Deep Learning Approaches for Cardiac MRI Segmentation: A Narrative Review of Methods and Clinical Applications

Brahmleen Papneja
Queen's University
brahmleen.papneja@queensu.ca

Patrick Romanescu
Queen's University
patrick.romanescu@queensu.ca

Shrika Vejandla
Queen's University
shrika.vejandla@queensu.ca

***Abstract*—Cardiac magnetic resonance imaging (CMR) is the reference standard for noninvasive assessment of cardiac structure, function, and tissue characterization, but manual segmentation remains time-intensive and variable. We systematically reviewed deep learning methods for cardiac MRI segmentation published from 2015 to 2026 to evaluate architectural progress, real-world robustness, and barriers to clinical translation. Searches of Scopus, MEDLINE, Web of Science, and PubMed identified 1,769 records; 278 met eligibility criteria, and 51 studies were selected for detailed synthesis. Deep learning models, particularly U-Net derivatives and newer transformer- and state-space-based approaches, achieved strong benchmark performance, often with Dice coefficients above 0.90. However, performance frequently declined under domain shift, motion artifact, rare pathology, and multi-center deployment. Overall, benchmark success has outpaced clinical readiness. Future progress will depend less on architectural complexity and more on pathological diversity, quality control integration, external validation, and prospective outcome-based studies.**

I. Introduction

Cardiac magnetic resonance imaging (CMR) has emerged as the gold standard for sensitive assessment of cardiac pathology, providing superior soft tissue contrast to echocardiography and comprehensive tissue characterization without ionizing radiation [1, 2]. Quantitative CMR-derived biomarkers, including left ventricular ejection fraction, global longitudinal strain, and fibrosis burden from late gadolinium enhancement (LGE), inform risk stratification and guide therapeutic decision-making in conditions such as cardiomyopathies, myocarditis, ischemic heart disease, and congenital heart disease and serve as surrogate endpoints of ventricular function in clinical trials [4, 5]. However, manual segmentation of CMR imaging remains clinically necessary, despite requiring 20–30 minutes per case and exhibiting substantial inter-observer variability that can influence diagnostic assessment [6]. This labor-intensive workflow limits scalability in both research and clinical practice as imaging volumes increase and multi-parametric CMR protocols, including cine function assessment, parametric mapping (T1/T2), perfusion imaging, and late gadolinium enhancement, become routine. Cardiac MRI segmentation presents unique technical challenges. Substantial anatomical variability exists across patient populations, while pathological remodeling from myocardial infarction, cardiomyopathy, and valvular disease alters expected morphology and tissue characteristics [4, 6, 7]. Multi-sequence protocols including cine, LGE, T1/T2 parametric mapping, and phase contrast imaging each present distinct contrast properties and imaging artifacts, such as hyperintense signals mimicking fibrosis and motion-related ghosting [2, 10]. Inter-observer variability can result in measurement differences of 10–15% for ventricular volumes and dependent metrics such as ejection fraction, and up to 20% for myocardial mass [6]. Class imbalance further complicates segmentation, as myocardial pathology (e.g., fibrosis) may represent less than 0.2% of pixels in LGE imaging [6, 8]. These challenges make cardiac MRI segmentation uniquely complex and establish the imperative for robust automated solutions. Deep learning marked a paradigm shift from traditional approaches. Early implementations of deep learning in the context of CMR segmentation, including atlas-based registration and hand-engineered feature classifiers, demonstrated limited generalization across protocols and pathologies [1]. However, the 2015 first-generation U-Net architecture achieved unprecedented accuracy on sparsely labeled training data and established the foundational encoder-decoder framework with skip connections that undermines contemporary approaches to CMR segmentation [13]. Subsequent innovations on the U-Net framework, such as the incorporation of residual connections, attention mechanisms, and multi-scale feature fusion, have driven rapid improvements in sensitivity, with recent models achieving Dice coefficients exceeding 0.90 on standardized benchmarks such as the Automated Cardiac Diagnosis Challenge (ACDC) [6, 12]. The proliferation of open datasets and community challenges has accelerated progress and established reproducible evaluation frameworks [5, 12].

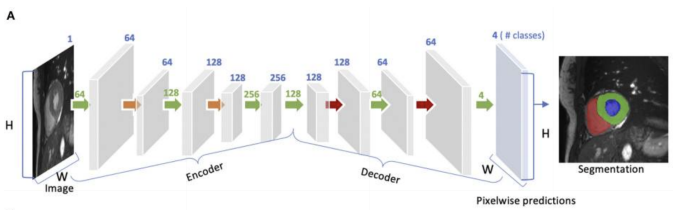

Fig. 1. U-Net architecture for cardiac MRI segmentation. The contracting path (left) progressively reduces spatial dimensions while increasing feature channels. The expansive path (right) recovers spatial resolution through upsampling and concatenation with skip connections. Adapted from Chen et al.

Despite strong benchmark performance, clinical translation remains challenging. Pathology-specific applications have emerged for myocardial infarction, fibrosis, and cardiomyopathies [4,6-7], yet multi-center validation studies often reveal performance degradation on unseen data distributions [5, 7]. Persistent issues include limited generalization across scanner manufacturers and protocols [10], sensitivity to domain shift [5, 10], inadequate quality control [12], and critically, limited prospective validation linking automated metrics to clinical outcomes [5]. Motion artifacts from free-breathing acquisitions substantially degrade performance [10,11], while annotation scarcity limits development for rare pathologies, motivating semi-supervised and cross-modality learning strategies [8 - 10]. This gap underscores the need for systematic evaluation of robustness, reliability, and clinical utility beyond technical accuracy.

Deep learning models for cardiac MRI segmentation achieve strong performance on benchmark datasets but struggle in real clinical settings. Models trained on curated data often fail when faced with different scanners, imaging protocols, rare pathologies, and motion artifacts. Additionally, current research is fragmented across different technical approaches without clear understanding of what works in practice. This review systematically evaluates deep learning methods for cardiac MRI segmentation, from basic architectures to clinical applications, to identify what progress has been made and what barriers remain before these tools can be reliably used in everyday clinical practice.

II. Methodology

A comprehensive literature search was conducted across Scopus, MEDLINE, Web of Science, and PubMed using Medical Subject Headings (MeSH) and keywords organized into four domains: methods, anatomy, imaging modalities, and applications (Table 1). Boolean operators were applied to combine terms within and across domains. The search included peer-reviewed publications from January 2015 through January 2026.

TABLE I
SEARCH STRATEGY KEYWORDS AND MESH TERMS ORGANIZED BY CONCEPTUAL DOMAIN.

Domain	Terms
Methods	Deep learning, CNN, U-Net, transformers, attention mechanisms, ResNet, residual networks, diffusion models, GAN, machine learning, artificial intelligence
Anatomy	Cardiac, heart, myocardium, ventricle, atrium, cardiomyopathy
Imaging	MRI, CMR, cardiovascular magnetic resonance, cardiac magnetic resonance, cine MRI, 4D MRI
Application	Segmentation, delineation, contour, boundary detection

Note: Terms within each domain were combined using OR operators; domains were combined using AND operators.

Studies were included if they: (1) involved human subjects undergoing cardiac MRI with segmentation of cardiac structures (ventricles, atria, myocardium, or whole heart); (2) employed deep learning methods including CNNs, U-Net variants, transformers, attention mechanisms, GANs, or hybrid architectures for 2D, 3D, or 4D segmentation of cine, late gadolinium enhancement, or parametric mapping sequences; (3) reported quantitative performance metrics (Dice coefficient, Hausdorff distance, Jaccard index, sensitivity, specificity, or volumetric error) or qualitative evaluation with methodological detail; and (4) were original research articles in English. Exclusions included animal-only or phantom-only studies, non-cardiac anatomy, classical machine learning without deep learning, pure classification tasks, lack of performance validation, and review articles or abstract-only submissions.

Two reviewers (BP, PR) independently screened 1,769 articles. After removing 693 duplicates, 1,076 studies underwent title and abstract screening, 632 advanced to full-text review, and 278 met eligibility criteria. From this pool, 51 studies were selected for detailed synthesis based on methodological rigor, thematic relevance to foundational architectures, pathology-oriented segmentation, annotation-efficient learning, emerging models, and clinical translation, and representation of diverse technical and clinical approaches. Discrepancies were resolved through consensus discussion.

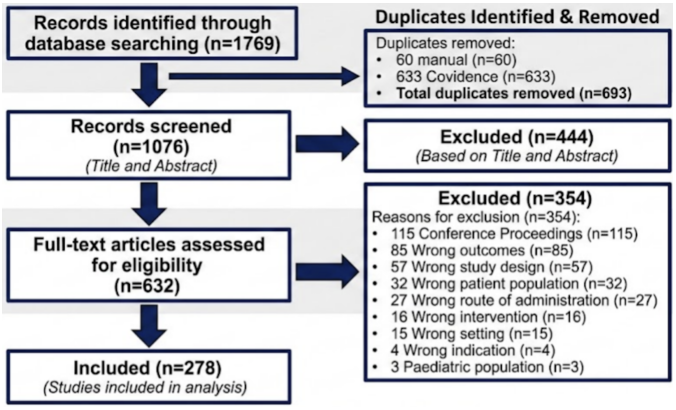

Fig. 2. PRISMA flow diagram showing literature selection process. From 1,769 records to 278 eligible studies, with 51 selected for final synthesis.

III. Results and Discussion

A. Core Supervised Segmentation in Cine Cardiac MRI

1) Architectural Evolution of Supervised Models: Deep learning architectures for cardiac MRI segmentation have evolved through iterative refinement of the U-Net framework, targeting limitations in feature extraction, spatial encoding, and boundary delineation [14]. The original 2D U-Net established strong frame-level performance but lacked mechanisms for exploiting inter-slice continuity or temporal dependencies across cardiac phases [15]. Key innovations included: dilated convolutions to expand receptive fields without reducing resolution [16]; residual connections to enable deeper architectures and improve gradient flow [17, 18]; self-attention mechanisms

for long-range spatial dependencies, with emerging JEPA frameworks emphasizing recurrent cardiac structures over pixel-level variations [17, 19]; and multi-scale feature aggregation through dilated inception modules and cross-attentive skip connections, particularly beneficial for challenging regions like RV apical segments [15, 20]. Dimensional modeling strategies remain debated. While 2D networks offer efficiency, 3D architectures (including nnU-Net) improve spatial coherence at computational cost [14, 22]. 4D temporal modeling has demonstrated improved inter-phase consistency, reducing LV myocardial volume mismatches from 4.0% to 2.6% in one study [14]. Geometry-informed optimization through curvature-aware loss functions and probabilistic noise correction modules enhances boundary quality, though performance gains on well-curated benchmarks have diminished as baseline models approach ceilings imposed by annotation variability [15, 20].

2) Benchmark Performance and Population-Scale Validation: Supervised models achieve Dice coefficients exceeding 0.90 for LV blood pool and 0.85 for LV myocardium on ACDC, with RV segmentation surpassing 0.90 in recent implementations [15-21]. Hausdorff distances approach expert inter-observer variability [16, 18]. However, overlap metrics may not capture clinically significant volumetric errors affecting ejection fraction and stroke volume [23].

The NAKO study, evaluating nnU-Net segmentation in 29,908 participants, revealed that despite strong performance, 17% of cases required quality review and 9.2% were excluded [22]. Time-volume curve analysis proved most effective for quality discrimination, emphasizing that physiological plausibility constraints provide critical validation beyond geometric overlap [22].

Pathology-specific studies expose generalizability limitations. In hypertrophic cardiomyopathy, automated segmentation correlated strongly with manual references ($r > 0.84$) but showed wider RV parameter limits of agreement [23]. Models trained on acquired heart disease underperform on congenital variations, reflecting insufficient pathological diversity rather than architectural deficiencies [23]. JEPA-based frameworks demonstrate improved generalization to morphologically diverse pediatric cohorts, underscoring the importance of learning recurrent anatomical structures [23]. Performance varies across cardiac phases, with end-systolic frames exhibiting lower accuracy due to smaller cavities, increased complexity, and greater morphological variability [20, 23]. Training on all cardiac phases rather than only ED/ES improved temporal coherence but yielded only modest functional metric improvements [14]. Stroke volume concordance between ventricles (biases <1 mL, ICC >0.93) indicates adequate reliability, though wider RV limits of agreement reflect persistent challenges [14, 23]. External validation consistently demonstrates performance degradation under domain shift. Models achieving Dice >0.95 often experience 2-5% reductions on external cohorts despite similar acquisition parameters [22]. Scanner hardware, protocol, positioning, coil configuration, and reconstruction algorithm differences introduce systematic biases. Models trained on standardized datasets struggle with extreme anatomical variants, severe dysfunction, and rare congenital malformations, motivating exploration beyond conventional supervised learning [22, 23].

B. Pathology-Oriented and Multi-Sequence Segmentation

1) Infarction, Scar, and Fibrosis in LGE Imaging: LGE imaging segmentation faces extreme class imbalance (pathological tissue $<0.2\%$ of image), irregular borders, and intensity overlap between infarcted and viable myocardium [24, 25]. Semi-supervised and adversarial approaches leverage unlabeled data through pseudo-label generation, with boundary mining architectures enhancing foreground-background distinction [26].

Cascaded architectures address class imbalance through hierarchical localization. Two-stage models first localize the LV, then refine MI territories and microvascular obstruction via morphological post-processing and texture-semantic encoding [25]. Similar multi-stage strategies for RV infarction employ centroid-guided extraction and coarse-to-fine refinement [24].

GAN-based augmentation synthesizes realistic scar patterns, with cGANs effectively augmenting rare infarct morphologies [27]. Self-supervised Siamese networks with contrastive objectives generate multi-scale representations outperforming single-scale approaches [28]. Transformer models with Pre-LN decoders and CIoU/L1 losses achieve higher precision than DETR for spatially heterogeneous scar patterns [28].

Persistent challenges include boundary ambiguity from partial volume effects, pathophysiologic variability (STEMI vs. non-STEMI replacement fibrosis; diffuse interstitial changes in pulmonary hypertension), and limited reproducibility affecting infarct burden quantification and risk stratification [24-26]. Multi-center deployment reliability remains unproven [24-28].

2) Multi-Parametric and Complex Disease Modeling: Multi-parametric frameworks integrate cine, LGE, T1/T2 mapping to enable simultaneous segmentation of edema, scar, and structural abnormalities, mirroring clinical radiologist workflows [29, 30]. Early fusion models combined up to five sequences with explicit encoding pathways for incomplete inputs [29]. Cine-based models predict scar/edema without contrast, offering benefits for patients with gadolinium contraindications, though cross-disease-stage validation is needed [30].

State-space Mamba modules capture long-range dependencies for subtle edema-scar boundaries [31]. Hybrid radiomic-deep learning approaches combine texture features (Joint Entropy, Run Length Non-Uniformity) with U-Net segmentation, outperforming standalone deep learning for MI detection [32]. T1/T2 mapping segmentation across 1,009 patients achieved inter-observer-level accuracy, with multi-feature classifiers outperforming single-threshold methods [33].

In light-chain cardiac amyloidosis, transformer models with contrastive pretraining outperformed Mayo staging for prognostic stratification, with attention analyses revealing integration of subtle global remodeling patterns beyond visible amyloid deposition [34].

C. Reducing Annotation Burden and Improving Generalization

1) Semi- and Weakly-Supervised Learning: Pixel-level annotation requirements create major deployment barriers [35]. Scribble-based supervision reduces annotation time by more than 0.95 but historically produces irregular contours [37]. Feature decomposition methods separate class-specific and class-agnostic regions, enabling discriminative boundary learning from sparse supervision, with lightweight networks achieving competitive performance on ACDC and MMs benchmarks [36]. However, pseudo-label quality depends critically on initial predictions, and teacher-student frameworks risk iteratively reinforcing errors if training is insufficient [38].

Multi-stage training separates labeled and unlabeled data streams to reduce gradient interference [38]. Uncertainty-aware modules dynamically adjust pseudo-label confidence thresholds, improving robustness but introducing hyperparameter sensitivity [38]. Student-teacher frameworks with meta pseudo-labeling enable adaptive label generation, reducing systematic bias in fixed architectures [37, 39]. Combining weak and strong augmentation encourages consistency across perturbations, improving generalization [39].

Multidimensional consistency constraints with shared encoders, separate decoders, and pyramid boundary features achieved 91.2% Dice on ACDC and 89.7% on multi-center MMs data [40]. Performance drops significantly for small or abnormal structures where pseudo-labels are less accurate [41]. Shape priors and ensemble methods with graph cut refinement reduce Hausdorff errors and improve volumetric consistency [41]. Frequency-domain pseudo-label mixing with partial supervision constraints improves boundary delineation, achieving competitive performance with reduced labeled data, though gaps remain for complex or diseased anatomy [35].

2) Cross-Modality and Data Augmentation Strategies: LGE annotation scarcity motivates cross-modality approaches using abundant bSSFP data [42, 43]. Image-to-image translation networks generate synthetic LGE training data while preserving anatomical structure, enabling biventricular segmentation without LGE labels [42, 43]. Translation quality is critical; distribution mismatch reduces performance if synthetic images inadequately represent pathology-specific enhancement patterns [42].

Mixup-based masking at image and feature levels increases scribble annotation diversity [37]. Adversarial discriminators refine pseudo-label contours using unpaired anatomical masks, enforcing anatomically plausible shapes based on population priors [37]. Augmentation-driven consistency in student-teacher frameworks with dual augmentation (weak/strong) improves cross-site robustness, particularly when labeled data is limited, suggesting synergistic rather than independent contributions [39]. Deformable modeling via learned energy functions from localized anatomical features enables competitive performance with fewer training volumes [44]. However, these methods require accurate initialization and degrade when disease-related remodeling alters expected anatomy [44].

Persistent challenges include overfitting to synthetic/augmented distributions that inadequately represent rare pathologies or unusual acquisition conditions [42, 43]. External multi-center validation remains inconsistently reported, limiting generalizability assessment [39, 44]. While promising, annotation-efficient methods require further validation before routine clinical translation [41, 44].

D. Emerging Architectures and Foundation Models

1) State-Space, Transformer, and Prompt-Based Models: Mamba-based architectures provide linear-complexity sequence modeling with global receptive fields, avoiding quadratic scaling of self-attention while preserving fine anatomical details in high-resolution cardiac MRI [45, 46]. Dual-stream architectures combine visual state-space and convolutional branches for discriminative features in small pathological regions like scar and edema [47, 48]. Frequency-domain reasoning separates low-frequency structural components from high-frequency boundary details, addressing boundary ambiguity arising from attenuation during feature extraction [45].

Foundation model adaptation shows promise. cineCMR-SAM extends SAM with temporal-spatial attention and text/box prompts, achieving Dice 0.94 (LV) and 0.86 (myocardium) on multi-center data, with robustness across aortic stenosis severity levels and ventricular remodeling patterns [49]. DINOv2-based encoders with multi-scale fusion achieved 0.923 Dice for left atrium segmentation [50].

Domain generalization frameworks combine Vision Mamba with bidirectional state-space modeling and dual-normalization to align features across modalities without target domain supervision [46]. VEBlock modules integrate sequence modeling with non-local attention for cross-modality consistency, though generalization to complex congenital disease, advanced cardiomyopathies, and post-surgical variants remains unclear, as training typically includes only common pathologies [46].

2) Robustness Under Motion and Domain Shift: The CMRx-Motion challenge revealed substantial performance degradation under respiratory motion artifacts, with LV ejection fraction errors exceeding 10% and myocardial mass estimates degrading $>$15% under severe motion [51]. Segmentation-guided diffusion models for motion-corrupted image restoration achieved 12-18% SNR and 15-22% CNR improvements over unconditional baselines, demonstrating interdependence between restoration and segmentation [52].

SI2CRL framework models causal image formation structure, treating frequency-domain phase information as anatomical content while applying amplitude interventions to simulate distribution shifts [53]. Contrastive causal decoupling and adversarial purification remove domain-specific correlations, achieving 3-7% Dice improvements on unseen domains across cross-site, cross-modality, and cross-sequence tasks [53]. Active shape models with deep learning feature extraction provide complementary robustness through parametric shape priors [54].

Critical gaps persist. CMRxMotion provides controlled artifact variations, but clinical motion patterns are more diverse and unpredictable [51]. Diffusion restoration adds computational complexity and risks generating inaccurate

anatomy if conditioning signals are imperfect [52]. Causal methods require broader pathological validation to confirm generalization beyond training domains [53]. Prospective validation under uncontrolled acquisition conditions, diverse pathologies, and different scanner hardware remains limited [49, 51, 53].

E. Clinical Deployment and Real-World Validation

Automated segmentation deployment requires quality control systems to detect silent failures; anatomically plausible but inaccurate contours affecting downstream measurements [60]. QCResUNet produces subject-level quality predictions and voxel-level error maps, identifying outliers and highlighting challenging regions (RV apex, atrial boundaries, papillary muscles) [60]. This dual-level approach enables automated alerts for human review and targeted correction rather than full re-segmentation [60].

CT-guided refinement addresses MRI's lower spatial contrast in detailed regions (sinus of Valsalva, ascending aorta) through non-rigid registration of CT-derived labels [61]. While preserving anatomically plausible boundaries and valve geometry, this approach did not improve Dice accuracy versus nnU-Net alone, highlighting misalignment between anatomical realism and overlap metrics [61]. Label transfer across patients introduces registration errors, particularly for unusual anatomy (bicuspid valve, aneurysm, post-surgical reconstruction) [61].

The most compelling clinical evidence links AI-derived measurements to outcomes. A multicenter, multivendor study (1,808 patients) validated against invasive right heart catheterization showed automated LV stroke volume correlated more strongly with hemodynamics than manual analysis (r=0.72 vs. 0.68), with high interstudy repeatability (ICC 0.79-0.99) and non-inferiority to manual measurements [8]. In 3,487 pulmonary hypertension patients, automated RV volumes, RVEF, and LV stroke volume were independently associated with mortality, demonstrating prognostic value [8]. External validation across 40 multivendor studies from 32 centers provides rare generalizability evidence [8].

Critical limitations persist: small sample sizes in prospective/disease-specific studies (3-50 patients) reduce statistical power and limit failure mode detection [1]; disease-specific models lack multi-center validation [2, 3]; quality control systems remain unintegrated into complete deployment pipelines [6]; cross-modality refinement lacks large-scale validation [7]. Large prospective trials evaluating segmentation-driven decisions in real procedural settings are needed to bridge the benchmark-to-bedside gap [8].

IV. Conclusion

The development of deep learning-based cardiac MRI segmentation has progressed from foundational U-Net architectures through emerging foundation models and prospective clinical applications, yet a consistent pattern persists across all examined areas: strong benchmark performance does not reliably translate into clinical robustness, and architectural complexity alone cannot compensate for insufficient pathological diversity in training data. Supervised models achieve Dice coefficients exceeding 0.90 on standard benchmarks yet degrade under external validation, struggling further with rare pathologies, end-systolic frames, and vendor-specific variability. Annotation-efficient methods demonstrate competitive performance with reduced labeling requirements, though pseudo-label quality and robustness on structurally abnormal regions remain inconsistent. Emerging state-space and prompt-based architectures offer efficiency advantages but largely retain conventional encoder-decoder designs and have not demonstrated meaningful gains in the edge cases that matter most clinically. Robustness to respiratory motion and domain shift remains inadequately addressed, quality control systems have not been integrated into complete deployment pipelines, and outcome-based validation, while the most compelling evidence of real utility, remains limited in scale and disease diversity. The most important next steps are not architectural: expanding training data to reflect pathological heterogeneity, integrating quality control into deployment pipelines, and conducting large prospective trials under real acquisition conditions are what the field most needs. Federated learning, uncertainty quantification, and explainability frameworks will be critical enablers of this transition. Until these gaps are addressed, the distance between benchmark success and bedside implementation will remain the defining challenge of the field.

V. Acknowledgements

This research was supported by QMIND and Queen's University, whose support were instrumental to the completion of this work. Their commitment to fostering student-led research and innovation is sincerely appreciated.

The author also acknowledges Liam McNamara and Soheil Chavoshi for their contributions to the early ideation and initial stages of implementation during the development of this project.

The opportunity to present this work at CUCAI is gratefully recognized, and appreciation is extended to the organizers for providing a platform for scholarly exchange.

References

[1] S. E. Petersen et al., "Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort," *J. Cardiovasc. Magn. Reson.*, 2017.

[2] J. Qiu et al., "MyoPS-Net: Myocardial pathology segmentation with flexible combination of multi-sequence CMR images," *Med. Image Anal.*, vol. 84, p. 102694, 2023, doi: 10.1016/j.media.2022.102694.

[3] S. Li et al., "Segmentation of the Left Ventricle and Its Pathologies for Acute Myocardial Infarction After Reperfusion in LGE-CMR Images," *IEEE Trans. Med. Imaging*, vol. 44, no. 11, pp. 4323–4334, 2025, doi: 10.1109/TMI.2025.3573706.

[4] S. Koehler et al., "Deep Learning-based Aligned Strain from Cine Cardiac MRI for Detection of Fibrotic Myocardial Tissue in Patients with Duchenne Muscular Dystrophy," *Radiol. Artif. Intell.*, vol. 7, no. 3, p. e240303, 2025, doi: 10.1148/ryai.240303.

[5] S. Alabed et al., "Validation of Artificial Intelligence Cardiac MRI Measurements: Relationship to Heart Catheterization and Mortality Prediction," *Radiology*, vol. 305, no. 1, pp. 68–79, 2022, doi: 10.1148/radiol.212929.

[6] C. Xu et al., "Segmentation of the right ventricular myocardial infarction in multi-centre cardiac magnetic resonance images," *Med. Image Anal.*, vol. 109, p. 103911, 2026, doi: 10.1016/j.media.2025.103911.

[7] L. Zhang et al., "FDDSeg: Unleashing the power of scribble annotation for cardiac MRI images through feature decomposition distillation," *IEEE J. Biomed. Health Inform.*, vol. 29, no. 1, pp. 285–296, 2025, doi: 10.1109/JBHI.2024.3404884.

[8] C. Li, Z. Zheng, and D. Wu, "Shape-aware adversarial learning for scribble-supervised medical image segmentation with a MaskMix Siamese network," *Bioengineering*, vol. 11, no. 11, p. 1146, 2024, doi: 10.3390/bioengineering11111146.

[9] W. Wang et al., "Cross-modality LGE-CMR segmentation using image-to-image translation based data augmentation," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 20, no. 4, pp. 2367–2375, 2023, doi: 10.1109/TCBB.2022.3140306.

[10] K. Wang et al., "Extreme cardiac MRI analysis under respiratory motion: Results of the CMRxMotion challenge," *Med. Image Anal.*, vol. 109, p. 103883, 2026, doi: 10.1016/j.media.2025.103883.

[11] V. Kesavan, M. Noga, and K. Punithakumar, "Segmentation-Guided Diffusion for Free-Breathing Cardiac Magnetic Resonance Image Restoration," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2025, doi: 10.1109/EMBC58623.2025.11254491.

[12] P. Qiu et al., "QCResUNet: Joint subject-level and voxel-level segmentation quality prediction," *Med. Image Anal.*, vol. 107, pt. A, p. 103718, 2026, doi: 10.1016/j.media.2025.103718.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv:1505.04597, 2015.

[14] A. C. Ogier et al., "Cardiac Function Assessment with Deep-Learning-Based Automatic Segmentation of Free-Running 4D Whole-Heart CMR," *J. Cardiovasc. Magn. Reson.*, p. 102677, 2025, doi: 10.1016/j.jocmr.2025.102677.

[15] Y. Du et al., "CASNet: curvature-aware cardiac MRI segmentation with multi-scale and attention-driven encoding for enhanced risk-oriented structural analysis," *Front. Med. (Lausanne)*, vol. 12, p. 1688872, 2025, doi: 10.3389/fmed.2025.1688872.

[16] F. Ahmad, W. Hou, J. Xiong, and Z. Xia, "Fully automated cardiac MRI segmentation using dilated residual network," *Med. Phys.*, vol. 50, no. 4, pp. 2162–2175, 2023, doi: 10.1002/mp.16108.

[17] Y. Z. Li et al., "RSU-Net: U-net based on residual and self-attention mechanism in the segmentation of cardiac magnetic resonance images," *Comput. Methods Programs Biomed.*, vol. 231, p. 107437, 2023, doi: 10.1016/j.cmpb.2023.107437.

[18] D. Abdelrauof, M. Essam, and M. Elattar, "Light-Weight Localization and Scale-Independent Multi-gate UNET Segmentation of Left and Right Ventricles in MRI Images," *Cardiovasc. Eng. Technol.*, vol. 13, no. 3, pp. 393–406, 2022, doi: 10.1007/s13239-021-00591-2.

[19] Z. Fu et al., "TF-Unet: An automatic cardiac MRI image segmentation method," *Math. Biosci. Eng.*, vol. 19, no. 5, pp. 5207–5222, 2022, doi: 10.3934/mbe.2022244.

[20] S. Dong et al., "DeU-Net 2.0: Enhanced deformable U-Net for 3D cardiac cine MRI segmentation," *Med. Image Anal.*, vol. 78, p. 102389, 2022, doi: 10.1016/j.media.2022.102389.

[21] P. M. Full et al., "Cardiac Magnetic Resonance Imaging in the German National Cohort (NAKO): Automated Segmentation of Short-Axis Cine Images and Post-Processing Quality Control," *J. Cardiovasc. Magn. Reson.*, p. 101958, 2025, doi: 10.1016/j.jocmr.2025.101958.

[22] S. Angiras et al., "Automated Deep Learning Based Cardiac Quantification in Hypertrophic Cardiomyopathy: A Comparative Study with Manual Segmentation," *Acta Med. Litu.*, vol. 32, no. 2, pp. 333–347, 2025, doi: 10.15388/Amed.2025.32.2.8.

[23] W. Y. Chai et al., "A Deep Learning-Based Fully Automated Cardiac MRI Segmentation Approach for Tetralogy of Fallot Patients," *J. Magn. Reson. Imaging*, vol. 63, no. 1, pp. 264–276, 2026, doi: 10.1002/jmri.70113.

[24] C. Xu et al., "Segmentation of the right ventricular myocardial infarction in multi-centre cardiac magnetic resonance images," *Med. Image Anal.*, vol. 109, p. 103911, 2026, doi: 10.1016/j.media.2025.103911.

[25] S. Li et al., "Segmentation of the Left Ventricle and Its Pathologies for Acute Myocardial Infarction After Reperfusion in LGE-CMR Images," *IEEE Trans. Med. Imaging*, vol. 44, no. 11, pp. 4323–4334, 2025, doi: 10.1109/TMI.2025.3573706.

[26] C. Xu et al., "BMAnet: Boundary Mining With Adversarial Learning for Semi-Supervised 2D Myocardial Infarction Segmentation," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 1, pp. 87–96, 2023, doi: 10.1109/JBHI.2022.3215536.

[27] D. R. P. R. M. Lustermans et al., "Optimized automated cardiac MR scar quantification with GAN-based data augmentation," *Comput. Methods Programs Biomed.*, vol. 226, p. 107116, 2022, doi: 10.1016/j.cmpb.2022.107116.

[28] Y. Ding et al., "DE-MRI myocardial fibrosis segmentation and classification model based on multi-scale self-supervision and transformer," *Comput. Methods Programs Biomed.*, vol. 226, p. 107049, 2022, doi: 10.1016/j.cmpb.2022.107049.

[29] J. Qiu et al., "MyoPS-Net: Myocardial pathology segmentation with flexible combination of multi-sequence CMR images," *Med. Image Anal.*, vol. 84, p. 102694, 2023, doi: 10.1016/j.media.2022.102694.

[30] W. Ding et al., "CineMyoPS: Segmenting Myocardial Pathologies From Cine Cardiac MR," *IEEE Trans. Med. Imaging*, vol. 44, no. 12, pp. 5027–5037, 2025, doi: 10.1109/TMI.2025.3586254.

[31] W. Zhang et al., "U-shaped network combining dual-stream fusion mamba and redesigned multilayer perceptron for myocardial pathology segmentation," *Med. Phys.*, vol. 52, no. 6, pp. 4567–4584, 2025, doi: 10.1002/mp.17812.

[32] W. Xu and X. Shi, "Integrating radiomic texture analysis and deep learning for automated myocardial infarction detection in cine-MRI," *Sci. Rep.*, vol. 15, no. 1, p. 24365, 2025, doi: 10.1038/s41598-025-08127-7.

[33] A. B. Popescu et al., "Deep learning-based segmentation of T1 and T2 cardiac MRI maps for automated disease detection," *Eur. Radiol.*, 2025, doi: 10.1007/s00330-025-12069-z.

[34] S. Wang et al., "Predicting prognosis of light-chain cardiac amyloidosis by magnetic resonance imaging and deep learning," *Eur. Heart J. Cardiovasc. Imaging*, vol. 26, no. 11, pp. 1771–1781, 2025, doi: 10.1093/ehjci/jeaf248.

[35] W. Li et al., "Weakly semi-supervised cardiac MRI segmentation with frequency-domain pseudo label dynamic mixed supervision and partial-dice constraints," *Neural Networks*, vol. 196, p. 108329, 2026, doi: 10.1016/j.neunet.2025.108329.

[36] L. Zhang et al., "FDDSeg: Unleashing the power of scribble annotation for cardiac MRI images through feature decomposition distillation," *IEEE J. Biomed. Health Inform.*, vol. 29, no. 1, pp. 285–296, 2025, doi: 10.1109/JBHI.2024.3404884.

[37] C. Li, Z. Zheng, and D. Wu, "Shape-aware adversarial learning for scribble-supervised medical image segmentation with a MaskMix Siamese network: A case study of cardiac MRI segmentation," *Bioengineering (Basel)*, vol. 11, no. 11, p. 1146, 2024, doi: 10.3390/bioengineering11111146.

[38] J. Liu et al., "Semi-supervised medical image segmentation based on multi-stage iterative training and high-confidence pseudo-labeling," *Biomed. Phys. Eng. Express*, vol. 11, no. 5, 2025, doi: 10.1088/2057-1976/adf3b7.

[39] S. M. K. Hasan and C. Linte, "STAMP: A self-training student-teacher augmentation-driven meta pseudo-labeling framework for 3D cardiac MRI image segmentation," *Med. Image Underst. Anal.*, vol. 13413, pp. 371–386, 2022, doi: 10.1007/978-3-031-12053-4_28.

[40] H. Cui et al., "Advancing cardiac MRI multi-structure segmentation: A semi-supervised multidimensional consistency constraint learning network," *Med. Phys.*, vol. 52, no. 7, p. e17805, 2025, doi: 10.1002/mp.17805.

[41] F. Guo et al., "Cardiac MRI segmentation with sparse annotations: Ensembling deep learning uncertainty and shape priors," *Med. Image Anal.*, vol. 81, p. 102532, 2022, doi: 10.1016/j.media.2022.102532.

[42] W. Wang et al., "Cross-modality LGE-CMR segmentation using image-to-image translation based data augmentation," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 20, no. 4, pp. 2367–2375, 2023, doi: 10.1109/TCBB.2022.3140306.

[43] X. Yu et al., "Cardiac LGE MRI segmentation with cross-modality image augmentation and improved U-Net," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 2, pp. 588–597, 2023, doi: 10.1109/JBHI.2021.3139591.

[44] H. R. Torres et al., "Deep-DM: Deep-driven deformable model for 3D image segmentation using limited data," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 12, pp. 7287–7299, 2024, doi: 10.1109/JBHI.2024.3440171.

[45] G. Ren et al., "CFG-MambaNet: Contextual and Frequency-Guided Mamba Network for medical image segmentation," *NPJ Digit. Med.*, 2026, doi: 10.1038/s41746-026-02393-z.

[46] Y. Shi et al., "Vision Mamba Empowered by Dynamic Domain Generalization for Cross-Modality Medical Segmentation," *J. Imaging Inform. Med.*, 2025, doi: 10.1007/s10278-025-01687-0.

[47] S. Li et al., "An enhanced visual state space model for myocardial pathology segmentation in multi-sequence cardiac MRI," *Med. Phys.*, vol. 52, no. 6, pp. 4355–4370, 2025, doi: 10.1002/mp.17761.

[48] W. Zhang et al., "U-shaped network combining dual-stream fusion mamba and redesigned multilayer perceptron for myocardial pathology

segmentation," *Med. Phys.*, vol. 52, no. 6, pp. 4567–4584, 2025, doi: 10.1002/mp.17812.
[49] Z. Chen et al., "Cine cardiac magnetic resonance segmentation using temporal-spatial adaptation of prompt-enabled segment-anything-model: a feasibility study," *J. Cardiovasc. Magn. Reson.*, vol. 27, no. 1, p. 101909, 2025, doi: 10.1016/j.jocmr.2025.101909.
[50] B. Kundu et al., "Multi-Scale Feature Fusion with Image-Driven Spatial Integration for Left Atrium Segmentation from Cardiac MR Images," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2025, doi: 10.1109/EMBC58623.2025.11253637.
[51] K. Wang et al., "Extreme cardiac MRI analysis under respiratory motion: Results of the CMRxMotion challenge," *Med. Image Anal.*, vol. 109, p. 103883, 2026, doi: 10.1016/j.media.2025.103883.
[52] V. Kesavan, M. Noga, and K. Punithakumar, "Segmentation-Guided Diffusion for Free-Breathing Cardiac Magnetic Resonance Image Restoration," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2025, doi: 10.1109/EMBC58623.2025.11254491.
[53] W. Liu et al., "Spectrum intervention based invariant causal representation learning for single-domain generalizable medical image segmentation," *Med. Image Anal.*, vol. 105, p. 103741, 2025, doi: 10.1016/j.media.2025.103741.
[54] D. O. Medley, C. Santiago, and J. C. Nascimento, "Deep Active Shape Model for Robust Object Fitting," *IEEE Trans. Image Process.*, 2019, doi: 10.1109/TIP.2019.2948728.
[55] A. P. Neofytou et al., "Deep learning-based prospective slice tracking for continuous catheter visualization during MRI-guided cardiac catheterization," *Magn. Reson. Med.*, vol. 94, no. 4, pp. 1626–1634, 2025, doi: 10.1002/mrm.30574.
[56] S. Koehler et al., "Deep Learning-based Aligned Strain from Cine Cardiac MRI for Detection of Fibrotic Myocardial Tissue in Patients with Duchenne Muscular Dystrophy," *Radiol. Artif. Intell.*, vol. 7, no. 3, p. e240303, 2025, doi: 10.1148/ryai.240303.
[57] W. Wang et al., "Segmentation of the Left Atrium in Cardiovascular Magnetic Resonance Images of Patients with Myocarditis," *J. Vis. Exp.*, no. 221, 2025, doi: 10.3791/68664.
[58] J. Baraboo et al., "Deep learning based automated left atrial segmentation and flow quantification of real time phase contrast MRI in patients with atrial fibrillation," *Int. J. Cardiovasc. Imaging*, vol. 41, no. 6, pp. 1197–1208, 2025, doi: 10.1007/s10554-025-03407-9.
[59] T. Yoganathan et al., "Automated cardiac MRI analysis for robust profiling of heart failure models in mice," *Sci. Rep.*, vol. 16, no. 1, p. 1158, 2025, doi: 10.1038/s41598-025-30810-y.
[60] P. Qiu et al., "QCResUNet: Joint subject-level and voxel-level segmentation quality prediction," *Med. Image Anal.*, vol. 107, pt. A, p. 103718, 2026, doi: 10.1016/j.media.2025.103718.
[61] H. Oda, M. Wakamori, and T. Akita, "Refining cardiac segmentation from MRI volumes with CT labels for fine anatomy of the ascending aorta," *Radiol. Phys. Technol.*, vol. 18, no. 3, pp. 734–745, 2025, doi: 10.1007/s12194-025-00926-x.
[62] S. Alabed et al., "Validation of Artificial Intelligence Cardiac MRI Measurements: Relationship to Heart Catheterization and Mortality Prediction," *Radiology*, vol. 305, no. 1, pp. 68–79, 2022, doi: 10.1148/radiol.212929.