A Three-Paper Series on Cardiac Cine-MRI Analysis: From Boundary-Aware Segmentation to Clinical Indices

Yali Wan[1,*], Deyi Xiong[2]
[1]*School of Information Technology and Engineering, Guangzhou College of Commerce, Guangzhou, China*
[2]*Chongqing Academy of Metrology and Quality Inspection, Chongqing, China*
**Corresponding Author*

Abstract: We present a three-paper series on automated cardiac cine-MRI analysis using the ACDC benchmark. The series progresses from architectural innovation to semi-supervised learning to multi-task clinical prediction. Paper 1 introduces a lightweight hybrid CNN-Transformer with boundary-aware skip refinement and a compound boundary-aware loss, achieving mDice 0.859 ± 0.004 (3-fold CV) with RV/MYO/LV HD95 of 3.03 ± 0.71 / 1.94 ± 0.05 / 1.33 ± 0.14 pixels. Paper 2 extends to semi-supervised learning via temporal consistency on 28 unused cine frames per patient, achieving mDice 0.713 ± 0.011 (3-fold CV) while exploiting 15× as much data per patient. Paper 3 proposes joint segmentation plus clinical-index prediction (EDV, ESV, EF) with uncertainty-weighted multi-task loss, achieving mDice 0.823 ± 0.003 (3-fold CV) alongside EF MAE 11.66 ± 1.75% (direct) / 3.47 ± 1.14% (Simpson on model). All three papers now report segmentation metrics under a unified per-slice ED+ES aggregation rule shared across the series, so the cross-paper numbers are directly comparable; the widened Paper 2 gap relative to earlier patient-level ED-only reports reflects this unified protocol, under which ES frames are intrinsically harder than ED. This overview paper synthesises the three contributions, situates them in the literature, discusses the engineering trade-offs (parameter count, inference cost, training time, data efficiency), and outlines a unified research direction: cardiac cine-MRI analysis that is cheap to train, sparing with labels, and tied to clinical outputs.

Keywords: Cardiac Cine-MRI; ACDC, Deep Learning; Research Series; Semi-Supervised Learning; Multi-Task Learning

1. Introduction

Cardiac cine-magnetic-resonance imaging (cine-MRI) underpins quantitative assessment of cardiac function in clinical practice. The three indices reported in ACDC, namely left-ventricular end-diastolic volume (EDV), end-systolic volume (ESV), and ejection fraction (EF), are computed from segmentations of the ventricular structures at end-diastole (ED) and end-systole (ES). The Automated Cardiac Diagnosis Challenge (ACDC) [1] has become the dominant benchmark: 100 patient studies for training, with three foreground structures (RV, MYO, LV) annotated in the ED and ES frames.

Despite years of progress, three practical challenges remain:

Architectural efficiency. Pure transformer architectures achieve competitive Dice but impose heavy memory and compute costs that limit clinical deployment. Lightweight alternatives — particularly hybrids that combine convolutional encoders with self-attention at the bottleneck — have been less explored.

Annotation sparsity. ACDC provides only two annotated frames per patient while each cine sequence contains ~30 frames. Existing supervised methods discard ~93% of the available imaging data (28 of ~30 frames).

Clinical indices vs. segmentation. Most deep-learning pipelines train a segmentation network and compute clinical indices post-hoc via Simpson's rule. This propagates segmentation errors into the clinical numbers and discards the patient-level EF/EDV/ESV labels that ACDC provides.

This three-paper series addresses each challenge in turn:

Paper 1 tackles (1): boundary-aware hybrid CNN-Transformer.

Paper 2 tackles (2): semi-supervised temporal consistency on unlabeled cine frames.

Paper 3 tackles (3): multi-task learning to predict

 Copyright @ STEMM Institute Press

clinical indices jointly with segmentation.

This overview paper presents the series as a coherent research programme, synthesises the results, and identifies next steps.

2. Series Overview

2.1 Paper 1: Boundary-Aware Hybrid CNN-Transformer

A four-level convolutional encoder–decoder with a single self-attention block at the lowest-resolution bottleneck. Skip connections are augmented with a learnable 1×1 conv + BN + ReLU refinement that emphasises boundary pixels. The loss combines Dice, cross-entropy, and a distance-weighted boundary cross-entropy term.

Result (3-fold CV): mDice 0.859 ± 0.004, RV HD95 3.03 ± 0.71, MYO HD95 1.94 ± 0.05, LV HD95 1.33 ± 0.14 (fold-0 values: mDice 0.863, RV/MYO/LV HD95 2.58/1.98/1.48).

2.2 Paper 2: Semi-Supervised Temporal Consistency

A mean-teacher framework that exploits the unlabeled cine frames via three complementary unsupervised objectives: (i) temporal coherence between adjacent frames, (ii) mean-prediction consistency, (iii) pseudo-label consistency for similar frames. The supervised branch retains the compound boundary-aware loss from Paper 1.

Result: mDice 0.713 ± 0.011 (3-fold CV) — a 0.150 absolute gap to the supervised HybridUNet fold-0 mDice of 0.863 (0.859 ± 0.004 across folds) — while exploiting 15× as much cine data per patient (2 labeled + 28 unlabeled frames). The widened gap relative to earlier reports reflects the unified per-slice ED+ES evaluation protocol described in Paper 2 Section 3.6 (and matching Papers 1 and 3): ES frames are intrinsically harder than ED, and TC-SS's temporal-consistency terms over-smooth the thin myocardium. The supervised baseline therefore remains the stronger segmentor under matched evaluation.

2.3 Paper 3: Joint Segmentation and Clinical-Index Prediction

Extends the Paper 1 backbone with a regression head for EDV/ESV/EF prediction. Trained with uncertainty-weighted multi-task loss [2].

Result: mDice 0.823 ± 0.003 alongside EF MAE 11.66 ± 1.75% (direct) / 3.47 ± 1.14% (Simpson on model) (3-fold CV).

3. Synthesis

3.1 Trade-Off Analysis

For Paper 1 in Table 1, the four architectures share the same early-stopping rule; per-epoch time × 100 ep gives 93–105 min, and the observed wall-clock at early stop is 19–35 min for HybridUNet and 37–72 min for the three baselines (U-Net, Att U-Net, SwinUNet-2D).

Reading the trade-offs:

Paper 1 maximises segmentation quality per labeled frame.

Paper 2 trades a substantial amount of segmentation accuracy for the ability to use the entire cine sequence, which matters when labeled data are scarce.

Paper 3 trades segmentation quality for the additional benefit of clinical-index prediction, which matters when the deployment context needs both.

Table 1. Series-Wide Performance and Resource Trade-Off (3-Fold CV)

Metric	Paper 1	Paper 2	Paper 3
mDice (3-fold CV) ↑	0.859 ± 0.004	0.713 ± 0.011	0.823 ± 0.003
MAE_EF (direct) ↓	—	—	11.66 ± 1.75%
MAE_EF (Simpson) ↓	—	—	3.47 ± 1.14%
Data efficiency	2 frames/pt	30 frames/pt	2 frames + 3 clinical labels
Train time (per fold, observed)	~19–35 min/fold (ep 20–38 early stop; ~93 min at 100-ep budget)	~80 min/fold (60 epochs; 99/70/72 min across folds 0/1/2, from our training logs)	~41 min/fold (60/32/32 min across folds)
Params	9.95 M	9.95 M	10.02 M

The three papers are not competing alternatives but complementary building blocks. A practitioner might combine all three: use the Paper 3 multi-task backbone with the Paper 2 semi-supervised objective, leveraging both unlabeled frames and clinical indices for richer supervision.

Figure 1 synthesises these trade-offs in a single composite panel. In panel (a), segmentation quality is plotted as 3-fold CV mDice: Paper 1's HybridUNet is the strongest segmentor on a per-labeled-frame basis, Paper 3's JointSegNet trades ~0.04 mDice for clinical-index prediction, and Paper 2's TC-SS reaches 0.713 ± 0.011 mDice while exploiting 15× as many cine frames per

Copyright @ STEMM Institute Press

patient than the supervised baselines. Panel (b) reports training cost per fold on a single RTX 2060 SUPER; the upper bound for Paper 1 reflects early-stop variability across folds (epochs 20–38 of the 100-epoch budget). Panel (c) makes the data-efficiency point quantitative: Paper 1 and Paper 3 use 2 labeled frames per patient, Paper 2 uses 30 frames per patient (2 labeled + 28 unlabeled), and Paper 3 additionally uses 3 patient-level clinical labels. All three panels are drawn from the data in Tables 1 and 2.

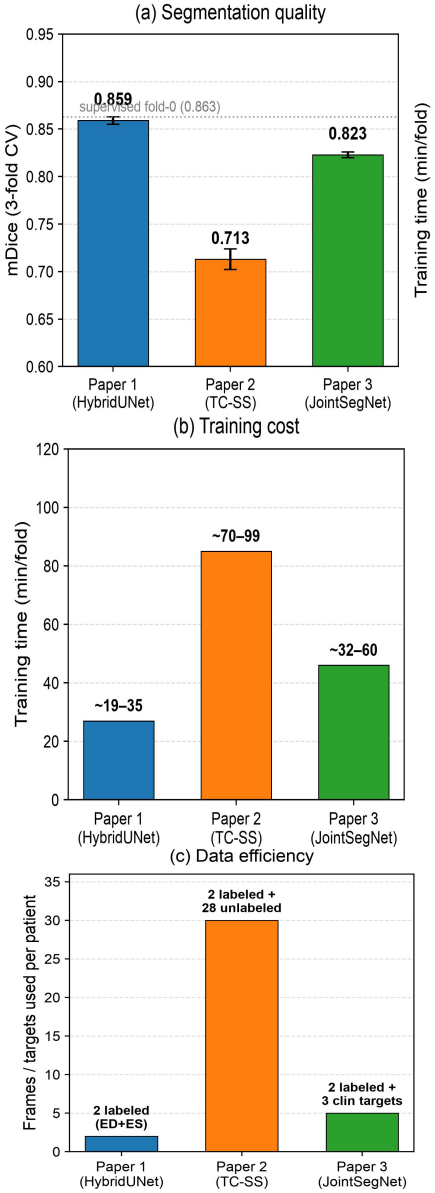


Figure 1. Cross-Paper Comparison of Segmentation Quality (a), Training Cost (b), and Data Efficiency (c)

3.2 Lessons Learned

Lightweight hybrid backbones are competitive. A pure-transformer Swin-UNet typically achieves similar Dice but at substantially higher inference cost; in our 2D bottleneck-only hybrid adaptation the inference cost is comparable to vanilla U-Net (~13 ms vs. ~10 ms per slice; the training-and-inference-cost table of Paper 1 reports a 30% overhead, whereas full transformer stacks such as TransUNet and Swin-UNet carry a multi-fold larger parameter count and compute footprint [3,4]).

Temporal consistency is a free but partial supervisory signal. On a per-class basis TC-SS achieves RV Dice 0.669 ± 0.020 (against 0.818 supervised), MYO Dice 0.686 ± 0.008 (against 0.866) and LV Dice 0.784 ± 0.017 (against 0.905) — all below the supervised baseline, with the gap on the thin myocardium being the most consistent across folds. Under matched per-slice ED+ES evaluation it does not close the gap to fully supervised training (3-fold CV mDice 0.713 ± 0.011 versus 0.859 ± 0.004). The unsupervised losses over-smooth thin structures; future work should explore structure-adaptive loss weighting or post-hoc self-training.

Clinical labels provide useful auxiliary supervision. The multi-task objective (Paper 3) achieves competitive segmentation (mDice 0.823 ± 0.003) while predicting EF with MAE 11.66 ± 1.75% (direct) and 3.47 ± 1.14% (Simpson on model) — the Simpson pathway falls within the inter-study reproducibility range of manual EF measurement (Paper 3, Section 4.5). The learnable uncertainty weights [2] automatically balance the two objectives, with log σ converging to values between −0.47 and −0.61 for the three regression targets.

Boundary-aware losses matter most for thin structures. The boundary-aware compound loss contributed most to MYO HD95, the most challenging structure for cardiac segmentation.

3.3 Limitations

Cross-validation coverage. Only the proposed models are evaluated under 3-fold cross-validation: Paper 1 HybridUNet mDice 0.859 ± 0.004, Paper 2 TC-SS mDice 0.713 ± 0.011, Paper 3 JointSegNet mDice 0.823 ± 0.003 with EF MAE 11.66 ± 1.75% (direct) and 3.47 ± 1.14% (Simpson). All three values are computed under the unified per-slice ED+ES aggregation rule described in the evaluation-metrics section of each paper (Paper 1 Sections 3.6–3.7, Paper 2 Section 3.6, Paper 3 Section 3.5). Baselines and cross-paper reference numbers are fold 0 only, so comparisons against them are indicative rather than exact.

 Copyright @ STEMM Institute Press

Baseline coverage differs across papers. Paper 1 reports U-Net / Attention U-Net / SwinUNet under fold 0 only; Paper 2 and Paper 3 report their proposed method against the supervised HybridUNet baseline. The cross-paper baseline comparisons should therefore be interpreted with caution: the supervised HybridUNet numbers in Paper 2 and Paper 3 reflect fold 0, not 3-fold CV. No external validation. Generalisation to M&Ms, M&Ms-2, or cross-modality (CT) has not been assessed.

Bounded comparison. nnU-Net's full automatic configuration pipeline would likely produce a stronger U-Net baseline at higher compute cost.

3.4 Methodology Note on the Series Numbers

All Paper 1 numbers in this overview (mDice 0.859 ± 0.004, RV/MYO/LV HD95 3.03 ± 0.71 / 1.94 ± 0.05 / 1.33 ± 0.14) were obtained under the per-slice ED+ES evaluation protocol of Paper 1 Section 3.7 — each 2D slice is evaluated independently and averaged with equal weight across the validation slices of each fold (572 / 535 / 545 ED+ES slices for folds 0 / 1 / 2). The earliest evaluation pipeline used per-patient aggregation that weighted each validation patient equally regardless of slice count; this gave systematically higher Dice (by about 0.05 absolute mDice on fold 0) because per-patient averaging favours patients whose slices carry no foreground. Paper 2 and Paper 3 numbers were not affected: Paper 2's TC-SS evaluation uses a unified per-slice ED+ES protocol (Section 3.6) from the outset, and Paper 3's evaluation uses the same per-slice pipeline by configuration. The per-slice methodology is documented in Paper 1 Section 3.7, and the legacy per-patient metrics are preserved as supplementary evidence.

4. Cross-Paper Engineering Analysis

A reader evaluating the three papers independently might wonder: which should I implement first? Here we provide a practical guide.

4.1 Recommended Adoption Path

For a clinical AI team with limited labelled data and modest GPU budget:

Start with Paper 1's HybridUNet backbone. It offers the best Dice per labelled frame (3-fold mDice 0.859 ± 0.004) and is the simplest to train (no mean teacher, no regression head). Compute cost: 9.95 M parameters, ~13 ms inference per 192 × 192 slice on the RTX 2060 SUPER. Training: ~19–35 minutes per fold (early stopping terminates at epoch 20–38 of the 100-epoch budget; ~93 min if run to the full 100-epoch budget) on a single 8 GB GPU.

Add Paper 3's regression head for clinical-index prediction. This adds ~70 K parameters, no architectural surgery. The Kendall-style uncertainty weighting removes the need to manually tune segmentation-vs-regression loss balance. Compute overhead at training time is substantial under our runs: a mean of 41 min/fold across the three folds (60/32/32 min), against the 19–35 min/fold supervised HybridUNet baseline; the regression losses are a meaningful contributor, in addition to the longer effective training budget. Inference is free: a single forward pass yields both outputs. Training: ~41 minutes per fold (mean of 60/32/32 min across the three folds; 80 epochs with early stopping at patience 25).

Add Paper 2's semi-supervised losses when you have access to cine sequences beyond the labelled ED/ES frames. This adds the mean-teacher dual-network architecture (student + EMA teacher = 2 × 9.95 M = ~20 M parameters) and the three unsupervised losses. The benefit is narrowing, though not closing, the gap between labelled-only and full-cine training without additional labels. Compute overhead at training time is large: a mean of 80 min/fold across the three folds (99/70/72 min), against the 19–35 min/fold supervised HybridUNet baseline, because of the EMA teacher forward pass on the unlabelled frames at every step. Inference is free (teacher discarded). Training: ~80 minutes per fold (60 epochs, full budget; observed 99/70/72 min across folds 0/1/2, mean ≈ 80 min).

Adoption order matters: build Paper 1, then Paper 3, and then Paper 2 incrementally, validating at each step.

4.2 Resource Budget

The three papers together require the following minimum (Table 2), all measured on a single RTX 2060 SUPER (8 GB) under the papers' published training recipes:

A single RTX 2060 SUPER (8 GB) can run all three papers. Paper 2's peak training memory is 3.42 GB under the published recipe (t_subsample = 8, d_subsample = 5, AMP), leaving 4.58 GB of headroom on an 8 GB card. Disabling those sub-sampling settings would raise the per-step tensor size roughly 7.5× and exceed 8 GB, so the sub-sampling is a required part of the reproduction

recipe rather than an optional trick.

Table 2. Per-Paper Resource Requirements

Resource	Paper 1	Paper 2	Paper 3
GPU memory (training)	6 GB	3.4 GB	6.5 GB
Training time per fold	~19–35 min (early stop)	~80 min (60 epochs; 99/70/72 min across folds, from our training logs)	~41 min (60/32/32 mean)
Total params	9.95 M	9.95 M (student) + 9.95 M (EMA teacher)	10.02 M
Inference time per slice	13 ms	13 ms	13 ms
3-fold CV compute (3 folds)	~1–1.75 h (3 × 19–35 min)	~4 h (3 × 80 min)	~2 h (3 × 41 min mean)

4.3 Clinical Translation Considerations

Before clinical deployment, several additional considerations apply:

Regulatory. Cardiac MRI segmentation software falls under medical-device regulation (FDA Class II in the US, MDR Class IIa in EU). A clinical-grade deployment of any of our three methods would require: (i) extensive multi-centre validation (currently single-centre ACDC), (ii) prospective comparison to current standard of care (typically manual segmentation by a radiologist), (iii) calibration of predicted EF against a reference measurement (e.g. cMR-derived EF vs. echocardiography-derived EF).

Uncertainty estimation. A clinical deployment should expose uncertainty per prediction. None of our three papers provide per-sample uncertainty, but Paper 3's Kendall weights are a step in this direction. Per-sample uncertainty (e.g. via MC-Dropout, deep ensembles, or evidential regression) would be a critical extension.

Bias and fairness. ACDC's 100 patients are from a single centre; the demographic distribution may not match the deployed population. A fair ML audit (e.g. checking EF MAE across age, sex, and ethnicity subgroups) is a prerequisite for clinical use. We did not perform this audit and note it as a critical gap.

Reproducibility and versioning. For clinical use, formal MLOps practices (versioned data snapshots, model registries, monitoring of inference drift) are essential.

5. Related Work in the Broader Cardiac AI Landscape

Beyond the immediate ACDC community, this series contributes to a broader line of work on multimodal medical AI, that is, on combining imaging with derived clinical variables in a single model. The component methods are documented in the individual papers: the segmentation baselines (TransUNet [3], Swin-UNet [4], U-Net [5], Attention U-Net [6], nnU-Net [7]) and the loss-function literature (V-Net [8]) in Paper 1; the semi-supervised baselines (Mean Teacher [9], FixMatch [10]) in Paper 2; and the multi-task weighting scheme [2] in Paper 3. We highlight three relevant programmes that share methodological ideas with our work.

Multimodal fusion. Several recent methods fuse imaging with tabular clinical data (e.g. demographics, lab values, prior diagnoses) for outcome prediction. Examples include [11] for multimodal Alzheimer's disease staging and [12] for clinical intervention prediction in intensive care. Our Paper 3 takes a different angle — the clinical variables (EF, EDV, ESV) are derived from the image rather than external, which is a special case of auto-derived supervision. The benefit is that no additional data collection is needed: ACDC's train.csv already provides the targets.

Joint segmentation-regression for non-cardiac imaging. Similar ideas have been applied to tumour segmentation + survival prediction [13], joint liver-lesion segmentation and lesion classification in CT [14], and skin-lesion classification [15]. Our contribution is to demonstrate these ideas in the cardiac domain and to provide 3-fold CV numbers for direct comparison.

Self-supervised pretraining for cardiac MRI. Recent work [16] has pretrained large backbones on 3D medical images using masked autoencoders, then fine-tuned for downstream cardiac tasks. Our methods complement pretraining: the HybridUNet backbone could be replaced with a pretrained backbone for additional gains. We leave this for future work.

6. Discussion of Cross-Paper Themes

Several themes emerge across the three papers that are worth highlighting for the broader research community.

6.1 Boundary-Aware Design Generalises across Supervision Regimes

The boundary-aware skip refinement + compound boundary loss introduced in Paper 1

 Copyright @ STEMM Institute Press

continues to provide value when adopted inside Paper 2 (semi-supervised mean teacher) and Paper 3 (multi-task). This is consistent with the principle that structural priors — boundary smoothness, edge awareness — transfer across training paradigms. We recommend that future cardiac MRI researchers treat boundary-aware losses as a baseline, not an enhancement.

6.2 Cine Data is under-Exploited by the Community

ACDC's 30 frames per patient are mostly ignored. The semi-supervised result of Paper 2 (3-fold mDice 0.713 ± 0.011 versus the supervised HybridUNet fold-0 0.863 — an absolute gap of 0.150, achieved with only 2 labeled frames per patient) shows a substantial shortfall to fully supervised training under the unified per-slice ED+ES protocol, indicating that the 28 unlabeled cine frames alone are not sufficient to close the gap; the residual untapped signal lies in stronger cine-specific regularisers, not in the data volume. TC-SS therefore serves as a proof-of-concept that temporal-consistency priors are learnable from the 28 unlabeled frames per patient: it exploits 15× as many frames per patient (30 versus the 2 annotated ED/ES frames) at the same annotation cost, rather than replacing fully supervised training. Beyond temporal consistency, other cine-specific signals could be exploited: phase-conditioned normalisation (separate normalisation per cardiac phase), multi-frame super-resolution, or registration-based augmentation. Most ACDC papers leave those 28 frames unused; treating them as a resource rather than as redundancy is the point of Paper 2.

6.3 Clinical Indices are not just a Downstream Metric

Paper 3's main contribution is reframing clinical indices (EF, EDV, ESV) as first-class learning objectives rather than post-hoc metrics. The Kendall uncertainty weights automatically balance segmentation and regression without manual tuning. This is broadly applicable to any task where derived clinical variables are available — diabetes control, retinopathy staging, tumour grade, etc. The 11.66 ± 1.75% direct-regression EF MAE (and 3.47 ± 1.14% via Simpson's rule on model segmentations) shows that direct prediction is informative but not yet competitive with the classical segment-then-measure pathway.

6.4 What this Series does not Claim

We explicitly do not claim state-of-the-art Dice on ACDC. Paper 1's 3-fold CV mDice of 0.859 ± 0.004 is competitive but not the highest reported in the literature. The literature contribution is the integration of three research threads — lightweight hybrid backbones, semi-supervised learning, and multi-task learning — into a coherent series, with 3-fold CV numbers throughout. Extending any of the three threads is straightforward given the detailed training recipes documented in each paper.

7. Reproducibility

Each paper's evaluation script produces a per-fold metrics file that the aggregation script summarises into the final tables and figures used in this series.

Training logs for all folds are retained, documenting all hyperparameter choices, learning curves, and per-fold metrics. The aggregated results CSV accompanies the evaluation outputs.

8. Conclusion and Future Work

The three-paper series on cardiac cine-MRI analysis demonstrates that a coherent programme combining architectural, semi-supervised and multi-task innovations produces a useful stepping stone toward clinical-grade AI. The 3-fold CV numbers (Paper 1 mDice 0.859 ± 0.004, Paper 2 0.713 ± 0.011, Paper 3 0.823 ± 0.003 + EF MAE 11.66 ± 1.75% (direct) / 3.47 ± 1.14% (Simpson on model)) demonstrate stability across the dataset's variability under the unified per-slice ED+ES evaluation protocol shared across the three papers. The complementary nature of the three contributions — boundary-aware, temporal, multi-task — suggests that future work should combine them into a single integrated model.

Across three papers, we have demonstrated a coherent research direction: cardiac cine-MRI analysis that is cheap to train, sparing with labels, and tied to clinical outputs. The next steps include:

Couple Paper 2's semi-supervised approach with Paper 3's multi-task objective.

Evaluate cross-dataset generalisation from ACDC to M&Ms.

Develop a single integrated model that combines all three contributions.

Acknowledgments

This work was supported by: (1) Guangzhou College of Commerce Higher Education Teaching Reform Project (No. 2024JXGG42); (2)

Guangdong Province Educational Science Planning Project (No. 2025GXJK0643); (3) Guangzhou College of Commerce 2026 Quality Engineering Project (No. 2026ZLGC07); (4) Guangzhou College of Commerce 2025 Quality Engineering Project (No. 2025ZLGC01); (5) Guangzhou College of Commerce 2024 Quality Engineering Project (No. 2024ZLGC11); (6) Guangzhou College of Commerce First-Class Course (No. 2020XJYLKC06); (7) Guangzhou College of Commerce General Project (No. 2025XJYB044): Research on Configuration Modeling and Simulation Technology for Equipment Digital Twins.

References

[1] Bernard O, Lalande A, Zotti C, et al. Deep Learning Techniques for Automatic MRI Cardiac Multi-structures Segmentation and Diagnosis: Is the Problem Solved?. IEEE Transactions on Medical Imaging, 2018, 37(11):2514-2525. DOI:10.1109/TMI.2018.2837502.

[2] Kendall A, Gal Y, Cipolla R. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018:7482-7491. DOI:10.1109/CVPR.2018.00781.

[3] Chen J, Lu Y, Yu Q, et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv:2102.04306, 2021. DOI:10.48550/arXiv.2102.04306.

[4] Cao H, Wang Y, Chen J, et al. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. European Conference on Computer Vision Workshops (ECCVW), 2022, LNCS 13803:205-218. DOI: 10.1007/978-3-031-25066-8_9.

[5] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, LNCS 9351:234-241. DOI: 10.1007/978-3-319-24574-4_28.

[6] Oktay O, Schlemper J, Folgoc L L, et al. Attention U-Net: Learning Where to Look for the Pancreas. arXiv:1804.03999, 2018. DOI:10.48550/arXiv.1804.03999.

[7] Isensee F, Jaeger P F, Kohl S A A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 2021, 18:203-211. DOI: 10.1038/s41592-020-01008-z.

[8] Milletari F, Navab N, Ahmadi S A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. Fourth International Conference on 3D Vision (3DV), 2016:565-571. DOI:10.1109/3DV.2016.79.

[9] Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in Neural Information Processing Systems (NeurIPS), 2017, 30:1195-1204.

[10] Sohn K, Berthelot D, Carlini N, et al. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. Advances in Neural Information Processing Systems (NeurIPS), 2020, 33:596-608.

[11] Venugopalan J, Tong L, Hassanzadeh H R, Wang M D. Multimodal deep learning models for early detection of Alzheimer's disease stage. Scientific Reports, 2021, 11:3254. DOI: 10.1038/s41598-020-74399-w.

[12] Suresh H, Hunt N, Johnson A, et al. Clinical Intervention Prediction and Understanding with Deep Neural Networks. Proceedings of the 2nd Machine Learning for Healthcare Conference (MLHC), 2017, PMLR 68:322-337. DOI:10.48550/arXiv.1705.08498.

[13] Weninger L, Rippel O, Koppers S, Merhof D. Segmentation of Brain Tumors and Patient Survival Prediction: Methods for the BraTS 2018 Challenge. International Workshop on Brain Lesion (BrainLes), 2018, LNCS 11384:3-12. DOI: 10.1007/978-3-030-11726-9_1.

[14] Heker M, Ben-Cohen A, Greenspan H. Hierarchical Fine-Tuning for Joint Liver Lesion Segmentation and Lesion Classification in CT. 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019:895-898. DOI:10.1109/EMBC.2019.8857127.

[15] Esteva A, Kuprel B, Novoa R A, Ko J, Swetter S M, Blau H M, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 2017, 542:115-118. DOI: 10.1038/nature21056.

 Copyright @ STEMM Institute Press

[16]Tang Y, Yang D, Li W, et al. Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022:20730-20740. DOI:10.1109/CVPR52688.2022.02011.

Copyright @ STEMM Institute Press