Grenze International Journal of Engineering and Technology, June Issue

A Comparative Study of Intracranial Aneurysm Diameter Measurement Approaches

Prasanna U M[1], B S Harish[2] and C K Roopa[3]
[1-3]Department of Information Science and Engineering, JSS Science and Technology University, JSS TI Campus, Mysore 570006, Karnataka State, India
Email: pachu.um@gmail.com

***Abstract*— Intracranial Aneurysm are balloon like structures in the brain arteries. These are formed at weak locations in the arteries, like at vessels forks and junctions. Blood get accumulated in these balloon like structures and are risk of rupture. Intracranial aneurysm rupture lead to subarachnoid hemorrhage, wherein the blood flows in space surrounding the brain. This could be a fatal and life threatening situation. It could lead to death of the patients in couple of months or lead to permanent physical, cognitive or physiological disability. Early detection and measuring the size of the aneurysm, would be beneficial to the patient. Physicians could diagnose the severity, analyze the rupture risk and plan appropriate treatment. In the current work, we have explored 8 approaches to measure the max diameter size of the aneurysm. Dice and IoU are commonly used metrics for measuring the quality of the segmentation overlap between ground truth mask and predicted mask. But both the metrics are insensitive to boundary errors. Meaning, we might get a good dice and IoU score even when the edge alignment is poor. For a proper diameter measurement between ground truth mask and predicted mask, we have considered 3 alternative metrics, HD95(Hausdorff Distance), ASD(Average Surface Distance) and Boundary IoU. A Unet based model is used segmenting the aneurysm from 2D MIP images. The 2D MIP images are generated from TOF-MRA from the ADAM dataset. Segmentations which have Dice and IoU >70% are selected for experimentation. HD95, ASD and Boundary IoU are computed for these selected segmented masks. Based on the scores of these three boundary based metrics, the images are categorized as Excellent, Good, Moderate and Poor boundary match. We found that there were no excellent category images where the segmentation ground truth and predicted mask matched very well. About 8% of the data were good boundary match. 84% were in moderate boundary match and 8% in poor boundary match. These images are used for measuring the diameter of the predicted mask and ground truth mask.**

***Keywords*— HD95, ASD, Boundary IoU, BoundingBox, MaxFeret, PCA, EllipseFit, MinBoundingRect, ImageMoments, Geodesic, DT_MaxChord.**

I. Introduction

Intracranial aneurysm are balloon like dilations in the arteries of the brain. There are risk of them rupturing leading subarachnoid hemorrhage [1], which could be mortal or could lead to severe morbidity [2]. One of the common treatment is surgical clipping. This make it essential to measure the diameter of the aneurysm. Other parameters that affect the treatment are the morphology and the growth rate.

Grenze ID: 01.GIJET.12.2.16_25

Therefore segmenting aneurysm accurately from any modality like MRA, CTA or DSA is very crucial. Additionally measuring the maximum diameter of the aneurysm is very essential. Deep learning model have show good segmentation results on some popular publicly available datasets and also on private datasets. The dice score range between 0.8 to 0.9 [5,6,8]. But even when segmented mask extends the boundary outside slightly or slightly lesser and falls inside the boundary of the aneurysm, the Dice and IoU might be still showing a good score. This could lead to incorrect size measurement. For aneurysm which are smaller in size this error in diameter measurement could be very large, which could impact the clinical decision.
In this work, there are primarily try to analyze two issues. (A) Which are the different geometric approaches for measuring diameter and how do they impact the clinical decision. (B) Which metrics apart from Dice and IoU help in understanding segmentation quality beyond area overlap.
In this work, eight diameter measurement approaches are explored and compared on segmentation mask predicted by the Unet model inferenced on ADAM dataset. The HD95, ASD and Boundary IoU are computed across multiple dilation tolerances with cumulative normalized surface distance (NSD) curves. A custom grading framework, A to F is defined, to interpret the metrics to a single interpretable score.

II. Literature Survey

A. Aneurysm Segmentation with Deep Learning

There has been an ongoing research in the area of automated aneurysm segmentation. From early research approaches like, level-set deformable models [3] and random forest pipelines [4] , the research has progress to fully convolutional networks [7], nnU-Net [8], and attention-gated architectures [9]. The ADAM challenge [10] provides the primary dataset for researchers. Top models achieve Dice > 0.80 on TOF-MRA, though performance degrades significantly for small aneurysms (< 4 mm) [10]. There are not much research done on boundary precision of the segmented masks.

B. Diameter Measurement in Medical Imaging

The maximum diameter of the aneurysm is one of the key parameter which is used in clinical practices for managing aneurysms [11]. Radiologists measure it as the Feret caliper distance on MIP images. Automated size estimation approaches include bounding-box diagonals [12], ellipse fitting from region moments [13], PCA projection [14], rotating-calipers minimum bounding rectangles [15], and medial-axis skeleton analysis [16]. In the current work, we have measured the diameter of intracranial aneurysm using 8 such approaches and compared them.

C. Boundary Precision Metrics

Hausdorff Distance (HD) and its robust 95th-percentile variant (HD95) are standard boundary error metrics [17]. Average Surface Distance (ASD) provides a more stable cumulative measure [18]. Boundary IoU (BIoU), introduced for instance segmentation evaluation [19], directly measures contour alignment quality through dilated boundary band overlap. The Normalized Surface Distance (NSD) at varying tolerances gives a cumulative view of boundary precision [20]. Thoug boundary metrics provide good information there are less papers which utilize them in aneurysm segmentation.

III. Methodology

The experiment contains multiple steps. The workflow of the experiment is as show in Fig 1

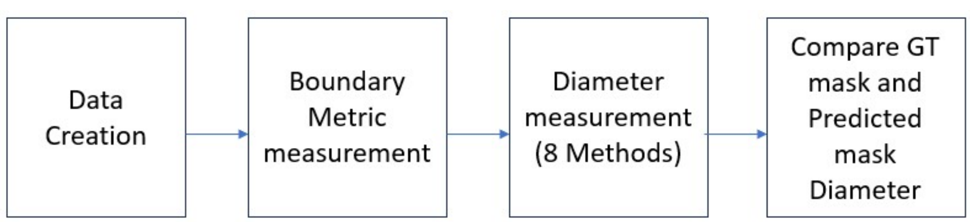


Fig 1: Workflow in the experiment

A. Dataset Creation

The data for the experiment is created from ADAM dataset. The steps to create the data is as explained below.
Step 1: 2D MIP generation

The TOF-MRA images in the dataset are converted to 2D MIP images. MIP are generated along standard angles, Axial, Coronal and Sagittal. Apart from these, the original TOF-MRA is rotated at 11 oblique angles [15, 30, 45, 60,75, 90, 105,120, 135, 150, 165] and MIP is taken along the axial, coronal and sagittal plane of the rotated MRA.

Step 2: Training UNet Model

Uniform patches are cropped from each MIP and it used for training an UNet model for detecting and segmenting aneurysms.

Step 3: Inferencing

The trained model is used for inferencing on cropped patches from a set of test MIP data.

Step 4: Shortlisting high dice score images

Resultant segmented mask, which have dice score >70 are selected and used for the experiment.

Fig2 below shows the distribution of the selected images and the Dice score range

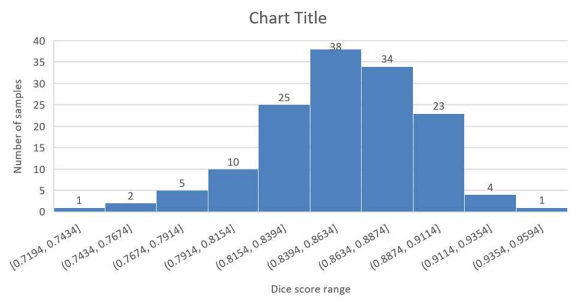


Fig 2: Dice score range and number of samples in each range.

A total of 143 TOF MIP patches are considered for the experiment, wherein the dice score ranges from 71.94% to 95.94%. Out of these ~84% of the data have the dice in the range 81.5% to 91.4%. The patches contain 4 information. (i) The original TOF MIP Patch, (ii) The ground truth GT binary mask, (iii) the predicted segmented mask and (iv) the full brain MIP reprojection of the location of the aneurysm. A sample patch image is shown below in Fig 3.

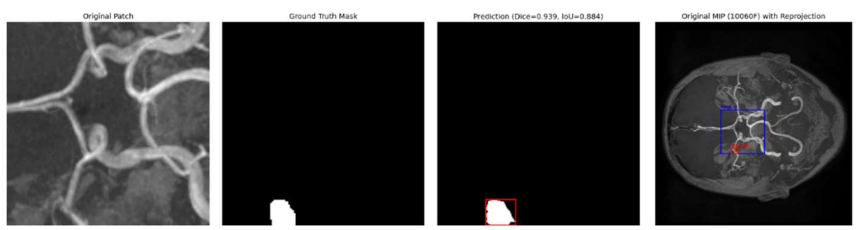


Fig 3: Sample patch image used in the experiment with Dice 93.9%

B. Boundary Metrics

It has been observed that though the Dice and IoU give information on the overlap, there could be cases where the boundary of ground truth may not match with the predicted mask boundary. When measuring the diameter of the mask, improper boundary match, could lead to improper measurements. Which could impact the clinical analysis and also surgery. Hence it becomes essential to measure the boundary mismatch using suitable metrics. We have considered three metrics to measure this mismatch. Fig 4 shows a sample of GT and Predicted mask overlapped and also the boundary mismatch

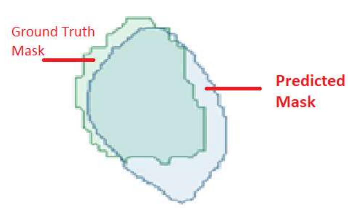


Fig 4: Sample ground truth and predicted mask overlapped

In this work, 3 boundary metrics are considered.

- Hausdroff distance, which measures the maximum distance from one surface to another. In our case, the two surfaces are ground truth mask and predicted mask. A higher Hausdroff distance indicates a larger mismatch. Figure 4 (a) shows a sample of Hausdroff distance measurement.
- Average surface distance, which measures the average of all the distance from one surface to another across all points. A lower ASD indicates lesser mismatch. Figure 4 (b) shows a sample of ASD metric computed.
- Boundary IoU, which measures the overlapping boundary regions of the ground truth and predicted mask. Figure 5 (b) shows a sample of Boundary IoU

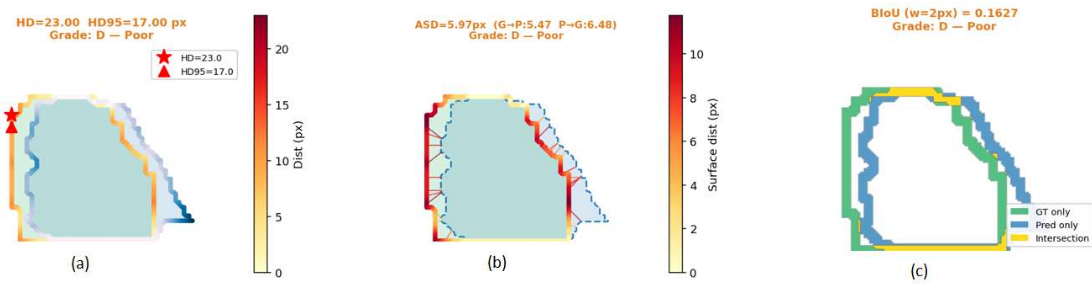


Fig 5: Hausdroff distance (a), ASD (b), and BIoU (c) for the sample image shown in Fig 3

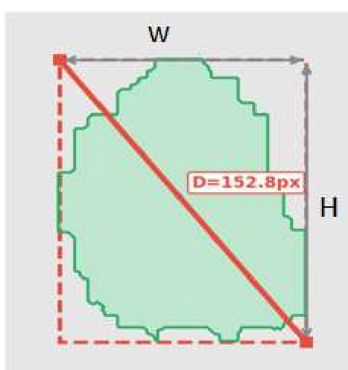
As you can see in the Fig 3 and Fig 5, even when though the Dice score is very high 93.9%, the boundary metrics indicates a bad boundary match. This is crucial for medical images, wherein anomaly in medical domain need to be very precisely segmented. The boundary metric is computed for all the segregated images.

C. Diameter measurement

In the experiment, we have considered 8 diameter measurement approaches. Each of the approach is explained below with samples image

D. Bounding box diagonal

The bounding box (or axis-aligned bounding box, AABB) is the smallest rectangle with sides parallel to the image axes that contains all mask pixels. The diameter is estimated as the diagonal of this rectangle the longest distance that fits within the box corners. Fig 6 shows a sample bounding box for a blob. Fig 7 shows the bounding box diagonal for the sample image in Fig 3



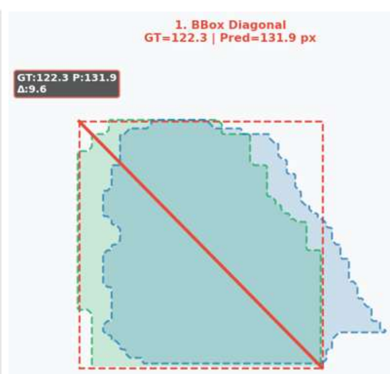


Fig 6: Sample blob and the bounding box diagonal

Fig 7: Bounding box diagonal for sample in Fig 3

$$H = r\_\max - r\_\min \ (Height\ of\ bounding\ box) \quad \text{Eq-1}$$

$$W = c\_\max - c\_\min \ (Width\ of\ bounding\ box) \quad \text{Eq-2}$$

$$D_{bbox} = \sqrt{H^2 + W^2} \quad \text{Eq-3}$$

Where

r_max and r_min are maximum and minimum rows in the foreground pixel in the blob image

c_max and c_min are maximum and minimum columns in the foreground pixel in the blob image

E. Max Feret Diameter / Caliper

The Feret diameter [11] is the perpendicular distance between two parallel tangent lines on opposite sides of the blob, just like using a physical caliper. The maximum Feret diameter is computed by rotating the caliper through all angles and taking the maximum. For a convex shape, this equals the longest Euclidean distance between any two boundary points. Fig 8 and Fig 9 shows the sample max ferret diameter for sample blob and also for the image in Fig 3

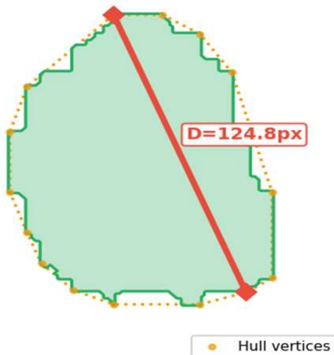


Fig 8 Sample blob and Max ferret diameter/ caliper

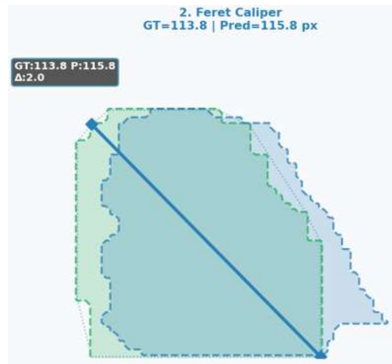


Fig 9 Max Feret/caliper for sample in Fig 3

$$mask = \{ (r_{i,} c_i) : mask(r_{i,} c_i) = 1 \qquad \text{Eq 4}$$

$$H = ConvexHull\ (r_{i,} c_i) \qquad \text{Eq 5}$$

$$||p - q|| = \sqrt{(r_p - r_q)^2 (c_p - c)^2} \qquad \text{Eq 6}$$

$$D_{feret= max_{p,q \in vertices(H)}} ||p - q|| \qquad \text{Eq 7}$$

Where $r_{i,}$ is row index and c_i Colum index

Where $||p - q||$ is the euclidean distance between two points p and q

F. PCA Major Axis

Principal Component Analysis (PCA) treats the binary mask as a 2D point cloud and finds the direction of maximum variance the axis along which the data is most spread out. The mask diameter is the total extent of the point cloud projected onto this principal axis. This is statistically robust because every pixel contributes equally, and it is invariant to rotation

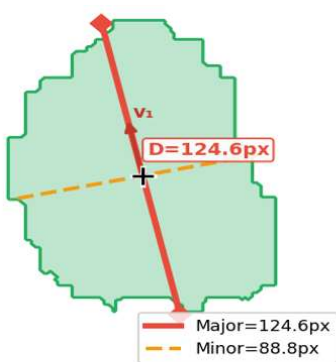
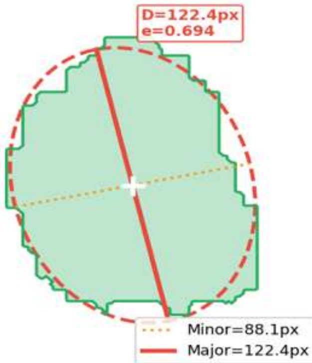


Fig 10 Sample blob and PCA

Fig 11 PCA for sample in Fig 3

G. Eclipse fit major axis

Rather than working directly with the boundary, this approach fits the ellipse that has the same area and second-order spatial moments as the binary region. The major axis of this equivalent ellipse is used as the diameter estimate.



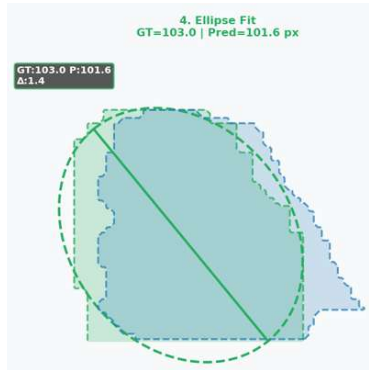
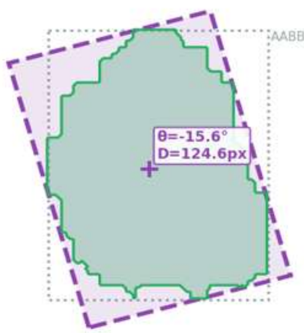


Fig 12 Sample blob and Eclipse fit

Fig 13 Eclipse fit for sample in Fig 3

H. Minimum bounding rectangle (Rotating calipers)

The Minimum-Area Bounding Rectangle (MABR) [15] is the tightest-fitting rectangle of any orientation that encloses the shape. Unlike the axis-aligned bounding box (which can be very loose for tilted shapes), the MABR is rotation-invariant and provides a much tighter upper bound. The Rotating Calipers algorithm (Toussaint, 1983) computes it efficiently in O(k) after convex hull construction



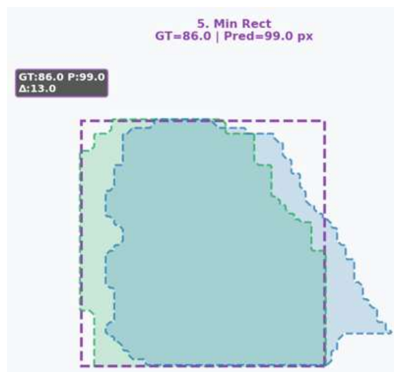
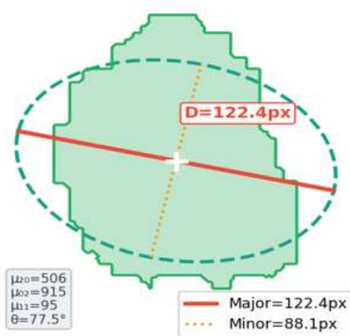


Fig 14 Sample blob and Minimum bounding rectangle

Fig 15 Minimum bounding rectangle for sample in Fig 3

I. Image moments major axis

Image moments are weighted averages of pixel intensities. Second-order central moments form an inertia tensor that describes how mass (pixels) is distributed around the centroid. The eigenvalues of this tensor give the semi-axis lengths of the equivalent inertia ellipse. This is analytically identical to the Ellipse Fit approach but computed directly via OpenCV moments, which provides a fully analytical formula and the orientation angle as a natural byproduct.



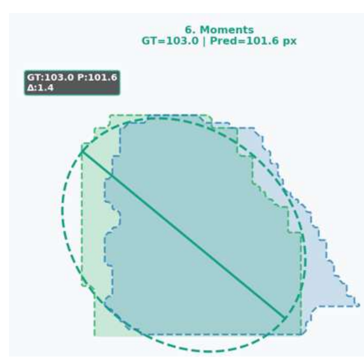


Fig 14 Sample blob and Image Moments

Fig 15 Image moments for sample in Fig 3

J. Geodesic skeleton diameter

Considering the irregular boundary of the aneurysm, a straight line may sometimes not be a good measure of the biggest diameter. Hence considering the medial axis could be beneficial. This leads us to geodesic skeleton approach. In geodesic skeleton approach, the medial axis is first identified and then the longest shortest path is computed. This is the geodesic diameter.

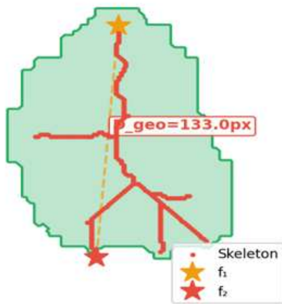


Fig 16 Sample blob and Geodesic skeleton

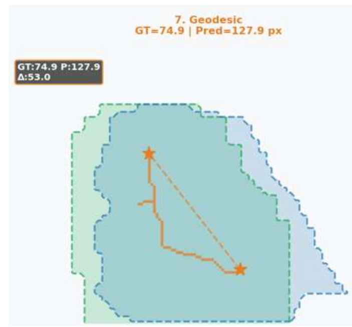


Fig 17 Geodesic skeleton for sample in Fig 3

K. DT Max-Chord

In the DT Max chord approach, first the largest circle that fits the entire blob is computed. This biggest circle touches the boundary of the blob. The center of the circle is the deepest interior point in the blob. Then a chord drawn passing through the center and touches the boundary of the blob. The entire length of this chord is calculated. The chord is then rotated all along the circle upto 1800 each time measuring the length of the chord. The longest length is the DT max chord.

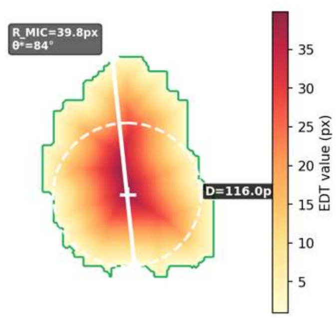


Fig 18 Sample blob and DT-Max chord

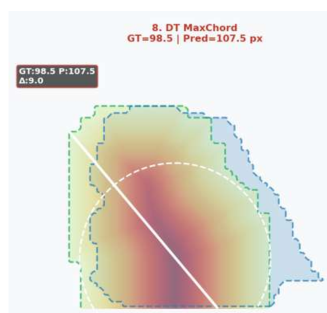


Fig 19 DT Max chord for sample in Fig 3

L. Quality grading framework

For each of the data point, all the five boundary metrics are computed. Each of the boundary metric range are different and based on the value we could grade them as excellent, good, moderate or poor boundary match. For each of the image in the dataset, the 5 boundary metric is measured and the average of all the 5 scores is computed. Based on this average score, the images could be grouped into 5 categories. These five categories are named A to F. Based on the average score, they are mapped to the category.

$$Score\ overall = \left(\frac{1}{5}\right) * [\,GS(Dice) + GS(IoU) + GS(HD95) + GS(ASD) + GS(BIoU)]$$

Overall grade: A if score >= 4.5, B if >= 3.5, C if >= 2.5, D if >= 1.5, else F.

Table 1 shows the different grade that is described above. Based on the metric computed they are grouped in to the below range.

TABLE I: CUSTOM GRADING FRAMEWORK

Metric	Grade A (Excellent)	Grade B (Good)	Grade C (Moderate)	Grade D (Poor)
Dice	>= 0.90	>= 0.80	>= 0.70	>= 0.60
IoU	>= 0.82	>= 0.67	>= 0.54	>= 0.40

HD95 (px)	<= 3.0	<= 6.0	<= 12.0	<= 20.0
ASD (px)	<= 1.5	<= 2.5	<= 5.0	<= 8.0
BIoU(w=2)	>= 0.65	>= 0.45	>= 0.25	>= 0.10

IV. Results

A. Diameter Measurement Comparison

For each of the image in the dataset, diameter are measured for both the ground truth mask and also for the predicted mask. Table 2, provides statistical information, like mean, min, max and difference between the Predicted and ground truth measurements.

TABLE II: COMPARISON OF THE 8 DIAMETER MEASUREMENT ALGORITHM

Approach	Mean GT	Std GT	Min GT	Max GT	Mean Pred	Std Pred	Mean |Diff|	Std |Diff|	Mean Ratio
BoundingBox	95.68	65.95	12.21	286.94	103.86	81.68	26.10	79.49	2.280
MaxFeret	79.71	57.17	9.22	283.03	84.65	71.10	25.47	73.75	2.536
PCA	78.46	56.56	9.00	283.02	83.24	70.18	25.12	72.80	2.529
EllipseFit	75.62	53.06	10.06	253.31	83.19	68.02	22.81	69.82	2.378
MinBoundingRect	74.16	54.25	9.00	283.00	79.86	65.70	23.16	68.80	2.449
ImageMoments	75.62	53.06	10.06	253.31	83.19	68.02	22.81	69.82	2.378
Geodesic	74.60	59.28	2.00	295.32	76.25	76.23	33.88	79.59	9.028
DT_MaxChord	78.03	55.75	10.00	283.50	83.03	67.03	24.05	69.59	2.317

TABLE III: RECOMMENDATION FOR EACH OF THE MEASUREMENT ALGORITHM

Approach	Recommendation
BoundingBox	This method can be used for initial screening only as it overestimates the measurement
MaxFeret	This is the recommended measurement and also mostly used in medical measurements
PCA	This could be the choice, if the input image has noise as, this method is robust to noise
EllipseFit	This could be best suited when aneurysms are round or oval perfectly
MinBoundingRect	This can be used when the aneurysms are tilted
ImageMoments	Cross-validation for ellipse fit
Geodesic	After MaxFeret this could be considered for irregular blobs
DT_MaxChord	This could be the third approach which could be considered .

B. Boundary Precision Metrics

From the experiments conducted, we have observed that even though dice score is good, the boundary metrics are low for most of the images in the dataset. The table 3 shows the percentage of images under each custom grade defined above.

TABLE III: GRADING OF THE 143 IMAGES ACROSS DIFFERENT BOUNDARY METRIC AND OVERALL SCORE

Grade	Dice	IoU	HD95	ASD	BIoU w2	Overall
A	17 (8%)	16 (7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
B	116 (53%)	116 (53%)	1 (0%)	2 (1%)	1 (0%)	12 (6%)
C	10 (5%)	11 (5%)	61 (28%)	61 (28%)	23 (11%)	120 (55%)
D	0 (0%)	0 (0%)	126 (58%)	75 (34%)	108 (50%)	11 (5%)
F	75 (34%)	75 (34%)	30 (14%)	80 (37%)	86 (39%)	75 (34%)

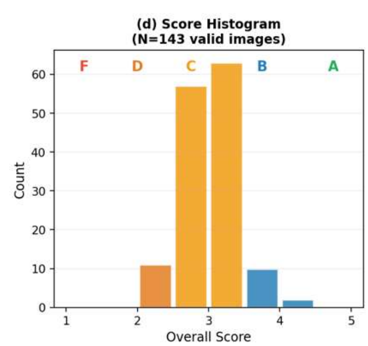


Fig 20: Histogram of the overall score

V. Discussion

A. Dice vs. Boundary Metrics: Systematic Divergence

In this experiment done, it is observed that many images which are rated as moderate to good when we use Dice score, the same images fall in the category of poor to failing when we use the boundary metrics like HD95 and ASD. The reason for this diversion in the measurement is due to the approach used by the metric for measurement. Dice uses a pixel overlap approach. In dice computation, for instance if there consider a 9000 pixel aneurysm, if there is a 15 pixel shift in the boundary, then the overall dice might show a very less deviation. But the same 15 pixel shift, when measured using HD95, all the 15 pixels are an increase in the deviation. Which is large for HD95. This shows that, if we are interested in segmentation, which needs accurate boundary too, then we need to use metric like HD95 and ASD.

B. Diameter Approach Selection

Base on the table 3, Feret diameter is recommended as the first preference for measuring diameter. It is also the preferred measurement by radiologists. Geodesic skeleton diameter is also very suitable as aneurysms are of irregular shape. PCA should be considered when the input image has noise since PCA approach is robust to noisy conditions. DT Max Chord could be more suitable for fusiform kind of aneurysm than the saccular aneurysm. Ellipse fit, moments are secondary measurements which could be considered. Bounding box and minbounding box can be used when one needs a rough estimation of the diameter rather than accurate measurement.

C. Boundary-Aware Training

Based on the experiment results, it is evident that for medical related segmentation, boundary based metrics have to be considered. It is very essential, in cases where shape of the lesion is very important and an incorrect measurement of geometrical features could lead to incorrect medical analysis. In future, trainings, it is recommended to consider the below for a better segmentation. (a) boundary loss [21], which is essential for minimizing the error in boundary mismatch, (b) clDice [22] for preserving skeleton, (c) Hausdorff Distance Loss, for measuring HD and (d) active contour regularization for smoother boundaries. Augmenting the above metrics with dice score during training would improve the HD95 and ASD thereby improving the segmentation with good boundary match with respect to ground truth.

D. Limitations

The major limitation include (1) single-institution dataset; (2) 2D MIP analysis vs full 3D; (3) single annotator for GT; (4) pixel spacing assumed uniform; (5) DT-MaxChord sensitivity for eccentric shapes; (6) grading thresholds are empirically motivated and may require domain-specific calibration.

VI. Conclusion

We presented a comprehensive evaluation framework for intracranial aneurysm segmentation integrating eight diameter measurement approaches, three boundary precision metrics, and a unified letter-grade quality scoring system. Our key findings are: (i) Dice scores systematically overestimate boundary quality for all subjects and angles studied; (ii) the novel DT-MaxChord approach provides an MIC-anchored diameter estimate complementary to the clinical Feret standard; (iii) HD95 and ASD expose systematic over-segmentation at the aneurysm periphery that is invisible to volume-overlap metrics. These results motivate boundary-aware training and multi-metric evaluation as standard practice in aneurysm segmentation research.

References

[1] Rinkel GJ, et al. Prevalence and risk of rupture of intracranial aneurysms. Stroke. 1998;29(1):251-256.
[2] Molyneux AJ, et al. International Subarachnoid Aneurysm Trial (ISAT). Lancet. 2002;360(9342):1267-1274.
[3] Firouzian A, et al. Intracranial aneurysm segmentation in 3D CT angiography. Med. Phys. 2011;38(5):2216-2225.
[4] Suniaga S, et al. Computer-aided detection of intracranial aneurysms in MRA. MICCAI. 2012.
[5] Sichtermann T, et al. Deep learning-based detection of intracranial aneurysms. Clin Neuroradiol. 2019;29(3):591-599.
[6] Yang X, et al. Automated detection of intracranial aneurysms using an ensemble of deep learning models. Neuroradiology. 2021;63(12):1939-1946.
[7] Cicek O, et al. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. MICCAI. 2016:424-432.
[8] Isensee F, et al. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18(2):203-211.

[9] Schlemper J, et al. Attention gated networks: Learning to leverage salient regions in medical images. Med Image Anal. 2019;53:197-207.
[10] Timmins KM, et al. Comparing methods of detecting and segmenting unruptured intracranial aneurysms on TOF-MRAs. NeuroImage Clin. 2021;24:101995.
[11] Wiebers DO, et al. Unruptured intracranial aneurysms: Natural history, clinical outcome, and risks of surgical and endovascular treatment. Lancet. 2003;362(9378):103-110.
[12] Zuluaga MA, et al. Learning from only positive and unlabelled data to detect lesions in vascular CT images. MICCAI. 2011.
[13] Van der Walt S, et al. scikit-image: Image processing in Python. PeerJ. 2014;2:e453.
[14] Dryden IL, Mardia KV. Statistical Shape Analysis. Wiley; 1998.
[15] Toussaint G. Solving geometric problems with the rotating calipers. IEEE MELECON. 1983.
[16] Lee TC, Kashyap RL, Chu CN. Building skeleton models via 3-D medial surface axis thinning algorithms. CVGIP. 1994;56(6):462-478.
[17] Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. IEEE TPAMI. 1993;15(9):850-863.
[18] Styner M, et al. Framework for the statistical shape analysis of brain structures using SPHARM-PDM. Insight J. 2006;1071.
[19] Cheng B, Girshick R, Dollar P, et al. Boundary IoU: Improving object-centric image segmentation evaluation. CVPR. 2021.
[20] Nikolov S, et al. Clinically applicable segmentation of head and neck anatomy for radiotherapy. Med Image Anal. 2021;67:101819.
[21] Kervadec H, et al. Boundary loss for highly unbalanced segmentation. Med Image Anal. 2021;67:101851.
[22] Shit S, et al. clDice - A novel topology-preserving loss function for tubular structure segmentation. CVPR. 2021.
[23] [Additional references as required]