A Hybrid Multi-Stage Framework For Intracranial Aneurysm Detection From Cta Images: A Comparative Study With End-To-End Deep Learning

Noor Hussain Ali[1*], Nuha Jameel Ibrahim[2]

[1]College of Computer Science, University of Technology, cs.24.03@grad.uotechnology.edu.iq
[2]College of Computer Science, University of Technology, Nuha.J.Ibrahim@uotechnology.edu.iq

Abstract
Intracranial aneurysm detection from brain CTA images remains challenging due to the small size of vascular abnormalities and the limited availability of expert annotations. This paper proposes a hybrid multi-stage framework and presents a direct comparison with a baseline end-to-end deep learning classifier. The framework integrates image preprocessing, synthetic mask generation, attention-based segmentation, hybrid feature extraction, and independent classification.
Synthetic vessel masks generated using vesselness filters and morphological operations were employed as surrogate labels to support training. An Attention U-Net enhanced with CBAM and optimized using Tversky loss achieved reliable segmentation performance, with a Dice score of 0.8798 and an IoU of 0.8056, enabling accurate localization of aneurysm regions even in low-contrast cases.
For classification, deep features extracted from ResNet50 were combined with handcrafted structural and textural descriptors and classified using an XGBoost model. This hybrid strategy achieved an accuracy of 98.46% and eliminated false-negative aneurysm cases, outperforming the baseline end-to-end ResNet50 classifier.
The results demonstrate that separating segmentation, feature extraction, and classification while integrating deep and structural information leads to improved diagnostic reliability and clinically safer predictions compared to conventional CNN-based approaches.

Keywords: Brain CTA; Intracranial aneurysm detection; Attention U-Net; Hybrid feature fusion; End-to-end comparison; XGBoost.

1- Introduction

Intracranial aneurysm (IA) is a potentially fatal cerebrovascular condition that is highly difficult to diagnose. Its rupture commonly leads to aneurysmal subarachnoid hemorrhage (aSAH), a life-threatening neurological emergency with extremely high morbidity and mortality. According to reported data, the mortality rate following rupture ranges between 30% and 50%, making prompt and accurate detection a crucial factor in preventing rebleeding and further neurological deterioration [1].
Despite significant advances in computed tomography angiography (CTA) and magnetic resonance imaging (MRI), the effective detection of intracranial aneurysms particularly small or morphologically irregular lesions remains challenging and time-consuming. Image interpretation requires substantial radiological expertise and is often affected by observer fatigue, which has increased the demand for reliable computer-aided diagnostic systems to support clinical decision-making [2].
In line with epidemiological surveys, unruptured intracranial aneurysms (UIAs) are estimated to affect approximately 2–6% of the adult population worldwide, with meta-analytic studies reporting an average prevalence of 3.2% [3]. This relatively high prevalence further highlights the importance of accurate and early detection strategies.
In recent years, artificial intelligence (AI) and deep learning (DL) techniques have emerged as powerful tools in medical image analysis, particularly for improving diagnostic efficiency and accuracy in complex pathologies such as intracranial aneurysms. Convolutional neural networks (CNNs), especially U-Net and its variants, have

demonstrated strong performance in anatomical segmentation tasks. However, IA detection remains difficult due to the small size of lesions, low contrast, and visual similarity to other intracranial abnormalities (e.g., tumors), necessitating the use of attention-enhanced and localization-aware models [4].

Moreover, the scarcity of expert-annotated datasets and the pronounced class imbalance where aneurysmal regions occupy only a small fraction of the image pose additional challenges for training robust and generalizable deep learning models [5]. These limitations often lead to unstable predictions and increase the risk of missed detections when end-to-end CNN classifiers are used directly for diagnosis.

Importantly, while many existing studies focus on improving overall classification accuracy, limited attention has been paid to the comparative reliability of such models against strong baseline approaches, particularly in terms of false-negative errors. In safety-critical applications such as intracranial aneurysm screening, false negatives represent a major clinical risk, as missed aneurysms may lead to catastrophic outcomes. This highlights the need for systematic comparative evaluation between conventional end-to-end CNN classifiers and more robust hybrid designs that explicitly prioritize diagnostic safety.

Building upon these challenges and opportunities, this study seeks to develop a hybrid deep learning framework capable of enhancing the automated detection and classification of intracranial aneurysms, while simultaneously evaluating its performance against a strong baseline model. The proposed approach aims to support radiological decision-making by improving diagnostic reliability and reducing clinically critical false-negative cases, thereby contributing to safer and more effective clinical workflows.

2. Related Work

In this paper, the proposed method is connected to several numerous prior arts, which Several studies have investigated aneurysm detection and segmentation using deep learning, proposing diverse architectures and preprocessing strategies to improve diagnostic accuracy

Shahzad et al. (2020) Proposed an ensemble CNN for CTA in aSAH that performs joint detection and segmentation; on an independent test set (185 patients/215 aneurysms) the system achieved 87% sensitivity, 0.42 FP/scan, Dice 0.80 for aneurysms >30 mm^3, and 96% sensitivity, Dice 0.87, 0.14 FP/scan for >100 mm^3 highlighting strong performance at clinically relevant sizes. [6].

Zhu et al. (2022) addressed data scarcity by training a 3D U-Net for CTA-based aneurysm segmentation under small-sample regimes, reporting solid overlap scores (e.g., Dice) and boundary quality (Hausdorff distance) while analyzing vessel-level recall. Their results support the feasibility of volumetric segmentation pipelines even when annotations are limited, provided careful training and validation design[2].

Yuan et al. (2022) introduced DCAU-Net, a Dense Convolutional Attention U-Net that fuses CBAM with dense connectivity to better integrate spatial–semantic cues in MRA segmentation. The model demonstrated strong class-balance and size-robust behavior, with reported IoU ≈ 0.79, Precision ≈ 0.85, and F1 ≈ 0.87, illustrating how attention and dense feature reuse can elevate baselines in aneurysm delineation[5].

Di Noto et al. (2023) view a TOF-MRA aneurysm detection trained with weak sphere labels (≈4× faster to annotate) and anatomy-informed priors, reporting 83% sensitivity at 0.8 FP/patient on an internal set (284 subjects) and 68% sensitivity at 2.5 FP/patient on a public challenge; data/code/weights released, underscoring reproducibility. [7].

Ham et al. (2023) combined skeleton-guided 3D patches with a 3D U-Net and an auxiliary classifier to counter class imbalance in TOF-MRA. The approach attained Accuracy = 0.910 (internal) and 0.883 (external ADAM validation), evidencing clinically meaningful

generalization and showing that topology-aware sampling plus multi-task learning can boost IA detection/segmentation performance[8].
Hlavata et al. (2024) evaluated CNN-based automated IA classification on CT images, reporting near-98% Accuracy with well-balanced Precision/Recall/F1 across classes. By benchmarking multiple backbones (e.g., VGG/ResNet families), the study highlights transfer-learning baselines for centers lacking MRA/DSA, and frames classification as a strong triage step preceding targeted segmentation[4].
Hou et al. (2025) developed MGLIA-Net, integrating MobileNetV2 depthwise-separable convolutions into a GLIA-style U-Net for CTA segmentation. Across two datasets, the method improved upon the baseline, with reported Accuracy of 55.9% and 73.1%, respectively demonstrating that lightweight separable convolutions can strengthen aneurysm-focused architectures while controlling computational cost[9].

Tabel (1): summary of related work

Ref No.	Author (Year)	Dataset	Techniques	Limitation	Performance Metrics
[6]	Shahzad (2020)	CTA	Ensemble CNN (detection + segmentation)	Low sensitivity for small aneurysms	Sensitivity 87%, Dice 0.80 (>30 mm^3); Sensitivity 96%, Dice 0.87 (>100 mm^3)
[2]	Zhu (2022)	CTA	3D U-Net vs V-Net / Res-UNet	Limited labels (single-center)	Dice 0.818 ± 0.100
[5]	Yuan (2022)	MRA/ADAM	DCAU-Net (Dense blocks + CBAM)	High complexity, limited cross-center robustness	Dice ≈ 0.75, IoU ≈ 0.70, F1 ≈ 0.87
[7]	Di Noto (2023)	TOF-MRA	CNN + weak labels + priors	Drop in external performance	Sensitivity 83% (0.8 FP), 68% (2.5 FP challenge)
[8]	Ham (2023)	TOF-MRA	3D U-Net + skeleton patches + aux classifier	High computational cost	Accuracy 0.910 (internal), 0.883 (external)
[4]	Hlavata (2024)	CT	CNN (VGG/ResNet backbones)	No lesion localization	Accuracy ≈ 98%, F1 ≈ 97–98%
[9]	Hou (2025)	CTA	MGLIA-Net (MobileNetV2 + U-Net)	CTA-only, small aneurysm limitations	Accuracy 55.9% → 73.1% (two sets)

3. Research Gap, Objectives, Contribution

3.1 Research Gap

Despite the significant progress achieved in deep learning based intracranial aneurysm detection, several important research gaps remain:

1. Most existing approaches rely on large, manually annotated datasets and fully supervised training strategies, which are often difficult to obtain in real clinical settings.
2. Small and low-contrast aneurysms continue to pose a major challenge, leading to missed detections even in advanced deep learning models.
3. Severe class imbalance persists, as aneurysmal regions typically occupy only a small fraction of the image, negatively affecting model generalization and sensitivity.
4. Many studies depend exclusively on deep features extracted by convolutional neural networks, while underutilizing handcrafted structural descriptors that may provide complementary and clinically interpretable information.
5. In addition, classification is frequently performed directly using end-to-end CNN architectures without systematic comparison to strong baseline models, making it difficult to assess diagnostic reliability, particularly with respect to false-negative errors in safety-critical clinical applications.

3.2 Objectives

This paper aims to:

1. Develop a reliable automated framework for intracranial aneurysm detection using brain CTA images.
2. Improve vessel and aneurysm segmentation performance by employing an Attention U-Net model guided by synthetically generated vessel masks.
3. Extract and combine deep features with handcrafted structural and textural descriptors to better capture the morphological characteristics of aneurysms.
4. Perform a comparative evaluation between a direct CNN-based baseline classifier and the proposed hybrid framework to quantify the impact of feature fusion and external classification on diagnostic reliability.
5. Enhance diagnostic performance and robustness by employing an independent XGBoost classifier trained on fused feature representations.

3.3 Contributions

This paper makes the following contributions:

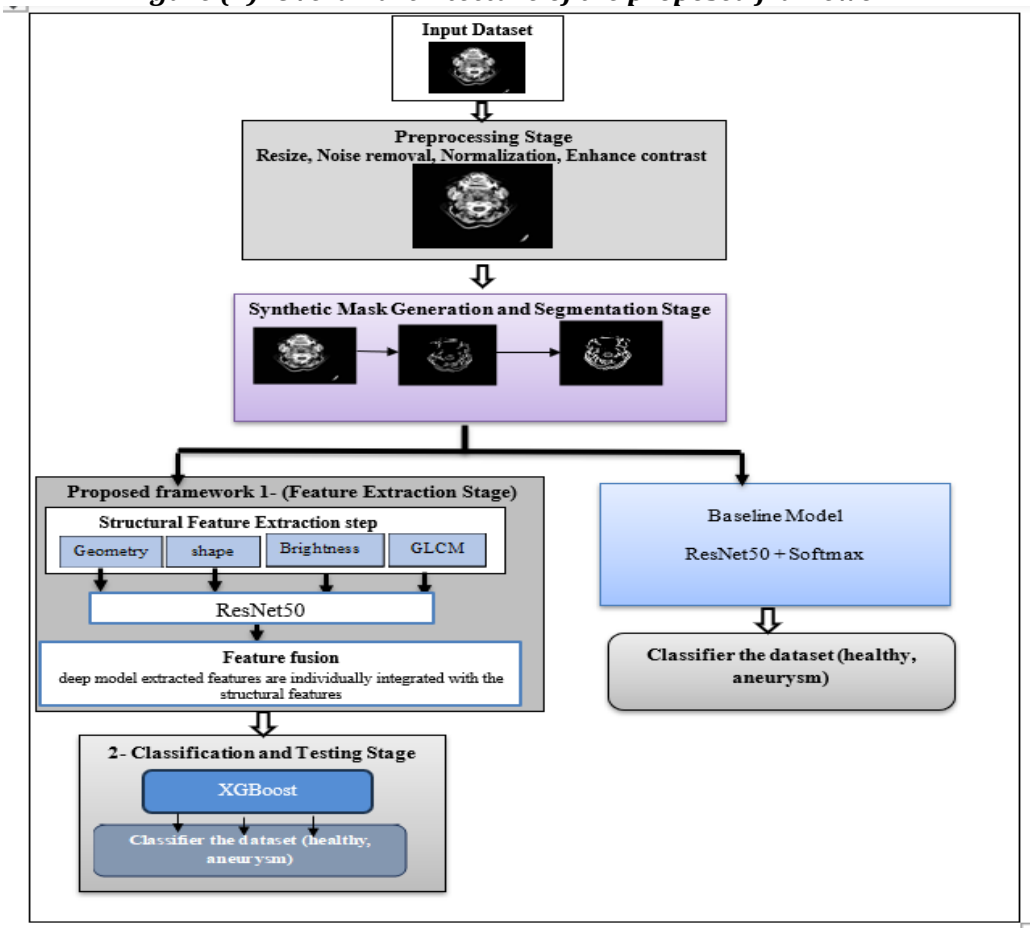
1. A hybrid segmentation strategy that utilizes synthetically generated vessel masks derived from classical image-processing techniques (e.g., Frangi filtering and CLAHE) to support an Attention U-Net architecture enhanced with CBAM and optimized using Tversky loss.
2. The integration of deep features extracted from ResNet50 with handcrafted structural descriptors related to shape, size, texture (GLCM), and intensity, resulting in a richer and more interpretable feature representation.
3. A systematic comparative study between a strong end-to-end CNN baseline and the proposed hybrid framework, demonstrating that direct CNN classification may lead to clinically critical false-negative cases.
4. An independent classification stage based on XGBoost trained on fused features, which improves diagnostic stability and eliminates false-negative predictions compared to relying solely on CNN-based classifiers.

3. Methodology

This section presents the complete framework developed for preprocessing brain CTA images and performing aneurysm detection, feature extraction, and final classification. The proposed methodology is structured as a multi-stage pipeline consisting of four main stages: (1) data preprocessing, (2) image segmentation, (3) feature extraction and feature

fusion, and (4) aneurysm classification using an XGBoost model. Figure 1 illustrates the overall architecture and workflow of the proposed system.

Figure (1): Overall architecture of the proposed framework



3.1 Data preprocessing

Data preprocessing was performed to standardize image dimensions and enhance vascular structures prior to segmentation. All CTA images were initially resized to 320 × 320 pixels. Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to improve contrast, followed by Gaussian filtering to reduce noise while preserving anatomical boundaries. Pixel intensity values were rescaled to the range [0, 255] to ensure consistency across the dataset.

To increase data variability and mitigate class imbalance, data augmentation was applied to the training set using the Keras ImageDataGenerator. Augmentation operations included small rotations (±10°), horizontal and vertical shifts (±5%), and horizontal flipping. Each original image was augmented to generate three additional variants, improving model robustness and generalization.

3.2 segmentation

The segmentation stage consists of two main components: (i) synthetic vessel mask generation and (ii) aneurysm segmentation using an Attention U-Net model. This stage is shared by both the baseline and the proposed frameworks, serving as a common preprocessing step to localize vascular structures and generate region-of-

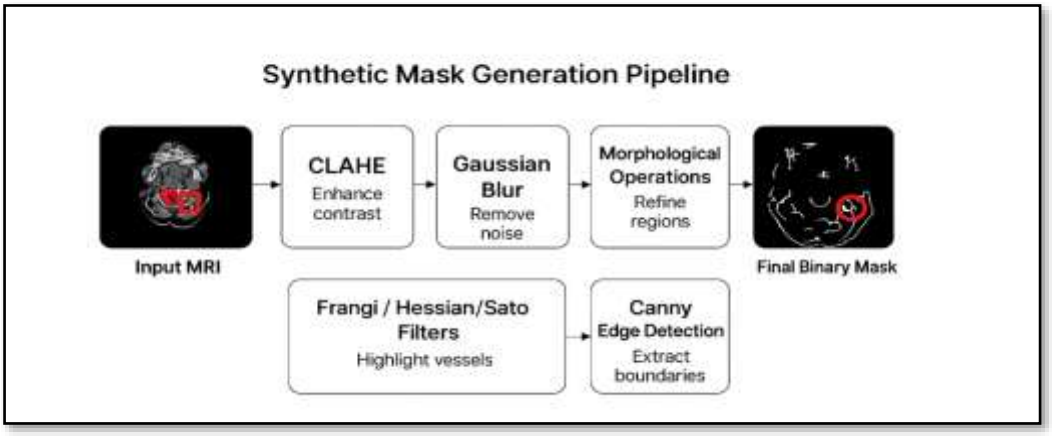
interest representations prior to classification. The subsequent classification strategy represents the main point of divergence between the two approaches.

3.2.1 Synthetic Mask Generation

Due to the absence of expert-annotated ground truth masks, synthetic vessel masks were generated using a classical image-processing pipeline (Figure 2). The process began with contrast enhancement using CLAHE followed by Gaussian blurring for noise reduction. Morphological operations and Otsu thresholding were then applied to obtain an initial vessel representation.

To enhance tubular vascular structures, multiscale vesselness filters including Frangi, Hessian, and Sato were employed. Final refinement steps consisted of morphological closing, white top-hat transformation to emphasize fine vessels, and Canny edge detection to delineate vessel boundaries. The resulting binary masks were used as surrogate training labels for the Attention U-Net and as a shared structural input for both the baseline and proposed classification pipelines.

Figure (2): Flowchart of the synthetic vessel mask generation pipeline.



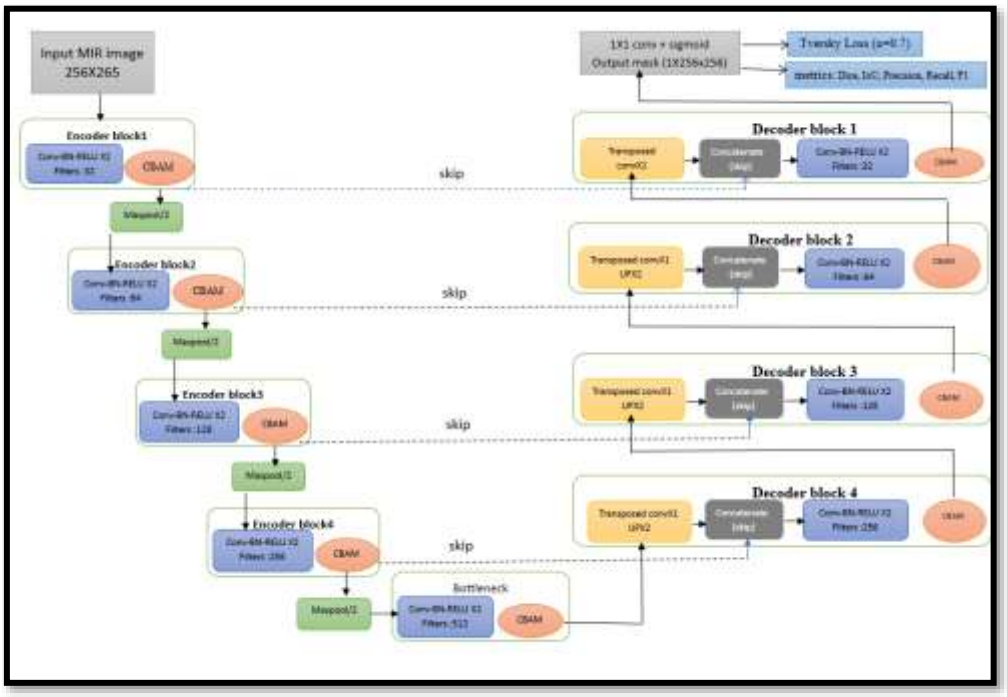
3.2.2 Segmentation

To enhance segmentation accuracy and address class imbalance, an Attention U-Net architecture augmented with Convolutional Block Attention Modules (CBAM) was employed. The attention mechanism improves feature representation by adaptively emphasizing informative spatial and channel-wise features.

The network was trained on 256 × 256 preprocessed CTA slices using synthetically generated vessel masks as surrogate labels. Training was performed for 30 epochs using the Adam optimizer and Tversky loss (batch size = 8). Skip connections combined with CBAM-enhanced encoder–decoder blocks helped preserve spatial consistency and improve vessel delineation, as illustrated in Figure 3.

The resulting segmentation masks were subsequently used as a shared input for both the baseline and the proposed classification pipelines, ensuring a fair and consistent comparative evaluation.

Figure (3): Overall architecture of the proposed Attention U-Net for aneurysm segmentation.



3.3 Feature Extraction

Feature extraction represents the main point of divergence between the baseline and the proposed frameworks. In the baseline design, classification relies solely on deep features extracted from the Global Average Pooling (GAP) layer of a ResNet50 model trained on the segmented aneurysm regions.

In contrast, the proposed framework augments these deep representations with a set of handcrafted structural features to capture complementary vascular characteristics. These features include morphological descriptors (area, perimeter, circularity, and elongation), texture measures derived from the Gray-Level Co-occurrence Matrix (GLCM), and intensity-based features reflecting local brightness variations.

The handcrafted features were concatenated with the ResNet50 deep feature vectors to form a unified hybrid feature space. This hybrid representation combines the strong discriminative power of deep learning with clinically interpretable structural information, enabling a more robust and reliable characterization of aneurysm patterns. By enriching the feature space beyond purely CNN-based embeddings, the proposed approach enhances diagnostic stability and reduces sensitivity to missed detections when compared to the baseline feature-only representation.

3.4 Classification

To perform the final classification, the proposed framework employs an XGBoost classifier rather than relying on conventional end-to-end CNN classifiers such as softmax or fully connected layers. XGBoost is well suited for heterogeneous and tabular feature representations and effectively handles class imbalance and non-linear feature interactions.

In the baseline framework, classification is performed directly using CNN-based decision layers operating solely on deep feature embeddings. In contrast, the proposed approach trains XGBoost on a fused feature set combining deep ResNet50 representations with handcrafted structural and textural descriptors. This independent classification stage

provides improved decision stability and robustness, leading to enhanced diagnostic reliability compared to direct CNN-based classification.
Algorithm 1 summarizes the main steps of the proposed aneurysm detection framework.

Algorithm1: Hybrid Aneurysm Detection proposed system
Input: Brain CTA images
Output: Classification result (Aneurysm / Healthy)
Begin
Step1: (Preprocessing)
- Resize images (320×320), apply CLAHE + Gaussian filter, normalize intensities.
- Augment training set with small rotations, shifts, and flips.

Step2: (segmentation)
Step (2.1): Synthetic Mask Generation
- Enhance vessels using CLAHE + Gaussian + Otsu thresholding.
- Apply morphological ops + Frangi/Hessian/Sato filters.
- Refine with white top-hat + Canny edges → binary mask labels.

Step (2.2): Segmentation
- Train Attention U-Net with CBAM using synthetic masks and Tversky loss.
- Generate predicted masks for all images.

Step3: Feature Extraction
Step (3.1): Deep features: extract embeddings via ResNet50 (GAP layer).
Step (3.2): structural features: compute geometry (area, perimeter, circularity), texture (GLCM), and intensity descriptors.
Step (3.3): Fuse features by concatenation into a unified vector.

Step 4: Classification
- Train XGBoost on fused features.
- Predict aneurysm status for new inputs.

End

3.5 Experimental Setup, Reporting, and Reproducibility

Dataset and Augmentation. The dataset comprised 552 brain CTA slices (300 healthy and 252 aneurysm-positive), collected from two publicly available sources: Kaggle [17] and the Roboflow platform [18], consistent with datasets reported in prior studies (e.g., Hlavata et al., 2024). To mitigate class imbalance and enhance data variability, augmentation was applied only to the training set, generating three additional variants per image using small rotations (±10°), horizontal and vertical shifts (±5%), and horizontal flipping. After augmentation, the training set contained 1,715 images. Identical datasets and augmentation strategies were used for both the baseline and proposed frameworks to ensure fair comparison.

Data Splitting Strategy. Data were stratified into training (70%), validation (15%), and testing (15%) subsets. Where subject identifiers were available, patient-level separation was enforced to prevent data leakage across splits.

Hardware and Software Environment.All experiments were conducted on a Windows 11 Pro (64-bit) workstation equipped with an AMD Ryzen 7 7435HS CPU and 32 GB RAM, without GPU acceleration. The software environment included Python 3.9.21, TensorFlow/Keras 2.12.0, OpenCV 4.11.0, NumPy 1.26.4, and XGBoost 2.0.3. A fixed random seed (42) was applied across all libraries to ensure reproducibility.

Training Configuration. Input images were resized to 256×256 pixels for segmentation. The Attention U-Net with CBAM was trained for 30 epochs using a batch size of 8, the

Adam optimizer, and Tversky loss (α = 0.7, β = 0.3), with early stopping based on validation Dice and model checkpointing. For classification, deep features extracted from the ResNet50 backbone were combined with handcrafted morphological, textural, and intensity descriptors and fed into an XGBoost classifier configured with 1,200 boosting rounds, a learning rate of 0.01, and a maximum depth of 6.

Baseline Models. For fair benchmarking, two baseline models were implemented under identical preprocessing, segmentation, and data-splitting conditions. First, a standard U-Net was trained for segmentation comparison. Second, a direct ResNet50-based classifier was employed using only deep feature embeddings derived from the segmented regions, without handcrafted feature fusion or an external classifier. This design enables a controlled comparison, where differences in performance can be attributed to feature representation and classification strategy rather than data or preprocessing variations.

Evaluation Metrics and Reporting. Segmentation performance was evaluated on the held-out test set using Dice coefficient, Intersection over Union (IoU), Precision, Recall, F1-score, and Accuracy. Classification performance was assessed using Accuracy, Precision, Recall, F1-score, Area Under the Curve (AUC), and confusion matrix analysis, with particular emphasis on false-negative and false-positive rates due to their clinical relevance. All preprocessing steps, augmentation strategies, training configurations, software versions, and random seeds were documented, and trained models, checkpoints, and execution scripts were archived to support full experimental reproducibility.

4. Results

The experimental results compare the proposed hybrid framework with a baseline ResNet50 end-to-end classifier under identical preprocessing and segmentation conditions.

For segmentation, the shared Attention U-Net with CBAM achieved a Dice coefficient of 0.8798 on the validation set, demonstrating accurate localization of aneurysm regions.

In classification, the baseline ResNet50 model achieved an accuracy of 90%, while the proposed feature-fusion framework with XGBoost improved performance to 98.46% accuracy and 100% recall.

Notably, the proposed approach eliminated false-negative aneurysm cases, highlighting its superior diagnostic reliability.

4.1 Metric Definitions

Evaluation metrics were employed to quantitatively assess the performance of both segmentation and classification tasks. For segmentation evaluation, overlap-based measures were used to quantify the agreement between the predicted masks and the reference masks.

The Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) are defined as:

$$Dice = \frac{2 \mid P \cap G \mid}{\mid P \mid + \mid G \mid}$$

$$IoU = \frac{\mid P \cap G \mid}{\mid P \cup G \mid}$$

where P and G represent the predicted and ground truth segmentation masks, respectively. These metrics are widely adopted in medical image segmentation to evaluate boundary accuracy and lesion localization quality.

For classification evaluation, standard performance metrics were derived from the confusion matrix , which summarizes true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions. The metrics are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
$$Precision = \frac{TP}{TP + FP}$$
$$Recall\ (Sensitivity) = \frac{TP}{TP + FN}$$
$$Specificity = \frac{TN}{TN + FP}$$
$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

In medical diagnosis, particular emphasis is placed on recall (sensitivity) and false-negative errors, as missed aneurysm cases may lead to severe clinical consequences. Therefore, these metrics were carefully analyzed when comparing the baseline and proposed frameworks.

4.2 Quantitative Evaluation

To objectively evaluate the performance of the proposed framework, quantitative metrics were reported for both segmentation and classification tasks. For segmentation, the Dice coefficient, Intersection over Union (IoU), Precision, Recall, and F1-score were used to assess the accuracy of aneurysm region localization, particularly for small and irregular vascular structures. For classification, Accuracy, Balanced Accuracy, Precision, Recall (Sensitivity), F1-score, Area Under the ROC Curve (AUC), and Average Precision (AP) were employed to provide a comprehensive assessment of diagnostic reliability.

Table 2 summarizes the quantitative performance of the proposed framework in both segmentation and classification tasks. The results demonstrate strong segmentation accuracy and highly reliable classification performance, with particular emphasis on sensitivity, which is critical in clinical aneurysm detection.

Table (2): Quantitative performance of the proposed framework in segmentation and classification

Task	Metric	Value
Segmentation	Dice (Validation)	87.98%
	IoU	80.56%
	Precision	85.01%
	Recall	93.85%
	F1-score	88.31%
Classification	Accuracy	98.46%
	Balanced Accuracy	98.28%
	Precision	97.30%
	Recall (Sensitivity)	100%
	F1-score	98.63%
	AUC	98.18%
	AP (Avg. Precision)	98.39%

To clearly demonstrate the contribution of the proposed framework, a direct quantitative comparison with a baseline ResNet50 end-to-end classifier is presented in Table 3. Both models were evaluated under identical preprocessing, segmentation, and data-splitting conditions to ensure a fair and unbiased comparison.

Table (3): Quantitative comparison between baseline and proposed framework

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
Baseline ResNet50 (End-to-End)	90.00	90.00	90.00	90.00	—
Proposed Hybrid Framework	98.46	97.30	100	98.63	98.18

The comparison in Table 3 confirms that the proposed hybrid framework significantly outperforms the baseline model across all classification metrics. In particular, the proposed approach eliminated false-negative aneurysm cases, resulting in perfect sensitivity, which is a crucial requirement for safe and reliable clinical decision-making.

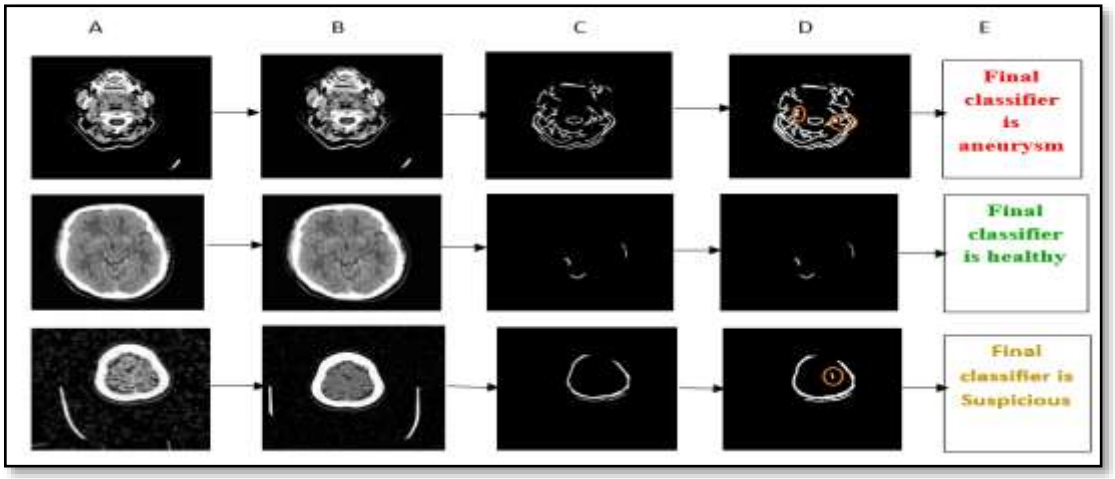
4.3 Visual Comparison

Figure 4 illustrates the visual pipeline of the proposed framework using three representative cases with different diagnostic outcomes. Each row corresponds to a specific scenario: an aneurysm-positive case (top row), a healthy case (middle row), and a challenging suspicious case with weak structural characteristics (bottom row).

The columns represent the sequential processing stages: (A) raw input CTA image, (B) preprocessed image after CLAHE and Gaussian filtering, (C) synthetic vessel mask generated using vessel enhancement filters and thresholding, (D) predicted mask obtained by the Attention U-Net with CBAM, and (E) final classification result produced by the feature-fusion strategy and XGBoost classifier.

The suspicious case highlights the strength of the proposed framework in handling borderline patterns that are often misclassified by end-to-end CNN baselines. By jointly analyzing deep representations and handcrafted structural features, the proposed system provides a more reliable decision in cases where vascular appearance alone is insufficient, thereby reducing the risk of false-negative diagnoses.

Figure (4): Visual pipeline of the proposed framework showing raw input, preprocessing, synthetic and predicted masks, and final classification for aneurysm, healthy, and suspicious cases.



4.4 Comprehensive Evaluation of the Proposed Framework

The evaluation of the proposed framework is presented in a sequential manner, beginning with the segmentation stage and followed by the classification stage, along with a direct visual comparison to a baseline model.

Figure (5) illustrates the training and validation curves of the Attention U-Net with CBAM for the segmentation task. The Dice coefficient steadily increases and stabilizes at a high value, while the loss curve decreases smoothly, indicating stable convergence and reliable learning of aneurysm regions.

Following segmentation, the classification performance is analyzed through a direct comparison between the proposed framework and a baseline ResNet50 end-to-end classifier. Figure (6) presents the classification training curves, where the proposed framework (top) demonstrates faster convergence, higher accuracy, and lower loss compared to the baseline model (bottom), reflecting improved stability and generalization.

The diagnostic reliability of both approaches is further examined using confusion matrices shown in Figure (7). The proposed framework (left) correctly identifies all aneurysm-positive cases, resulting in zero false-negative predictions. In contrast, the baseline classifier (right) produces multiple false-negative cases, indicating missed aneurysm diagnoses.

Overall, this sequential analysis confirms that while the segmentation stage performs reliably, the primary performance improvement is achieved at the classification stage. By separating feature extraction from decision making and integrating structural features with deep representations, the proposed framework significantly enhances diagnostic reliability and reduces clinically critical false-negative errors compared to the baseline approach.

Figure (5): Training and validation curves of the proposed Attention U-Net with CBAM. The Dice coefficient and accuracy show stable convergence, while the loss curve decreases smoothly, indicating robust and reliable segmentation performance without overfitting.

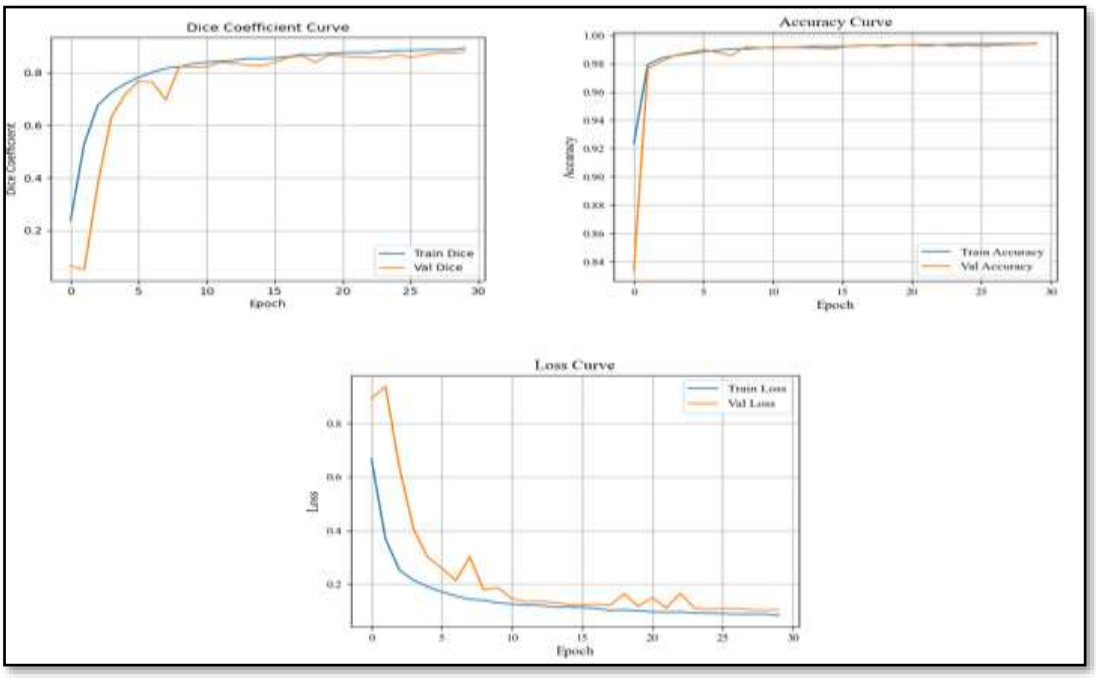


Figure (6): Classification curves comparing the proposed framework (top) and the baseline ResNet50 model (bottom), highlighting improved convergence and accuracy of the proposed approach.

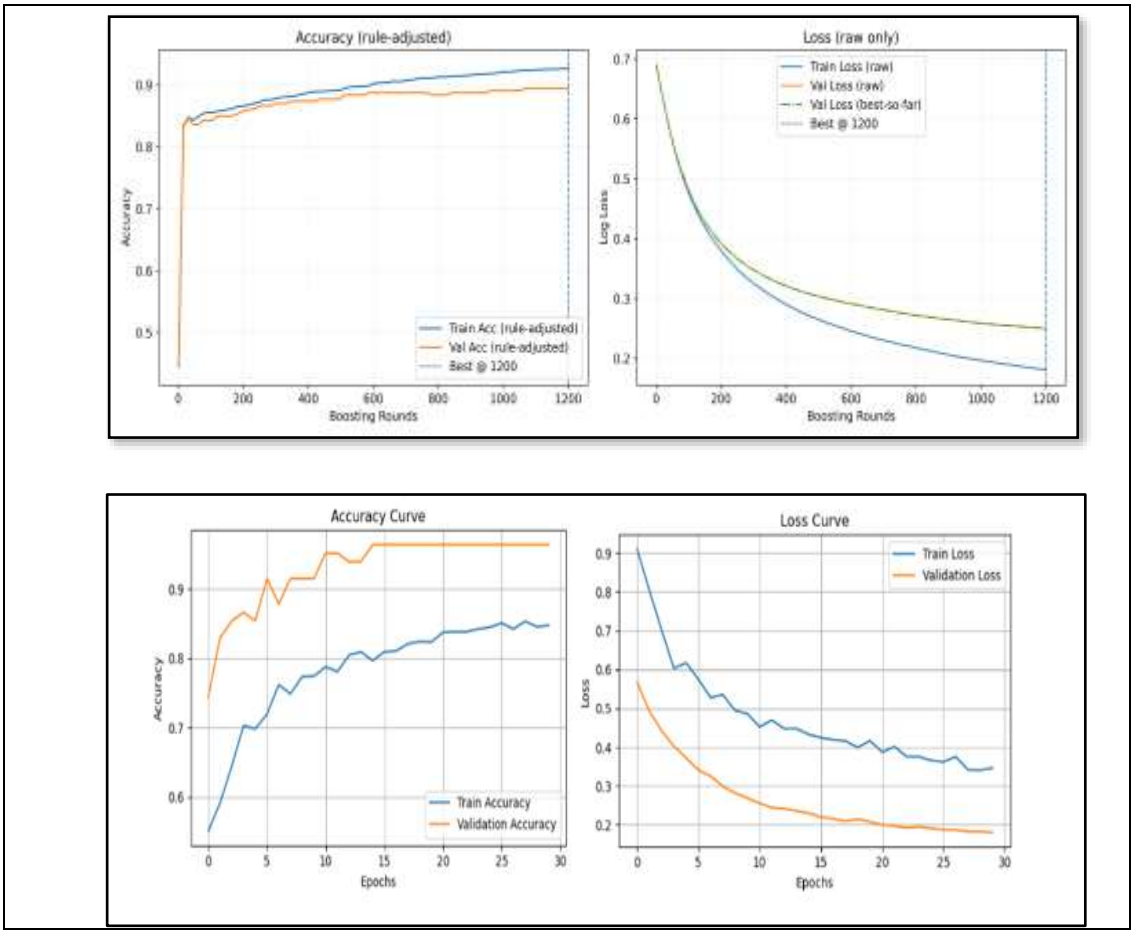
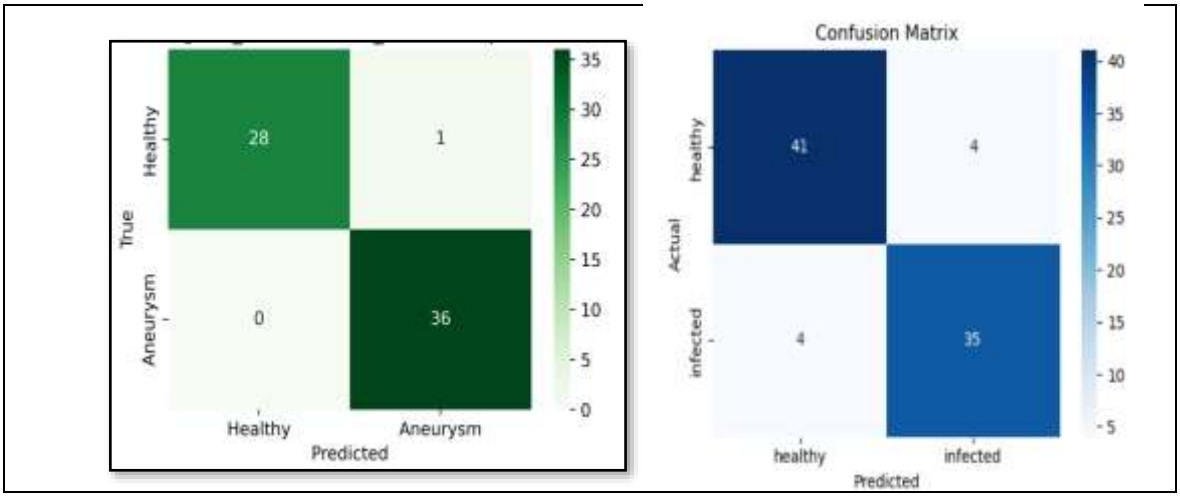


Figure (7): Confusion matrices of the proposed framework (left) and the baseline model (right), highlighting the superior diagnostic reliability of the proposed approach.



5. Discussion

The proposed framework was designed as a multi-stage pipeline in which each component plays a specific role in improving intracranial aneurysm detection. The preprocessing stage enhanced vascular contrast and reduced noise, providing more

reliable inputs for subsequent stages. In the absence of expert annotations, synthetic vessel masks were employed as a practical alternative for training the segmentation model, although this choice remains an inherent limitation compared to true ground-truth labels.

Segmentation was performed using an Attention U-Net enhanced with CBAM and optimized with Tversky loss. As shown in Figure (5), the segmentation model achieved stable convergence with high Dice and Recall values, indicating effective localization of aneurysm regions, including low-contrast and irregular structures. Importantly, this segmentation stage was identical for both the proposed framework and the baseline approach, ensuring a fair comparison in the downstream classification task.

The main performance gain of the proposed framework emerges at the classification stage. By combining handcrafted structural features (shape, size, texture, and intensity) with deep features extracted from ResNet50, a richer and more informative feature representation was obtained. When these fused features were classified using XGBoost, the proposed framework achieved an accuracy of 98.5% with zero false-negative aneurysm cases. In contrast, the baseline ResNet50 end-to-end classifier, trained under the same conditions but without feature fusion or an external classifier, showed lower accuracy and multiple false-negative predictions, as illustrated in Figures (6) and (7).

This comparison demonstrates that relying solely on end-to-end CNN classification may be insufficient for clinically sensitive tasks such as aneurysm detection. Separating feature extraction from decision-making and incorporating interpretable structural information significantly improves diagnostic reliability, particularly by reducing false-negative errors, which are critical in medical applications.

Despite these promising results, several limitations should be acknowledged. The dataset size remains relatively small, the segmentation relies on synthetic masks, and the experiments were conducted on 2D image slices rather than full 3D volumes. Future work will focus on validating the framework on larger multi-center datasets, extending the approach to 3D architectures, and incorporating more accurate expert annotations to further improve robustness and generalizability.

Finally, the comparative analysis with recent studies summarized in Table (4) indicates that the proposed hybrid framework achieves competitive or superior performance in both segmentation and classification tasks. These findings highlight the advantage of integrating attention-based segmentation with hybrid feature fusion and an independent classifier, offering a more reliable and clinically meaningful solution for intracranial aneurysm detection.

Table (4): summary comparison with recent studies.

Ref	Author (Year)	Dataset	Techniques	Limitations	Performance Metrics
[1]	Shahzad (2020)	CTA	Ensemble CNN for detection + segmentation	Sensitivity for small aneurysms needs improvement	Sensitivity 87%, Dice 0.80, FP/scan 0.42
[3]	Yuan (2022)	MRA/ADAM	DCAU-Net (Dense blocks + CBAM)	Heavy model; limited generalization	Dice = 0.7455, IoU ≈ 0.70
[4]	Di Noto (2023)	TOF-MRA	CNN with weak labels + anatomical priors	Drop in generalization (public vs. in-house)	Sensitivity 83% (in-house); 68% (challenge)

[5]	Ham (2023)	TOF-MRA (internal 154; external 113/ADAM)	3D U-Net + auxiliary classifier	Medium cohort; patch-based design	Accuracy (int.) 0.910; (ext.) 0.833
[6]	Huayta (2024)	CTA (classification)	CNNs incl. ResNet50/101/152, VGG16	No lesion localization	Accuracy > 98%, Precision/Recall/F1 ≈ 97–98%
	Proposed Work (2025)	CTA (560 slices; healthy + aneurysm)	Attention U-Net + CBAM + Tversky loss for segmentation; handcrafted + ResNet50 deep features; XGBoost classifier	Small dataset; synthetic masks; 2D slices	Dice = 0.8795, IoU = 0.8056 (seg.); Accuracy = 98.5% (classif.); No false negatives

6. Conclusion and Future Work

6.1 Conclusion

This paper investigated intracranial aneurysm detection through a structured multi-stage framework and a direct comparison with a baseline end-to-end ResNet50 model. The segmentation stage, based on Attention U-Net with CBAM and Tversky loss, provided consistent and accurate localization of aneurysm regions, confirming its suitability as a reliable preprocessing step for subsequent diagnosis.

The main contribution of this paper appears at the classification stage. Unlike the baseline approach, which relied solely on end-to-end deep learning, the proposed framework separated feature extraction from decision making. By combining deep features with handcrafted structural information and using an independent XGBoost classifier, the system achieved higher diagnostic stability, improved overall accuracy, and avoided false-negative aneurysm predictions.

These findings indicate that improving classification reliability, rather than increasing network complexity, plays a critical role in aneurysm detection. The proposed framework therefore offers a more balanced and clinically safer alternative to conventional CNN-based pipelines.

6.2 Future Work

Future research will focus on:

1- increasing the dataset size and including multi-center data for better generalization.
2- developing or adopting 3D models to capture volumetric information more effectively.
3- improving the quality of annotations and reducing dependence on synthetic masks.
4- exploring additional handcrafted and deep feature combinations.
5- evaluating the system in real clinical environments to assess practical applicability.

Funding
None
Acknowledgment
None

Conflicts of interest

The author declares no conflict of interest

REFERENCES

[1] T. Di Noto, M. Artzi, O. Aizenstein, *et al.*, "Towards Automated Brain Aneurysm Detection in TOF-MRA: Open Data, Weak Labels, and Anatomical Knowledge," *Neuroinformatics*, vol. 21, pp. 21–34, 2023,

[2] Z. Zhu, X. He, W. Chen, *et al.*, "Deep learning-based recognition and segmentation of intracranial aneurysms under small sample size," *Frontiers in Physiology*, vol. 13, Art. no. 1084202, Dec. 2022, doi: 10.3389/fphys.2022.1084202.

[3] B.-s. Kim, "Unruptured Intracranial Aneurysm: Screening, Prevalence and Risk Factors," *Neurointervention*, vol. 16, pp. 201–203, 2021, doi:10.5469/neuroint.2021.00451

[4] Hlavata *et al*, "Automated Method for Intracranial Aneurysm Classification Using Deep Learning," *Sensors*, vol. 24, Art. no. 4556, 2024, doi: 10.3390/s24144556.

[5] W. Yuan *et al.*, "DCAU-Net: Dense convolutional attention U-Net for segmentation of intracranial aneurysm images," *Visual Computing for Industry, Biomedicine and Art*, vol. 5, Art. no. 9, 2022, doi: 10.1186/s42492-022-00110-6.

[6] R. Shahzad *et al.*, "Fully automated detection and segmentation of intracranial aneurysms in subarachnoid hemorrhage on CTA using deep learning," *Scientific Reports*, vol. 10, Art. no. 21799, 2020, doi: 10.1038/s41598-020-78384-1.

[7] C. O'Rourke, S. Glover, R. Guerrero, *et al.*, "Towards Automated Brain Aneurysm Detection in TOF-MRA: Open Data, Weak Labels, and Anatomical Knowledge," *Neuroinformatics*, vol. 21, pp. 21–34, 2023, doi: 10.1007/s12021-022-09597-0

[8] S. Ham, J. Seo, J. Yun, *et al.*, "Automated detection of intracranial aneurysms using skeleton-based 3D patches, semantic segmentation, and auxiliary classification," *Scientific Reports*, vol. 13, Art. no. 12018, 2023, doi: 10.1038/s41598-023-38106-2.

[9] W. Hou *et al.*, "CTA image segmentation method for intracranial aneurysms based on MGLIA net," *Scientific Reports*, vol. 15, Art. no. 10593, 2025, doi: 10.1038/s41598-025-95143-2.

[10] F. Shahoveisi, H. T. Gorji, S. Shahabi, S. Hosseinirad, S. Markell, and F. Vasefi, "Application of image processing and transfer learning for the detection of rust disease," *Scientific Reports*, vol. 13, p. 5133, 2023.

[11] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, no. 29, pp. 1–28, Aug. 2015.

[12] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.

[13] H. K. Abdul Atheem, I. T. Ali, and F. A. Al Alawy, "A comprehensive analysis of deep learning and swarm intelligence techniques to enhance vehicular ad-hoc network performance," *Journal of Soft Computing and Computer Applications*, vol. 1, no. 1, Art. 1004, pp. 1–12, 2024. doi: [10.70403/3008-1084.1004]

[14] Z. S. Mahdi, R. M. Zaki, A. K. Farhan, and N. Majma, "Development of a hybrid methodology of deep learning and machine learning for lung nodule detection in medical computed tomography images," *Journal of Soft Computing and Computer Applications*, vol. 1, no. 2, Art. 1011, pp. 1–12, 2024. doi: [10.70403/3008-1084.1011]

[15] E. Talib, A. S. Jamil, N. F. Hassan, and M. E. Rana, "The robust digital video watermarking methods: A comparative study," *Journal of Soft Computing and Computer Applications*, vol. 1, no. 1, Art. 1002, pp. 1–10, 2024. doi: [10.70403/3008-1084.1002]

[16] S. S. Hamad, N. J. Ibrahim, and M. N. Fadhil, "Texture features with Resnet50V2 for face spoofing detection," *AIP Conference Proceedings*, vol. 3264, no. 1, Art. 030033, pp. 1–11, Mar. 2025. doi: [10.1063/5.0260733].
[17]Kaggle Dataset. (n.d.). Brain MRI aneurysm detection. Retrieved from: https://www.kaggle.com
[18]Roboflow Dataset. (n.d.). *Tumor, cancer, and aneurysm detection*. Retrieved from: https://universe.roboflow.com/discoverai/tumor-cancer-aneurysm-detection
[19] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale Vessel Enhancement Filtering," in *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98*, 1998, pp. 130–137.
[20] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
[21] L. G. Nyúl, J. K. Udupa, and X. Zhang, "New Variants of a Method of MRI Scale Standardization," *IEEE Transactions on Medical Imaging*, vol. 19, no. 2, pp. 143–150, Feb. 2000.
[22] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017.
[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, Las Vegas, NV, USA, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
[24] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. ICML*, Long Beach, CA, USA, 2019, pp. 6105–6114.
[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015.