



OPEN ACCESS

EDITED BY

Naizhuo Zhao,
McGill University Health Centre, Canada

REVIEWED BY

Chen Li,
Northeastern University, China
Xing Wu,
Chongqing University, China

*CORRESPONDENCE

Shaoguo Cui
✉ csg@cqu.edu.cn
Yongmei Li
✉ lymzhang70@163.com

†These authors have contributed equally to this work

RECEIVED 18 November 2024

ACCEPTED 11 April 2025

PUBLISHED 28 April 2025

CITATION

Han Y, Wang H, Luo Q, Wang J, Zeng C, Zheng Q, Dai L, Wei Y, Zhu Q, Lin W, Cui S and Li Y (2025) Automatic segmentation and volume measurement of anterior visual pathway in brain 3D-T1WI using deep learning.

Front. Med. 12:1530361.

doi: 10.3389/fmed.2025.1530361

COPYRIGHT

© 2025 Han, Wang, Luo, Wang, Zeng, Zheng, Dai, Wei, Zhu, Lin, Cui and Li. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Automatic segmentation and volume measurement of anterior visual pathway in brain 3D-T1WI using deep learning

Yongliang Han^{1†}, Haixiang Wang^{2†}, Qi Luo¹, Jingjie Wang¹, Chun Zeng¹, Qiao Zheng¹, Linqun Dai¹, Yiqiu Wei¹, Qiyuan Zhu¹, Wenlong Lin³, Shaoguo Cui^{3*} and Yongmei Li^{1*}

¹Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China, ²College of Physics and Electronic Engineering, Chongqing Normal University, Chongqing, China, ³College of Computer and Information Science, Chongqing Normal University, Chongqing, China

Objective: Accurate anterior visual pathway (AVP) segmentation is vital for clinical applications, but manual delineation is time-consuming and resource-intensive. We aim to explore the feasibility of automatic AVP segmentation and volume measurement in brain T1-weighted imaging (T1WI) using the 3D UX-Net deep-learning model.

Methods: Clinical data and brain 3D T1WI from 119 adults were retrospectively collected. Two radiologists annotated the AVP course in each participant's images. The dataset was randomly divided into training ($n = 89$), validation ($n = 15$), and test sets ($n = 15$). A 3D UX-Net segmentation model was trained on the training data, with hyperparameters optimized using the validation set. Model accuracy was evaluated on the test set using Dice similarity coefficient (DSC), 95% Hausdorff distance (HD95), and average symmetric surface distance (ASSD). The 3D UX-Net's performance was compared against 3D U-Net, Swin UNet Transformers (UNETR), UNETR++, and Swin Soft Mixture Transformer (Swin SMT). The AVP volume in the test set was calculated using the model's effective voxel volume, with volume difference (VD) assessing measurement accuracy. The average AVP volume across all subjects was derived from 3D UX-Net's automatic segmentation.

Results: The 3D UX-Net achieved the highest DSC (0.893 ± 0.017), followed by Swin SMT (0.888 ± 0.018), 3D U-Net (0.875 ± 0.019), Swin UNETR (0.870 ± 0.017), and UNETR++ (0.861 ± 0.020). For surface distance metrics, 3D UX-Net demonstrated the lowest median ASSD (0.234 mm [0.188–0.273]). The VD of Swin SMT was significantly lower than that of 3D U-Net ($p = 0.008$), while no statistically significant differences were observed among other groups. All models exhibited identical HD95 (1 mm [1–1]). Automatic segmentation across all subjects yielded a mean AVP volume of $1446.78 \pm 245.62 \text{ mm}^3$, closely matching manual segmentations (VD = 0.068 ± 0.064). Significant sex-based volume differences were identified ($p < 0.001$), but no age correlation was observed.

Conclusion: We provide normative values for the automatic MRI measurement of the AVP in adults. The 3D UX-Net model based on brain T1WI achieves high accuracy in segmenting and measuring the volume of the AVP.

KEYWORDS

magnetic resonance imaging, optic nerve, deep learning, convolutional neural network, medical image processing

1 Introduction

The anterior visual pathway (AVP), comprising the optic nerve, optic chiasm, and optic tract, transmits visual stimuli from the retina to the lateral geniculate nuclei (1). Various benign and malignant diseases can affect the AVP, causing acute swelling, chronic axonal loss, and atrophy (2). Accurate segmentation is crucial for clinical applications such as disease diagnosis, treatment planning, and monitoring disease progression. Successful image-guided surgery requires precise delineation of key neuro-logical and vascular structures during preoperative planning (3, 4). Modeling structural changes in the AVP throughout disease progression is important for characterizing neuropathic diseases (5). However, accurate segmentation and measurement of AVP are challenging due to their long pathway, the complexity of neighboring anatomical structures, and the high precision required for clinical use (6).

Manual delineation of AVP structures is time-consuming, resource-intensive, and prone to subjectivity. Automated quantification of AVP location and volumetrics would enable larger, more robust studies and potentially improve assessment accuracy compared to manual methods. Although significant progress has been made in AVP segmentation, traditional atlas-based methods vary in effectiveness for smaller structures like the optic nerve and chiasm, with Dice similarity coefficients (DSC) ranging from 0.39 to 0.78 (7, 8). Recently, convolutional neural networks (CNNs) have achieved significant advancements in MRI-based visual pathway segmentation. Mansoor et al. (6) proposed a deep learning approach that combines prior shape models with morphological features for AVP segmentation. Zhao et al. (9) improved segmentation performance by integrating a 3D fully convolutional network with a spatial probabilistic distribution map, effectively addressing issues of low contrast and blurred boundaries. Xie et al. (10) developed cranial nerves tract segmentation (CNTSeg), a multimodal deep learning network that accurately segments five major cranial nerve tracts—including CN II (optic nerve), CN III (oculomotor nerve), CN V (trigeminal nerve), and CN VII/VIII (facial and vestibulocochlear nerves)—by fusing T1-weighted images, fractional anisotropy (FA) images, and fiber orientation distribution function (fODF) peaks. However, due to the complexity near the optic chiasm and artifacts caused by magnetic susceptibility, AVP segmentation results remain unsatisfactory. Pravata et al. (11) manually segmented high-resolution MRI data from 24 healthy subjects to obtain biometric indicators such as volume, length, cross-sectional area, and ellipticity of the AVP, establishing a foundation for standardized analysis. Nevertheless, manual segmentation is time-consuming

and affects the reproducibility of results. The small sample size lacks age- and gender-specific standard data, and the ellipticity calculations may be overestimated. Further research is needed to refine automated segmentation techniques and develop more robust, standardized analysis methods for the AVP.

To enable broad application of the AVP segmentation model in clinical practice (non-GPU cluster environments), we selected the lightweight 3D UX-Net architecture (12) for automatic segmentation of the AVP in brain 3D- T1-weighted imaging (T1WI). As an optimized Swin Transformer variant, this volumetric convolutional neural network achieves computational efficiency through three key innovations: (a) replacement of multilayer perceptron (MLP) blocks with pointwise convolutions to reduce parameter count; (b) strategic minimization of normalization and activation layers to streamline processing; (c) implementation of large-kernel volumetric depthwise convolutions that expand the global receptive field while maintaining memory efficiency. These architectural modifications collectively enable robust 3D medical image segmentation with 34% fewer trainable parameters than standard Swin Transformer implementations (12, 13). To evaluate the segmentation performance of the 3D UX-Net model, we compared it with the 3D U-Net, Swin UNet TRansformers (UNETR), UNETR++, and Swin Soft Mixture Transformer (Swin SMT) models (14–17). This comparative framework allows comprehensive evaluation of 3D UX-Net's performance across model efficiency, anatomical detail preservation, and clinical applicability metrics. Our approach outperforms state-of-the-art methods in accuracy and robustness. Additionally, we report normative values for AVP volume in adult MR imaging, providing objective measurements for radiologists and ophthalmologists.

2 Materials and methods

2.1 Study population

This single-center study was approved by the institutional review board, with a waiver of informed consent due to its retrospective nature. Conducted in accordance with the latest version of the Declaration of Helsinki, the study evaluated patients with non-specific neurological symptoms reported between January and December 2022 in our institution's picture archiving and communication system (PACS). Inclusion criteria were patients aged 18 to 60 who underwent brain 3D-T1WI during this period. Exclusion criteria included patients with intracranial

or orbital tumors, craniocerebral or ocular trauma, deformity or infection, spherical refractive errors worse than $-6D$ (18), or poor image quality.

The final cohort comprised 119 participants (61 females, 58 males; mean age 45.52 ± 12.18 years). Participants were randomly divided into training (74.8%, $n = 89$), validation (12.6%, $n = 15$), and test (12.6%, $n = 15$) cohorts.

2.2 MR examination

Images were acquired using a single 3.0T Siemens Skyra MRI scanner (Erlangen, Germany) with a 32-channel head coil. Patients were instructed to close their eyes and avoid eye movement during scanning. The 3D-T1WI sequence parameters were: TR = 2,000 ms, TE = 2.48 ms, flip angle = 8° , layer thickness = 1 mm, no gap, voxel size = $0.898 \times 0.898 \times 1 \text{ mm}^3$, field of view = $20 \text{ cm} \times 20 \text{ cm}$, matrix size = 256×224 , with a total of 192 layers.

2.3 Data annotations

We converted the Digital Imaging and Communications in Medicine (DICOM) format T1-weighted images of the skull cross-section to the Neuroimaging Informatics Technology Initiative (NIfTI) format. To establish a reproducible, MRI-compatible landmark, we defined the specific range of the AVP based on prior anatomical knowledge (19, 20). This range extends from the optic nerve head to the lateral geniculate body—excluding the portion surrounded by cerebrospinal fluid—and encompasses the entire area of the bilateral optic nerves, optic chiasm, and bilateral optic tracts. Using ITK-SNAP 3.9.0 software,¹ two neuroradiology trainees (Y.H., Q.L.) labeled each AVP layer by layer on the 3D-T1WI according to the defined anatomical landmarks, under the constant supervision of an experienced neuroradiologist (C.Z.). Accuracy was enhanced through iterative segmentation revisions until consensus with the senior neuroradiologist was reached. Figure 1 illustrates the overall scheme of the proposed method.

2.4 Segmentation model

We selected the 3D UX-Net as the segmentation network model for AVP because long-distance dependencies are crucial for accurately determining its absolute and relative positions. By analyzing the entire image's structure and morphology, the network can swiftly identify and locate the AVP. Mainstream medical segmentation models like 3D U-Net have limited receptive fields and struggle to efficiently capture long-distance dependencies. In contrast, the 3D UX-Net, a lightweight volumetric convolutional network, expands the effective receptive field by using depth-wise convolutions with large kernel (LK) sizes ($7 \times 7 \times 7$). This approach enhances the network's ability to capture long-distance spatial dependencies, which is essential for accurately segmenting structures rich in positional information that connect distant areas,

such as nerve fibers. Additionally, the model's performance is improved with fewer normalization and activation layers, reducing the number of parameters. Figure 2 provides an overview of 3D UX-Net, highlighting convolutional blocks as the encoder's backbone. At the same time, the structures of the 3D U-Net, Swin UNETR, UNETR++, and Swin SMT models were constructed with reference to previous studies (14–17).

2.5 Training of the segmentation model

We utilized an NVIDIA GeForce RTX 3090Ti 24GB GPU, with software including Python 3.10, PyTorch 1.12.0, MONAI 1.0.0, OpenCV, NumPy, and SimpleITK. The AdamW optimizer and DiceCELoss function were employed for training. Image preprocessing parameters were set to a size of $16 \times 448 \times 256$ (z, y, x), with automatic window width and level adjustments. A total of 119 datasets were randomly divided into training ($n = 89$), validation ($n = 15$), and test ($n = 15$) sets in an 8:1:1 ratio. The T1WI and the labeled files were input to segment the entire AVP. Training parameters included a batch size of 1, a learning rate of 0.0001, and 60 epochs, with validation performed after each training epoch.

2.6 Model evaluation

Evaluate the segmentation effectiveness using manually annotated labels as the benchmark for assessing the 3D UX-Net model's performance on AVP. Evaluation metrics include the DSC, 95% Hausdorff distance (HD95), average symmetric surface distance (ASSD), and volume difference (VD) (21–25) (Equations 1–4). The calculation formulas are as follows:

$$DSC(V_{gt}, V_{seg}) = \frac{2|V_{gt} \cap V_{seg}|}{|V_{gt}| + |V_{seg}|}, \quad (1)$$

Where V_{gt} represents the ground truth and V_{seg} represents the automatic segmentation result.

$$HD95(A, B) =$$

$$percent_{95}\{\max\{\sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(b, a)\}\}, \quad (2)$$

Where the $percent_{95}$ refers to the 95th percentile, A and B are point sets where A is the predicted segmentation and B is the ground truth segmentation. $d(a, b)$ measures the distance between point a and point b .

$$ASSD(A, B) = \frac{1}{|A| + |B|} \left(\sum_{a \in A} \min_{b \in B} d(a, b) + \sum_{b \in B} \min_{a \in A} d(b, a) \right), \quad (3)$$

Where A and B are point sets representing the predicted segmentation and the true segmentation, respectively, and $d(a, b)$ denotes the distance from point a to point b .

$$VD = \frac{|TP_{seg} - TP_{gt}|}{TP_{gt}}, \quad (4)$$

Where TP_{seg} is the true positive of the predicted segmentation and TP_{gt} is the true positive of the true segmentation.

¹ <http://www.itksnap.org/pmwiki/pmwiki.php>

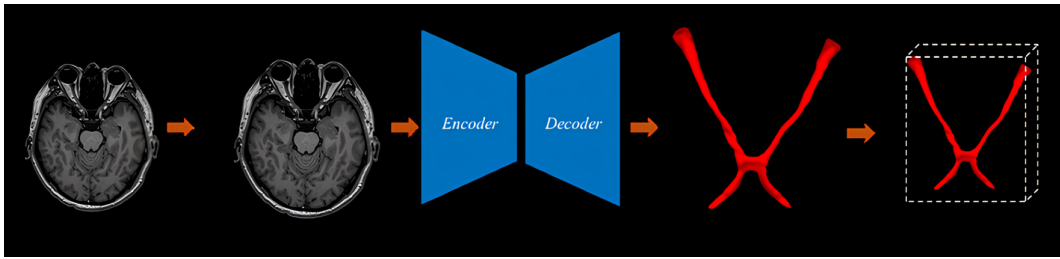


FIGURE 1
The proposed method encompasses data collection, preprocessing and labeling, deep learning predictions, result presentation, and volume calculation.

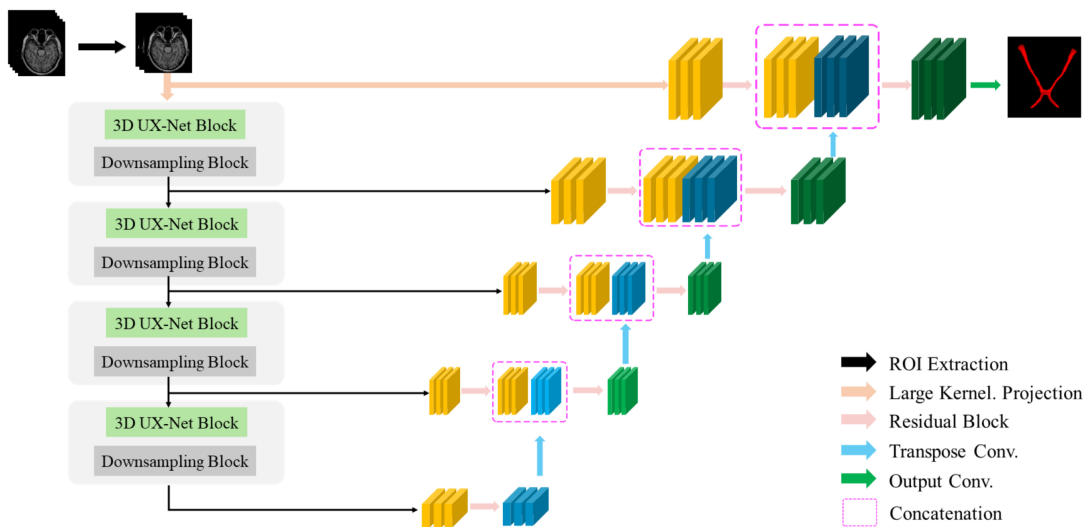


FIGURE 2
3D UX-Net employs a convolutional encoder backbone where large-kernel convolutions generate patch embeddings, with each stage's downsampling block expanding channel-wise context.

TABLE 1 Segmentation performance comparison of 3D UX-Net, 3D U-Net, Swin UNETR, UNETR++, and Swin SMT on the test set.

Model architecture	DSC (Mean $\pm$ SD)	HD95 (mm) (Median [IQR])	ASSD (mm) (Median [IQR])	VD (Median [IQR])
3D U-Net	0.875 $\pm$ 0.019	1 [1-1]	0.404 [0.265–0.463]	0.045 [0.010–0.072]
Swin UNETR	0.870 $\pm$ 0.017	1 [1-1]	0.307 [0.244–0.470]	0.066 [0.019–0.101]
UNETR++	0.861 $\pm$ 0.020	1 [1-1]	0.338 [0.260–0.422]	0.072 [0.028–0.114]
Swin SMT	0.888 $\pm$ 0.018	1 [1-1]	0.248 [0.210–0.312]	0.047 [0.003–0.126]
3D UX-Net	0.893 $\pm$ 0.017	1 [1-1]	0.234 [0.188–0.273]	0.054 [0.034–0.076]

Data are presented as mean $\pm$ SD or median [IQR]. SD, standard deviation; IQR, interquartile range; DSC, Dice similarity coefficient; HD95, Hausdorff distance 95%; ASSD, average symmetric surface distance; VD, volume difference. Bold values indicate the best performance in segmentation metrics (DSC, ASSD, VD).

Data augmentation through geometric transformations (flipping) and photometric adjustments (brightness/contrast/hue variation) enhanced training diversity. We systematically evaluated test-set AVP segmentation accuracy across five architectures: 3D UX-Net, 3D U-Net, Swin UNETR, UNETR++, and Swin SMT. Computational complexity was quantitatively assessed via parameter counts, FLOPs, and inference time for all models. Furthermore, the AVP segmentation performance of the 3D UX-Net model was benchmarked against state-of-the-art

segmentation outcomes documented in previous studies (9, 10, 26–28).

2.7 Volume measurement

We calculated the overall voxel volume (number of voxels $\times$ voxel size) from both the manually outlined and 3D UX-Net automatically segmented labels as the AVP volume, and then derived the average AVP volume for all participants based

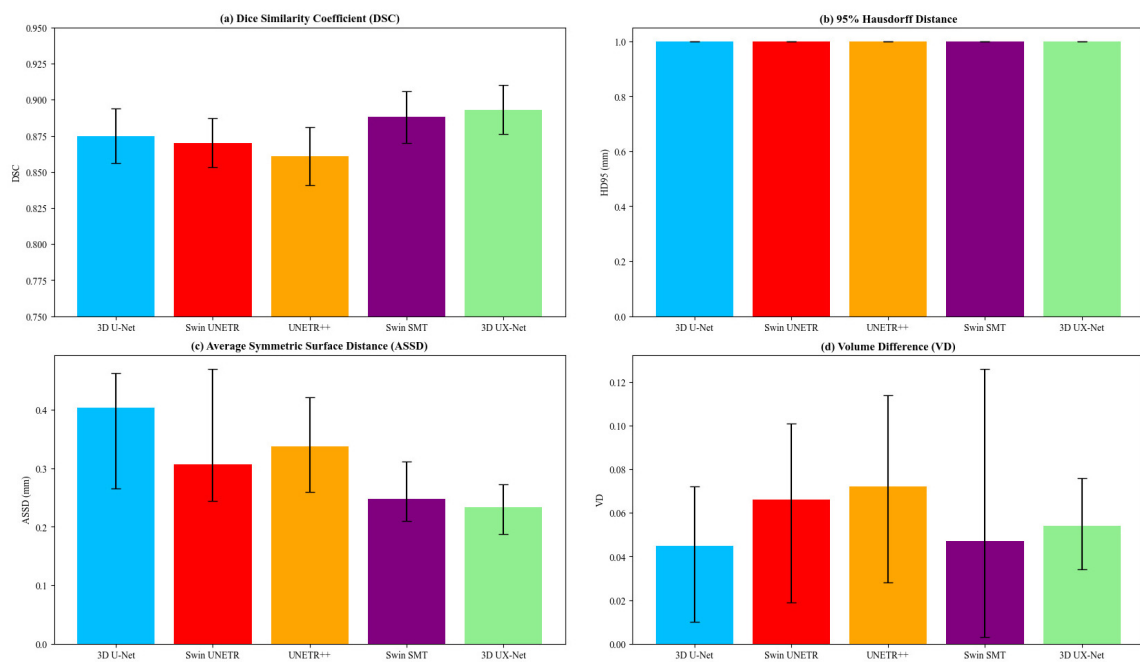


FIGURE 3 Comparative analysis of segmentation performance across 3D UX-Net, 3D U-Net, Swin UNETR, UNETR++, and Swin SMT on the test set, with quantitative evaluations including (a) Dice similarity coefficient, (b) 95% Hausdorff distance, (c) average symmetric surface distance, and (d) volume difference.

TABLE 2 Cross-model comparison of computational complexity.

Model architecture	Parameters (M)	FLOPs (G)	Inference time (s)
3D U-Net	85.43	2139.34	0.36 ± 0.39
Swin UNETR	92.96	1485.11	0.35 ± 0.39
UNETR++	40.48	912.23	0.36 ± 0.31
Swin SMT	73.05	886.03	0.37 ± 0.71
3D UX-Net	52.39	1325.73	0.33 ± 0.58

Results were presented as mean ± SD. SD, standard deviation; FLOPs, floating point operations per second. Bold values indicate the best performance in lightweight metrics (Parameters, FLOPs, Inference Time).

on the results of the 3D UX-Net automatic segmentation. Model comparison employed VD analysis, calculating absolute differences between automated and manual measurements across the test set.

2.8 Statistical analysis

Statistical analyses were performed using SPSS (v26.0) and Python (v3.10). A *P*-value less than 0.05 was considered statistically significant. Normal distribution of all continuous variables was examined using the *Shapiro-Wilk test*. Normally distributed variables were presented as mean ± standard deviation (SD), skewed distribution variables were described using median and interquartile range (IQR). Use *Levene’s test* to check if different groups have the same variance. If the data meets the criteria for normal distribution and homogeneity of variance, conduct a *one-way ANOVA*. If not, use the *Kruskal-Wallis test*. Post hoc tests were

conducted using the *Dunn test*, with *Bonferroni correction* applied for multiple comparisons. The two-sided paired *Wilcoxon signed-rank test* was employed to compare AVP volumes between 3D UX-Net and manual segmentation predictions in the test cohort. The mean and standard deviation of AVP volumes were calculated to establish the 95% reference range for this population. When grouped by gender, statistical analysis was conducted using a *two-sample t-test*. For age groups differing by 10 years, the 95% reference intervals of AVP volume were calculated for each group, and trends were illustrated using a line chart.

3 Results

3.1 Segmentation results

The proposed 3D UX-Net achieved state-of-the-art DSC of 0.893 ± 0.017 , outperforming the 3D U-Net (0.875 ± 0.019), Swin UNETR (0.870 ± 0.017), UNETR++ (0.861 ± 0.020), and Swin SMT (0.888 ± 0.017). Morphological analysis further confirmed its advantages, demonstrating the lowest ASSD of 0.234 mm [0.188–0.273], which significantly outperformed 3D U-Net ($p = 0.003$). All models exhibited identical median HD95 values of 1 mm (IQR = 1–1), indicating consistent control of extreme boundary errors across architectures. For surface distance accuracy, the 3D UX-Net achieved the lowest (0.234 mm [0.188–0.273]), significantly outperforming the 3D U-Net (0.404 mm [0.265–0.463], $p = 0.003$) and UNETR++ (0.338 mm [0.260–0.422], $p = 0.021$).

In terms of volume measurement, the Swin SMT exhibited the lowest median VD (0.047 [0.003–0.126]); however, its wider IQR suggested potential instability. No significant VD differences were

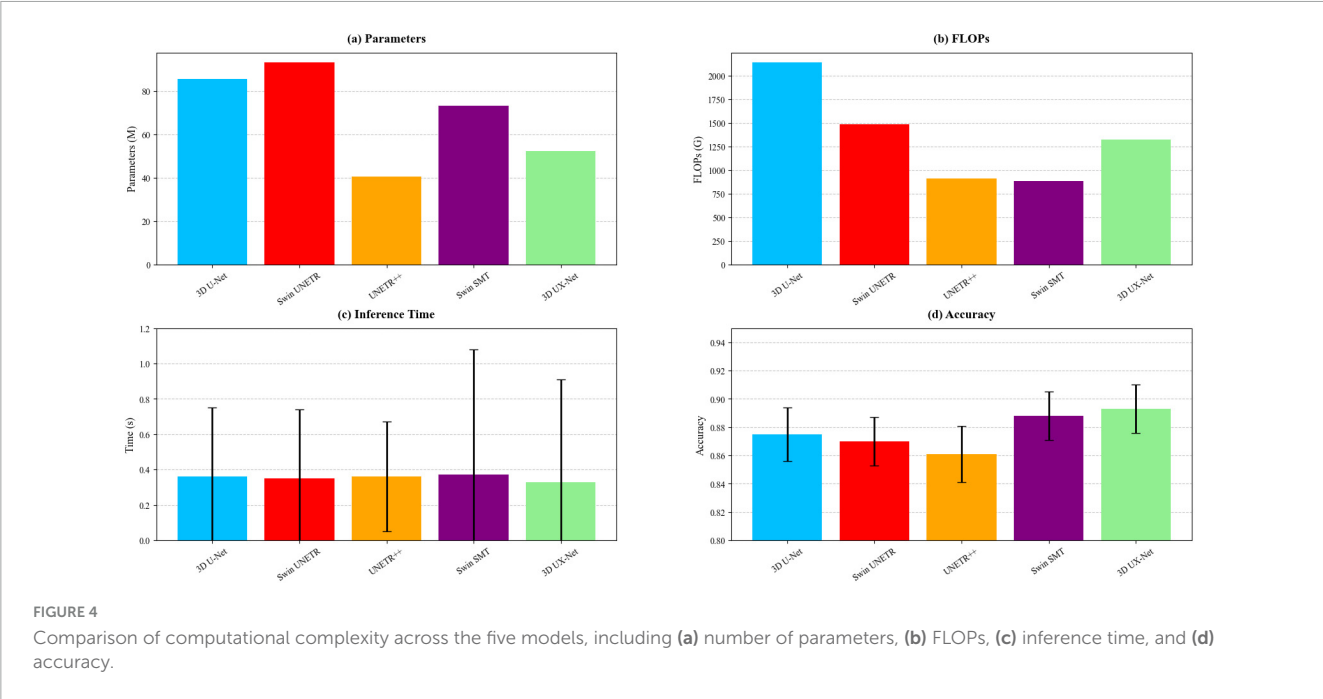


TABLE 3 The 3D UX-Net model performance for AVP segmentation on the test subjects.

Subjects	DSC	HD95 (mm)	ASSD (mm)	VD	Volume (mm ³)
HC001	0.894	1.0	0.237	0.054	1414.44
HC002	0.886	1.0	0.273	0.076	1504.94
HC003	0.915	1.0	0.159	0.034	1417.05
HC004	0.917	1.0	0.176	0.007	1777.66
HC005	0.867	1.41	0.428	0.093	1566.76
HC006	0.906	1.0	0.183	0.037	1351.95
HC007	0.896	1.0	0.263	0.061	1877.91
HC084	0.899	1.0	0.204	0.020	1723.63
HC085	0.897	1.0	0.219	0.035	1505.58
HC086	0.880	1.0	0.229	0.054	1405.34
HC087	0.895	1.0	0.259	0.067	1594.11
HC088	0.897	1.0	0.235	0.045	1428.13
HC089	0.913	1.0	0.188	0.007	1438.54
HC132	0.860	1.0	0.346	0.241	1419.66
HC133	0.872	1.0	0.548	0.182	1292.73
Mean/median	0.893 ± 0.017	1 [1-1]	0.234 [0.188–0.273]	0.054 [0.034–0.076]	1514.56 ± 165.52

Within-group statistical data are shown in bold and presented as mean ± SD or median [IQR]. SD, standard deviation; IQR, interquartile range; DSC, Dice similarity coefficient; HD95, Hausdorff distance 95%; ASSD, average symmetric surface distance; VD, volume difference; HC, healthy control.

observed between the 3D U-Net (0.045 [0.010–0.072]) and 3D UX-Net (0.054 [0.034–0.076], $p = 0.27$) (Table 1; Figure 3). The 3D UX-Net balances accuracy and efficiency with 52.39M parameters, a 38.6% reduction from 3D U-Net (85.43M), while outperforming Transformer-based Swin UNETR (92.96M) and Swin SMT (73.05M). Although UNETR++ has the fewest parameters (40.48M), its accuracy remains subpar (Table 1). In computational efficiency, 3D UX-Net (1325.73G FLOPs) reduces computational load by 38% versus 3D U-Net and is 10.7% more

efficient than Swin UNETR. Swin SMT achieves the lowest FLOPs (886.03G) but shows higher accuracy variability. All models show comparable inference times, though 3D UX-Net and Swin SMT exhibit greater latency fluctuations than UNETR++. Table 2 and Figure 4 delineates the comparative analysis of computational complexity parameters across the five architectures. The 3D UX-Net model exhibited minimal volumetric discrepancy in AVP predictions (1514.56 ± 165.52 mm³) compared to ground truth annotations, with detailed per-subject

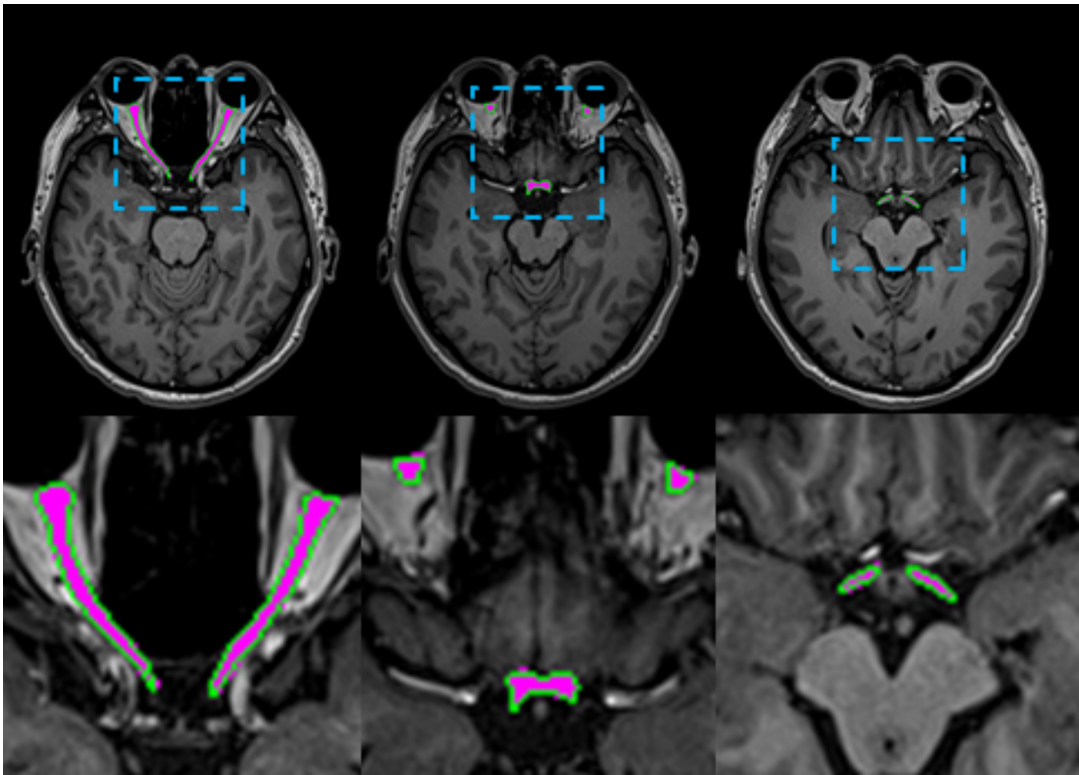


FIGURE 5
Visual comparisons of AVP segmentation examples are shown. The predicted AVP contours are overlaid in green on a representative axial T1-weighted MRI of the brain from one test subject, with manual annotations overlaid in magenta for reference. The inset provides a close-up comparison of the predicted values.

TABLE 4 Comparison of our method with the AVP segmentation methods reported in the literature.

Methods	Year	No. of subjects	MRI sequence	Voxel size (mm ³)	DSC	HD95 (mm)	ASSD (mm)
3D FCN + SPDM. (Zhao et al.)	2019	93	T1w	1 × 1 × 1	0.85 ± 0.01	4.64 ± 1.41	0.37 ± 0.04
3D U-Net (Ai et al.)	2020	93	T1w	1 × 1 × 1	0.86 ± 0.01	3.56 ± 1.89	0.34 ± 0.05
TPSN (Li et al.)	2021	102	T1w + FA	1.25 × 1.25 × 1.25	0.85 ± 0.02	2.33	0.16
3D U-Net (van Elst et al.)	2022	40	T2w	0.25 × 0.25 × 0.7/ 0.27 × 0.27 × 0.3	0.84	0.64	0.14
CNTSeg (Xie et al.)	2023	102	T1w + FA + fODF peaks	1.25 × 1.25 × 1.25	0.82 ± 0.02	–	–
3D UX-Net (ours)	–	119	T1w	0.90 × 0.90 × 1	0.893 ± 0.017	1 [1–1]	0.234 [0.188–0.273]

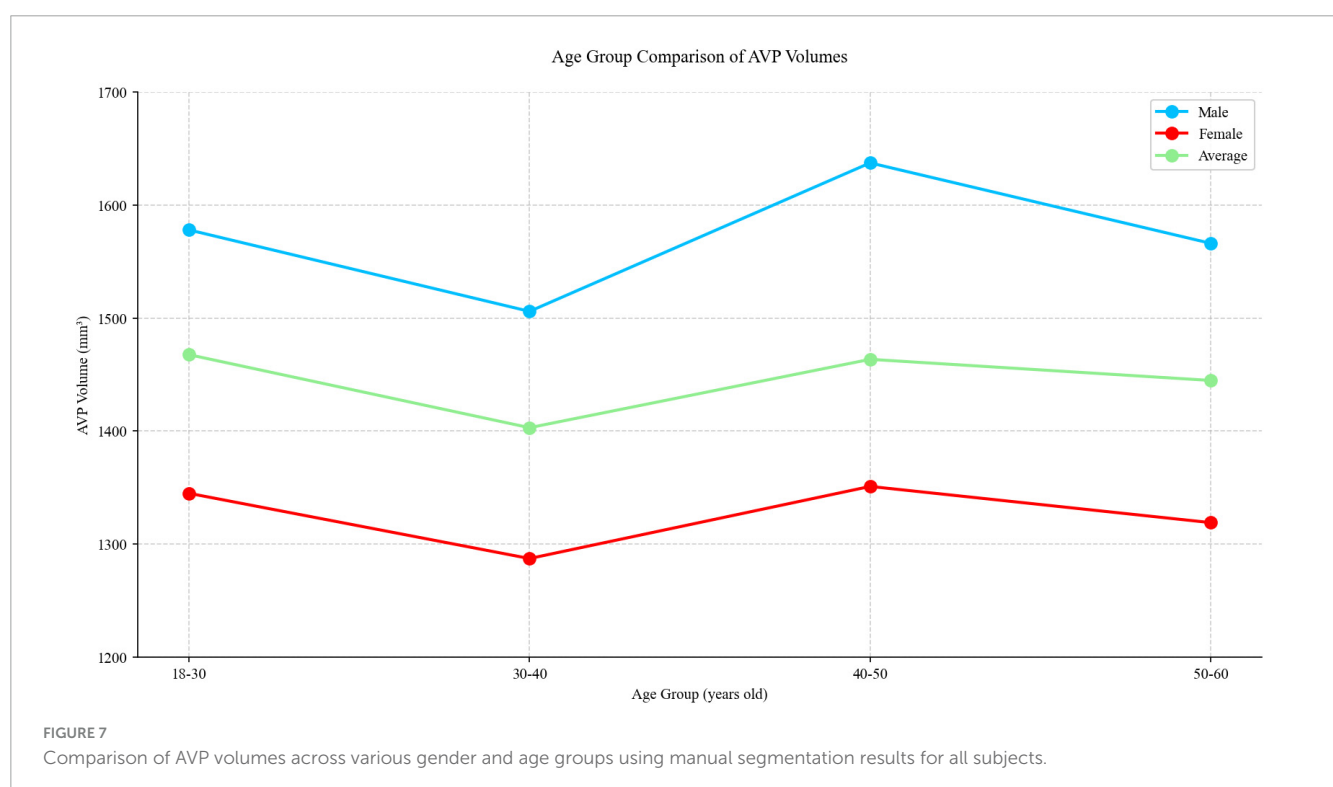
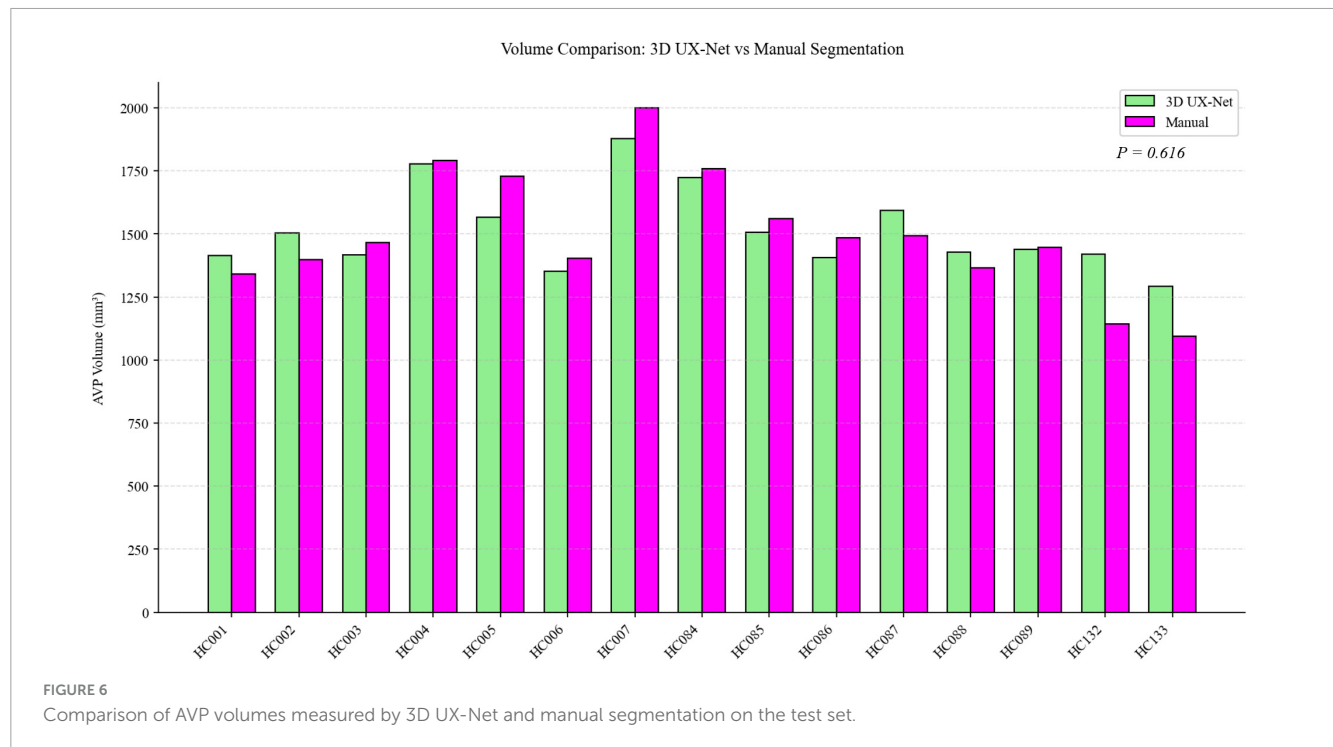
Data are presented as mean ± SD or median [IQR]. SD, standard deviation; IQR, interquartile range; DSC, Dice similarity coefficient; HD95, Hausdorff distance 95%; ASSD, average symmetric surface distance; FCN, fully convolutional network; SPDM, spatial probabilistic distribution map; CNTSeg, cranial nerves tract segmentation; FA, fractional anisotropy; fODF, fiber orientation distribution function; TPSN, two parallel stages network. Bold values indicate the best performance in AVP segmentation.

performance metrics documented in Table 3. As illustrated in Figure 5, automated contours generated by 3D UX-Net showed high anatomical concordance with manual delineations, achieving submillimeter boundary accuracy (0.23 mm) that meets precision requirements for image-guided surgical interventions.

Comparative analysis with prior studies addressing analogous AVP segmentation tasks revealed that the 3D UX-Net architecture achieves statistically significant improvements in segmentation accuracy, as empirically demonstrated in Table 4.

3.2 Volume measurement results

In the test set, the detailed results of applying the 3D UX-Net model to calculate individual AVP volumes are presented in Table 3. The average volume of the 3D UX-Net automatically segmented AVP across all subjects was 1446.78 ± 245.62 mm³, with an average VD of 0.068 ± 0.064 compared to manual labeling. A 2-tailed Wilcoxon signed-rank test showed no significant difference (*p* = 0.616) in AVP volumes predicted by 3D UX-Net and manual



segmentation on the test cohort. The volumes measured by 3D UX-Net and manual methods are visualized in Figure 6.

3.3 Volume comparison

We calculated the AVP volume for all subjects using the automatic segmentation results from 3D UX-Net and compared

the volumes between different sex and age groups. Male subjects exhibited significantly larger AVP volumes than females (Males: $1572.33 \pm 242.90 \text{ mm}^3$ vs. Females: $1327.41 \pm 181.29 \text{ mm}^3$; $p < 0.001$), despite comparable age distributions between groups ($p = 0.63$). No significant association emerged between age and AVP volume across the studied cohort. The complete statistical breakdown is visualized in Figure 7 and tabulated in Table 5.

TABLE 5 Volume distribution of the anterior visual pathway across different age groups.

Age (years old)	Gender	Number	Volume (mean $\pm$ SD mm ³)	95%IC (mm ³)	P-value
18-30	Male	10	1578.03 $\pm$ 278.51	(1378.80, 1777.26)	0.035*
	Female	9	1344.80 $\pm$ 136.44	(1239.92, 1449.68)	
	Total	19	1467.55 $\pm$ 56.83	(1348.15, 1586.96)	
30-40	Male	9	1505.87 $\pm$ 297.14	(1277.47, 1734.27)	0.079
	Female	8	1287.03 $\pm$ 164.16	(1149.79, 1424.27)	
	Total	17	1402.89 $\pm$ 63.53	(1268.21, 1537.56)	
40-50	Male	11	1637.36 $\pm$ 186.11	(1512.33, 1762.39)	0.001*
	Female	17	1350.85 $\pm$ 189.35	(1248.87, 1452.83)	
	Total	28	1463.41 $\pm$ 44.90	(1371.28, 1555.54)	
50-60	Male	28	1566.12 $\pm$ 237.63	(1473.97, 1658.26)	0.000*
	Female	27	1318.82 $\pm$ 193.94	(1242.10, 1395.54)	
	Total	55	1444.71 $\pm$ 33.55	(1377.45, 1511.98)	

*The differences are statistically significant.

4 Discussion

In this study, we developed and validated an automated framework for AVP segmentation and volume measurement in brain 3D T1WI. The 3D UX-Net accurately, robustly, and quickly detects AVP contours and quantifies volumes, demonstrating a higher mean DSC than the 0.82 reported in Xie et al. (10). Compared to the 3D U-Net, Swin UNETR, UNETR++, and Swin SMT, the 3D UX-Net emerged as the most robust architecture, achieving state-of-the-art DSC and ASSD while maintaining competitive volumetric consistency. Despite comparable HD95 values across all models, 3D UX-Net's superior surface distance accuracy and stability suggest enhanced capability for fine-grained anatomical segmentation. These results highlight the importance of balancing spatial precision with volumetric reliability in medical image analysis. Our method's reliability was confirmed in an independent cohort and was compared to manual segmentation, which is labor-intensive and impractical in busy clinical settings. Our approach offers rapid predictions with accuracy comparable to manual results, providing timely AVP contour and volume data for radiologists. Additionally, we offer a reference range for adult AVP volume, useful for diagnosing AVP-related diseases. We also analyzed AVP volume variations across different age groups and genders using manual segmentation.

Variability in AVP measurements is likely influenced by factors such as MRI sequences, slice thickness, measurement locations, and measurement methods. While some researchers prefer T1-weighted images for visualizing the optic nerve (29, 30), others advocate for standard or novel T2-weighted sequences that may better distinguish the optic nerve from cerebrospinal fluid (31, 32). We chose to segment high-resolution T1WI due to their near isotropic voxel size, avoiding image resolution degradation during multiplanar reconstruction. Additionally, this sequence is included in our institution's standard radiological acquisition. We anticipate

no significant differences in AVP measurements using 3D-T1WI sequences across centers employing MRI systems from different manufacturers.

Previously, studies have segmented orbital structures in CT images using semi-automatic or automatic methods for diagnostic imaging and surgical planning. Pravata et al. (11) proposed a deep learning-guided partitioned shape model for anterior visual pathway segmentation, initially utilizing a marginal space deep-learning stacked autoencoder to locate the pathway, then combining a novel partitioned shape model with an appearance model to guide segmentation (6). Zhao et al. (9) introduced a method for visual pathway segmentation in MRI based on a 3D FCN combined with a Spatial Probabilistic Distribution Map (SPDM), which represents the probability that a voxel belongs to a specific tissue by summing all manual labels in the training dataset. Incorporating SPDM effectively overcomes issues of low contrast and blurry boundaries, achieving improved segmentation performance. Harrigan et al. (33) enhanced the multi-atlas segmentation method by introducing a technique for quantitatively measuring the optic nerve and cerebrospinal fluid sheath, demonstrating its potential to differentiate patients with optic nerve atrophy or hypertrophy from healthy individuals. Aghdasi et al. (34) defined a volume of interest (VOI) encompassing the desired structures using anatomical prior knowledge, employing it for rapid localization and effective segmentation of orbital structures—globes, optic nerves, and extraocular muscles—in CT images. This method is precise, efficient, requires no training data, and its intuitive pipeline enables adaptation to other structures. In our study, we used the 3D UX-Net network structure to train the segmentation model in two steps using a coarse-to-fine method. Compared to previous studies on similar AVP segmentation tasks, the 3D UX-Net architecture significantly improves segmentation accuracy. Additionally, we observed that all models had a median HD95 of 1 mm with an IQR of [1-1], showing only a small number of outliers around 1.4 mm. Possible explanations include: a. All subjects were free of optic nerve-related disorders, and anatomical variations in the AVP are smaller in healthy states compared

to pathological conditions; b. While most AVP boundaries are distinct due to high tissue contrast, partial volume effects from adjacent structures in rare cases may contribute to elevated HD95 values.

Comparative analysis of computational complexity among the 3D U-Net, Swin UNETR, UNETR++, Swin SMT, and 3D UX-Net reveals that 3D UX-Net achieves the optimal accuracy-efficiency balance in parameter efficiency and computational precision, providing a viable solution for deploying high-accuracy real-time segmentation systems in clinical settings. However, its FLOPs still lag significantly behind Swin SMT, revealing potential for further compression of computational redundancy.

Several studies have shown that age and ethnicity influence optic nerve size (35, 36). However, our findings revealed no clear correlation between age and optic nerve volume, possibly due to variations such as different examination techniques and post-processing programs. Inglese et al. (37) included controls with a mean age of 37 years old from an Italian population, whereas our study's controls had an older mean age of 46 years old from an Asian population. We believe these factors may also affect optic nerve volume, which requires verification through an expanded sample size.

Mncube and Goodier (38) showed that the transverse diameter of the optic chiasm is the most reliable measure of the AVP in adults because it is the largest structure and easiest to measure. However, due to the irregular morphology of the AVP, measurements of the optic chiasm may not capture all its features. Our comprehensive volumetric assessment addressed this limitation. In our test set, two subjects exhibited lower segmentation accuracy and significantly greater volume differences than others. This may be attributed to the tortuous course of the optic nerve within the orbital segment of these subjects, along with a longer segment parallel to the ophthalmic artery (39).

This study has several limitations. Firstly, it relies on single-center data, necessitating future validation in broader clinical settings with larger datasets. Secondly, our training cohort included only subjects with normal AVP images, excluding related diseases such as congenital malformations, inflammation, and tumors. We intend to apply this method for automated segmentation and volume measurement of AVP and for differential diagnosis of inflammatory demyelinating diseases. Lastly, we only used the 3D-T1WI sequence to measure the AVP's total volume. Future research could include T2-FLAIR and DTI sequences to improve accuracy and reliability. Additionally, refining measurement metrics such as segmented volume, length, and cross-sectional area could provide more comprehensive evaluation data.

5 Conclusion

In this study, we introduced the 3D UX-Net for segmenting AVP on brain 3D-T1WI, which outperformed the 3D U-Net, Swin UNETR, UNETR++, and Swin SMT. Our convolutional neural network, trained and validated on MRI scans from

119 healthy individuals, achieved an average DSC of 0.893 for AVP segmentation with a volume error under 1%. Using the AVP volume from the normal study group as a benchmark, we can quantitatively evaluate the volume difference between a given MR image and the baseline, automatically identifying abnormalities when differences exceed the normal distribution. This comparison provides radiologists with valuable information within a clinically practical timeframe, aiding in the detection of subtle AVP lesions. The approach's high accuracy and computational efficiency make it suitable for quantifying and distinguishing optic nerve-related diseases.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving humans were approved by the Ethics Committee of the First Affiliated Hospital of Chongqing Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because Patient consent was waived due to the retrospective nature of the data collection for this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

YH: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. HW: Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Writing – review & editing. QL: Data curation, Methodology, Resources, Writing – review & editing. JW: Data curation, Writing – review & editing. CZ: Data curation, Writing – review & editing. QZ: Data curation, Writing – review & editing. LD: Data curation, Writing – review & editing. YW: Data curation, Investigation, Writing – review & editing. QZ: Formal Analysis, Validation, Writing – review & editing. WL: Investigation, Resources, Software, Writing – review & editing. SC: Conceptualization, Formal Analysis, Methodology, Project administration, Resources, Supervision, Visualization, Writing – review & editing. YL: Conceptualization, Formal Analysis, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by the Foundation of the First Affiliated Hospital of Chongqing Medical University (grant no. PYJJ2020-06) and the Natural Science Foundation of Chongqing, China (grant no. CSTB2022NSCQ-M SX0973).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Romano N, Federici M, Castaldi A. Imaging of cranial nerves: A pictorial overview. *Insights Imaging*. (2019) 10:33. doi: 10.1186/s13244-019-0719-5
- Jäger H, Miszkil K. Pathology of the optic nerve. *Neuroimaging Clin N Am*. (2008) 18:243–50. doi: 10.1016/j.nic.2007.10.001
- De Fauw J, Ledsam J, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. (2018) 24:1342–50. doi: 10.1038/s41591-018-0107-6
- Dolz J, Leroy H, Reyns N, Massotier L, Vermandel M. A fast and fully automated approach to segment optic nerves on MRI and its application to radiosurgery. *Proceedings of the IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. Brooklyn, NY: (2015).
- Carrozzi A, Gramegna L, Sighinolfi G, Zoli M, Mazzatenta D, Testa C, et al. Methods of diffusion MRI tractography for localization of the anterior optic pathway: A systematic review of validated methods. *Neuroimage Clin*. (2023) 39:103494. doi: 10.1016/j.nicl.2023.103494
- Mansoor A, Cerrolaza J, Idrees R, Biggs E, Alsharid M, Avery R, et al. Deep learning guided par-titioned shape model for anterior visual pathway segmentation. *IEEE Trans Med Imaging*. (2016) 35:1856–65. doi: 10.1109/TMI.2016.2535222
- Harrigan R, Panda S, Asman A, Nelson K, Chaganti S, DeLisi M, et al. Robust optic nerve segmentation on clinically acquired computed tomography. *J Med Imaging*. (2014) 1:34006. doi: 10.1117/1.JMI.1.3.034006
- Gensheimer M, Cmelak A, Niermann K, Dawant B. Automatic delineation of the optic nerves and chiasm on CT images. *SPIE Med Imaging*. (2007) 41:651216. doi: 10.1117/12.711182
- Zhao Z, Ai D, Li W, Fan J, Song H, Wang Y, et al. Spatial probabilistic distribution map based 3D FCN for visual pathway segmentation. In: Zhao Y, Barnes N, Chen B, Westermann R, Kong X, Lin C editors. *ICIG 2019. Lecture Notes in Computer Science*. (Vol. 11902), Cham: Springer (2019). p. 509–18. doi: 10.1007/978-3-030-34110-7_42
- Xie L, Huang J, Yu J, Zeng Q, Hu Q, Chen Z, et al. CNTSeg: A multimodal deep-learning-based network for cranial nerves tract segmentation. *Med Image Anal*. (2023) 86:102766. doi: 10.1016/j.media.2023.102766
- Pravatà E, Diociai A, Navarra R, Carmisciano L, Sormani M, Roccatagliata L, et al. Biometry extraction and probabilistic anatomical atlas of the anterior Visual Pathway using dedicated high-resolution 3-D MRI. *Sci Rep*. (2024) 14:453. doi: 10.1038/s41598-023-50980-x
- Lee H, Bao S, Huo Y, Landman B. 3D UX-Net: A Large Kernel Volumetric ConvNet Modernizing Hierarchical Transformer for Medical Image Segmentation. (2023). Available online at: <https://arxiv.org/abs/2209.15076v4> (accessed March 2, 2023).
- Wang X, Wang X, Yan Z, Yin F, Li Y, Liu X, et al. Enhanced choroid plexus segmentation with 3D UX-Net and its association with disease progression in multiple sclerosis. *Mult Scler Relat Disord*. (2024) 88:105750. doi: 10.1016/j.msard.2024.105750
- Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. (2015). Available online at: <https://arxiv.org/abs/1505.04597v1> (accessed May 18, 2015).
- Hatamizadeh A, Nath V, Tang Y, Yang D, Roth H, Xu D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. (2022). Available online at: <https://arxiv.org/abs/2201.01266v1> (accessed January 4, 2022).
- Shaker A, Maaz M, Rasheed H, Khan S, Yang M, Khan F. UNETR++: Delving into Efficient and Accurate 3D Medical Image Segmentation. (2024). Available online at: <https://arxiv.org/abs/2212.04497> (accessed on 4 May 2024 (v3)).
- Plotka S, Chrabaszcz M, Biecek P. Swin SMT: Global Sequential Modeling in 3D Medical Image Segmentation. (2024). Available online at: <https://arxiv.org/abs/2407.07514> (accessed on 10 Jul 2024).
- Flitcroft D, He M, Jonas J, Jong M, Naidoo K, Ohno-Matsui K, et al. IMI-defining and classifying myopia: A proposed set of standards for clinical and epidemiologic studies. *Invest Ophthalmol Vis Sci*. (2019) 60:M20–30. doi: 10.1167/iovs.18-25957
- Radunovic M, Vukcevic B, Radojevic N, Vukcevic N, Popovic N, Vuksanovic-Bozarc A. Morphometric characteristics of the optic canal and the optic nerve. *Folia Morphol*. (2019) 78:39–46. doi: 10.5603/FM.a2018.0065
- Selhorst J, Chen Y. The optic nerve. *Semin Neurol*. (2009) 29:29–35. doi: 10.1055/s-0028-1124020
- Zhang Z, Sabuncu M. Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels. (2018). Available online at: <https://arxiv.org/abs/1805.07836> (accessed November 28, 2018).
- Crum W, Camara O, Hill D. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans Med Imaging*. (2006) 25:1451–61. doi: 10.1109/TMI.2006.880587
- Kraft D. Computing the hausdorff distance of two sets from their signed distance functions. *Int J Comput Geometry Appl*. (2020) 30:19–49. doi: 10.48550/arXiv.1812.06740
- Yeghiazaryan V, Voiculescu I. Family of boundary overlap metrics for the evaluation of medical image segmentation. *J Med Imaging*. (2018) 5:15006. doi: 10.1117/1.JMI.5.1.015006
- Yeghiazaryan V, Voiculescu I. *An Overview of Current Evaluation Methods used in Medical Image Segmentation*. Oxford: University of Oxford (2015).
- Ai D, Zhao Z, Fan J, Song H, Qu X, Xian J, et al. Spatial probabilistic distribution map-based two-channel 3D U-net for visual pathway segmentation. *Pattern Recogn Lett*. (2020) 138:601–7. doi: 10.1016/j.patrec.2020.09.003
- Li S, Chen Z, Guo W, Zeng Q, Feng Y. Two parallel stages deep learning network for anterior visual pathway segmentation. In: Gyori N, Hutter J, Nath V, Palombo M, Pizzolato M, Zhang F editors. *Computational Diffusion MRI. Mathematics and Visualization*. Cham: Springer (2021). p. 279–90. doi: 10.1007/978-3-030-73018-5_22
- van Elst S, de Bloeme C, Noteboom S, de Jong M, Moll A, Görcke S, et al. Automatic segmentation and quantification of the optic nerve on MRI using a 3D U-Net. *J Med Imaging*. (2023) 10:034501. doi: 10.1117/1.JMI.10.3.034501
- Pineles S, Demer J. Bilateral abnormalities of optic nerve size and eye shape in unilateral amblyopia. *Am J Ophthalmol*. (2009) 148:551–557.e2. doi: 10.1016/j.ajo.2009.05.007

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

30. Avery R, Mansoor A, Idrees R, Biggs E, Alsharid M, Packer R, et al. Quantitative MRI criteria for optic pathway enlargement in neurofibromatosis type 1. *Neurology*. (2016) 86:2264–70. doi: 10.1212/WNL.0000000000002771
31. Yiannakas M, Toosy A, Raftopoulos R, Kapoor R, Miller D, Wheeler-Kingshott CAM. MRI acquisition and analysis protocol for in vivo intraorbital optic nerve segmentation at 3T. *Invest Ophthalmol Vis Sci*. (2013) 54:4235–40. doi: 10.1167/iovs.13-12357
32. Tosun M, Güngör M, Uslu H, Anık Y. Evaluation of optic nerve diameters in individuals with neurofibromatosis and comparison of normative values in different pediatric age groups. *Clin Imaging*. (2022) 85:83–8. doi: 10.1016/j.clinimag.2022.02.021
33. Harrigan R, Plassard A, Mawn L, Galloway R, Smith S, Landman B. Constructing a statistical atlas of the radii of the optic nerve and cerebrospinal fluid sheath in young healthy adults. *Proc SPIE Int Soc Opt Eng*. (2015) 9413:941303. doi: 10.1117/12.2081887
34. Aghdasi N, Li Y, Berens A, Harbison R, Moe K, Hannaford B. Efficient orbital structures segmentation with prior anatomical knowledge. *J Med Imaging*. (2017) 4:34501. doi: 10.1117/1.JMI.4.3.034501
35. Seider M, Lee R, Wang D, Pekmezci M, Porco T, Lin S. Optic disk size variability between African, Asian, white, Hispanic, and Filipino Americans using Heidelberg retinal tomography. *J Glaucoma*. (2009) 18:595–600. doi: 10.1097/IJG.0b013e3181996f05
36. Girkin C, McGwin G, Sinai M, Sekhar G, Fingeret M, Wollstein G, et al. Variation in optic nerve and macular structure with age and race with spectral-domain optical coherence tomography. *Ophthalmology*. (2011) 118:2403–8. doi: 10.1016/j.ophtha.2011.06.013
37. Inglese M, Rovaris M, Bianchi S, La Mantia L, Mancardi G, Ghezzi A, et al. Magnetic resonance imaging, magnetisation transfer imaging, and diffusion weighted imaging correlates of optic nerve, brain, and cervical cord damage in Leber's hereditary optic neuropathy. *J Neurol Neurosurg Psychiatry*. (2001) 70:444–9. doi: 10.1136/jnnp.70.4.444
38. Mncube S, Goodier M. Normal measurements of the optic nerve, optic nerve sheath and optic chiasm in the adult population. *SA J Radiol*. (2019) 23:1772. doi: 10.4102/sajr.v23i1.1772
39. Juenger V, Cooper G, Chien C, Chikermane M, Oertel F, Zimmermann H, et al. Optic chiasm measurements may be useful markers of anterior optic pathway degeneration in neuromyelitis optica spectrum disorders. *Eur Radiol*. (2020) 30:5048–58. doi: 10.1007/s00330-020-06859-w