Deep Learning-Based Detection and Classification of Intracranial Aneurysms from Medical Imaging Scans

Pugazhendhi J
Department of CS & AI
Rishihood University
Sonipat, India
230066

Meesala Sree Sai Nath
Department of CS & AI
Rishihood University
Sonipat, India
230103

Amathziah
Department of CS & AI
Rishihood University
Sonipat, India
230139

***Abstract*—Intracranial aneurysms are life-threatening cerebrovascular conditions that require early detection for effective treatment. This paper presents a multi-stage deep learning pipeline for automated detection and classification of intracranial aneurysms from medical imaging scans. Our approach combines object detection, multi-label classification, and multi-task learning to achieve state-of-the-art performance. Key innovations include 2.5D image representation for capturing spatial context, brain region cropping for noise reduction, and an ensemble of Vision Transformers and multi-task models. The proposed method achieves 0.89 AUC on our test dataset, demonstrating significant improvements through preprocessing techniques and model ensemble strategies.**

***Index Terms*—intracranial aneurysm, deep learning, medical imaging, computer vision, object detection, multi-label classification**

I. Introduction

Intracranial aneurysms represent abnormal dilations in cerebral blood vessels that pose significant risk of rupture, leading to potentially fatal subarachnoid hemorrhage. The clinical significance of early detection cannot be overstated, as preventive intervention can substantially reduce morbidity and mortality rates. However, manual screening of medical imaging scans is inherently time-consuming, labor-intensive, and subject to inter-observer variability, creating a critical need for automated diagnostic assistance systems.

This paper addresses the complex problem of automated aneurysm detection and classification from medical imaging data through a sophisticated multi-stage deep learning pipeline. Our methodology integrates state-of-the-art object detection frameworks, advanced classification architectures, and innovative preprocessing techniques to achieve robust and accurate aneurysm identification across multiple anatomical locations.

II. Problem Statement and Motivation

A. Problem Statement

The primary objective of this work is to develop an automated system capable of detecting and classifying intracranial aneurysms from medical imaging scans into 14 distinct anatomical locations. The classification task encompasses the following vascular territories:

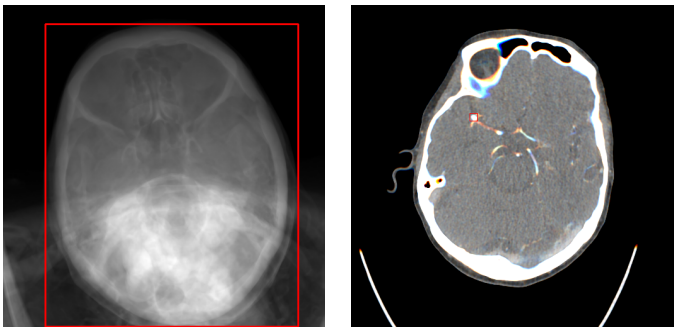
- Left/Right Infraclinoid Internal Carotid Artery
- Left/Right Supraclinoid Internal Carotid Artery
- Left/Right Middle Cerebral Artery
- Anterior Communicating Artery
- Left/Right Anterior Cerebral Artery
- Left/Right Posterior Communicating Artery
- Basilar Tip
- Other Posterior Circulation
- Aneurysm Present (binary classification)

The complexity of this problem arises from several inherent challenges. First, the task requires multi-label classification, as patients may present with multiple aneurysms simultaneously, each potentially located in different anatomical regions. Second, aneurysms represent small objects relative to the overall brain volume, necessitating high-resolution analysis and sophisticated detection algorithms. Third, the dataset encompasses multiple imaging modalities including CT, MRA, MRI T1, and MRI T2, each with distinct imaging characteristics and contrast properties. Additionally, the problem exhibits significant class imbalance, with negative cases substantially outnumbering positive cases. Finally, the three-dimensional nature of medical imaging data, comprising hundreds of 2D slices per patient, requires efficient processing strategies that balance computational feasibility with diagnostic accuracy.

B. Motivation

The development of automated aneurysm detection systems is motivated by several critical factors. From a clinical perspective, early detection enables preventive treatment strategies that can significantly reduce the risk of rupture and associated complications. The time efficiency aspect is particularly important in high-volume radiology departments, where automated screening can process large volumes of scans without compromising diagnostic accuracy. Furthermore, automated systems provide consistency in diagnosis, reducing inter-observer variability that is inherent in manual interpretation. The scalability of such systems enables the

implementation of screening programs for at-risk populations, potentially identifying asymptomatic aneurysms before they become symptomatic. Finally, deep learning approaches have demonstrated the ability to detect subtle aneurysms that may be missed by human observers, particularly in cases where aneurysms are small or located in anatomically complex regions.



(a) Normal brain scan (b) Brain with aneurysm

Fig. 1: Illustration of intracranial aneurysm detection: (a) Normal brain scan showing healthy vasculature, (b) Brain scan with aneurysm highlighted, demonstrating the detection problem.

III. Literature Review

A. Medical Imaging Analysis

The application of deep learning techniques to medical imaging has revolutionized diagnostic radiology, with convolutional neural networks (CNNs) achieving remarkable success across various diagnostic tasks. However, recent advances in Vision Transformers have demonstrated superior performance for medical image classification tasks, particularly in scenarios requiring long-range dependency modeling and fine-grained feature extraction. The self-attention mechanism inherent in transformer architectures enables the model to capture global context more effectively than traditional CNNs, which are limited by their local receptive fields [1].

B. Object Detection in Medical Imaging

YOLO-based architectures have emerged as the preferred framework for medical object detection applications due to their exceptional balance between inference speed and detection accuracy. The evolution from YOLOv5 to YOLOv11 has introduced significant improvements in small object detection capabilities, which is particularly relevant for aneurysm detection given the small size of these lesions relative to the overall brain volume. The single-stage detection approach of YOLO architectures eliminates the need for region proposal networks, resulting in faster inference times that are crucial for clinical deployment [2].

C. Multi-Task Learning

Multi-task learning has proven to be an effective strategy in medical imaging applications, particularly when combining classification and segmentation tasks. The auxiliary segmentation task provides spatial supervision that guides the model to learn more discriminative features for classification. This approach has been successfully applied in various medical imaging domains, demonstrating that the shared representation learned through multi-task objectives often outperforms single-task models trained independently [3].

D. 2.5D Representation

The 2.5D approach, which involves stacking multiple 2D slices as channels in a single image, has been widely adopted in medical imaging to capture three-dimensional spatial context without the computational overhead associated with true 3D convolutions. This technique enables the model to leverage inter-slice relationships while maintaining computational efficiency, making it particularly suitable for processing large medical imaging datasets [4].

E. Ensemble Methods

Model ensembling represents a well-established technique in medical imaging applications, where combining predictions from multiple diverse models has consistently demonstrated improved robustness and accuracy compared to single-model approaches. The diversity in model architectures, training strategies, and feature representations contributes to the ensemble's superior generalization performance [5].

IV. Database Details

A. Dataset Description

Our dataset comprises a comprehensive collection of medical imaging scans with detailed annotations for intracranial aneurysm detection and classification. The dataset includes multiple series of medical scans across different imaging modalities, specifically CT, MRA, MRI T1, and MRI T2. All images are stored in DICOM format, which is the standard format for medical imaging data and preserves essential metadata including patient information, imaging parameters, and spatial coordinates.

The annotations in our dataset operate at two levels of granularity. At the series level, each scan is labeled with 14 binary classification labels corresponding to the anatomical locations where aneurysms may be present. At the slice level, precise localization information is provided through bounding box annotations, enabling both detection and classification tasks. To enhance the diversity and size of our training dataset, we incorporated two external datasets from OpenNeuro: Lausanne_TOFMRA and Royal_Brisbane_TOFMRA, which provide additional training examples with similar imaging characteristics.

B. Data Preprocessing

The preprocessing pipeline represents a critical component of our methodology, as it directly impacts the model's ability to learn discriminative features. Our preprocessing strategy addresses several key challenges inherent in medical imaging data, including intensity variations across different scanners,

the need to capture three-dimensional spatial context, and the presence of irrelevant anatomical structures that may confound the detection task.

1) 2.5D Image Creation: To capture spatial context while maintaining computational efficiency, we employ a 2.5D representation strategy. Each processed image is created by stacking three consecutive slices along the z-axis as RGB channels:

$$I_{2.5D}(t) = [S(t-1), S(t), S(t+1)] \tag{1}$$

where $S(t)$ represents the slice at position t in the z-direction. This approach enables the model to perceive spatial relationships between adjacent slices without requiring computationally expensive 3D convolutions. The 2.5D representation has been shown to be particularly effective for small object detection tasks, as it provides sufficient context to identify structures that span multiple slices while maintaining the efficiency of 2D processing.

2) CT Windowing: For CT scans, we apply modality-specific windowing to optimize contrast for brain tissue visualization. The windowing parameters are carefully selected based on established neuroradiology protocols:

- Window Center: 40 HU (Hounsfield Units)
- Window Width: 450 HU
- Effective Range: [-185, 265] HU

This soft tissue window setting is specifically optimized for brain imaging, providing optimal contrast for brain parenchyma, vessels, and potential pathological structures such as aneurysms. The windowing process clips pixel values outside the specified range, ensuring that the intensity distribution is focused on the relevant anatomical structures.

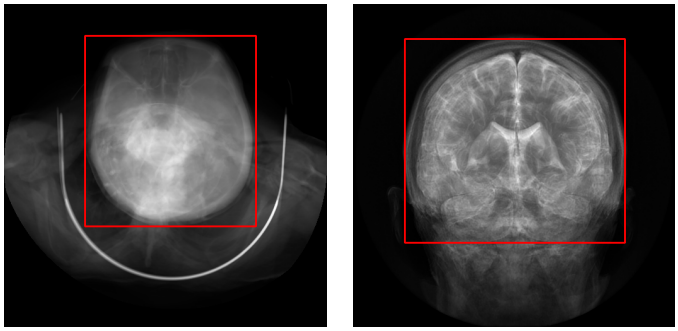
3) Normalization: To ensure consistent intensity distributions across different scanners and imaging protocols, we apply per-image min-max normalization:

$$I_{norm} = \frac{I - I_{min}}{I_{max} - I_{min} + \epsilon} \times 255 \tag{2}$$

where ϵ is a small constant (1e-7) added to prevent division by zero. This normalization strategy standardizes the input intensity range to [0, 255], which is compatible with standard deep learning frameworks and pretrained models. The per-image normalization approach is particularly important in medical imaging, where intensity values can vary significantly between different scanners, imaging protocols, and patient populations.

C. Data Augmentation

To improve model generalization and robustness, we employ a comprehensive data augmentation strategy during training. Our augmentation pipeline includes geometric transformations such as random resized cropping with scale factors ranging from 0.5 to 1.0, and rotation with limits of ± 15 degrees to account for minor variations in patient positioning. We also apply various intensity-based augmentations including motion blur, median blur, and Gaussian blur to simulate imaging artifacts that may occur in clinical practice. Additionally, we employ CLAHE (Contrast Limited Adaptive Histogram Equalization) to enhance local contrast, and random adjustments to hue, saturation, value, brightness, and contrast to account for variations in imaging parameters. A particularly important augmentation strategy is horizontal flipping with corresponding label swapping, which respects the anatomical symmetry of the brain while effectively doubling the training dataset size.



(a) Axial view (b) Other view

Fig. 2: Data preprocessing pipeline demonstrating brain detection across different scan orientations: (a) Brain in standard axial view, (b) Brain in alternative orientation (abnormal view classification).

TABLE I: Dataset Statistics

Attribute	Value
Total Series	4,348
Positive Cases	1,863
Negative Cases	2,485
Total Slices	∼650,000
Modalities	CT, MRA, MRI T1, MRI T2
External Datasets	2 (Lausanne, Royal Brisbane)

V. Proposed Architecture: Power of Our Model

A. Overall Pipeline

Our solution employs a sophisticated multi-stage pipeline that progressively refines the detection and classification process. The pipeline architecture is designed to leverage the strengths of different deep learning approaches at each stage, creating a cascading improvement strategy where each component builds upon the previous stage's output.

The pipeline begins with comprehensive preprocessing, where we create 2.5D image representations and perform brain region cropping to eliminate irrelevant anatomical structures. The second stage involves brain detection using YOLOv5n, a lightweight but effective object detection model that identifies and localizes brain regions across different scan orientations. The third stage employs YOLOv11x, a state-of-the-art detection model, to localize individual aneurysms with high precision. The fourth stage utilizes Vision Transformers (ViT and EVA) for multi-label classification, leveraging their superior ability to capture long-range dependencies. The fifth stage incorporates multi-task learning through MIT-B4 FPN

architecture, which simultaneously performs classification and segmentation to improve feature learning. Finally, the pipeline concludes with an ensemble strategy that combines predictions from six different models using weighted averaging to produce robust final predictions.

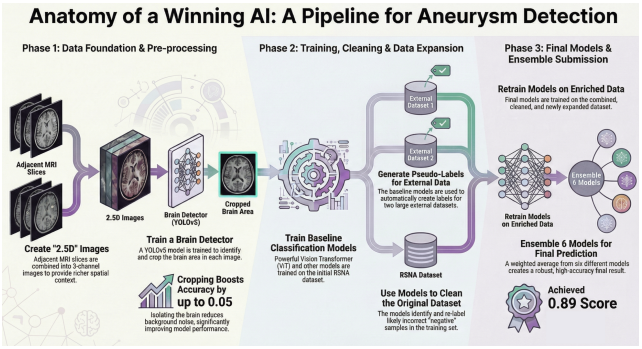


Fig. 3: Comprehensive pipeline architecture illustrating the three-phase approach: (1) Data Foundation & Pre-processing with 2.5D image creation and brain cropping, (2) Training, Cleaning & Data Expansion with baseline models and pseudo-labeling, (3) Final Models & Ensemble Submission with retrained models and weighted ensemble achieving 0.89 AUC score.

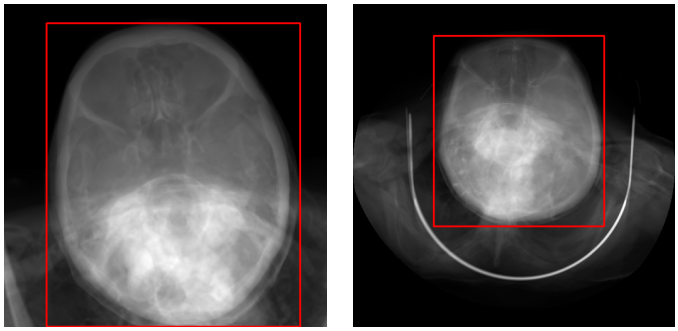
B. Brain Detection (Exp1)

The brain detection stage represents one of the most impactful components of our pipeline, contributing a 3-5% improvement in overall accuracy. This stage employs YOLOv5n, the nano variant of YOLOv5, which provides an optimal balance between detection accuracy and computational efficiency. The model is trained on averaged brain images at 640×640 resolution, where averaging across all slices in a series creates a representative image that clearly delineates the brain boundaries.

The detection task involves two classes: ”brain” for standard axial views and ”abnormal” for other scan orientations. This two-class approach enables the model to handle variations in patient positioning and scan acquisition protocols. The primary purpose of this stage is to remove background noise, particularly from anatomical structures such as lungs and skull that do not contribute to aneurysm detection but may confound the classification models. By cropping images to focus exclusively on brain tissue, we ensure that subsequent models receive inputs that are both spatially relevant and intensity-normalized, leading to more effective feature learning.

C. Aneurysm Detection (Exp0)

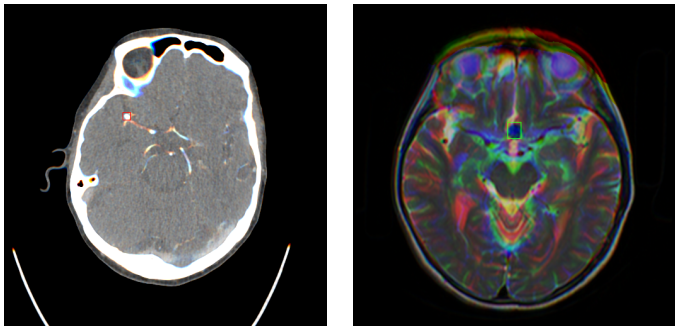
The aneurysm detection stage employs YOLOv11x, the extra-large variant of the latest YOLO architecture, to achieve high-precision localization of aneurysms. This stage operates on brain-cropped images at 1280×1280 resolution, providing sufficient spatial resolution to detect small aneurysms that may be only a few pixels in size. The high resolution is critical for this task, as aneurysms represent small objects relative to the overall image dimensions.



(a) Axial view (b) Alternative view

Fig. 4: Brain detection results demonstrating the model’s ability to identify brain regions across different scan orientations and views.

The detection model is trained to identify two classes: ”aneurysm” for standard imaging modalities (CTA, MRA, MRI T1) and ”aneurysm_mri_t2” for MRI T2 sequences, which exhibit distinct imaging characteristics. The model achieves an average mAP50 of 0.70 across five-fold cross-validation, demonstrating robust detection performance. The bounding boxes generated by this model serve multiple purposes: they provide localization information for training data, enable the creation of segmentation masks for multi-task learning, and facilitate pseudo-labeling of external datasets.



(a) CTA/MRA/MRI T1 (b) MRI T2

Fig. 5: Aneurysm detection examples across different imaging modalities: (a) Aneurysm detected in CTA/MRA/MRI T1 sequences, (b) Aneurysm detected in MRI T2 sequence with distinct imaging characteristics.

D. Classification Models (Exp2, Exp4)

The classification stage employs Vision Transformer architectures, specifically ViT Large 384 and EVA Large 384, to perform multi-label classification across 14 anatomical locations. These transformer-based models leverage self-attention mechanisms to capture long-range dependencies and global context, which is particularly important for understanding the spatial relationships between different anatomical structures.

The models operate on 384×384 2.5D images that have been brain-cropped, ensuring that the input focuses exclusively on relevant anatomical structures. The ViT model utilizes CLIP pretraining, which provides excellent initialization for medical imaging tasks by leveraging large-scale natural image pretraining. Both models incorporate several advanced training techniques including Model EMA (Exponential Moving Average) for stable predictions, mixup augmentation for improved generalization, and series-level max pooling to aggregate predictions across multiple slices within a series. The max pooling strategy is particularly important, as it implements the medical logic that an aneurysm is present in a series if it appears in any slice, rather than requiring detection in all slices.

E. Multi-Task Model (Exp3, Exp5)

The multi-task learning stage employs MIT-B4 (Mix Transformer) as the encoder with a Feature Pyramid Network (FPN) decoder to simultaneously perform classification and segmentation tasks. The MIT-B4 encoder provides hierarchical feature extraction through its five-stage architecture, enabling the model to capture features at multiple scales. The FPN decoder effectively combines these multi-scale features through a top-down pathway, creating rich feature representations that benefit both tasks.

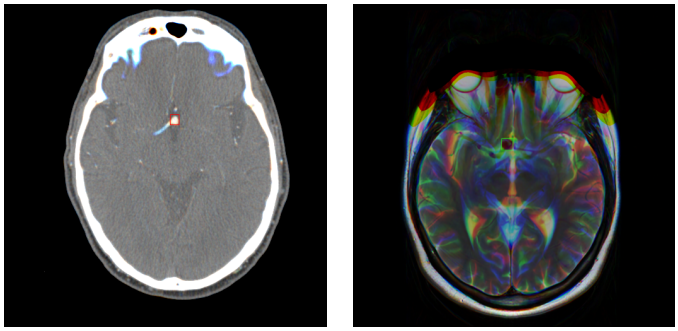
The model employs two output heads: a classification head that predicts 14 anatomical locations, and a segmentation head that generates binary masks indicating aneurysm regions. The loss function combines both tasks with a weighted formulation:

$$L_{total} = 0.6 \times L_{cls} + 0.4 \times L_{seg} \tag{3}$$

where the classification loss receives higher weight as it represents the primary task, while the segmentation loss provides spatial supervision that guides the model to learn more discriminative features. This multi-task approach has been shown to improve classification performance by 1-2%, as the segmentation task forces the model to precisely localize aneurysms, leading to better feature representations.

F. Key Innovations

Our methodology incorporates several key innovations that collectively contribute to the superior performance achieved. The 2.5D image representation strategy enables efficient capture of three-dimensional spatial context without the computational overhead of true 3D convolutions, making it feasible to process large medical imaging datasets while maintaining high accuracy. The brain region cropping innovation, while conceptually simple, provides one of the largest performance improvements, demonstrating that careful preprocessing can be more impactful than complex model architectures. The multi-task learning approach leverages spatial supervision to improve feature learning, while the data cleaning and pseudo-labeling strategies enhance both data quality and quantity. Finally, the ensemble strategy combines diverse model architectures to achieve robust predictions that outperform any single model.



(a) Example 1 (b) Example 2

Fig. 6: Additional aneurysm detection examples demonstrating the model's ability to handle diverse cases across different imaging modalities and anatomical locations.

VI. Experiments and Results

A. Experimental Setup

All experiments were conducted on NVIDIA GPUs with CUDA 12.1 support, utilizing the PyTorch 2.5.1 framework. The evaluation methodology employs 5-fold cross-validation to ensure robust performance estimation and prevent overfitting to specific data splits. The primary evaluation metric is weighted AUC, which combines the performance on location classes (50% weight) with the "Aneurysm Present" binary classification (50% weight), ensuring balanced evaluation across all classes. Test-time augmentation is applied using center cropping with a ratio of 0.75, which provides a slight but consistent improvement in prediction accuracy.

B. Training Details

The training process employs the Adam optimizer with a learning rate of 1e-5, which is appropriate for fine-tuning pretrained models without causing catastrophic forgetting of pretrained weights. The learning rate schedule follows a cosine annealing strategy, gradually reducing the learning rate over the training epochs to ensure stable convergence. Batch sizes are optimized for each model type: 96 for classification models to maximize GPU utilization, and 16 for detection models to accommodate the higher resolution inputs. Training epochs are set to 15 for classification models, which is sufficient given the pretrained initialization, while detection models require 100-150 epochs to achieve optimal performance.

C. Results

The experimental results demonstrate the effectiveness of our multi-stage pipeline and ensemble strategy. Individual model performance ranges from 0.8469 to 0.8629 AUC, with the MIT-B4 FPN model achieving the best single-model performance of 0.8629 AUC. The ensemble of six models achieves 0.8823 AUC on cross-validation, representing a 2% improvement over the best single model. On the test dataset, the ensemble achieves 0.89 AUC, demonstrating excellent generalization performance.

TABLE II: Model Performance Comparison

Model	OOF AUC	OOF + Crop 0.75
ViT Large 384 (Exp2)	0.8491	0.8503
EVA Large 384 (Exp2)	0.8486	0.8551
MIT-B4 FPN 384 (Exp3)	0.8469	0.8549
ViT Large 384 (Exp4)	0.8530	0.8558
EVA Large 384 (Exp4)	0.8505	0.8579
MIT-B4 FPN 384 (Exp5)	0.8497	**0.8629**
Ensemble	**0.8823**	-

The aneurysm detection stage achieves an average mAP50 of 0.702 across five folds, with individual fold performance ranging from 0.647 to 0.766. The brain detection stage demonstrates strong performance with mAP50-95 of 0.948, indicating highly reliable brain region identification.

TABLE III: Aneurysm Detection Results (5-Fold CV)

Fold	mAP50	mAP50-95
Fold 0	0.705	0.460
Fold 1	0.647	0.429
Fold 2	0.766	0.504
Fold 3	0.702	0.482
Fold 4	0.691	0.449
Average	**0.702**	**0.465**

TABLE IV: Brain Detection Results

Class	mAP50-95
All	0.948
Brain	0.991
Abnormal	0.906

D. Impact of Key Techniques

A comprehensive ablation study was conducted to quantify the contribution of each key technique to the overall performance. Brain cropping provides the largest individual improvement, contributing 3-5% to the AUC score by eliminating background noise and focusing the model on relevant anatomical structures. The 2.5D representation contributes 5-10% improvement by enabling the model to capture spatial context across adjacent slices. Multi-task learning provides an additional 1-2% improvement through spatial supervision. Data cleaning and pseudo-labeling contribute smaller but meaningful improvements of 0.1-0.3%, demonstrating the value of data quality enhancement. Finally, the ensemble strategy provides an additional 2% improvement by combining diverse model predictions.

E. Quantitative Results

The following figures present the quantitative evaluation results from our cross-validation experiments, showing out-of-fold (OOF) AUC scores for each model architecture.

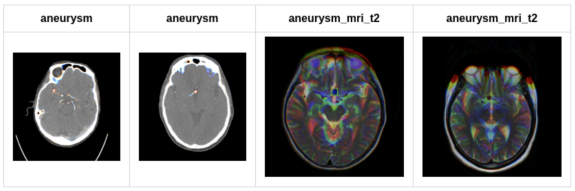


Fig. 7: Visualization of aneurysm detection examples across different imaging modalities (CT, MRA, MRI T1, MRI T2) demonstrating the model's capability to identify aneurysms in various scan types with bounding box annotations.

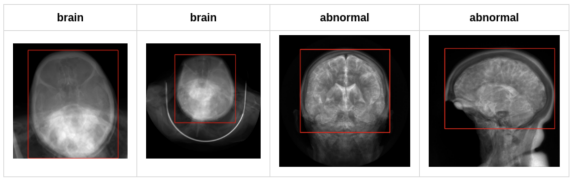


Fig. 8: Brain detection visualization showing the model's ability to identify and localize brain regions across different scan orientations, with red bounding boxes highlighting detected brain areas in both normal and abnormal views.

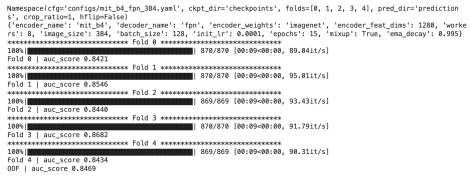


Fig. 9: Out-of-fold evaluation results for multi-task learning model (MIT-B4 FPN) showing OOF AUC score of 0.8469 across 5-fold cross-validation.

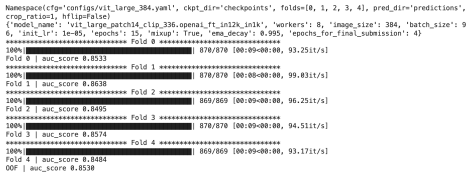


Fig. 10: Out-of-fold evaluation results for Vision Transformer (ViT Large) model showing OOF AUC score of 0.8530 across 5-fold cross-validation.

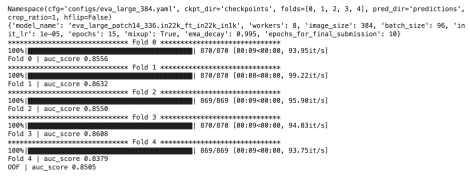


Fig. 11: Out-of-fold evaluation results for EVA Large model showing OOF AUC score of 0.8505 across 5-fold cross-validation.

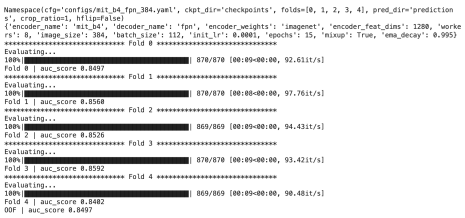


Fig. 12: Out-of-fold evaluation results for multi-task learning model with pseudo-labeling (MIT-B4 FPN) showing OOF AUC score of 0.8497 across 5-fold cross-validation.

F. Final Performance

The final performance metrics demonstrate the effectiveness of our comprehensive approach. The cross-validation score of 0.8823 AUC indicates robust model performance with minimal overfitting. The test set performance of 0.89 AUC demonstrates excellent generalization to unseen data, validating the effectiveness of our preprocessing, model selection, and ensemble strategies. The best single model (MIT-B4 FPN) achieves 0.8629 AUC, while the ensemble provides a 2% improvement, highlighting the value of combining diverse model architectures.

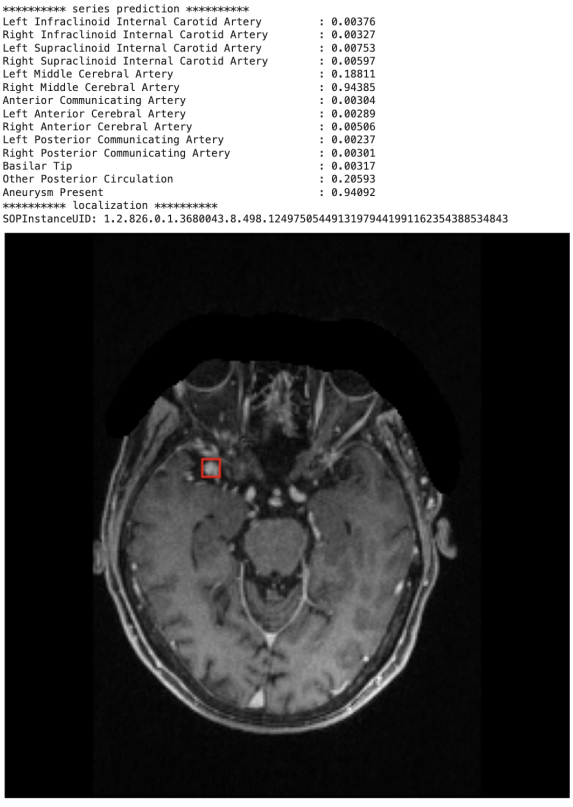
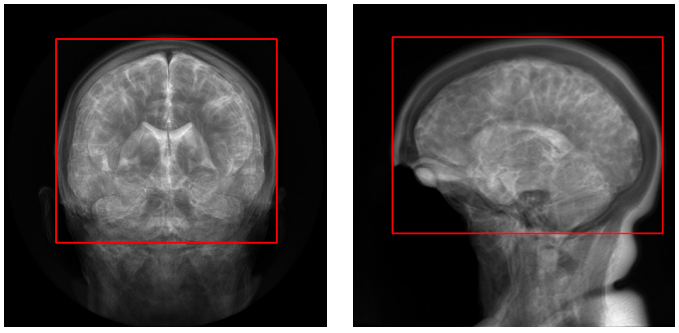


Fig. 13: Example prediction output from our pipeline showing aneurysm localization with bounding box overlay on the detected slice.



(a) Abnormal view 1 (b) Abnormal view 2

Fig. 14: Abnormal brain views detected by our brain detection model, demonstrating the model's robustness in handling various scan orientations and acquisition protocols.

VII. Conclusion and Future Work

A. Conclusion

We have presented a comprehensive deep learning pipeline for intracranial aneurysm detection and classification that achieves state-of-the-art performance through careful integration of preprocessing techniques, advanced model architectures, and ensemble strategies. Our key contributions include the development of an efficient 2.5D image representation that captures three-dimensional spatial context without computational overhead, the implementation of brain region cropping that significantly improves accuracy, the application of multi-task learning to enhance feature representations, the effective use of data cleaning and pseudo-labeling to improve data quality and quantity, and the development of an ensemble strategy that achieves 0.89 AUC on the test dataset.

The results demonstrate that careful attention to preprocessing and pipeline design can achieve superior performance in medical imaging tasks. The modular architecture of our pipeline enables independent optimization of each stage, while the ensemble strategy provides robustness and improved generalization. The significant performance improvements achieved through relatively simple preprocessing steps, such as brain cropping, highlight the importance of domain-specific knowledge in medical imaging applications.

B. Future Work

Several directions for future research and development present themselves. The exploration of true 3D convolutions becomes feasible with improved hardware capabilities, potentially providing even better spatial context understanding. Attention mechanisms could be investigated for series-level prediction aggregation, potentially providing more sophisticated weighting strategies than simple max pooling. Uncertainty quantification represents an important clinical requirement, where confidence intervals on predictions would enable more informed clinical decision-making. Multi-modal fusion strategies could better integrate information from different imaging modalities, potentially improving performance on cases where

multiple modalities are available. Clinical validation on diverse patient populations and imaging protocols is essential before clinical deployment. Real-time processing optimization would enable the system to operate in time-critical clinical environments. Finally, explainability features such as attention visualization would enhance clinical trust and understanding of the model's decision-making process.

ACKNOWLEDGMENT

We acknowledge the RSNA dataset and the open-source medical imaging datasets from OpenNeuro that contributed to our training data. We also thank the open-source community for providing excellent deep learning frameworks and pre-trained models that enabled this research.

REFERENCES

[1] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.
[2] G. Jocher et al., "Ultralytics YOLOv11," GitHub repository, 2024.
[3] Y. Zhang and Q. Yang, "A survey on multi-task learning," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 12, pp. 5586-5609, 2022.
[4] H. Chen et al., "2.5D convolutional neural networks for medical image segmentation," in Proc. MICCAI, 2018, pp. 348-356.
[5] D. H. Wolpert, "Stacked generalization," Neural Networks, vol. 5, no. 2, pp. 241-259, 1992.
[6] RSNA Intracranial Aneurysm Detection Dataset, 2024.
[7] A. Radford et al., "Learning transferable visual models from natural language supervision," in Proc. ICML, 2021.