Baseline Method of the MVAA 2026 Challenge for Multimodal Mitral Valve Segmentation

Bo Deng[1,*] and Jieyun Bai[1,2,3]

1 The First Affiliated Hospital of Jinan University, Jinan University, Guangzhou, China

2 Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand

3 School of Information Science and Technology, Jinan University, Guangzhou, China

** Corresponding author: Bo Deng (dbd2002@foxmail.com)*

Abstract. Mitral valve anatomy analysis is a clinically important yet technically challenging problem because the valve is thin, deformable, and observed through heterogeneous imaging modalities. The MVAA 2026 challenge evaluates automated mitral valve segmentation across cardiac computed tomography (CT), three-dimensional transesophageal echocardiography (3D TEE), and surgical video. This paper describes the baseline methods released for the challenge. The baseline intentionally combines reliable, reproducible components rather than highly specialized task-specific engineering: 3D U-Net models are used for volumetric CT and TEE segmentation, while a two-dimensional U-Net++ model with an EfficientNet encoder is used for video-frame segmentation. To reflect the limited annotation regime, the CT and surgical-video branches additionally use exponential-moving-average teacher models and confidence-filtered pseudo-label supervision on unlabeled data. All predictions are evaluated by the unified MVAA validation protocol using Dice similarity coefficient (DSC), Hausdorff distance (HD), and average surface distance (ASD). On the local validation split, the baseline obtains DSC, HD, and ASD scores of 0.6184, 9.6213, and 0.9594 for CT; 0.7451, 18.7673, and 1.6552 for 3D TEE; and 0.6440, 193.2962, and 24.7622 for surgical video. These results provide a transparent reference point for future multimodal and geometry-aware methods.

Keywords: Mitral Valve Anatomy · Multimodal Segmentation · Semi-supervised Learning · Baseline Method

1 Introduction

Mitral valve disease is a major component of valvular heart disease and often requires accurate assessment of annular geometry, leaflet morphology, and device-tissue relationships for diagnosis, intervention planning, intra-operative guidance, and follow-up [17, 24, 6]. In clinical practice, no single modality captures all relevant information. Cardiac CT provides high-resolution volumetric anatomy for pre-procedural planning, 3D transesophageal echocardiography (TEE) supports intra-operative functional assessment but suffers from acoustic shadowing and speckle, and surgical video offers direct visual confirmation

while introducing specular reflection, occlusion, camera motion, and changing surgical fields. Earlier mitral-valve-specific segmentation work, especially in 3D TEE, also shows that the anatomy is difficult to delineate even when a single modality is considered in isolation [18]. Automated segmentation of the mitral valve apparatus is therefore a natural but demanding benchmark for medical image analysis. The target structures are small, have ambiguous boundaries, and differ substantially across CT, ultrasound, and RGB video. These issues are amplified by limited expert annotations, especially for video and high-resolution volumetric data, a recurring challenge in medical image segmentation surveys and imperfect-dataset settings [12, 21]. Related cardiac studies further show the importance of standardized datasets, benchmarks, and anatomical modeling, including right atrial LGE-MRI resources [30, 2], digital-twin and electrophysiological analyses for atrial arrhythmia and antiarrhythmic drug assessment [25, 1, 11], and task-specific cardiovascular segmentation models such as CFVP-Net [4]. Deep segmentation networks such as U-Net [20], V-Net [16], and self-configuring pipelines such as nnU-Net [9] have become strong baselines in medical image segmentation, while large-scale CT systems such as TotalSegmentator demonstrate the value of robust volumetric pipelines across heterogeneous clinical scans [26]. Thus, a multimodal challenge needs a reproducible baseline that can expose modality-specific failure modes. The Mitral Valve Anatomy Analysis using Multimodal Imaging Data (MVAA 2026) challenge addresses this setting through three segmentation tasks: CT segmentation, 3D TEE segmentation, and surgical-video segmentation. The validation platform expects a single submission package containing predictions for all three tasks and reports a common set of metrics: DSC, HD, and ASD. In this paper, we describe the baseline implementation accompanying the challenge release. The baseline follows three design principles. First, it uses standard, widely available tools: MONAI [3] for 3D medical volumes and segmentation-models-pytorch [27] for video frames. Second, it treats unlabeled data as an explicit part of the problem by adding teacher-student pseudo-labeling to the CT and video branches. Third, it keeps the submission and evaluation path identical to the challenge platform, so that the reported numbers are a practical reference for participants. Recent segmentation research increasingly emphasizes promptable and foundation-model segmentation, including SAM [10], MedSAM [14], SAM 2 [19], and surgical-domain SAM variants [29]. These models are promising for MVAA, but they introduce prompts, adaptation choices, and engineering decisions that are hard to interpret as a minimal reference. The official baseline therefore answers a narrower question: what performance can be expected from conventional, reproducible segmentation pipelines under the released data regime? Our contributions are as follows. We present a complete baseline system for all MVAA 2026 modalities, including preprocessing, training, inference, and submission formatting. We describe supervised and semi-supervised training strategies that are intentionally conservative and reproducible. Finally, we report unified validation results and analyze the distinct limitations revealed by the three modalities.

Table 1. Task characteristics used by the baseline experiments.

Task	Modality	Training data	Validation data	Prediction target
Task 1	Cardiac CT	27 lab. + 1040 unlab. vols.	30 vols.	Binary MV mask
Task 2	3D TEE	105 labeled vols.	20 vols.	Classes 1 and 2
Task 3	Surgical video	180 lab. + 1379 unlab. frames	48 frames	Binary label 10 mask

2 CHALLENGE AND DATASET

MVAA 2026 is organized as a unified segmentation evaluation over three clinically connected modalities, summarized in Table 1. Each task requires segmentation masks and a JSON file mapping case identifiers to mask paths; missing masks, invalid shapes, or unreadable predictions are penalized by the evaluator. Although each modality uses a different baseline branch, all predictions are packaged and scored through one validation protocol.

Task 1 targets mitral-valve-related structures in cardiac CT and uses 27 labeled volumes, 1040 unlabeled volumes, and 30 validation cases. Task 2 targets mitral valve segmentation in 3D TEE, with 105 labeled training pairs and 20 validation cases; foreground metrics are averaged over two TEE classes. Task 3 targets mitral valve regions in surgical frames, with 180 labeled frames, 1379 unlabeled frames, and 48 validation frames. Ground-truth video labels are stored as compressed archives, and the evaluator treats the configured target label as foreground.

Although all tasks share the same metric family, their data characteristics differ substantially. CT and TEE require memory-conscious patch training and sliding-window inference, whereas surgical video is two-dimensional but more vulnerable to camera-dependent appearance changes and small foreground masks.

3 METHOD

In this section, we delineate the official MVAA baseline framework. The baseline consists of modality-specific branches coordinated by a common training, inference, and submission interface. Given a task identifier t, the framework activates branch Φ_t and produces a mask in the validation format. The CT and 3D TEE branches use MONAI 3D U-Nets, while the surgical-video branch uses a 2D U-Net++ architecture. All branches use AdamW optimization [13], cosine learning-rate decay, and deterministic output formatting.

3.1 Framework Overview

Because CT, 3D TEE, and surgical video differ in dimensionality, intensity statistics, annotation density, and output format, the baseline uses explicit task routing:

$$\hat{Y} = \Phi_t(X), \quad t \in \{\mathrm{CT}, \mathrm{TEE}, \mathrm{Video}\}, \tag{1}$$

where X is the input image or volume and $\hat{Y}$ is the predicted mask. This deterministic routing preserves a unified challenge interface while allowing each modality to use an appropriate backbone and preprocessing pipeline.

The volumetric branches share the same high-level encoder-decoder formulation: a 3D U-Net predicts dense voxel-wise logits and sliding-window inference reconstructs full-resolution masks. The video branch uses a 2D encoder-decoder network and frame-wise post-processing. CT and video additionally exploit unlabeled data through teacher-student pseudo-labeling, while TEE is trained fully supervised because the available training set is labeled.

3.2 Volumetric Branches for CT and 3D TEE

For Tasks 1 and 2, the baseline uses a residual 3D U-Net implemented in MONAI. The reported configuration uses the base preset for CT and the large preset for TEE. Both models receive single-channel volumes and output voxel-wise class logits. Training is patch-based with a default region of interest of $128 \times 128 \times 128$, and inference uses Gaussian-blended sliding windows to recover full-size masks. CT intensities are clipped to $[-1000, 1000]$ and scaled to $[0, 1]$, while TEE intensities are scaled from $[0, 255]$ to $[0, 1]$. Random flips and rotations are used in both branches; TEE further applies affine and intensity perturbations to improve ultrasound robustness.

Semi-supervised CT Branch The CT task includes a large unlabeled set. To use it without complicating the baseline, we adopt a mean-teacher strategy [23]. This choice follows the broader trend of semi-supervised medical segmentation methods that combine pseudo-labeling, consistency regularization, and uncertainty- or confidence-aware filtering [28]. A student model is optimized by gradient descent, and a teacher model is maintained as an EMA of the student weights. The teacher predicts pseudo-labels on weakly perturbed unlabeled crops, while the student is trained on stronger perturbations including Gaussian noise and voxel dropout. A warm-up period trains the model with labeled data only, after which the unlabeled loss is gradually introduced.

Fully Supervised 3D TEE Branch The TEE branch is trained fully supervised because all available training cases are labeled. The split follows the baseline convention of holding out the last 20 cases for validation during training. Class-balanced random crops sample foreground classes more frequently than background, which is important because the mitral structures occupy a small part of the volume. The best checkpoint is selected by a weighted validation quality score based on DSC, HD, and ASD. During prediction, sliding-window logits are converted to labels by argmax and saved as NIfTI files with the source image header.

3.3 Surgical Video Branch

For Task 3, the baseline uses U-Net++ with an EfficientNet-B4 encoder [22] and a single binary output channel. Frames are resized to 448×800 and normalized with ImageNet statistics. The supervised objective combines Dice loss and focal loss, which is better suited to sparse foreground masks than binary cross-entropy alone. The labeled frames are split by video rather than by frame, preventing adjacent frames from the same video from appearing in both training and validation sets. Foreground-balanced sampling increases the chance that frames with small valve regions contribute useful gradients.

The video branch also uses an EMA teacher. Weak and strong augmented views are generated for each unlabeled frame; the teacher predicts on the weak view and the student is trained on the strong view. Pseudo-labels are accepted using asymmetric foreground/background confidence thresholds and minimum-area constraints. During validation and inference, horizontal/vertical flip TTA is averaged and the final binarization threshold is selected from validation candidates. Unlike memory-based video segmentation systems [5], this branch treats frames independently, leaving temporal modeling for future submissions.

3.4 Objective Functions

The active loss is determined by the task branch. For labeled CT and TEE volumes, the supervised objective is a weighted sum of Dice loss and cross-entropy loss:

$$\mathcal{L}_{sup} = 0.8\,\mathcal{L}_{Dice} + 0.2\,\mathcal{L}_{CE}. \tag{2}$$

Dice loss emphasizes foreground overlap and is widely used for imbalanced medical segmentation [16], while cross-entropy stabilizes voxel-wise class learning.

For an unlabeled CT crop x_u, the teacher predicts class probabilities $p_T(c|x_u)$. A pseudo-label is accepted only when the maximum probability exceeds a confidence threshold τ:

$$\hat{y}_u = \arg\max_c p_T(c|x_u), \quad m_u = \mathbf{1}\left[\max_c p_T(c|x_u) \geq \tau\right]. \tag{3}$$

The unsupervised loss is the confidence-masked pseudo-label cross-entropy:

$$\mathcal{L}_u = \frac{1}{\sum_i m_i} \sum_i m_i\,\mathcal{L}_{CE}(p_S(x_u)_i, \hat{y}_{u,i}), \tag{4}$$

with zero loss when no voxel passes the threshold. The total CT objective is $\mathcal{L}_{CT} = \mathcal{L}_{sup} + \lambda_u \mathcal{L}_u$, where λ_u is ramped up after warm-up.

For surgical video, the supervised objective combines Dice loss and focal loss:

$$\mathcal{L}_{video} = 0.7\,\mathcal{L}_{Dice} + 0.3\,\mathcal{L}_{Focal}. \tag{5}$$

The semi-supervised video loss follows the same teacher-student principle, but operates on binary frame masks and uses foreground/background confidence thresholds to reduce all-background pseudo-label collapse.

3.5 Inference and Submission Formatting

During inference, volumetric predictions are produced by sliding-window aggregation, converted to labels using argmax, and saved as NIfTI files. Video predictions are generated frame by frame, averaged under flip-based TTA, thresholded, and saved as binary masks. The final submission is a JSON-indexed package mapping validation identifiers to prediction files, which is included as part of the baseline because formatting errors are penalized by the evaluator.

4 EXPERIMENTS AND RESULTS

4.1 Datasets

The baseline is trained and evaluated using the local MVAA 2026 split summarized in Table 1. The validation set contains 30 CT volumes, 20 3D TEE volumes, and 48 surgical-video frames. CT and video are evaluated as binary foreground segmentation tasks, whereas TEE contains two foreground classes averaged by the evaluator. Unlabeled data are treated as part of the benchmark: Task 1 provides 1040 unlabeled CT volumes, and Task 3 provides 1379 unlabeled surgical frames. The baseline uses them through conservative pseudo-label supervision, while Task 2 remains fully supervised.

4.2 Evaluation Metrics

The validation evaluator computes DSC, HD, and ASD for each task. Let X be the predicted foreground set and Y the ground-truth foreground set. DSC measures overlap [7]:

$$\mathrm{DSC}(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|}. \tag{6}$$

HD measures the maximum symmetric boundary deviation:

$$\mathrm{HD}(X, Y) = \max\{h(\partial X, \partial Y), h(\partial Y, \partial X)\}, \tag{7}$$

where $h(\cdot, \cdot)$ is the directed Hausdorff distance [8]. ASD is the mean symmetric surface distance, computed from the distances between boundary voxels or pixels [15]. For Task 2, foreground metrics are averaged over classes 1 and 2. For Task 3, binary prediction masks are compared against the target label foreground.

4.3 Implementation Details

All branches are implemented in PyTorch using the released baseline code. Table 2 summarizes the main configuration. The CT and TEE branches use mixed precision when CUDA is available, AdamW optimization, cosine decay, and up to 200 epochs. The video branch uses 200 epochs, a batch size of 6, a learning rate of 2×10^{-4}, a 20-epoch semi-supervised warm-up, and a 30-epoch ramp-up of the unlabeled loss. The reported results use the same output packaging path as the validation bundle.

Table 2. Main baseline configurations.

Task	Backbone	Training type	Key inference detail
CT	3D U-Net (base)	DiceCE + EMA	Sliding window, argmax
3D TEE	3D U-Net (large)	Supervised DiceCE	Sliding window, argmax
Video	U-Net++ / EfficientNet-B4	Dice-Focal + EMA	TTA + threshold search

Table 3. Unified MVAA validation results of the baseline. HD and ASD are measured in the evaluator's native spacing units.

Task	Cases	DSC ↑	HD ↓	ASD ↓
Task 1: CT	30	0.6184	9.6213	0.9594
Task 2: 3D TEE	20	0.7451	18.7673	1.6552
Task 3: Surgical video	48	0.6440	193.2962	24.7622

4.4 Results and Analysis

The unified validation results are shown in Table 3. The TEE branch achieves the highest DSC, indicating that the large 3D U-Net and class-balanced crops provide a strong supervised baseline for labeled ultrasound volumes. The CT branch obtains a lower DSC but the best HD and ASD, suggesting spatially compact predictions when foreground is detected. In contrast, the video branch obtains a comparable DSC to CT but much worse distance metrics, indicating boundary outliers or missing thin components.

4.5 Qualitative Results

Figure 1 shows representative predictions. The CT example captures the broad valve-related region but misses thin components. The TEE example shows stronger overlap, consistent with its higher DSC. The video example highlights the sensitivity of HD and ASD to spatial shifts and missing leaflet portions.

5 DISCUSSION

The results reveal complementary strengths and weaknesses. TEE achieves the best overlap score, suggesting that a large 3D U-Net with class-balanced crops is effective for fully labeled ultrasound volumes. However, its distance metrics remain non-trivial because thin leaflet boundaries are difficult to trace under speckle and acoustic artifacts. For CT, the lower DSC indicates that EMA pseudo-labeling does not fully compensate for the small labeled set, although the low HD and ASD suggest limited extreme boundary outliers.

The surgical-video task exposes the clearest gap between overlap and boundary quality. Its DSC is comparable to CT, but HD and ASD are substantially worse because distant false-positive pixels or missing thin leaflet components strongly affect surface distances in high-resolution frames. The current baseline also processes frames independently, leaving temporal consistency unused.

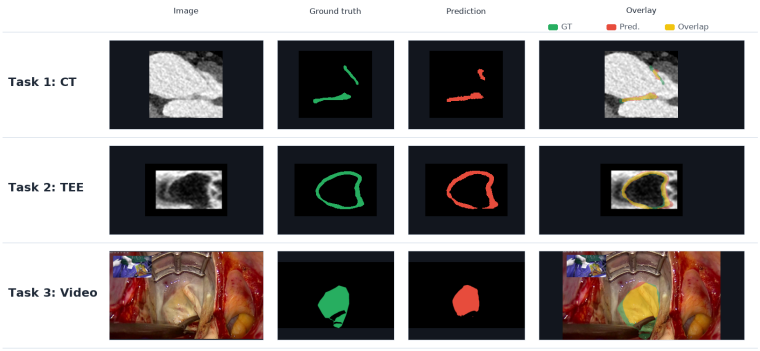

Fig. 1. Qualitative validation examples. Green denotes ground truth, red denotes prediction, and yellow denotes overlap. The three rows show CT, 3D TEE, and surgical-video examples.

Future work should consider video-level aggregation, optical-flow consistency, topology-aware losses, and cross-modal priors in which CT geometry regularizes TEE or video predictions. Promptable segmentation and video foundation models are promising, but they should be evaluated carefully because surgical video contains smoke, specular highlights, instruments, blood, and abrupt viewpoint changes that differ from natural-image pre-training data.

6 CONCLUSION

We presented the MVAA 2026 baseline for multimodal mitral valve segmentation across cardiac CT, 3D TEE, and surgical video. The baseline combines standard 3D U-Nets for volumetric segmentation, a 2D U-Net++ model for video frames, and conservative EMA pseudo-labeling for tasks with unlabeled data. Unified validation results show that the baseline is functional across all modalities while leaving substantial room for improved boundary accuracy, temporal consistency, and cross-modal anatomical modeling.

Acknowledgments

This work was supported in part by the Guangzhou Science and Technology Planning Project under Grant Nos. 2025B03J0127, 2024B03J1283, and 2024B03J1289.

References

1. Bai, J., Wang, W., Zhang, X., Lu, H., Zhang, H., Panfilov, A.V., Zhao, J.: Digital twin for sex-specific identification of class iii antiarrhythmic drugs based on in vitro measurements, computer models, and machine learning tools. PLOS Computational Biology **21**(7), e1013154 (2025)
2. Bai, J., Zhu, J., Chen, Z., Yang, Z., Lu, Y., Li, L., Li, Q., Wang, W., Zhang, H., Wang, K., et al.: A benchmark framework for the right atrium cavity segmentation from lge-mris. IEEE Transactions on Medical Imaging (2025)
3. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murray, B., Myronenko, A., Zhao, C., Yang, D., et al.: MONAI: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
4. Chen, Z., Bai, J., Li, S., Zhang, X., Lu, H.: Cfvp-net: Coarse-to-fine visual perception network for aortic dissection segmentation. In: 2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 5809–5816. IEEE (2025)
5. Cheng, H.K., Schwing, A.G.: XMem: Long-term video object segmentation with an atkinson-shiffrin memory model. In: European Conference on Computer Vision. pp. 640–658. Springer (2022)
6. Delgado, V., Ajmone Marsan, N., Bax, J.J.: Multimodality imaging in valvular heart disease. Nature Reviews Cardiology **16**(10), 567–584 (2019)
7. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)
8. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence **15**(9), 850–863 (1993)
9. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnu-net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021)
10. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollar, P., Girshick, R.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026 (2023)
11. Kulathilaka, A., Sharma, R., Kennelly, J., Li, N., Bai, J., Smaill, B.H., Trew, M.L., Fedorov, V.V., Zhao, J.: Structural determinants of re-entrant drivers in atrial fibrillation: insights from digital twins derived from 3d micrometre-resolution imaging of human heart. The Journal of physiology (2025)
12. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**, 60–88 (2017)
13. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019)
14. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**, 654 (2024)
15. Maurer, C.R., Qi, R., Raghavan, V.: A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. IEEE Transactions on Pattern Analysis and Machine Intelligence **25**(2), 265–270 (2003)
16. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision. pp. 565–571. IEEE (2016)

17. Otto, C.M., Nishimura, R.A., Bonow, R.O., Carabello, B.A., Erwin, J.P., Gentile, F., Jneid, H., Krieger, E.V., Mack, M., McLeod, C., et al.: 2020 acc/aha guideline for the management of patients with valvular heart disease. Circulation **143**(5), e72–e227 (2021)
18. Pouch, A.M., Wang, H., Takabe, M., Jackson, B.M., Gorman, J.H., Gorman, R.C., Yushkevich, P.A., Sehgal, C.M.: Fully automatic segmentation of the mitral leaflets in 3d transesophageal echocardiographic images using multi-atlas joint label fusion and deformable medial modeling. Medical Image Analysis **18**(1), 118–129 (2014)
19. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al.: SAM 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer (2015)
21. Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X.: Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Medical Image Analysis **63**, 101693 (2020)
22. Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th International Conference on Machine Learning. pp. 6105–6114 (2019)
23. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Advances in Neural Information Processing Systems. pp. 1195–1204 (2017)
24. Vahanian, A., Beyersdorf, F., Praz, F., Milojevic, M., Baldus, S., Bauersachs, J., Capodanno, D., Conradi, L., De Bonis, M., De Paulis, R., et al.: 2021 esc/eacts guidelines for the management of valvular heart disease. European Heart Journal **43**(7), 561–632 (2022)
25. Wang, W., Bai, J.: Assessment of proarrhythmic risk for class iii antiarrhythmic drug in atrium. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 5561–5568. IEEE (2024)
26. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., et al.: Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5), e230024 (2023)
27. Yakubovskiy, P.: Segmentation models pytorch. `https://github.com/qubvel/segmentation_models.pytorch` (2020)
28. Yang, X., Tian, J., Wan, Y., Chen, M., Chen, L., Chen, J.: Semi-supervised medical image segmentation via cross-guidance and feature-level consistency dual regularization schemes. Medical Physics **50**(7), 4269–4281 (2023)
29. Yue, W., Zhang, J., Hu, K., Xia, Y., Luo, J., Wang, Z.: SurgicalSAM: Efficient class promptable surgical instrument segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 6890–6898 (2024)
30. Zhu, J., Bai, J., Zhou, Z., Liang, Y., Chen, Z., Chen, X., Zhang, X.: Ras dataset: a 3d cardiac lge-mri dataset for segmentation of right atrial cavity. Scientific Data **11**(1), 401 (2024)