, the intermediate multi-scale head of YOLO11 (CSP/upsample fusions) is removed as it is complex. The goal is to combine the strengths of each model. We conducted a comparative study between YOLO11 standard version and our model using the same dataset, hyper parameters and metrics. After more experiment, the proposed model has a higher result than YOLO11 in all metrics. It achieved 99.4% as mAP@50 and 99.8% as precision. These results have shown that our framework is both resilient, reliable and could run on the low resource environment. For future work, we plan to explore additional architectural optimizations and extend validation to larger, multi-institutional datasets. Further development will focus on using this model in new datasets.

Keywords- Brain tumor detection, Efficient diagnosis, Real-time object detection, lightweight model detection

INTRODUCTION

Reference [1] described that Brain tumours stands as third most common cancer and the third leading cause of cancer related death within adolescents and young adults. The advanced evolution of artificial intelligence has more impacts in several domains such as medical domain. So, in their work [11] proposed an approach to segment the images using median filter, the Otsu method is used for automated segmentation and morphological operators for the filtering, the classification was carried out by CNN. Reference [20] introduced ShuffleNet, a lightweight CNN to detect a brain tumor with the BraTS 2013 dataset. The accuracy of their model is reported to be 92,5%. A U-Net model using BraTS 2013 dataset was suggested in [10], their model accomplished 93,4% as precision rate. In [7], they used SqueezeNet with BraTS 2017 dataset for diagnosing brain tumor and achieved 94.1% of accuracy. Reference [6] presented principal component analysis to reduce the features and used support vector machine (SVM) for the classification of multi-sequence magnetic resonance images. In [17], they developed K-means clustering with HSV (hue, saturation, value) colour features for the detection of the tumor and cyst using CT image registration. A combination of three method such as fuzzy c-means, Zernike moments and region growing algorithm were developed in [13] to detect the tumor. The fuzzy c-means was used for the MR image segmentation. Then, they applied Zernike moments to examine every tissue about the existence or not of the tumor. At the end, they located the tumor by using region growing algorithm. However, these previous approaches are either complex and require significant

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue IX September 2025

computer resources, or they are lightweight but use a small dataset. Face to this situation, we proposed an hybrid lightweight model has MobilNetV3Small as backbone and YOLO11n (YOLO11 default version) as detection component in order to overcome the previous problems. The rest of this paper is organized as follow: brief related work on the different brain tumor detection approach is described in section II. Section III presents materials and methods. Results and discussion about our approach are discussed in section IV and section V conclude this study.

Related Work

In this section, we discuss the existing literature about the visual method for detecting brain tumor using images dataset. Various approaches were existed to detect the tumor. The author in [9] used YOLOv7 with transfer learning (TL) to detected early brain tumor using MRI scan. This dataset contains 3 classes like glioma, Meningioma, and Pituitary. After experimentation, they achieved an 99,5% of accuracy.

In [14], the authors proposed a lightweight approach of diagnosing brain tumor using YOLOv5 with fine tuning technics. They used a RSNA-MICCAI dataset partitioning to train and test. Their result gave an accuracy 88%.

Reference [19] compared different pre-trained deep neural networks, i.e., Inceptionresnetv2, Inceptionv3, Xception, Resnet18, Resnet50, Resnet101, Shufflenet, Densenet201 and Mobilenetv2, to evaluate their performance in identifying and classifying different kinds of brain tumours. They added some new layers for each pre-trained deep neural network. They employed the brain tumor classification (MRI) dataset. Then, 80% of the data was used for training, and the remaining 20% was used for testing. For training pre-trained DL models, they used stochastic gradient descent (SGD) through TL, 0.01 learning rate and a 10-image minibatch size. In addition, each DL model was trained for 14 epochs to conduct the TL experiments for detecting and categorizing brain tumor types. Their experiment result shows that the hybrid TL based on Mobilnetv2 surpassed all other ones with an accuracy rate of 82,61%.

Single Shot Detector (SSD) was developed in [3] to diagnosis a brain tumor using endoscopic images. Their work has an accuracy of 82,7%.

The authors in [2] proposed Lightweight-CancerNet model. It is designed to detect brain tumor. Their framework utilizes MobileNet architecture as the backbone and NanoDet as the primary detection component. The proposed model has the ability to detect brain tumours with different image distortions. They have used two different publicly available datasets. One is the Multimodal Brain Tumor Segmentation Benchmark (BraTS) with 1,140 images and the other one is RSNA-MICCAI competition data having 400 images. They used horizontal and vertical flip augmentation to expand the dataset to 1,131 images. Their result gave a mean average precision (mAP) of 93.8% and an accuracy of 98%.

A method combines Harmony Search Optimisation (HSO) and Convolution Neural Networks (CNN) based on Deep Learning techniques proposed in [15]. They utilized CT and MRI images of brain tumours to assess the functioning of their system. The HSO approach was utilised to extract some information from MRI and CT images, and the CNN model was used to diagnosis brain cancers. The result showed that their model achieved an accuracy rate of 99.13% for both detection and classification tasks.

Reference [4] developed a method based on an improved fuzzy factor fuzzy local information C means (IFF-FLICM) segmentation and hybrid modified harmony search and sine cosine algorithm (MHSSCA) optimized extreme learning machine (ELM) for detecting and classifying brain tumor images. They used image Dataset-255 for their study. The SSIM and PSNR are used for segmentation qualities measures and sensitivity, specificity, and accuracy for detection and classification ones. The IFF-FLICM segmentation approach achieved a peak signal-to-noise ratio (PSNR) of 37.24 dB and a structural similarity index (SSIM) of 0.9823. The MHS-SCA-based ELM model achieved a sensitivity, speci0city, and accuracy of 98.78%, 99.23%, and 99.12%.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue IX September 2025

In [5], the authors developed a method that can automatically categorize brain tumours into various pathological categories. The model consists of a deep neural network and an image processing framework. It is divided into various phases, such as the mapping stage, the data augmentation stage, and the tumor discovery stage. Their framework is based on a DCNN and a SVM algorithm to perform generative model analysis on large datasets. The images within the dataset are equipped with a resolution of 512 × 1024 pixels and a slice that is 6 mm thick. The data was split into a pair of datasets: the training dataset and the test one. The first dataset was used for training the model while the testing one performed the feature extraction. They used data augmentation technique to reduce or avoid overfitting problem. Their model achieved a 99% accuracy rate and a sensitivity of 97.3%.

[11] presented a framework single shot multi-box detector (SSD) combined with MobileNetv2. The proposed

method used Mobile Net as baseline method in order to have a lightweight model and to obtain a best result in term of classification. The second part of their architecture is constituted by the auxiliary network. This later was introduced for final object detection. They used the MRI image database provided by Kaggle to train and evaluate their model. The dataset was formed by 250 MRI scans with the infected region that is tumor. The proposed brain tumor detection method showed 98% accuracy after 4000 epochs.

The authors in [8] introduced a method of object detection to ameliorate the accuracy of the Single Shot Multibox Detector (SSD). They proposed the feature map. The output was obtained by transforming the layout close to the classifier network: change the VGGNet with ResNet. However, current models are either complex and require significant computer resources, or they are lightweight but use a small dataset. To overcome these problems, in this paper, we presented an hybrid model for the detection of the tumor in MRI images. The proposed model is the fusion of the MobilNetV3Small and YOLO11n. We further incorporated the MobileNetv3Small neural network architecture within YOLO11 because it is lightweight for the feature extraction and object classification. Our model classifies and detects the objects faster.

Materials and Methods

In this section, we describe all steps involved to build the proposed approach. Our method is based on YOLO11n model combined with MobilNetV3Small as backbone.

Overview and motivation

Although YOLO11n has a relatively low number of parameters (2,693,000), close to that of MobileNetV3Small (2,542,856) in pytorch, its architecture remains complex. Furthermore, despite improved performance compared to other YOLO versions, YOLO11's inference time remains slower than that of MobileNetV3Small. The main difficulty lies in YOLO11's feature extractor, which, while performant, requires significant resources, limiting its use on mobile devices.

To reduce complexity and improve efficiency on mobile devices, the goal is to combine the strengths of each model, as MobileNetV3Small is a lightweight model capable of extracting rich features, particularly in the C4 and C5 layers, while YOLO11 is a performant model for object detection. And feature has the shape [batch_size, channel, height, width] where batch_size is the number of samples processed, channel is the number of channels (or color layers), and height/width are the dimensions.

Design choice: replace the backbone

After conducting a comparative analysis on the features extracted from each model, we noticed that the dimensions of the output tensors from each model are very different. MobileNetv3Small has [1, 576, 7, 7] and YOLO11 has [1, 3, 640, 640]. This disparity complicates direct combination without resorting to upsampling or downsampling operations, making the direct combination of YOLO11 and MobileNetv3 Small difficult. Therefore, we proceed by replacing YOLO11 backbone with MobileNetV3Small for feature extraction.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue IX September 2025

To achieve this, we create within the YOLO11 network a new model YOLOmobilenetv3, which consists of a new backbone (MobileNetV3Small), followed by a single conv layer (the neck) to adjust the channel count, then the same Detect class used by YOLO11 for the final prediction. The intermediate multi-scale head of YOLO11 (CSP/upsample fusions) is removed as it is complex. The structure of our model is depicted in Fig. 1.

Mobile NetV3Small backbone

We use depthwise separable convolutions (DWConv). Each 3×3 convolution is decomposed into a depthwise convolution (one filter per channel, isolated) followed by a 1×1 pointwise convolution that mixes channels; convolution is a linear operation sliding a filter over the image/features. The 3×3 convolutions (used in the backbone and head) extract local patterns. The 1×1 convolutions reduce or mix channels without altering spatial resolution. These Conv blocks are typically followed by normalization (BatchNorm) and activation. Batch Normalization stabilizes internal distributions and accelerates learning. This decomposition drastically reduces parameters and computations. MobileNetV3Small also integrates squeeze-and-excitation attention blocks and hardswish or SiLU activations for a good accuracy/cost trade-off.

We instantiate MobileNetV3Small pretrained on ImageNet, then extract its convolutional features. Iterating through MobileNetV3Small's layers produce three outputs at different scales:

p3: Output from an intermediate module (Layer 4) with ~40 channels and high resolution,

p4: Output from layer C8 (Layer 8) with 48 channels,

p5: Final output from layer C12 (Layer 12) with 576 channels.

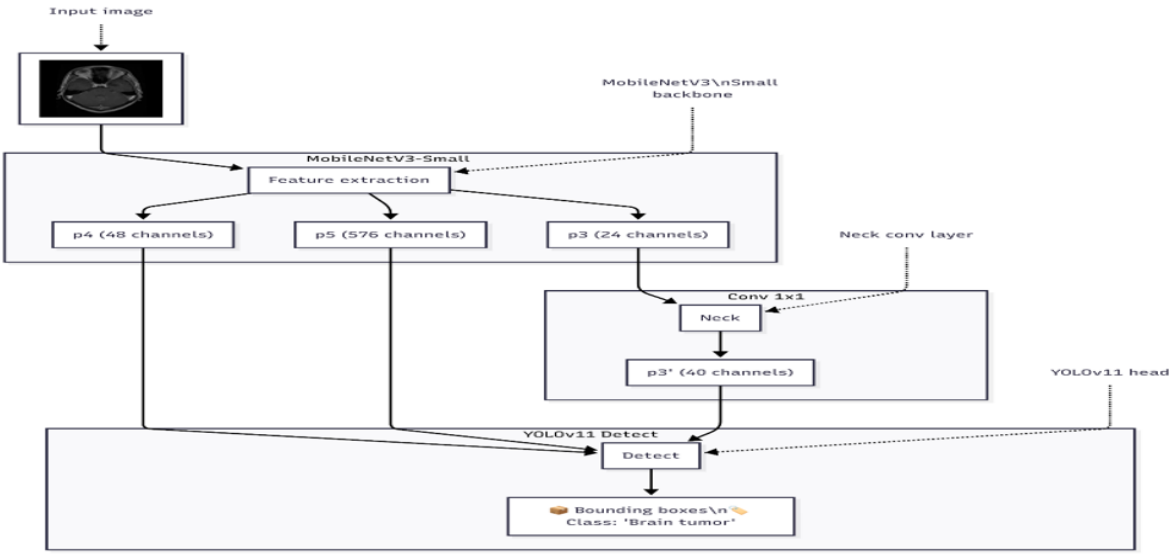
Each corresponds to a different resolution (lower than the previous) and serves as a detection scale (as in YOLO: small objects use high-resolution p3, medium objects use p4, etc.). If the highest resolution (p3) is missing, it is recreated by upsampling p4 with bilinear interpolation. Where upsampling is a process increasing the spatial size of a feature map, often via bilinear interpolation. In YOLO's backbone, deep maps are scaled for concatenation (upsampling×2). Upsampling only adds grid structure without altering fundamental semantic information. Each output's stride is implicit: the ratio between the input image size and the feature map size. This operational logic is represented in Fig. 2.

Intermediate Neck Layer

After feature extraction, the p3, p4, p5 feature maps must be standardized for the detection head. We therefore added a 1×1 convolution (called neck_conv) to p3, transforming its initial 24 channels to 40 channels. This pointwise convolution operation solely adjusts the channel dimension (to ensure the head receives the expected channels) without modifying spatial size. In YOLO11's original architecture, multiple convolutions and CSP modules were used to merge p3 with higher layers (via upsampling and concatenation). Here, this complex fusion is simplified: we retain only a single conv1×1 layer as the neck. This simplification avoids computational overhead while preserving sufficient spatial information for detection. The neck acts as an intermediate network ("neck network") that prepares the backbone's features for prediction. In many modern detectors, the neck incorporates feature pyramids (FPN/PAN, etc.) to blend scales. Here, our neck is extremely lightweight

Fig. 1 architecture of the proposed model

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue IX September 2025



(one conv layer), which is feasible because MobileNetV3Small already produces rich hierarchical feature maps.

Detection Head (Detect)

The one from YOLO11 remains unchanged here. It is an anchor-based implementation where each position in a p3, p4, or p5 feature map proposes multiple candidate boxes (anchors). According to [14] anchor-based is a strategy where box prediction relies on predefined anchors of different sizes and aspect ratios. For each cell and each anchor, the network predicts box adjustments and class probability. This approach (popularized by Faster R-CNN and adopted in YOLOv2/v3) allows covering a range of object dimensions. The anchors are then refined through regression.

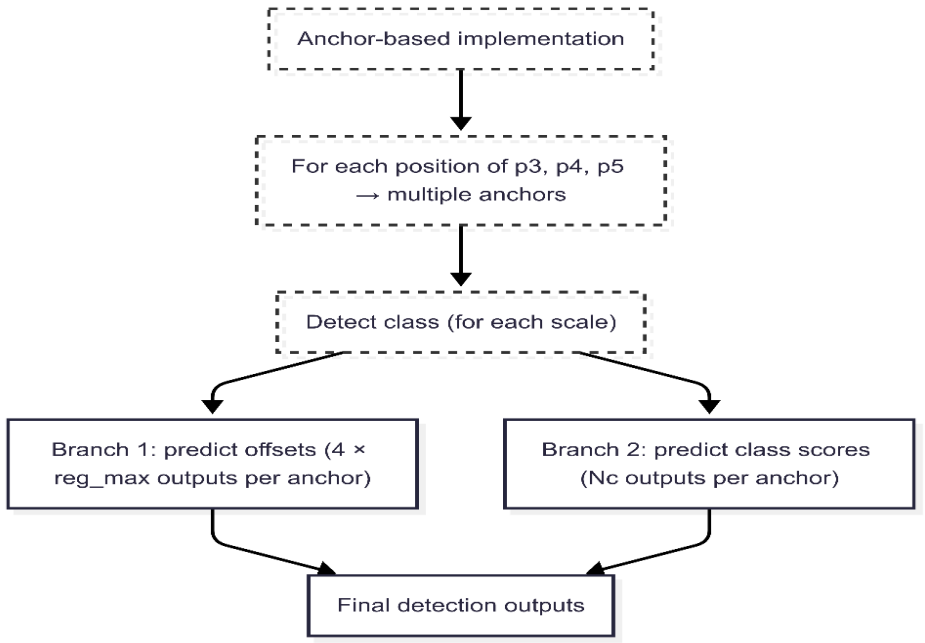
The Detect class uses two convolution branches for each scale: one to predict box offsets (4× reg_max outputs per anchor) and the other for class scores (Nc outputs per anchor),

Fig. 2 Working principle of our backbone

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue IX September 2025

then concatenates these results. During training, we apply Distribution Focal Loss (DFL) for box regression. Specifically, instead of directly predicting width/height, we predict a discrete distribution (with reg_max bins) for each side, improving localization accuracy (DFL learns the exact distribution of deviations). Anchors are dynamically generated based on each feature map's stride, then aligned with DFL predictions. Outputs are post-processed into final boxes (center + width/height) and probability scores (sigmoid for classes). During inference, strides (which, according to [7], are a convolution parameter defining the filter's step size): a stride > 1 reduces resolution (downsampling). For example, a stride of 2 in 3×3 convolution doubles the step size, dividing spatial dimensions by 2. We initialized it via a dummy. Its operational algorithm is shown in Fig. 3.

Fig. 3 Working algorithm of the detection head (Detect class)



Dataset Description

Fig. 4 Samples images within the dataset before pre-processing

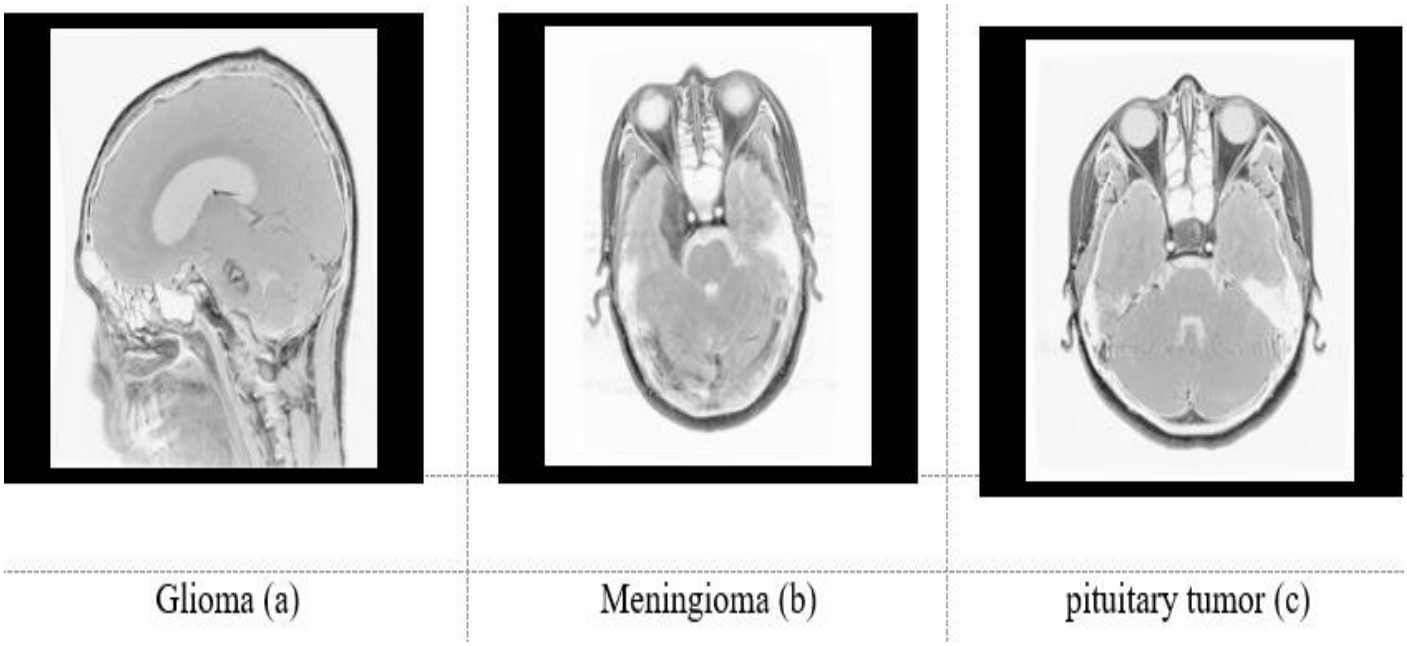
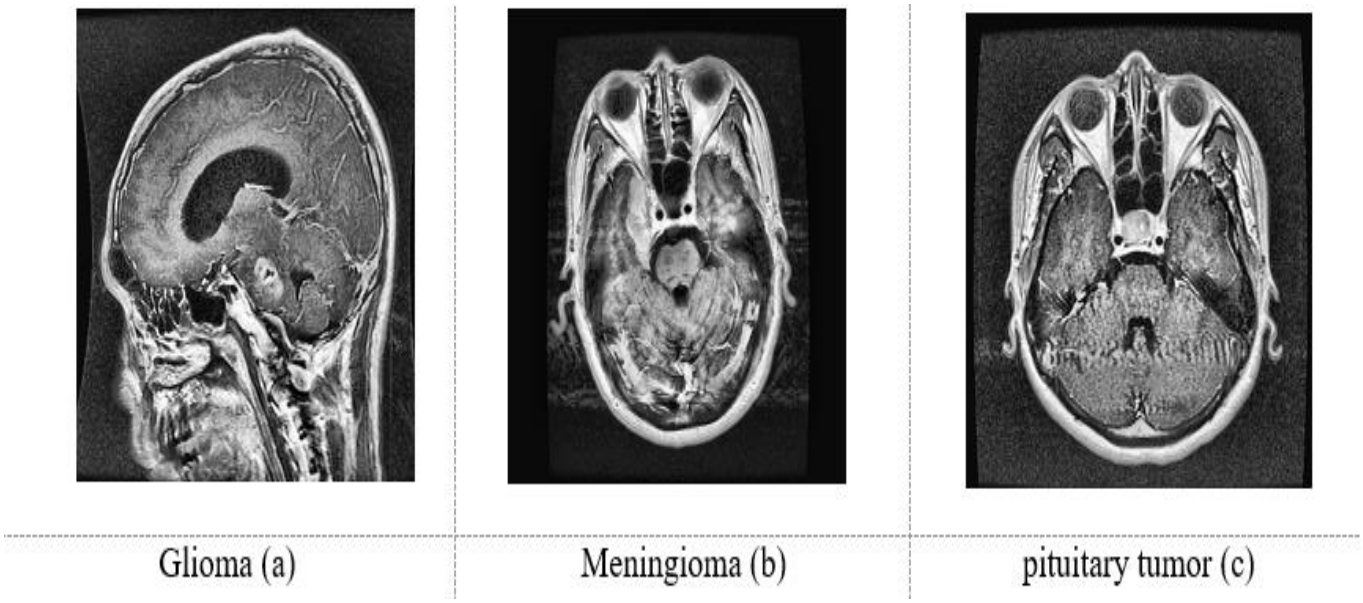


Fig. 5 Samples images within the dataset after pre-processing

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue IX September 2025



Our model has used the public dataset downloaded from [20]. This brain tumor dataset containing 3064 T1-weighted contrast-inhanced images from 233 patients with three kinds of brain tumor: meningioma (708 slices), glioma (1426 slices), and pituitary tumor (930 slices). It is a pre-labelled dataset but the target object properties [class_id x1 y1 x2 y2] in the labels files in which class_id is a numerical class identifier; first class has 1 as identifier. xi and yi are the planar coordinates on tumor border. These properties not follow the YOLO format [class_id center_x center_y width height] in which class_id is a numerical class identifier; first class has 0 as identifier. center_x and center_y are the bounding box center coordinates (normalized 0-1). width and height are the bounding box dimensions (normalized 0-1). To be compliance with the requirement of our model, some transformations were applied to these ones. Thus, 90% of 3064 images were used for train set and the rest was used for validation set. Then, some pre-processing must be applied to improve image quality. Thus, a contrast-limited adaptive histogram equalization (CLAHE) method combined with sharpening method were used for each image. Fig. 4 and Fig. 5 show sample images in the dataset before and after pre-processing.

Moreover, the weights of a neural network are trained using training images have a fixed size and have the same dimensions. The images in our dataset have 640×640 as resolution. In order to conform with the feature extractor requirement, all images must be resized into 224x224 pixels and normalized using MobilenetV3 built-in pre-processing function before passing through our feature extractor.

RESULTS AND DISCUSSION

We conducted our experimentation on Intel(R) Core (TM) i7-1255U CPU, 10 cores, 12 threads @ 2.30GHz, and 24 Gb RAM. We have been developed our system using python, torch, NumPy, Matplotlib, torchvision, OpenCV and PIL libraries. We used precision, recall, mAP@50 and mAP@[.50:.95] as metrics for evaluating our approach. The precision indicates the number of correct positive predictions from the total number of actual predictions classified by the model as positive. Recall corresponds to the score of true positive predictions to the instances that actually belong to the positive class. mAP@50 is the means of the average precision at Intersection over Union (IoU) and mAP@[.50:.95] indicate the mean Average Precision across all classes.

The proposed model is compared with YOLO11n using the same dataset. A lot of experiments were conducted during the training phase but the batch size of 8 and an epoch of 700 performed the best result. The Fig. 6 shows the curves of results.

From plot is seen that our model achieved the best result than YOLO11 model.

Fig. 6.a F1-Confidence Curve of YOLO11

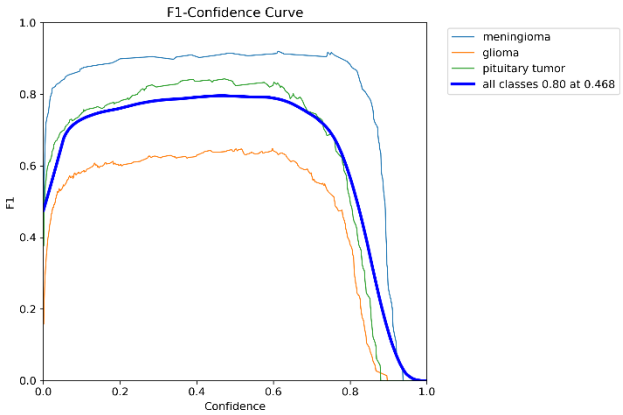


Fig. 6.b Precision-Recall Curve of YOLO11

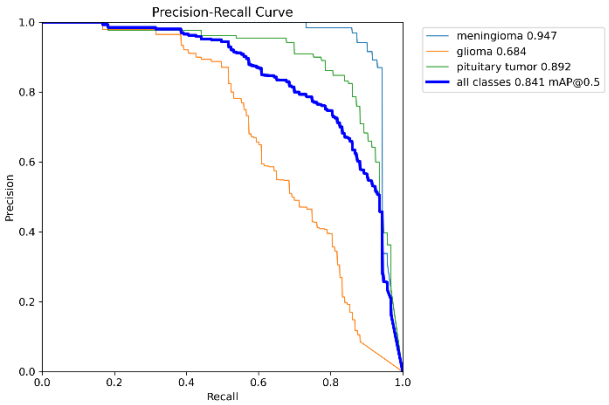


Fig. 6.c F1-Confidence Curve of our Model

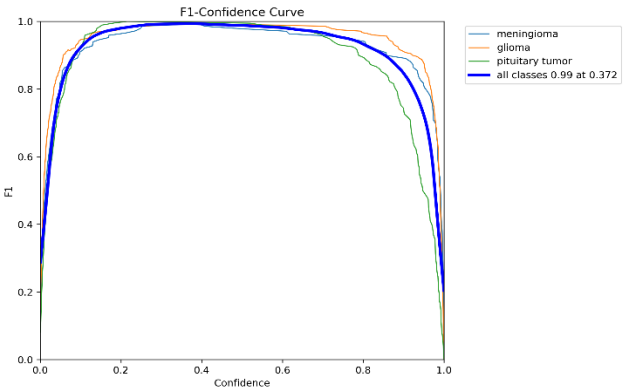
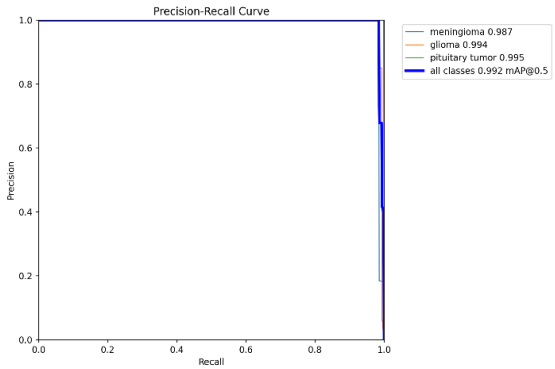


Fig. 6.d Precision-Recall Curve of our Model



ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue IX September 2025

To evaluate the performance of our YOLOv11-MobileNetV3-small, we first tested it on PyTorch using an inference script, meaning we performed detection on multiple images in a folder. The results are shown in Figure 7 below.

Fig. 7.a Detection of glioma

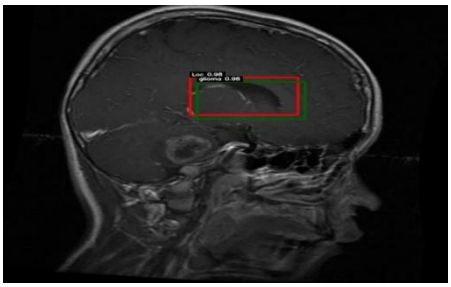


Fig. 7.b Detection of meningioma

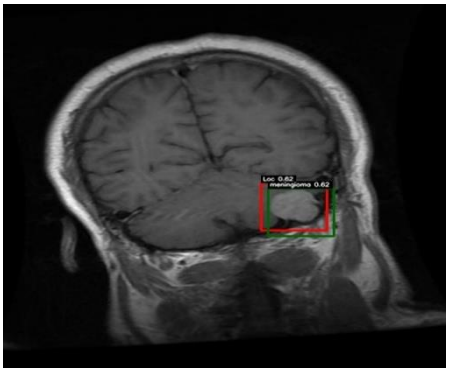
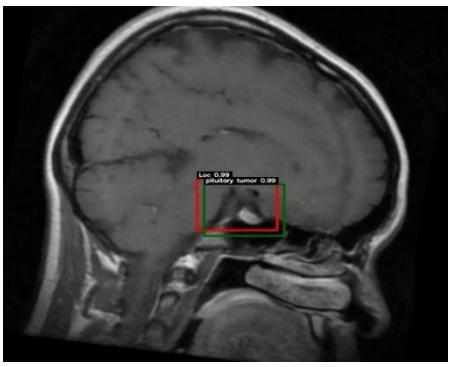


Fig. 7.c Detection of pituitary tumor



Then, we converted the model to TensorFlow Lite for use on a mobile platform. We imported the YOLOv11-MobileNetV3-small TFLite model into a mobile application we designed and performed detection on the mobile device, where the detection results along with the inference time are presented in Figure 8.

Fig. 8.a Detection of glioma on android device

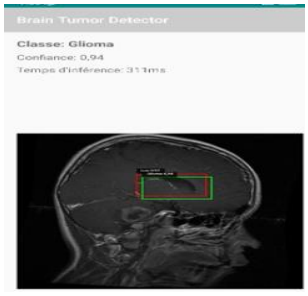


Fig. 8.b Detection of meningioma on android device

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue IX September 2025

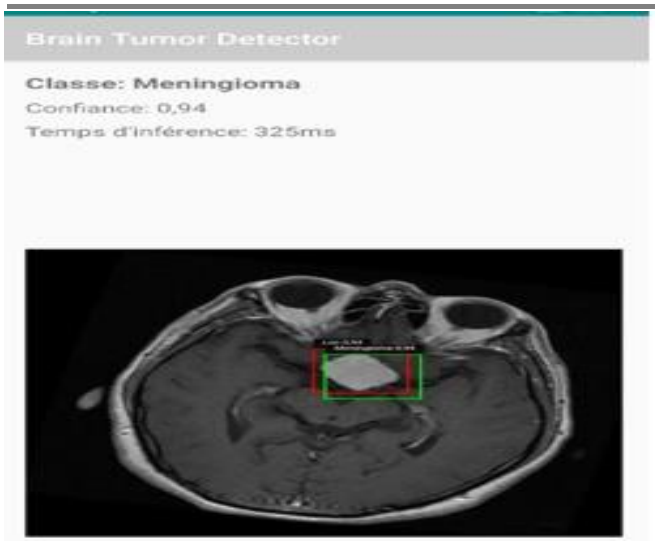


Fig. 8.c Detection of pituitary on android device

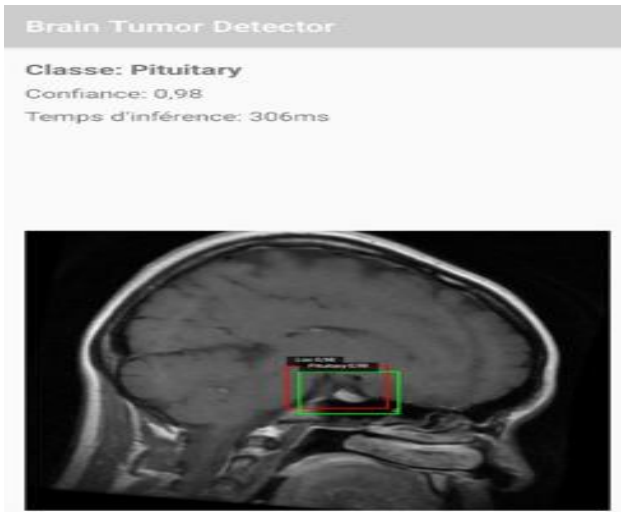


Table I below shows that the proposed model demonstrates higher performance than YOLO11 using the same dataset, obtaining a precision of 99.8%, a recall of 98.8%, a mAP@50 of 99.4% and mAP@ [.50:.95] of 98.4 %. Significantly, this surpasses all results attained by YOLO11n.

Table I: comparative results

Figure 7 shows sample detections produced by our model.

Model	Precision	Recall	mAP@50	mAP@[.50:.95
YOLO11n (default)	83%	77.4%	84.1%	57.4%
Proposed Model	99.8%	98.8%	99.4%	98.4%

CONCLUSION

In this work, we proposed a hybrid lightweight model for brain tumor detection by integrating MobileNetV3Small as a feature extraction with the YOLO11 detection framework. The model was rigorously evaluated and compared against the standard YOLO11 implementation across 700 training epochs. Our proposed architecture demonstrated superior performance across all key metrics, achieving a precision of 99.8% (compared to 83% for YOLO11), recall of 98.8% (versus 77.4%), mAP@50 of 99.4% (versus 84.1%),

and mAP@50-95 of 98.4% (versus 57.4%). These results represent a significant improvement in detection accuracy while maintaining computational efficiency suitable for low-resource environments.

The substantial performance gains confirm the effectiveness of our architectural modifications, particularly the replacement of the backbone and simplification of the neck network. This approach successfully reduces computational complexity while enhancing feature extraction capabilities for medical imaging applications.

For future work, we plan to explore additional architectural optimizations and extend validation to larger, multi-institutional datasets. Further development will focus on using this model in new datasets.

REFERENCES

1. Aaron Cohen-Gadol, (2024). Must-known brain tumor statistics. www.aaroncohen-gordol.com.
2. Asif Raza and Muhammad Javed Iqbal, (2025). Lightweight-CancerNet: a deep learning approach for brain tumor detection. In PeerJ Computer Science 11:e2670 DOI 10.7717/peerj-cs.2670.
3. Cen Q, Pan Z, Li Y, Ding H., (2019). Laryngeal tumor detection in endoscopic images based on convolutional neural network. In: IEEE 2nd International Conference on Electronic Information and Communication Technology. Piscataway: IEEE.
4. Dash et al, (2024). Brain Tumor Detection and Classification Using IFF-FLICM Segmentation and Optimized ELM Model, Journal of Engineering Volume 2024, Article ID 8419540, 24 pages https://doi.org/10.1155/2024/8419540.
5. Goodfellow, I., et al., (2016). Deep Learning. MIT Press.
6. Gunasundari and Selva Bhuvaneswari, (2025). A novel approach for the detection of brain tumor and its classification via independent component analysis. Scientific reports. 14 pages. https://doi.org/10.1038/s41598-025-87934-4.
7. Gupta et al., (2005). Support vector machine for optical diagnosis of cancer. Journal of Biomedical Optics 10(2), 024034.
8. https://www.kaggle.com/datasets/ahmedsorour1/mri-for-brain-tumor-with-bounding-boxes
9. Iandola FN, et al., (2016). SqueezeNet: a simple and efficient CNN for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 3511–3519.
10. J. Jeong, H. et al., (2017). Enhancement of SSD by concatenating feature maps for object detection. arXiv preprint arXiv:1705.09587.
11. Kumar A., (2023). Study and analysis of different segmentation methods for brain tumor MRI application. Multimedia Tools and Applications 82(5):7117–7139. DOI 10.1007/s11042-022-13636-.
12. Menze BH, et al., (2015). The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Transactions on Medical Imaging 34(10):1993–2024 DOI 10.1109/TMI.2014.2377694.
13. Naseer-u-Din et al. (2022). Brain tumor detection in MRI scans using single shot multibox detector. in Journal of Intelligent & Fuzzy Systems · March 2022 DOI: 10.3233/JIFS-219298.
14. P. K. Barik and R. C., (2019). Barik. Feed forwarded ct image registration for tumour and cyst detection using rigid transformation with hsv colour segmentation. International Journal of Computational Systems Engineering, 5(5-6):277–286.
15. P. Maji and S. Roy. SoBT-RFW, (2015). rough-fuzzy computing and wavelet analysis based automatic brain tumor detection method from MR images. Fundamenta Informaticae, 142(1-4):237–267.
16. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems, 28, 91–99
17. S. Naz and N. Kumar, (2019). An Efficient Brain Tumor Detection system using Automatic segmentation with Convolution Neural Network.
18. Shahariar et al., (2019). Automatic Human Brain Tumor Detection in MRI Image Using Template-Based K Means and Improved Fuzzy C Means Clustering Algorithm. In Big Data Cogn. Comput.
19. Shaik et al., (2024). Detection and Classification of Brain Tumor from MRI And CT Images using Harmony Search Optimization and Deep Learning. Journal of Artificial Intelligence Research & Advances, ISSN: 2395-6720 Volume 11, Issue 3, 2024 September–December DOI Journal: 10.37591/JoAIRA.

20. Shelatkar T, et al., (2022). Diagnosis of brain tumor using light weight deep learning model with fine-tuning approach. Computational and Mathematical Methods in Medicine 2022:1–9 DOI 10.1155/2022/2858845.
21. Sorour, A. (2024). MRI for Brain Tumor with Bounding Boxes. Kaggle.
22. T. Gupta, et al., (2017). Multi-sequential mr brain image classification for tumor detection. Journal of Intelligent & Fuzzy Systems, 32(5):3575–3583.
23. Ullah N, et al., (2022). An effective approach to detect and identify brain tumours using transfer learning. Applied Sciences 12(11):5645 DOI 10.3390/app12115645.
24. Zhang et al. (2017). ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. arXiv:1707.01083. https://doi.org/10.48550/arXiv.1707.01083