



ARTICLE

Automated Detection of Apical Rocking in Echocardiographic Videos Using Deep Learning

Fangyan Tian^{1,2,#}, Yu Liu^{1,3,#}, Yuman Li^{4,#}, Mingshan Li^{5,6}, Haiyan Chen^{1,3}, Guohui Zhou^{5,6}, Yi Guo^{5,6}, Yuanyuan Wang^{5,6} and Xianhong Shu^{1,3,*}

¹Department of Echocardiography, Zhongshan Hospital, Fudan University, Shanghai, China

²Department of Ultrasound Medicine, The Affiliated Hospital of Guizhou Medical University, Guiyang, China

³Shanghai Institute of Medical Imaging, Shanghai, China

⁴Department of Ultrasound Medicine, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

⁵College of Biomedical Engineering, Fudan University, Shanghai, China

⁶Key Laboratory of Medical Imaging Computing and Computer Assisted Intervention (MICCAI) of Shanghai, Shanghai, China

*Corresponding Author: Xianhong Shu. Email: shu.xianhong@zs-hospital.sh.cn

#These authors contributed equally to this work

Received: 22 June 2026; Accepted: 10 September 2026; Published: 21 October 2026

ABSTRACT: Apical rocking (AR) is a dynamic echocardiographic manifestation of left ventricular (LV) mechanical dyssynchrony, but routine assessment remains qualitative and reader-dependent. This retrospective two-center proof-of-concept study developed and externally tested a video-based deep-learning pipeline for automated AR detection in a selected cohort of 351 patients with heart failure undergoing cardiac resynchronization therapy (CRT) evaluation. The framework first applied a U-Net-based segmentation model with a shape-prior module to delineate the LV endocardium and suppress non-anatomic background information. Segmented video sequences were then classified using a convolutional neural network (CNN) with long short-term memory (LSTM), with a CNN+Transformer model evaluated as a comparator. The segmentation model achieved a mean Dice coefficient of 0.956 in the held-out test set. In the 56-patient external test cohort from the second center, the CNN+LSTM model achieved an accuracy of 0.91, sensitivity of 0.96, specificity of 0.87, and area under the receiver-operating characteristic curve of 0.95. These findings support the technical feasibility of segmentation-constrained spatiotemporal learning for detecting expert-defined visual AR. Because the classification cohort was enriched for AR and excluded poor-quality examinations, the results should not be interpreted as performance in an unselected screening population. The study did not evaluate CRT response or clinical outcomes. Prospective multicenter validation in consecutive, multi-vendor cohorts that include suboptimal-quality examinations and longitudinal outcomes is required before clinical implementation.

KEYWORDS: Apical rocking; cardiac resynchronization therapy; deep learning; echocardiography; left ventricular dyssynchrony; video analysis

1 Introduction

Heart failure (HF) is the common end-stage of many cardiovascular diseases and remains associated with high morbidity, mortality, and healthcare costs [1]. Cardiac resynchronization therapy (CRT) is an established treatment for selected patients with HF and conduction abnormalities and can improve cardiac function, exercise tolerance, hospitalization, and survival [2–4]. Nevertheless, a substantial proportion of

patients do not respond despite meeting conventional implantation criteria [5,6]. Because the therapeutic target of CRT is correction of mechanical dyssynchrony, considerable effort has been directed toward imaging markers that may refine patient evaluation beyond electrocardiographic criteria alone [7,8].

Apical rocking (AR) and septal flash (SF) are echocardiographic manifestations of left ventricular (LV) dyssynchrony. AR is characterized by early septal contraction followed by delayed posterolateral wall contraction, producing a side-to-side rocking motion of the apex [9]. Both AR and SF have been associated with CRT response and prognosis [10–14]. However, visual assessment remains qualitative and observer-dependent, making consistent recognition difficult for inexperienced readers. Quantitative approaches based on speckle-tracking echocardiography, including systolic stretch index and peak longitudinal displacement, have been developed to characterize these phenomena, but the requirement for dedicated post-processing limits routine use [12].

Deep learning (DL), particularly convolutional neural networks (CNNs), has become a powerful approach for medical image analysis [15–17]. In echocardiography, DL models have been developed for automated disease recognition, cardiac chamber segmentation, and detection of regional wall motion abnormalities [18–21]. Recent studies have extended video-based and multiview approaches to automated functional assessment and comprehensive echocardiographic interpretation [22,23]. Many earlier models, however, were developed using static images. AR depends on both the spatial pattern of contraction and the temporal sequence of motion throughout the cardiac cycle, making video-based analysis more appropriate. Qu et al. reported automatic recognition of SF in patients with complete left bundle branch block (LBBB) using an echocardiographic video-based framework [24]. In contrast, automated recognition of AR remains underexplored.

The aim of this study was to develop and validate a DL system for automated detection of AR from routine apical four-chamber echocardiographic videos. This approach was motivated by the clinical concept that correction of LV dyssynchrony may also correct the characteristic rocking motion of the apex [25]. The present proof-of-concept study evaluated agreement with expert-defined visual AR in a selected CRT-evaluation cohort; it was not designed to predict CRT response, reverse remodeling, hospitalization, or survival. We present this article in accordance with the Standards for Reporting Diagnostic Accuracy Studies (STARD) reporting checklist.

2 Materials and Methods

2.1 Study Population

This retrospective two-center proof-of-concept study analyzed echocardiographic data from patients with HF who underwent CRT evaluation between January 2016 and August 2022 at Zhongshan Hospital of Fudan University and the First Affiliated Hospital of Wenzhou Medical University. A total of 351 subjects were included: 100 for development and testing of the segmentation model and 251 for AR classification, comprising 114 patients with AR and 137 without AR (Fig. 1). The classification cohort was assembled for model development and evaluation and was enriched for AR relative to an unselected echocardiography workflow; it should therefore not be interpreted as a prevalence or screening cohort. The 251 classification cases were divided at the patient level into 156 training cases, 39 validation cases, and 56 external test cases, with the external test cohort drawn from the second center.

Inclusion criteria were based on contemporary CRT recommendations and included sinus rhythm, QRS duration ≥ 130 ms, left ventricular ejection fraction (LVEF) $\leq 35\%$ despite guideline-directed medical therapy, and New York Heart Association functional class II–IV [26]. Patients with atrial fibrillation, severe valvular heart disease, narrow QRS complexes, previous cardiac surgery, or poor echocardiographic image

quality were excluded. No prespecified ordinal image-quality score was recorded; consequently, performance could not be stratified by acoustic window, apical foreshortening, shadowing, or lateral-wall visibility. Demographic and electrocardiographic data were retrieved from institutional electronic medical records.

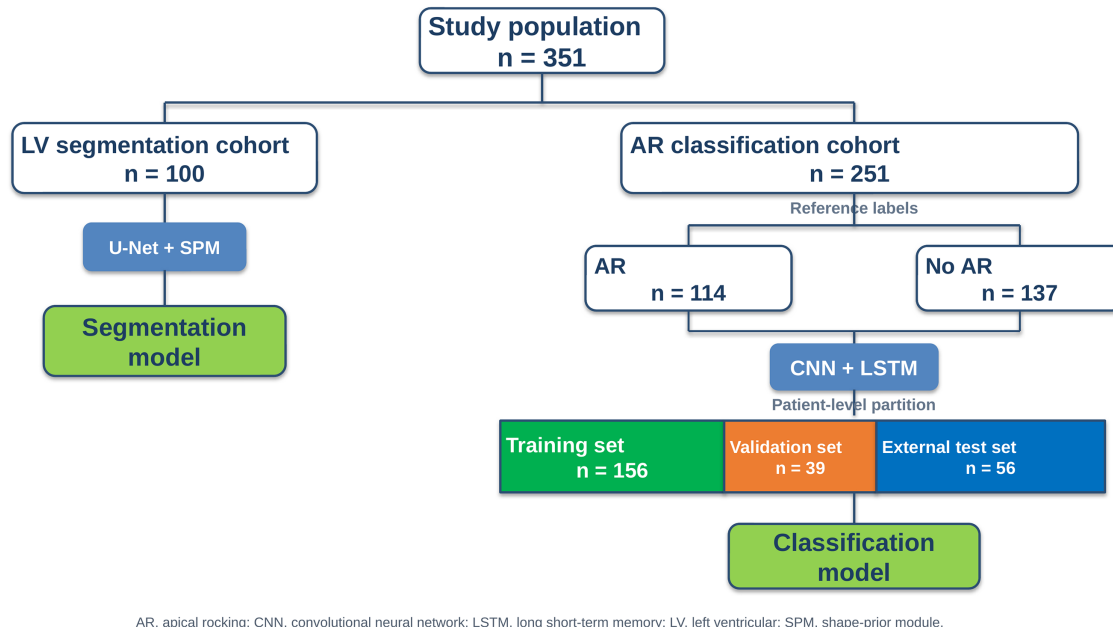


Figure 1: Study flow diagram. Of the 351 included subjects, 100 were allocated to development and testing of the U-Net with shape-prior-module left ventricular segmentation model, and 251 were allocated to apical-rocking classification. The classification cohort comprised 114 cases with apical rocking and 137 without apical rocking and was divided at the patient level into training ($n = 156$), validation ($n = 39$), and external test ($n = 56$) sets for development and evaluation of the convolutional neural network with long short-term memory classifier. AR, apical rocking; CNN, convolutional neural network; LSTM, long short-term memory; LV, left ventricular; SPM, shape-prior module.

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of Fudan University (No. FE22041R) and the Ethics Committee of the First Affiliated Hospital of Wenzhou Medical University (No. 2019-028). Written informed consent was obtained from all participants.

2.2 Echocardiography

2.2.1 Image Acquisition and Conventional Measurements

Two-dimensional transthoracic echocardiography was performed using GE Vivid E9 or E95 ultrasound systems equipped with an M5Sc transducer. All examinations were acquired with the patient in the left lateral decubitus position and with simultaneous electrocardiographic recording. Images were obtained in accordance with the 2015 American Society of Echocardiography recommendations for adult transthoracic echocardiography [27]. The frame rate was optimized and imaging depth minimized when possible. At least three cardiac cycles were recorded for each subject. Left ventricular end-diastolic volume, left ventricular end-systolic volume, and LVEF were measured using the biplane Simpson method.

2.2.2 Visual Assessment of Apical Rocking

AR was assessed visually by five echocardiographers with 5–15 years of experience. Two readers independently reviewed each examination. In cases of disagreement, adjudication was performed by a third reader or, when necessary, by group consensus. Consensus classification served as the reference standard for AR. The initial reader-specific labels and confidence scores were not retained in a form permitting retrospective calculation of pre-adjudication agreement or consensus strength. In the apical four-chamber view, AR was defined as early systolic septal contraction that displaced the apex toward the septum, followed by delayed lateral-wall contraction during ejection that displaced the apex toward the lateral wall while the septum stretched (Fig. 2).

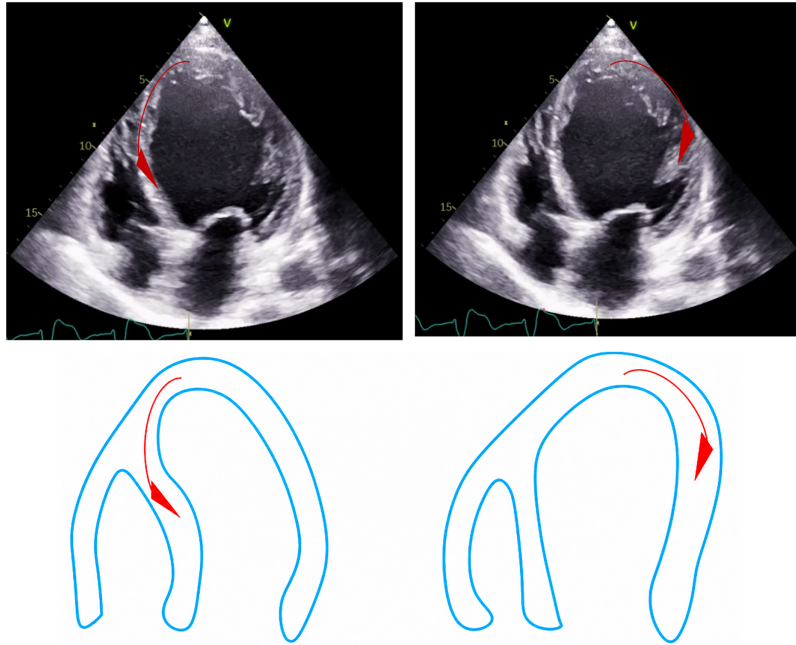


Figure 2: Echocardiographic images (upper panels) and corresponding schematics (lower panels) illustrating apical rocking in the apical four-chamber view. Early septal contraction (left) displaces the apex toward the septum, whereas delayed lateral-wall contraction later in systole (right) displaces the apex toward the lateral wall while the septum stretches. Red arrows indicate the direction of apical motion.

2.3 Deep-Learning Model Development

2.3.1 Data Preprocessing and Overall Framework

All echocardiographic videos were exported in Digital Imaging and Communications in Medicine (DICOM) format and de-identified before analysis. Videos were manually checked to ensure complete removal of patient identifiers. Regions outside the scan sector were masked, and extraneous markers that might confound the neural network were removed, as recommended for reducing non-anatomic shortcuts in echocardiographic classification tasks [28]. Each input video was verified to contain one complete cardiac cycle. Pixel intensities were normalized by the maximum pixel value, and frames were resized to 256×256 pixels using bilinear interpolation. For classification, 16 frames were sampled at equal intervals across each cardiac cycle.

The framework consisted of two steps: (I) LV endocardial segmentation using a U-Net-based model incorporating a shape-prior module (SPM); and (II) AR classification using temporal features extracted from segmented echocardiographic videos.

2.3.2 Left Ventricular Endocardial Segmentation

Manual endocardial contours were generated by echocardiographers using Labelme software in 100 HF subjects. These cases were divided into training and test sets in an 8:2 ratio. Segmentation was performed using a U-Net architecture. To improve contour accuracy, an SPM was incorporated into the U-Net. The SPM interacted with multi-scale features and refined the encoder output at different levels (Figs. 3 and 4).

The U-Net comprised a contracting path and an expansive path. The contracting path consisted of repeated 3×3 convolutions without padding, each followed by a rectified linear unit, together with 2×2 max-pooling layers for downsampling. At each downsampling step, the number of feature channels doubled. The expansive path performed upsampling followed by 2×2 up-convolution, halving the number of channels and concatenating the result with the corresponding cropped feature map from the contracting path. Two additional 3×3 convolutions were then applied. A final 1×1 convolution mapped the 64-channel feature representation to the desired output classes.

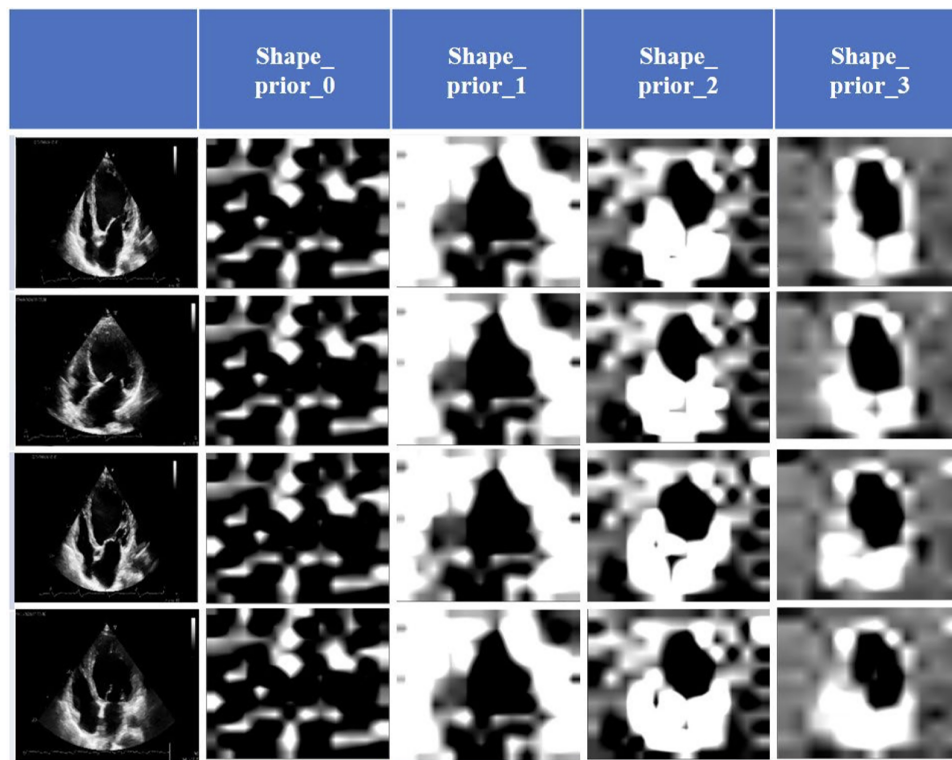


Figure 3: Multiscale interaction of the shape-prior module with encoder features. Columns show the input image and learned prior representations at successive scales.

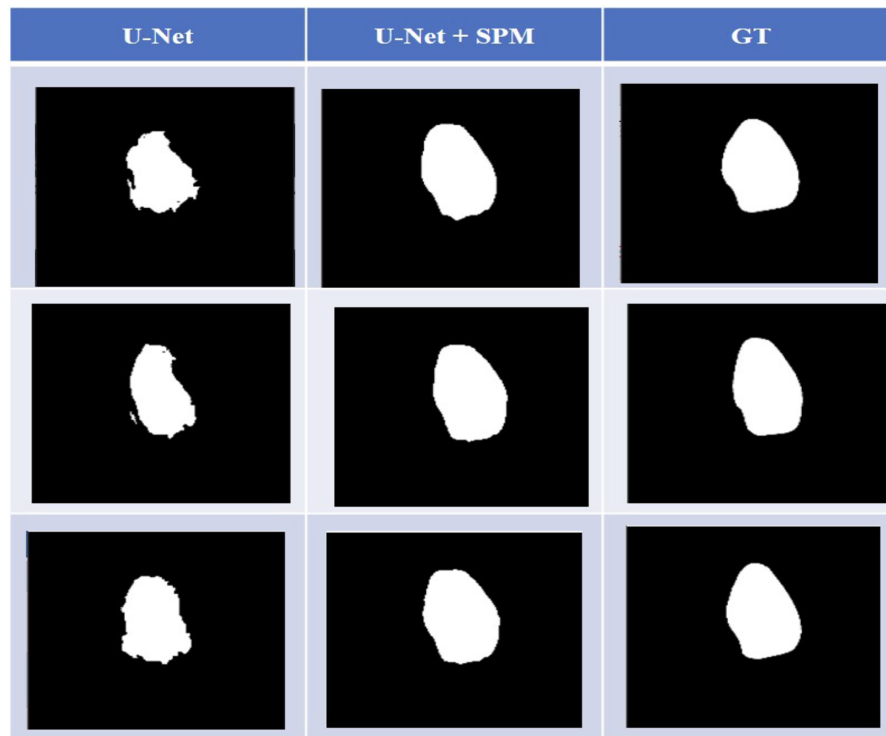


Figure 4: Representative binary masks produced by U-Net and U-Net with the shape-prior module compared with manual ground truth. The shape-prior-enhanced model produced smoother and more anatomically continuous contours. GT, ground truth; SPM, shape-prior module.

The SPM was used as a plug-in module to introduce explicit trainable shape information into the U-Net skip connections, following the general concept of learnable shape priors [29]. Four scale-specific ShapePriorBlocks were attached to the four encoder skip levels. In each block, the prior was a trainable tensor of size $1 \times 64 \times 1 \times 1$, initialized from a standard normal distribution and optimized end-to-end rather than derived from an atlas, mean contour, or statistical shape model. The tensor was expanded by bilinear interpolation to the spatial dimensions of the corresponding skip feature and refined by a two-layer 3×3 convolutional self-update module. The updated prior was concatenated with the skip feature and fused through a 1×1 convolution; the fusion output was then added residually to the original skip feature before transfer to the decoder. Each scale therefore retained its own learnable prior and provided scale-specific anatomic regularization. Segmentation performance was assessed using the Dice coefficient, Hausdorff distance, and Jaccard coefficient [30]. Dice and Jaccard coefficients were calculated for each evaluated predicted-reference binary-mask pair using $\text{Dice} = 2\text{TP}/(2\text{TP} + \text{FP} + \text{FN})$ and $\text{Jaccard} = \text{TP}/(\text{TP} + \text{FP} + \text{FN})$, respectively, and were summarized as mean $\pm$ standard deviation across evaluated mask pairs.

2.3.3 Apical Rocking Classification

The echocardiographic data of 251 AR and non-AR patients were divided at the patient level into 156 training cases, 39 validation cases, and 56 external test cases. The external test cohort consisted of cases from another center and was not used during model development. To reduce the influence of background noise, the trained segmentation model was first applied to all cardiac cycles to generate endocardial masks. Images were cropped to a square and resized to 256×256 pixels before prediction. Only the largest connected area was retained. The predicted masks were expanded by 10 pixels to include the myocardium and blood pool.

The masks were then multiplied by the normalized images to suppress irrelevant surrounding tissues. Sixteen frames were sampled at equal intervals from each cardiac cycle and used as input to the classification model. No alternative frame-count or early-systole-dense sampling ablation was performed.

The primary classification architecture was a cascaded CNN and long short-term memory (LSTM) network. A ResNet backbone was used to extract spatial features from individual frames, and the LSTM module was used to model temporal information across the video sequence (Fig. 5).

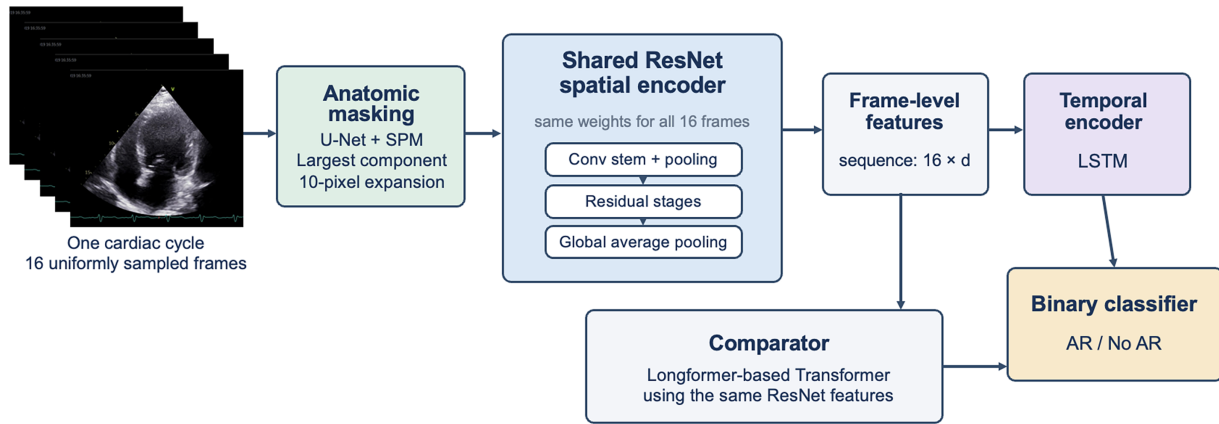


Figure 5: Apical-rocking classification pipeline. Sixteen uniformly sampled frames underwent U-Net with shape-prior-module anatomic masking. A shared ResNet encoded each frame, and a long short-term memory network aggregated the 16-frame feature sequence for binary apical-rocking/no-apical-rocking classification. A Longformer-based Transformer using the same ResNet features served as the comparator. AR, apical rocking; LSTM, long short-term memory; SPM, shape-prior module.

A cascaded CNN and Transformer model was also evaluated as a comparator. In this architecture, ResNet was again used for spatial feature extraction, and temporal modeling was performed using a Longformer-based Transformer module. The primary CNN+LSTM classifier used a single-channel ResNet-18 spatial encoder with global average pooling to generate a 512-dimensional feature vector for each frame. The resulting 16-frame feature sequence was processed by a two-layer unidirectional LSTM with a hidden dimension of 256 and a linear two-class output layer. The deep-learning implementation used PyTorch within a Python 3.6 workflow; NumPy, Pandas, and scikit-learn were used for data handling and evaluation. The exact historical PyTorch/CUDA versions and GPU configuration were not retained. A public reference implementation of the core two-stage pipeline is available at <https://github.com/iimperial/AR>. Model development used the training and validation sets only, and final diagnostic performance was assessed in the external test set.

2.4 Statistical Analysis

Continuous variables are presented as mean $\pm$ standard deviation, and categorical variables as counts and percentages. Between-group comparisons were performed using appropriate parametric or non-parametric tests for continuous variables and chi-square or Fisher exact tests for categorical variables. Diagnostic performance was evaluated using accuracy, sensitivity, specificity, precision (positive predictive value; PPV), negative predictive value (NPV), F1-score, and area under the receiver-operating characteristic curve (AUC). PPV, NPV, and F1-score were derived from the external-test confusion matrices; 95% confidence intervals for binomial proportions, including PPV and NPV, were calculated using the Wilson method, whereas F1-score was reported as a point estimate. Because PPV and NPV are prevalence-dependent, they

are interpreted only within the case-enriched external test cohort. The operating threshold was selected by maximizing Youden's index on the validation-set ROC curve and was then applied unchanged to the external test set; no cohort-specific threshold tuning was performed on the external test cohort. Calibration metrics, Brier scores, and decision-curve analysis were not reported because the deliberately case-enriched classification cohort did not reflect real-world AR prevalence; estimates derived from this artificial case mix would be strongly influenced by cohort composition and could mislead real-world probability-calibration and net-benefit interpretation. After model evaluation, the five discordant external-test cases were reviewed qualitatively by the clinical coauthors to identify the dominant reason for disagreement. This review was exploratory, *post hoc*, and not blinded. No formal pre-adjudication interobserver-agreement analysis was performed. Statistical analyses were performed using IBM SPSS Statistics version 22.0 and MedCalc version 12.5.0. A two-sided p value < 0.05 was considered statistically significant. The separate physician reader provided only binary AR/no-AR classifications and no continuous or ordinal confidence scores. Consequently, the reader assessment represented a single operating point rather than a threshold-varying diagnostic score, and a conventional physician ROC curve, physician AUC, and DeLong comparison were not retained in the revised analysis.

3 Results

3.1 Clinical Characteristics

A total of 351 subjects were included, comprising 100 HF subjects for the segmentation model and 251 HF patients for AR classification. In the classification cohort, 114 patients had AR and 137 did not. The mean age was 66 ± 7 years in the AR group and 65 ± 9 years in the non-AR group. Compared with the non-AR group, patients with AR had a higher proportion of female sex and a longer QRS duration. Left ventricular end-diastolic volume, left ventricular end-systolic volume, and LVEF did not differ significantly between groups (Table 1).

Table 1: Clinical characteristics of patients with and without apical rocking.

Variable	AR Group (n = 114)	Non-AR Group (n = 137)	p Value
Age (years)	66 ± 7	65 ± 9	0.361
Female sex, n (%)	52 (46%)	23 (17%)	<0.001
QRS duration (ms)	184 ± 20	178 ± 18	0.016
LVEDV (mL)	230 ± 67	227 ± 103	0.791
LVESV (mL)	173 ± 59	172 ± 92	0.961
LVEF (%)	26 ± 6	27 ± 8	0.281

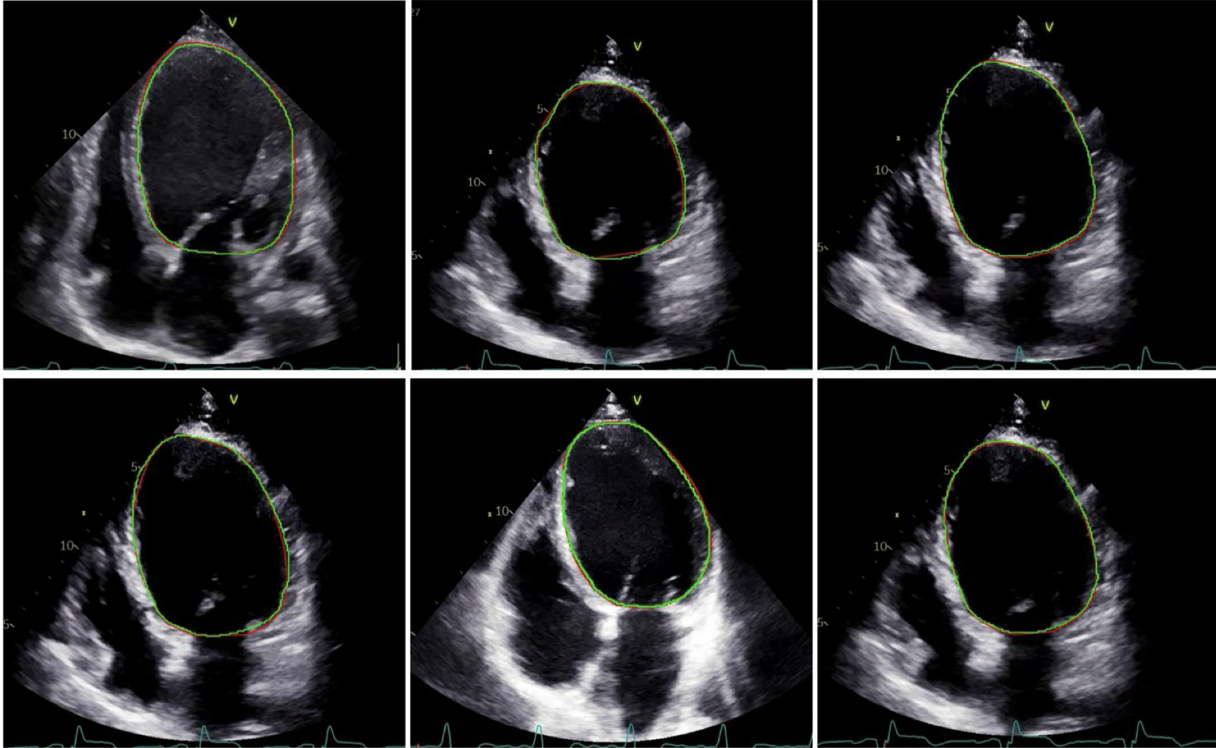
Note: AR, apical rocking; LVEDV, left ventricular end-diastolic volume; LVEF, left ventricular ejection fraction; LVESV, left ventricular end-systolic volume. p values are two-sided and are reported to three decimal places, with values <0.001 shown as <0.001 .

3.2 Left Ventricular Segmentation Performance

In the segmentation training set, the Dice coefficient was 0.994 ± 0.005 , the Hausdorff distance was 2.059 ± 0.534 mm, and the Jaccard coefficient was 0.991 ± 0.006 . In the held-out test set, the corresponding values were 0.956 ± 0.037 , 3.543 ± 0.924 mm, and 0.918 ± 0.053 , respectively, indicating high overlap between automated and manual LV endocardial segmentation (Table 2, Fig. 6).

Table 2: Performance of left ventricular segmentation.

Dataset	Dice Coefficient	Hausdorff Distance	Jaccard Coefficient
Training set	0.994 ± 0.005	2.059 ± 0.534 mm	0.991 ± 0.006
Held-out test set	0.956 ± 0.037	3.543 ± 0.924 mm	0.918 ± 0.053

**Figure 6:** Representative left ventricular segmentation results. Red contours indicate manual reference contours, and green contours indicate model predictions across examples with variable ventricular size and morphology.

3.3 Model Training and Classification Performance

Training and validation loss curves for the CNN+LSTM model are shown in Fig. 7. The model was trained for 200 epochs, after which the training and validation losses had stabilized.

The CNN+LSTM model achieved an accuracy of 0.97, sensitivity of 0.99, specificity of 0.94, and AUC of 0.98 in the training set. In the validation set, the corresponding values were 0.92, 0.91, 0.94, and 0.98. In the external test set, which included 26 AR and 30 non-AR cases, the model achieved an accuracy of 0.91, sensitivity of 0.96, specificity of 0.87, and AUC of 0.95 (Table 3). Based on the confusion matrix, the external-test accuracy was 51/56 (95% confidence interval [CI], 0.81–0.96), sensitivity was 25/26 (95% CI, 0.81–0.99), and specificity was 26/30 (95% CI, 0.70–0.95). The PPV was 25/29 (0.86; 95% CI, 0.69–0.95), the NPV was 26/27 (0.96; 95% CI, 0.82–0.99), and the F1-score was 0.91 (Table 4). The five discordant classifications comprised one false negative and four false positives. Exploratory *post hoc* review by the clinical coauthors indicated that the false-negative case was affected by an image-quality limitation, whereas all four false-positive cases showed nonspecific apical translation rather than the characteristic apical-rocking motion sequence.

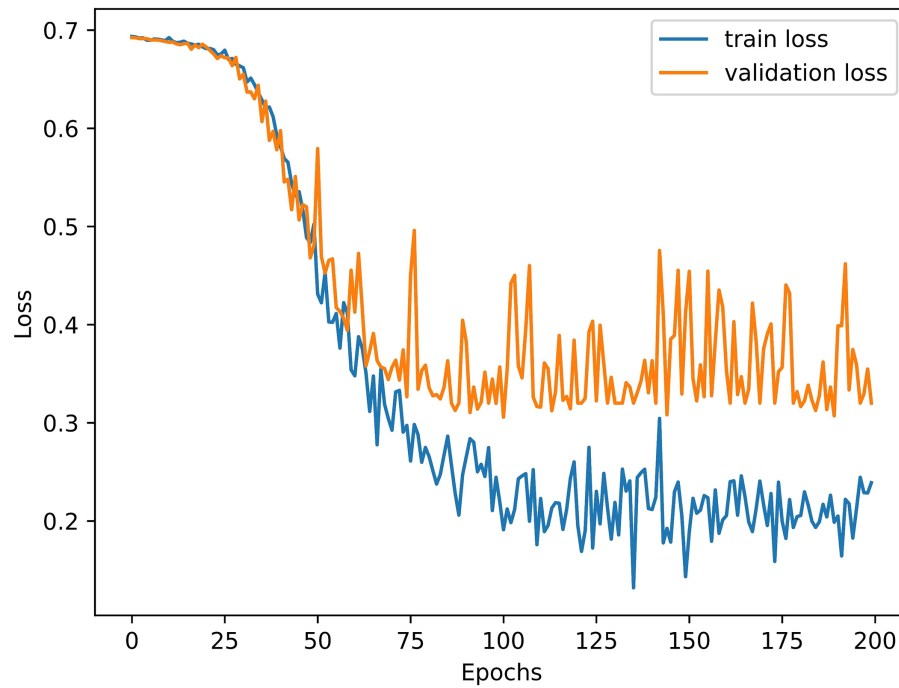


Figure 7: Training and validation loss curves over 200 epochs for the convolutional neural network and long short-term memory classification model.

Table 3: Performance of the CNN+LSTM classification model.

Dataset	Accuracy	Sensitivity	Specificity	AUC
Training set	0.97	0.99	0.94	0.98
Validation set	0.92	0.91	0.94	0.98
External test set	0.91	0.96	0.87	0.95

Note: AUC, area under the receiver-operating characteristic curve; CNN, convolutional neural network; LSTM, long short-term memory.

Table 4: Confusion matrix of the CNN+LSTM model in the external test set.

	Actual AR	Actual non-AR
Predicted AR	25	4
Predicted non-AR	1	26

Note: Accuracy = 0.91; precision/positive predictive value = 0.86 (95% CI, 0.69–0.95); negative predictive value = 0.96 (95% CI, 0.82–0.99); F1-score = 0.91. AR, apical rocking; CI, confidence interval; CNN, convolutional neural network; LSTM, long short-term memory.

The CNN+LSTM model was compared with a CNN+Transformer model in the external test set. The CNN+LSTM model achieved numerically higher overall accuracy than the CNN+Transformer model (0.91 vs. 0.88). The CNN+Transformer model achieved a sensitivity of 0.85, specificity of 0.90, and AUC of 0.93. Its PPV was 22/25 (0.88; 95% CI, 0.70–0.96), NPV was 27/31 (0.87; 95% CI, 0.71–0.95), and F1-score was 0.86 (Tables 3 and 5).

Table 5: Confusion matrix of the CNN+Transformer model in the external test set.

	Actual AR	Actual non-AR
Predicted AR	22	3
Predicted non-AR	4	27

Note: Accuracy = 0.88; precision/positive predictive value = 0.88 (95% CI, 0.70–0.96); negative predictive value = 0.87 (95% CI, 0.71–0.95); F1-score = 0.86. AR, apical rocking; CI, confidence interval; CNN, convolutional neural network.

4 Discussion

In this retrospective two-center proof-of-concept study, we developed a two-stage deep-learning pipeline for automated detection of apical rocking from apical four-chamber echocardiographic videos. The framework combined LV segmentation with temporal sequence classification and demonstrated promising diagnostic performance in a small external test cohort from a second center. These results establish technical feasibility within a selected CRT-evaluation cohort, but they do not establish screening performance, clinical effectiveness, or outcome prediction.

The potential value of the proposed approach is to standardize an assessment that is currently qualitative and experience-dependent, not to replace clinical judgment. AR and SF are established manifestations of dyssynchronous ventricular contraction [8,9]. They have been associated with CRT response, remodeling, and prognosis [10–12,31,32]. Recent studies continue to support the prognostic and treatment-response relevance of mechanical dyssynchrony in CRT recipients [13,14]. The current model, however, was trained to reproduce expert-defined visual AR only; it was not evaluated as a predictor of CRT response, reverse remodeling, hospitalization, or survival.

The present study also highlights the importance of analyzing echocardiographic video rather than isolated still frames. AR is an inherently dynamic phenomenon: the diagnostic pattern is defined by early septal motion followed by delayed lateral-wall contraction across systole. A static image cannot adequately represent this temporal sequence. By combining a CNN for frame-level spatial feature extraction with an LSTM for temporal modeling, the proposed architecture was designed to capture both ventricular morphology and motion evolution across the cardiac cycle. In the present dataset, this strategy slightly outperformed the Transformer-based comparator, which may reflect the relatively short video sequences used for classification. This should not be interpreted as general superiority of one temporal architecture over another. The use of spatiotemporal modeling is consistent with earlier motion-atlas approaches and more recent video-based echocardiographic artificial-intelligence systems [22,23,33–35].

Classification used fixed uniform sampling of 16 frames per cardiac cycle. Because AR is driven by brief early-systolic septal-to-lateral transitions, uniform downsampling may under-represent high-frequency, short-duration motion events. We did not perform an ablation across alternative frame counts or early-systole-dense sampling schemes; therefore, the optimal temporal sampling strategy for automated AR detection remains undetermined.

A further strength of the pipeline is the explicit segmentation step before classification. Echocardiographic videos contain non-anatomic information, including near-field artifacts, extracardiac tissue, sector boundaries, annotations, and variable image depth. Segmenting the LV and masking surrounding content encouraged the downstream classifier to focus on the anatomic region where AR occurs. The held-out segmentation results showed high overlap with manual contours. Nevertheless, poor-quality examinations were excluded, and no ordinal image-quality score was recorded. The present study therefore cannot establish robustness in suboptimal acoustic windows, apical foreshortening, shadowing, or inadequate lateral-wall visualization, all of which may propagate error through both segmentation and classification.

The external test cohort provides an initial estimate of transportability beyond the development data, but broader generalizability remains uncertain. The classification cohort was enriched for AR and non-AR cases rather than drawn consecutively from an unselected echocardiography population. Accordingly, the reported accuracy, sensitivity, specificity, and AUC characterize discrimination in this selected CRT-evaluation cohort and should not be interpreted as screening performance or workflow utility in routine echocardiography. Both centers used GE Vivid E9 or E95 systems from the same vendor family, so performance across other vendors, acquisition protocols, operators, and clinical settings remains unknown.

Consensus expert visual assessment reflects current clinical practice and is a reasonable reference standard for this proof-of-concept task, but it is not a perfect biological ground truth. Although two readers independently reviewed each examination before adjudication, the initial reader-specific labels and confidence scores were not retained in an analyzable form. Pre-adjudication agreement, probabilistic labels, and consensus strength therefore could not be reconstructed. This limitation prevents quantification of label uncertainty and precludes any claim that model-reference discrepancies fall within expected inter-reader variability.

The external test confusion matrix contained one false negative and four false positives. Exploratory *post hoc* review identified two failure patterns: the single false-negative case was affected by an image-quality limitation, whereas all four false-positive cases showed nonspecific apical translation rather than true AR. These findings suggest that residual image-quality limitations may reduce sensitivity and that global or nonspecific apical motion may be mistaken for the characteristic early-septal-to-delayed-lateral AR sequence. Because the review was not prespecified or blinded and included only five cases, these observations should be regarded as descriptive and hypothesis-generating. Future studies should use a prespecified blinded error-review protocol and evaluate whether explicit motion features or multiview information can distinguish true dyssynchronous rocking from nonspecific apical translation.

To improve reproducibility, we released a public reference implementation of the two-stage pipeline under the MIT License at <https://github.com/iimperial/AR>. The repository provides code for Digital Imaging and Communications in Medicine preprocessing, scan-sector masking and intensity normalization, uniform 16-frame sampling, U-Net with shape-prior-module segmentation, largest-connected-component and 10-pixel mask expansion, ResNet-18 with long short-term memory classification, inference and post-processing, and segmentation and classification evaluation. Clinical imaging data and trained model checkpoints are not publicly released, and the exact historical package versions and hardware configuration were not retained. Accordingly, the repository enables inspection and reimplementing of the core methodology but does not by itself permit exact numerical reproduction of the reported results.

From a workflow perspective, a prospectively validated model could potentially provide a consistent second read in laboratories evaluating mechanical dyssynchrony. The current study did not test whether model output changes reader confidence, reporting consistency, interpretation time, treatment decisions, or clinical outcomes. Nor did it establish incremental value beyond guideline-based CRT criteria. Longitudinal validation should link automated AR detection to implantation status, reverse remodeling, changes in LVEF, heart-failure hospitalization, and survival, while implementation studies should compare unaided and AI-assisted interpretation.

Model interpretability is also relevant for clinical adoption. Although the segmentation mask provides an anatomical constraint, the classification network remains only partially explainable. Visualization of salient temporal frames, attention to apical and septal regions, motion-vector analysis, or comparison with quantitative displacement curves could increase user trust and help identify failure modes. The operating threshold was selected by maximizing Youden's index on the validation-set ROC curve and was applied unchanged to the external test set. However, discrimination at a fixed operating point does not establish

clinical calibration. Because the classification cohort was deliberately enriched for AR and non-AR cases, its predicted-probability distribution does not reflect real-world prevalence. Calibration metrics, the Brier score, and decision-curve analysis were therefore not reported, because estimates derived from this artificial case mix would not be directly transportable to a prevalence-representative clinical population and could mislead real-world threshold selection. Prospective studies should retain patient-level probabilities and evaluate calibration, Brier score, and net benefit in consecutive prevalence-representative cohorts.

This study has several limitations. First, it was retrospective and based on a relatively modest sample size. Second, the external test set contained only 56 patients and used the same ultrasound vendor family as the development center. Third, poor-quality examinations were excluded and no ordinal image-quality scale was available. Fourth, the classification cohort was case-enriched and does not estimate performance in an unselected screening population. Fifth, the initial independent reader labels and reader confidence scores were unavailable, and the discordant-case review was exploratory, *post hoc*, not blinded, and limited to five cases. Sixth, although a public reference implementation is available, the clinical data, trained model checkpoints, exact historical software versions, and hardware configuration are not public, limiting exact numerical reproduction. Seventh, calibration metrics, Brier score, and decision-curve analysis were not reported because the case-enriched cohort was not prevalence-representative; prospective calibration and net-benefit evaluation are required before clinical use. Eighth, classification used fixed uniform sampling of 16 frames per cardiac cycle; because AR contains brief early-systolic septal-to-lateral transitions, this approach may under-sample short-duration motion events, and no alternative frame-count or phase-aware ablation was performed. Ninth, the study relied on a single echocardiographic view and did not integrate complementary markers such as SF, strain-derived dyssynchrony indices, electrocardiographic morphology, or scar assessment. Finally, the study evaluated expert-defined visual AR rather than CRT response or clinical outcomes.

Future investigations should proceed in three stages. Technical validation should use consecutive, multi-vendor, multicenter cohorts with prespecified image-quality strata, retained reader-level labels, patient-level model probabilities, validation-derived operating thresholds applied without retuning to external cohorts, versioned code releases with trained checkpoints and complete software environments, and prespecified blinded error analysis that explicitly distinguishes true AR from nonspecific apical translation. Prespecified temporal ablation should compare alternative frame counts (for example, 8, 16, and 32 frames), electrocardiogram- or phase-normalized sampling, and denser sampling around early systole while maintaining identical data splits. Calibration, Brier score, and decision-curve analysis should then be performed in prevalence-representative cohorts. Clinical validation should test whether automated AR detection adds incremental value to established CRT criteria and predicts reverse remodeling and long-term outcomes. Implementation studies should determine whether artificial-intelligence assistance improves inter-reader agreement, reporting consistency, reader confidence, and interpretation time without introducing automation bias.

5 Conclusions

This two-center proof-of-concept study demonstrates the technical feasibility of segmentation-constrained spatiotemporal deep learning for detecting expert-defined apical rocking from apical four-chamber echocardiographic videos. In the small external test cohort, the CNN+LSTM model achieved an accuracy of 0.91, sensitivity of 0.96, specificity of 0.87, PPV of 0.86, NPV of 0.96, F1-score of 0.91, and AUC of 0.95. These findings support further investigation of automated AR assessment as a reader-support method in patients undergoing CRT evaluation, but they do not establish performance in an unselected screening population, improved CRT selection, or prediction of therapeutic response and clinical

outcomes. Generalization is limited by the retrospective case-enriched design, exclusion of poor-quality examinations, small external cohort, lack of pre-adjudication agreement, exploratory *post hoc* discordant-case review, absence of prevalence-representative calibration and decision-curve analysis, fixed 16-frame sampling without temporal ablation, and limited exact reproducibility because clinical data and trained checkpoints are not publicly available despite release of reference code. Prospective validation in consecutive, multi-vendor cohorts—including suboptimal-quality acquisitions, prevalence-representative calibration, prespecified threshold evaluation, temporal-sampling ablation, and longitudinal CRT outcomes—is required before clinical implementation.

Acknowledgement: None.

Funding Statement: This work was supported by the National Natural Science Foundation of China (grant numbers 82471997, 82227803, 82260350), National Key Research and Development Program of China, Key Special Project “Frontier Biotechnology” (grant number 2024YFC3406702), Guizhou Medical University Affiliated Hospital Doctoral Research Start-up Fund Project (gyfybsky-2025-27), and Clinical Special Project of the Health Commission of Shanghai (202440047).

Author Contributions: The authors confirm contribution to the paper as follows: conception and design, Fangyan Tian, Yu Liu and Xianhong Shu; administrative support, Yi Guo, Yuanyuan Wang and Xianhong Shu; provision of study materials or patients, Fangyan Tian, Yu Liu, Yuman Li, Haiyan Chen and Xianhong Shu; collection and assembly of data, Fangyan Tian, Yu Liu, Yuman Li, Mingshan Li and Haiyan Chen; data analysis and interpretation, Fangyan Tian, Yu Liu, Yuman Li, Mingshan Li, Guohui Zhou, Yi Guo, Yuanyuan Wang and Xianhong Shu; manuscript writing, all authors. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The clinical echocardiographic data underlying this article are not publicly available because they contain potentially identifiable clinical imaging data and are not currently available for external sharing. A public reference implementation of the two-stage pipeline is available under the MIT License at <https://github.com/iiimperial/AR>. The repository contains code for preprocessing, U-Net with shape-prior-module segmentation, the convolutional neural network with long short-term memory model definition, inference and post-processing, and performance evaluation. Trained model checkpoints and clinical data are not included; consequently, the repository supports methodological inspection and reimplementations but not exact reproduction of the reported numerical results.

Ethics Approval: The study involved human participants and was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of Fudan University (No. FE22041R) and the Ethics Committee of the First Affiliated Hospital of Wenzhou Medical University (No. 2019-028). Written informed consent was obtained from all participants.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AR	Apical rocking
AUC	Area under the receiver-operating characteristic curve
CNN	Convolutional neural network
CRT	Cardiac resynchronization therapy
DICOM	Digital Imaging and Communications in Medicine
DL	Deep learning
HF	Heart failure
LBBB	Left bundle branch block
LSTM	Long short-term memory

LV	Left ventricular
LVEDV	Left ventricular end-diastolic volume
LVEF	Left ventricular ejection fraction
LVESV	Left ventricular end-systolic volume
NPV	Negative predictive value
PPV	Positive predictive value
SF	Septal flash
SPM	Shape-prior module

References

- Emmons-Bell S, Johnson C, Roth G. Prevalence, incidence and survival of heart failure: a systematic review. *Heart*. 2022;108(17):1351–60. doi:10.1136/heartjnl-2021-320131.
- Herweg B, Welter-Frost A, Vijayaraman P. The evolution of cardiac resynchronization therapy and an introduction to conduction system pacing: a conceptual review. *Europace*. 2021;23(4):496–510. doi:10.1093/europace/euaa264.
- Daubert C, Behar N, Martins RP, Mabo P, Leclercq C. Avoiding non-responders to cardiac resynchronization therapy: a practical guide. *Eur Heart J*. 2016;32:ehw270. doi:10.1093/eurheartj/ehw270.
- Zhu M, Wang Y, Cheng Y, Su Y, Chen H, Shu X. The value of non-invasive myocardial work indices derived from left ventricular pressure-strain loops in predicting the response to cardiac resynchronization therapy. *Quant Imaging Med Surg*. 2021;11(4):1406–20. doi:10.21037/qims-20-754.
- Bazoukis G, Hui JMH, Lee YHA, Chou OHI, Sfairopoulos D, Vlachos K, et al. The role of cardiac magnetic resonance in identifying appropriate candidates for cardiac resynchronization therapy—a systematic review of the literature. *Heart Fail Rev*. 2022;27(6):2095–118. doi:10.1007/s10741-022-10263-5.
- Bax JJ, Bleeker GB, Marwick TH, Molhoek SG, Boersma E, Steendijk P, et al. Left ventricular dyssynchrony predicts response and prognosis after cardiac resynchronization therapy. *J Am Coll Cardiol*. 2004;44(9):1834–40. doi:10.1016/j.jacc.2004.08.016.
- Villegas-Martinez M, Krogh MR, Andersen Ø.S, Sletten OJ, Wajdan A, Odland HH, et al. Tracking early systolic motion for assessing acute response to cardiac resynchronization therapy in real time. *Front Physiol*. 2022;13:903784. doi:10.3389/fphys.2022.903784.
- Calle S, Delens C, Kamoen V, De Pooter J, Timmermans F. Septal flash: at the heart of cardiac dyssynchrony. *Trends Cardiovasc Med*. 2020;30(2):115–22. doi:10.1016/j.tcm.2019.03.008.
- Voigt JU, Schneider TM, Korder S, Szulik M, Gürel E, Daniel WG, et al. Apical transverse motion as surrogate parameter to determine regional left ventricular function inhomogeneities: a new, integrative approach to left ventricular asynchrony assessment. *Eur Heart J*. 2009;30(8):959–68. doi:10.1093/eurheartj/ehp062.
- Stankovic I, Prinz C, Ciarka A, Daraban AM, Kotrc M, Aaronson M, et al. Relationship of visually assessed apical rocking and septal flash to response and long-term survival following cardiac resynchronization therapy (PREDICT-CRT). *Eur Heart J Cardiovasc Imaging*. 2016;17(3):262–9. doi:10.1093/ehjci/jev288.
- Beela AS, Ünlü S, Duchenne J, Ciarka A, Daraban AM, Kotrc M, et al. Assessment of mechanical dyssynchrony can improve the prognostic value of guideline-based patient selection for cardiac resynchronization therapy. *Eur Heart J Cardiovasc Imaging*. 2019;20(1):66–74. doi:10.1093/ehjci/jeu029.
- Mada RO, Lysyansky P, Duchenne J, Beyer R, Mada C, Muresan L, et al. New automatic tools to identify responders to cardiac resynchronization therapy. *J Am Soc Echocardiogr*. 2016;29(10):966–72. doi:10.1016/j.echo.2016.06.008.
- Layec J, Decroocq M, Delelis F, Appert L, Guyomar Y, Riolet C, et al. Dyssynchrony and response to cardiac resynchronization therapy in heart failure patients with unfavorable electrical characteristics. *JACC Cardiovasc Imaging*. 2023;16(7):873–84. doi:10.1016/j.jcmg.2022.12.023.
- Ahmed MA, Gercek M, Sommer P, Rudolph V, Dumitrescu D, Faber L, et al. Echocardiographic mechanical dyssynchrony predicts long-term mortality in patients with cardiac resynchronisation therapy. *Int J Cardiovasc Imaging*. 2024;40(1):35–43. doi:10.1007/s10554-023-02972-1.
- Davis A, Billick K, Horton K, Jankowski M, Knoll P, Marshall JE, et al. Artificial intelligence and echocardiography: a primer for cardiac sonographers. *J Am Soc Echocardiogr*. 2020;33(9):1061–6. doi:10.1016/j.echo.2020.04.025.

16. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60–88. doi:10.1016/j.media.2017.07.005.
17. Krittanawong C, Johnson KW, Rosenson RS, Wang Z, Aydar M, Baber U, et al. Deep learning for cardiovascular medicine: a practical primer. *Eur Heart J.* 2019;40(25):2058–73. doi:10.1093/eurheartj/ehz056.
18. Madani A, Ong JR, Tibrewal A, Mofrad MRK. Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *npj Digit Med.* 2018;1:59. doi:10.1038/s41746-018-0065-x.
19. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, et al. Fully automated echocardiogram interpretation in clinical practice. *Circulation.* 2018;138(16):1623–35. doi:10.1161/CIRCULATIONAHA.118.034338.
20. Kusunose K, Abe T, Haga A, Fukuda D, Yamada H, Harada M, et al. A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. *JACC Cardiovasc Imaging.* 2020;13(2):374–81. doi:10.1016/j.jcmg.2019.02.024.
21. Huang MS, Wang CS, Chiang JH, Liu PY, Tsai WC. Automated recognition of regional wall motion abnormalities through deep neural network interpretation of transthoracic echocardiography. *Circulation.* 2020;142(16):1510–20. doi:10.1161/CIRCULATIONAHA.120.047530.
22. Chen X, Yang F, Zhang P, Lin X, Wang W, Pu H, et al. Artificial intelligence-assisted left ventricular diastolic function assessment and grading: multiview versus single view. *J Am Soc Echocardiogr.* 2023;36(10):1064–78. doi:10.1016/j.echo.2023.07.001.
23. Holste G, Oikonomou EK, Tokodi M, Kovács A, Wang Z, Khera R. Complete AI-enabled echocardiography interpretation with multitask deep learning. *JAMA.* 2025;334(4):306. doi:10.1001/jama.2025.8731.
24. Qu M, Wang Y, Li H, Yang J, Ma C. Automatic identification of septal flash phenomenon in patients with complete left bundle branch block. *Med Image Anal.* 2022;82:102619. doi:10.1016/j.media.2022.102619.
25. Delgado V. Correct the left ventricular dyssynchrony, correct the rocking. *Eur Heart J Cardiovasc Imaging.* 2016;17(3):270–1. doi:10.1093/ehjci/jev324.
26. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J.* 2021;42(36):3599–726. doi:10.1093/eurheartj/ehab368.
27. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American society of echocardiography and the European association of cardiovascular imaging. *J Am Soc Echocardiogr.* 2015;28(1):1–39.e14. doi:10.1016/j.echo.2014.10.003.
28. Degerli A, Zabihi M, Kiranyaz S, Hamid T, Mazhar R, Hamila R, et al. Early detection of myocardial infarction in low-quality echocardiography. *IEEE Access.* 2021;9:34442–53. doi:10.1109/ACCESS.2021.3059595.
29. You X, He J, Yang J, Gu Y. Learning with explicit shape priors for medical image segmentation. *IEEE Trans Med Imaging.* 2025;44(2):927–40. doi:10.1109/TMI.2024.3469214.
30. Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson.* 2018;20(1):65. doi:10.1186/s12968-018-0471-x.
31. Stankovic I, Prinz C, Ciarka A, Daraban AM, Mo Y, Aarones M, et al. Long-term outcome after CRT in the presence of mechanical dyssynchrony seen with chronic RV pacing or intrinsic LBBB. *JACC Cardiovasc Imaging.* 2017;10(10):1091–9. doi:10.1016/j.jcmg.2016.08.015.
32. Michalski B, Stankovic I, Pagourelas E, Ciarka A, Aarones M, Winter S, et al. Relationship of mechanical dyssynchrony and LV remodeling with improvement of mitral regurgitation after CRT. *JACC Cardiovasc Imaging.* 2022;15(2):212–20. doi:10.1016/j.jcmg.2021.08.010.
33. Duchateau N, De Craene M, Piella G, Silva E, Doltra A, Sitges M, et al. A spatiotemporal statistical atlas of motion for the quantification of abnormal myocardial tissue velocities. *Med Image Anal.* 2011;15(3):316–28. doi:10.1016/j.media.2010.12.006.

34. Laumer F, Di Vece D, Cammann VL, Würdinger M, Petkova V, Schönberger M, et al. Assessment of artificial intelligence in echocardiography diagnostics in differentiating takotsubo syndrome from myocardial infarction. *JAMA Cardiol.* 2022;7(5):494–503. doi:10.1001/jamacardio.2022.0183.
35. Xiao Z, Su Y, Deng Z, Zhang W. Efficient combination of CNN and transformer for dual-teacher uncertainty-guided semi-supervised medical image segmentation. *Comput Methods Programs Biomed.* 2022;226:107099. doi:10.1016/j.cmpb.2022.107099.