Diagnostic utility of artificial intelligence for left ventricular scar identification using cardiac magnetic resonance imaging—A systematic review

Nikesh Jathanna, BMedSci, MRCP,*‡ Anna Podlasek, MD,†‡ Albert Sokol, MBBS,‡ Dorothee Auer, MD, PhD,†‡ Xin Chen, PhD,‡ Shahnaz Jamil-Copley, MRCP, PhD*‡

*From the *Trent Cardiac Centre, Nottingham City Hospital, Nottingham University Hospitals NHS Trust, Nottingham, United Kingdom, †NIHR Nottingham Biomedical Research Centre, Queen's Medical Centre, Nottingham, United Kingdom, and ‡University of Nottingham, Nottingham, United Kingdom.*

BACKGROUND Accurate, rapid quantification of ventricular scar using cardiac magnetic resonance imaging (CMR) carries importance in arrhythmia management and patient prognosis. Artificial intelligence (AI) has been applied to other radiological challenges with success.

OBJECTIVE We aimed to assess AI methodologies used for left ventricular scar identification in CMR, imaging sequences used for training, and its diagnostic evaluation.

METHODS Following PRISMA recommendations, a systematic search of PubMed, Embase, Web of Science, CINAHL, OpenDissertations, arXiv, and IEEE Xplore was undertaken to June 2021 for full-text publications assessing left ventricular scar identification algorithms. No pre-registration was undertaken. Random-effect meta-analysis was performed to assess Dice Coefficient (DSC) overlap of learning vs predefined thresholding methods.

RESULTS Thirty-five articles were included for final review. Supervised and unsupervised learning models had similar DSC compared to predefined threshold models (0.616 vs 0.633, $P = .14$) but had higher sensitivity, specificity, and accuracy. Meta-analysis of 4 studies revealed standardized mean difference of 1.11; 95% confidence interval -0.16 to 2.38, $P = .09$, $I^2 = 98\%$ favoring learning methods.

CONCLUSION Feasibility of applying AI to the task of scar detection in CMR has been demonstrated, but model evaluation remains heterogenous. Progression toward clinical application requires detailed, transparent, standardized model comparison and increased model generalizability.

KEYWORDS Artificial intelligence; Cardiac scar; Deep learning; Imaging – cardiac magnetic resonance imaging (MRI); Machine learning; Neural networks

(Cardiovascular Digital Health Journal 2021;2:S21–S29)

Background and objectives

Accurate identification of cardiac scar is growing in importance. Previous studies demonstrate that ventricular scar volume/mass is associated with ventricular arrhythmia episode risk[1] and response to medical and device therapies.[2] Descriptive delineation can improve electrophysiological procedural outcomes.[3] Subsequently, scar metrics are increasingly integrated into clinical practice as a decision aid guiding patient therapy.[4]

Cardiac magnetic resonance imaging (CMR) is the current gold standard for tissue characterization.[5] However, expert manual delineation is time-consuming, with significant inter-operator variability. Methods to improve objective quantification through means such as thresholding and full-width half-maximum can reduce time and improve reproducibility. However, outcome variability between these methods exists and none are preferentially recommended.[5,6] Moreover, these techniques are often solely based on signal intensity and may require further human postprocessing (Figure 1).

Artificial intelligence (AI) is a broad term encompassing a multitude of functions, including the undertaking of tasks normally requiring human intelligence. Subsets including machine learning (ML) and deep learning (DL) have been investigated as solutions to clinical challenges. The use of AI in general medical and cardiac imaging research has developed rapidly owing to technological advances and the data-rich environment of imaging. Models ranging from simple task automation to deep investigative associations have been developed.[7]

Classical AI methods typically make decisions based on hand-crafted rules or predefined thresholds, whereas ML models can "learn" rules and associations from patterns statistically discerned from datasets. DL is a subcategory of

Address reprint requests and correspondence: Dr Nikesh Jathanna, Cardiology Research Office, Nottingham City Hospital, Hucknall Rd, Nottingham, Nottinghamshire, NG5 1PB, England. E-mail address: Nikesh.jathanna@nhs.net.

2666-6936/

https://doi.org/10.1016/j.cvdhj.2021.11.005

KEY FINDINGS

- Artificial intelligence models have been shown to be feasible for the segmentation of cardiac structures including scar.
- There is no clear benefit of supervised or unsupervised learning models over predefined thresholding models. Random-effect analysis may suggest a benefit of learning models, but high heterogeneity exists.
- Comparison of models to identify superior methodologies remains challenging in humans owing to a lack of true scar "ground truth" as reference, with significant variation in assessment methods utilized and limited access to final Algorithm codes.
- Standardization of model evaluation, transparency in training/testing data, and highly generalizable models are recommended before transition to clinical practice can be considered.

ML that specifically uses multilayer artificial neural networks for decision-making. Training can be supervised or unsupervised.[8] The proposed advantage of AI tools for scar identification includes rapid precision-based undertaking of routine tasks utilizing the wealth of data available with CMR.

We undertook a targeted systematic review of publications including AI methods for the identification of left ventricular (LV) cardiac scar in CMR to answer the following questions:

(1) What AI methods are being employed for ventricular scar assessment in those with cardiac disease on CMR?
(2) What CMR sequences are being utilized?
(3) What are the diagnostic evaluations of these methods?

Methods

This review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations[9] (Supplemental Appendices A and B). No preregistration was undertaken.

Search strategy and data sources

We performed a systematic search up to June 2021 of PubMed, Embase, Web of Science, Cumulative Index to Nursing and Allied Health Literature (CINAHL), OpenDissertations, arXiv, and IEEE Xplore. Full Boolean criteria appear in Supplemental Appendix C.

Records were imported into Rayyan (Qatar Computing Research Institute, Doha, Qatar) and duplicates were manually removed.

Study selection

Articles underwent independent abstract screening for eligibility and full-text review by at least 2 reviewers (NJ, AP, AS), with conflicts resolved by a third.

Full-text articles evaluating AI algorithms for LV scar identification with CMR annotations were included. Exclusion criteria included animal-based studies, algorithms primarily utilizing predefined thresholds, absence of scar annotations, non-CMR imaging, letters to editors, editorials, abstract-only publications/posters, studies not in the English language, and non-LV scar studies.

Data extraction and analysis

We extracted following variables, where available: AI Algorithm characteristics, dataset characteristics, primary cardiac disease, CMR characteristics, ground-truth assessment, and evaluation measures.

Owing to the unique complexities of individual algorithms and to enable study comparison, methodologies were broadly categorized into supervised and unsupervised methods, with further subcategorization if appreciable in the main text. The definition of "fully automated" required no human interaction from image input to result output.

CMR sequence categorization accounted for nomenclature among different manufacturers. Dataset size was assessed through total caseload, with cases defined as individual CMR scans. Primary cardiac disease was broadly categorized, with mixed-etiology studies highlighted.

Where required, conversion to mean was undertaken.[10,11] Grouped means were calculated using Cochrane's formula. Welch test was used for comparison of means. Random-effect meta-analyses of continuous primary outcome (Dice coefficient) was expressed as standardized mean difference with a 95% confidence interval. Comparison between proposed and comparator predefined thresholding methods was undertaken. Analysis was visualized by a forest plot. The extent of between-study heterogeneity was assessed with the I^2 statistic. Funnel plots were used to assess publication bias. The ROBINS-I tool was used to evaluate the risk of bias of each study included in quantitative analysis by independent assessors and the robvis tool was used for visualization.[12] P values were 2-tailed, with values $<.05$ considered statistically significant. Extracted data are available in Supplemental Appendix C.

Results

Of 6156 results, 35 were included for qualitative analysis. Four contained predefined thresholding comparators for quantitative analysis (Figure 2).

Articles included were published between the years 2010 and 2021, with increasing publications in recent years, with the highest in 2020 (n = 8).

CMR dataset

Sequences and magnet strength

All studies used CMR; however, 1 study did not specify sequences.

All 34 articles describing CMR sequences utilized late gadolinium–enhanced short-axis imaging as a minimum,

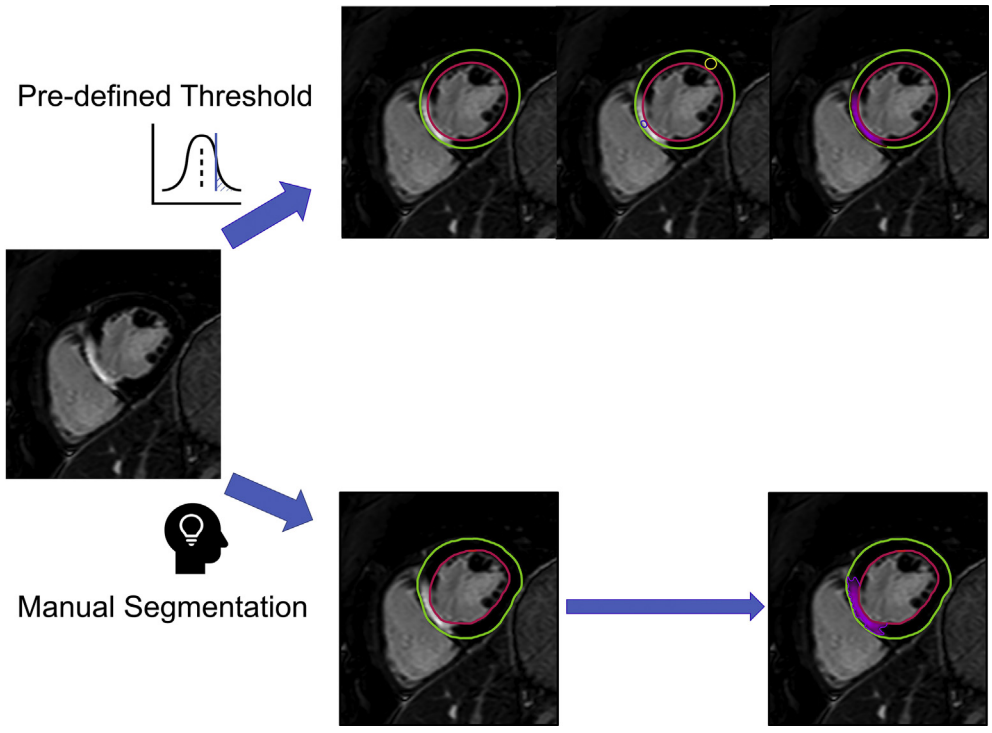

Figure 1 Simplified comparison of distinct cardiac magnetic resonance image segmentation methods. Far-left image demonstrates a short-axis view of transmural septal scar. In subsequent images, green represents epicardial border, magenta endocardial, and purple segmented scar for thresholding (*top*) and manual (*bottom*) techniques.

with most undertaken with 2D imaging and breath-holds (82.35%). Of the 6 of 34 (17.65%) utilizing 3D late gadolinium enhancement, 4 utilized 3D whole-heart imaging, with the remaining consisting of breath-hold sequences.

Additional sequences were only used in combination with 2D late gadolinium enhancement (25/28, 89.29%), with the most frequent being a variation of cine imaging.

Ten articles did not declare CMR magnet strength. Of the remaining, 12 used only 1.5 Tesla (T), 6 only 3 T, and 7 a combination.

Manufacturers

CMR manufacturer was declared in 25 of 35 articles: 59.3%, 25.9%, and 18.5% for Siemens, Philips, and General Electric, respectively. Only 5 studies used multiple manufacturers; 1 utilized all 3 manufacturers, 2 used Philips and Siemens, and 2 were unclear on designations.

Dataset size

Excluding postprocessing transformations, a total of 3856 human CMR studies in 32 of 35 articles were utilized, with a range of 3–1073 and a median (interquartile range) of 45 (24–143.5). Three articles had unclear datasets.

Nine studies employed augmentation to increase the number of examples for model training.

Cardiac diseases

The most common cardiac disease examined was ischemic heart disease (21/35); however, only 14 contained solely ischemic images. The remaining ischemic cohorts were combined with "normal" images (4), tetralogy of Fallot (1), unspecified (1), or a combination (1). Four studies assessed hypertrophic cardiomyopathy (11.4%) and 9 did not specify any disease etiology (25.7%). Intracardiac device presence was not declared.

AI methods

DL and ML were represented in 60% and 40%, respectively. Subcategorization of DL algorithms were convolutional neural networks based on 2D models (12/21), 3D models (4/21), or other/unspecified (5/21). Most DL approaches used a convolutional neural network with U-Net architecture. ML algorithms that referred to classical ML methods using hand-crafted features included 13 various models with at least 6 studies employing multiple models.

Only 4 papers utilized unsupervised primary methods—all classic ML. Twenty-four of 35 ML/DL methods were fully automated, with 9 of 11 nonautomated algorithms utilizing ML methods (Table 1).

Diagnostic evaluation

Ground truth

Ground-truth segmentations were all acquired from human delineation, 9 of which highlighted semiautomated thresholding techniques as segmentation aids. No histological confirmation was undertaken.

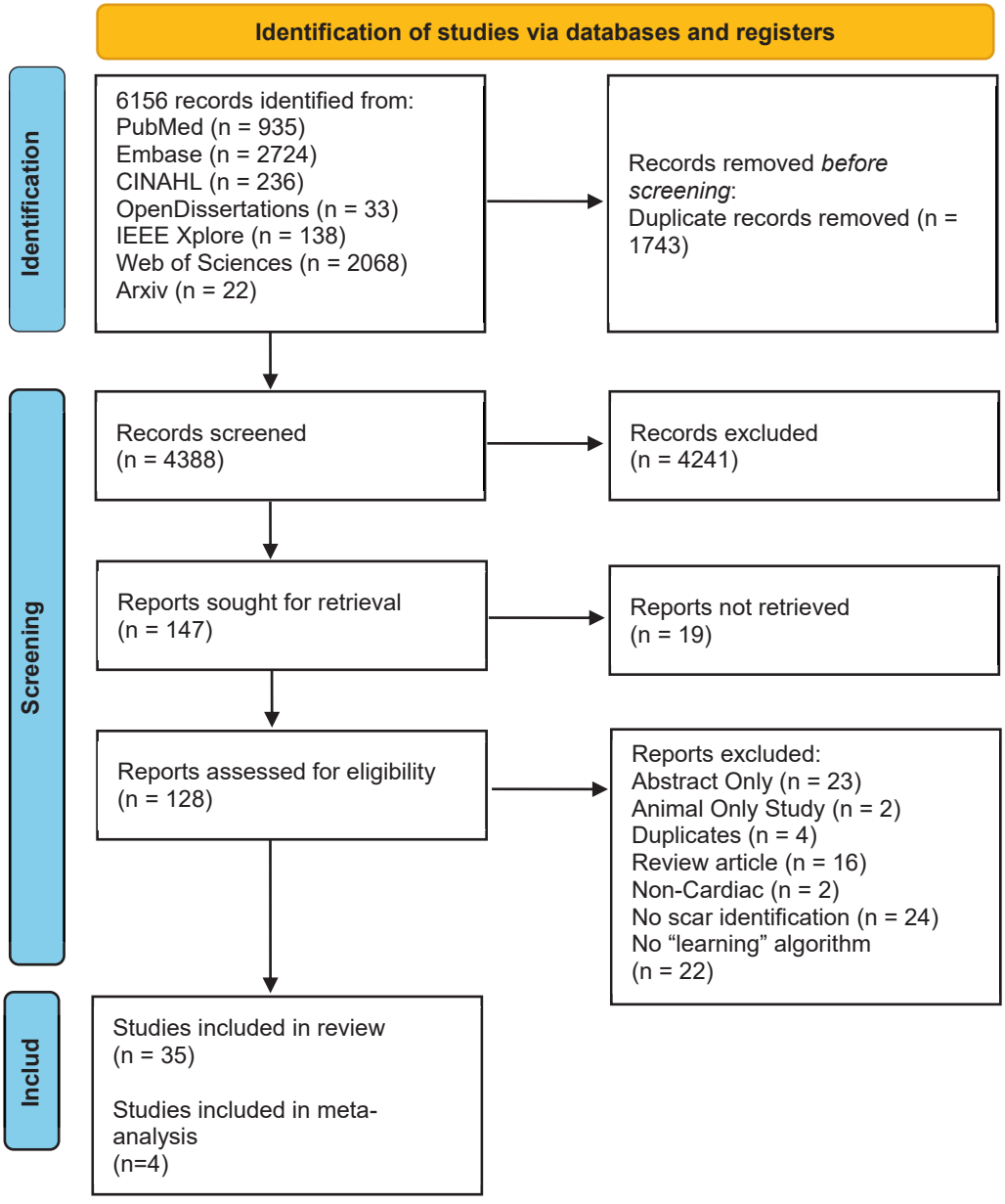

Figure 2 PRISMA flowchart of study selection process.

Evaluation metrics

All methods were compared to a minimum of human delineated ground truth. Four studies compared performance of proposed and predefined thresholding methods.[19,22,43,44]

Twenty-seven different reported evaluation metrics were utilized. Common metrics existed in 3 groups: overlap, distance, and volume metrics (Table 2).

DSC was the most common statistical metric in 54% of studies, followed by sensitivity (17.1%).

Tables 3 and 4 compare results of 5 evaluation methods in studies with comparable data. No statistically significant difference was seen in DSC between predefined threshold models compared to supervised and unsupervised learning methods (0.633 vs 0.616, $P = .14$), with unsupervised methods performing better than supervised (0.732 vs 0.599, $P < .05$). However, substantial dataset size variance is present.

Higher sensitivity, specificity, and accuracy were seen in proposed methods, namely supervised, compared to predefined thresholding with an associated lower Haussdorf distance.

In 4 studies with direct predefined thresholding vs supervised/unsupervised learning comparisons, weak evidence of a small effect with high heterogeneity was seen in learning models with higher DSC, standardized mean difference = 1.11; 95% CI -0.16 to 2.38, $P = .09$ (Figure 3). Visual inspection of funnel plots did not reveal asymmetry in studies, though sparse data existed (Figure 4). Studies had low (n = 2) and moderate (n = 2) risk of bias (Supplemental Appendix C).

Table 1 Summary of reviewed studies

Author, year of publication	*Supervised learning?*	Full Automation?	Final code available?	MRI sequences	Dataset condition cohort
Abramson, 2020[13]	Yes	Yes	No	Cine, 2D LGE	Ischemic
Brahim, 2020[14]	Yes	Yes	No	2D LGE	Mixed – Ischemic, healthy, not specified
Brahim, 2021[15]	Yes	Yes	No	2D LGE	Mixed – Ischemic, healthy
Brahim, 2021[16]	Yes	Yes	No	2D LGE	Mixed – Ischemic, healthy
Campello, 2020[17]	Yes	Yes	No	Cine, 2D LGE, T2	Not specified
Carminati, 2015[18]	No	No	No	2D LGE	Ischemic
Carminati, 2016[19]	No	No	No	2D LGE	Ischemic
De la Rosa, 2019[20]	Yes	Yes	No	Cine, 2D LGE	Mixed – Ischemic, healthy
Engblom, 2016[21]	Yes	Yes	No	2D LGE	Ischemic
Fadil, 2021[22]	Yes	Yes	Yes	Cine, 2D LGE, pre- & postcontrast T1, T2	Mixed – Ischemic, healthy, not specified
Fahmy, 2020[23]	Yes	Yes	Yes	Cine, 2D LGE	Hypertrophic cardiomyopathy
Fahmy, 2021[24]	Yes	Yes	Yes	2D LGE	Hypertrophic cardiomyopathy
Heidenreich, 2021[25]	Yes	Yes	Yes	2D LGE	Ischemic
Kotu, 2011[26]	Yes	Yes	No	2D LGE	Not specified
Kurzendorfer, 2018[27]	Yes	No	No	3D LGE	Not specified
Larroza, 2017[28]	Yes	No	Yes	Cine, 2D LGE	Ischemic
Larroza, 2018[29]	Yes	No	No	Cine, 2D LGE	Ischemic
Lau, 2018[30]	Yes	Yes	No	2D LGE	Not specified
Mantilla, 2015[31]	Yes	Yes	No	2D LGE	Hypertrophic cardiomyopathy
Merino-Caviedes, 2016[32]	Yes	No	No	Cine, 2D LGE	Hypertrophic cardiomyopathy
Metwally, 2010[33]	No	Yes	No	2D LGE	Not specified
Moccia, 2018[34]	Yes	Yes	No	2D LGE	Ischemic
Moccia, 2019[35]	Yes	Yes	No	2D LGE	Ischemic
Moccia, 2020[36]	Yes	Yes	No	Cine, 2D LGE	Ischemic
Morisi, 2015[37]	Yes	No	No	3D LGE	Not specified
Rajchl, 2014[38]	No	No	No	3D LGE (WH)	Mixed – Ischemic, Tetralogy of Fallot
Rukundo, 2020[39]	Yes	Yes	No	2D LGE	Not specified
Wang, 2011[40]	Yes	Yes	No	2D LGE	Ischaemic
Wang, 2020[41]	Yes	Yes	No	Not specified	Mixed – Ischemic, healthy
Zabihollahy, 2018[42]	Yes	No	No	3D LGE (WH)	Ischemic
Zabihollahy, 2019[43]	Yes	No	No	3D LGE (WH)	Ischemic
Zabihollahy, 2020[44]	Yes	Yes	No	3D LGE (WH)	Ischemic
Zhang Z, 2020[45]	Yes	Yes	No	Cine, 2D LGE, T2	Not specified
Zhang X, 2020[46]	Yes	Yes	No	Cine, 2D LGE, T2	Not specified
Zhuang, 2019[47]	Yes	No	No	Cine, 2D LGE, T2	Not specified

LGE = late gadolinium enhancement; WH = whole-heart imaging.

Discussion

In this article, we reviewed published literature surrounding methods for CMR scar segmentation. As expected, with technological advances, increasing numbers of studies using learning methods have been published internationally, highlighting the subject's global interest and clinical potential.

With the recent rapid expansion of AI, many state-of-the-art reviews exist covering current and potential applications for general cardiology.[7] Our review focuses on progress within 1 specific aspect of cardiology to provide clinically relatable interpretation. For this purpose, it is important to consider evaluation, practicality, and patient/scanner generalizability.

Evaluation

Our review has not demonstrated clear benefit of supervised/unsupervised learning vs predefined thresholding methods for LV scar segmentation methods. Unsupervised models

Table 2 Summary of evaluation metrics utilized

Reported evaluation metrics
Overlap
Dice coefficient, Jaccard index/intersection over union, Sensitivity specificity, Accuracy, Precision, F-score, Mean BF1, Recall, Segment overlap, Repeatability, True/false positive & negative
Distance
Haussdorf distance, Surface distance, Average contour distance, Root-mean-squared area
Volume
Left ventricular volume, Scar/infarct volume, Scar mass, Absolute volume difference ± normalization, Total volume error, Percentage volume error, Scar as myocardial percentage, Mean absolute error ± normalization, Left ventricular mass

performed better than other models when considering volume overlap (DSC), but the inverse was seen in sensitivity, specificity, and Haussdorf distance, with supervised methods outperforming others. However, limited comparisons and significant heterogeneity in evaluation methods was evident. In direct comparison from 4 studies, there was weak evidence favoring supervised/unsupervised learning methods. Furthermore, it is questionable whether the small differences demonstrated would translate to clinical benefit.

Three evaluation categories were present: overlap, distance, and volume metrics. Combinations of metrics are required to extensively assess segmentation models. Unfortunately, owing to absence of clear guidance regarding minimal and/or preferable metrics, many models described to date may not be directly comparable. The use of segmentation challenges mitigates some of these issues with standardized datasets and reporting measures but is associated with its own caveats.[48]

Moreover, studies predominantly compared their proposed model to initial ground-truth or other supervised models, with little comparison to various clinically utilized non–learning methods.

For integration into clinical practice, future research must compare standard clinical applications against novel methods through robust multimodal comparable metrics.

Practicality

A noteworthy benefit of the described AI methods is automaticity. Guidelines recommend objective quantification of cardiac structures and function.[5] Hence, full automation of scar assessment is desirable to reduce operator burden. Many existent predefined thresholding algorithms for scar quantification are applied once myocardial borders are delineated—a semiautomatic process.[5,6] Myocardial border annotation deviations may lead to scar misevaluation and this issue extends to all methodologies dependent on predefined ground-truth borders. Subsequently, published segmentation results may not be directly transferable clinically when reliant on border segmentation quality. Fully automated AI methods for both myocardial and scar segmentation may result in more reproducible and objective segmentation labels.

However, the black-box problem of DL decision-making processes being essentially noninterpretable remains a considerable issue. Until this problem is solved, human oversight and manual input for segmentation and correction remains essential to ensure patient safety in the clinical context, currently limiting the potential of a truly automatic approach outside the research environment. Unsupervised methods and rule-based ML algorithms have an advantage in this respect owing to relatively explainable methodologies, which may be important for clinical confidence.

Generalizability

Clinical use and external generalizability of models requires applicability to a variety of patients/scanners and is dependent on image or model complexity and training datasets.

Scar patterns vary across different cardiomyopathies. Ischemic scar arises from the subendocardium, with epicardial progression compared with a more heterogenous distribution in nonischemic cardiomyopathies.[49] With the high prevalence of ischemic heart disease and its associated significant mortality, predominance of ischemic scar models is understandable and clinically necessary.[50] Nonischemic etiologies have been investigated but are more sparse.[23,24] Of greater concern is that 25.7% of publications did not specify disease etiology.

Training datasets have significant impact on model performance. Hence, detailed descriptions of dataset disease etiologies are of great importance for clinical utility, especially as studies assessing model transferability across disease cohorts without retraining are sparse.

Metadata can vary between manufacturers and magnet strengths,[51] and models trained with specific data may not

Table 3 Comparison of reported Dice coefficient and Haussdorf distance for scar segmentation

	Dice coefficient			Haussdorf distance		
	Mean (SD)	Total test cases	No. of studies	Mean	Total test cases	No. of studies
Predefined threshold	0.633† (0.15)	306	4	37.973	82	2
Supervised and unsupervised learning	0.616† (0.256)	1125	13	18.135	230	3
Supervised learning	0.599‡ (0.264)	984	9	18.135	230	3
Unsupervised learning	0.732‡ (0.153)	141	4	-	-	-

†$P = .14$.
‡$P < .05$.

Table 4 Comparison of reported mean sensitivity, specificity, and accuracy for scar segmentation

	Sensitivity	Total test cases	No. of studies	Specificity	Total test cases	No. of studies	Accuracy	Total test cases	No. of studies
Predefined threshold	83.91	160	1	90.79	160	1	88.51	160	1
Supervised and unsupervised learning	91.4	882	4	97.11	870	3	92.99	1296	6
Supervised learning	95.09	520	2	98.95	520	2	94.03	776	4
Unsupervised learning	83.79	372	5	94.38	350	3	88.4	520	3

be generally applicable. Only a small subset of studies included multiple CMR manufactures to mitigate this risk.[13,21-23,30,39] Training data on various field strengths would allow greater generalizability clinically, but 1.5 T remains the current standard, with 3 T employed mainly by more experienced imaging centers.[5] Similarly, regional manufacturer predominance may exist.[52] Inclusion of multiple manufacturers and field strengths for generalizability is to be commended and should be actively sought; however, data suggest 1.5 T Siemens scans have the largest data support for clinical application.[52]

Small datasets are the main limitation for optimal Algorithm creation. Access to data remains a limitation to researchers explaining, in part, the large variation in dataset size range.

Reducing barriers to data accessibility may reduce training/testing data variability and improving model comparability. Collaborative sharing and access to such data is an important consideration.[21] Critically, labeled data are required to avoid time-consuming reanalysis and promote transparent result comparison. Other smaller datasets exist in the form of CMR challenges for more standardized model comparison. Further options to improve generalizability within existing datasets include augmentation through image transformations or using generative adversarial networks to produce synthetic images.[13,30]

Conclusion

Feasibility of applying AI to the task of scar segmentation in CMR has been demonstrated. Progression toward clinical application requires dataset transparency, evaluation, standardization, and model generalizability.

Funding Sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Disclosures

The authors have no conflicts of interest to declare.

Authorship

All authors attest they meet the current ICMJE criteria for authorship.

Patient Consent

Not applicable.

Ethics Statement

This review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations[9] (Supplemental Appendices A and B). No pre-registration was undertaken.

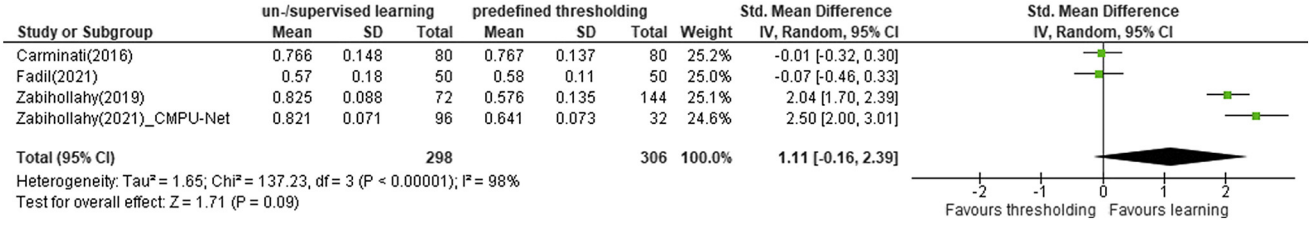

Figure 3 Forest plot of supervised and unsupervised learning vs predefined thresholding models.

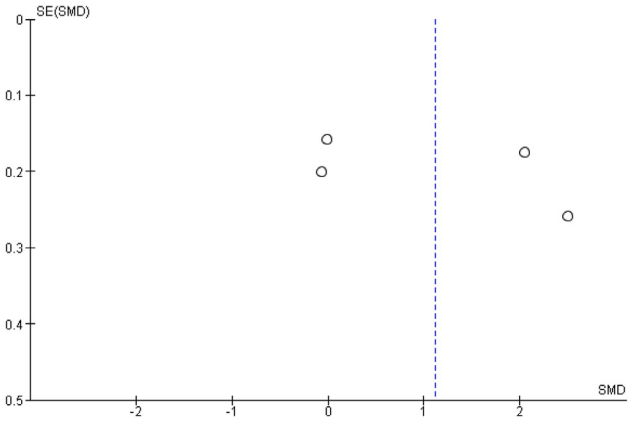

Figure 4 Funnel plot. Visual analysis suggests no bias, though data points are sparse. SE = standard error; SMD = standardized mean difference.

Appendix

Supplementary data

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.cvdhj.2021.11.005.

References

1. Tülümen E, Rudic B, Ringlage H, et al. Extent of peri-infarct scar on late gadolinium enhancement cardiac magnetic resonance imaging and outcome in patients with ischemic cardiomyopathy. Heart Rhythm 2021;18:954–961.
2. Taylor RJ, Umar F, Panting JR, Stegemann B, Leyva F. Left ventricular lead position, mechanical activation, and myocardial scar in relation to left ventricular reverse remodeling and clinical outcomes after cardiac resynchronization therapy: a feature-tracking and contrast-enhanced cardiovascular magnetic resonance study. Heart Rhythm 2016;13:481–489.
3. Andreu D, Penela D, Acosta J, et al. Cardiac magnetic resonance-aided scar dechanneling: influence on acute and long-term outcomes. Heart Rhythm 2017;14:1121–1128.
4. Cronin EM, Bogun FM, Maury P, et al. 2019 HRS/EHRA/APHRS/LAHRS expert consensus statement on catheter ablation of ventricular arrhythmias. J Interv Card Electrophysiol 2020;59:145–298.
5. Schulz-Menger J, Bluemke DA, Bremerich J, et al. Standardized image interpretation and post-processing in cardiovascular magnetic resonance - 2020 update: Society for Cardiovascular Magnetic Resonance (SCMR): Board of Trustees Task Force on Standardized Post-Processing. J Cardiovasc Magn Reson 2020;22:19.
6. Rajchl M, Stirrat J, Goubran M, et al. Comparison of semi-automated scar quantification techniques using high-resolution, 3-dimensional late-gadolinium-enhancement magnetic resonance imaging. Int J Cardiovasc Imaging 2015;31:349–357.
7. Quer G, Arnaout R, Henne M, Arnaout R. Machine learning and the future of cardiovascular care: JACC State-of-the-Art Review. J Am Coll Cardiol 2021;77:300–313.
8. Leiner T, Rueckert D, Suinesiaputra A, et al. Machine learning in cardiovascular magnetic resonance: basic concepts and applications. J Cardiovasc Magn Reson 2019;21:61.
9. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
10. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014;14:135.
11. Luo D, Wan X, Liu J, Tong T. Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Stat Methods Med Res 2018;27:1785–1805.
12. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919.
13. Abramson HG, Popescu DM, Yu R, et al. Anatomically-informed deep learning on contrast-enhanced cardiac MRI for scar segmentation and clinical feature extraction. arXiv 2020;1–15.
14. Brahim K, Qayyum A, Lalande A, Boucher A, Sakly A, Meriaudeau F. A 3D deep learning approach based on Shape Prior for automatic segmentation of myocardial diseases. 2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA) 2020;1–6.
15. Brahim K, Qayyum A, Lalande A, Boucher A, Sakly A, Meriaudeau F. A deep learning approach for the segmentation of myocardial diseases. 2020 25th International Conference on Pattern Recognition 2021;4544–4551.
16. Brahim K, Qayyum A, Lalande A, Boucher A, Sakly A, Meriaudeau F. A 3D network based shape prior for automatic myocardial disease segmentation in delayed-enhancement MRI. IRBM 2021;1:1–11.
17. Campello VM, Martín-Isla C, Izquierdo C, et al. Combining multi-sequence and synthetic images for improved segmentation of late gadolinium enhancement cardiac MRI. Lect Notes Comput Sci 2020;12009 LNCS:290–299.
18. Carminati MC, Boniotti C, Pepi M, Caiani EG. Quantification of myocardial viability in late gadolinium enhancement cardiac MRI. Comput Cardiol 2015;42:97–100.
19. Carminati MC, Boniotti C, Fusini L, et al. Comparison of image processing techniques for nonviable tissue quantification in late gadolinium enhancement cardiac magnetic resonance images. J Thorac Imaging 2016;31:168–176.
20. de la Rosa E, Sidibé D, Decourselle T, Leclercq T, Cochet A, Lalande A. Myocardial infarction quantification from late gadolinium enhancement MRI using top-hat transforms and neural networks. 2019. Available at http://arxiv.org/abs/1901.02911.
21. Engblom H, Tufvesson J, Jablonowski R, et al. A new automatic Algorithm for quantification of myocardial infarction imaged by late gadolinium enhancement cardiovascular magnetic resonance: experimental validation and comparison to expert delineations in multi-center, multi-vendor patient data. J Cardiovasc Magn Reson 2016;18:27.
22. Fadil H, Totman JJ, Hausenloy DJ, et al. A deep learning pipeline for automatic analysis of multi-scan cardiovascular magnetic resonance. J Cardiovasc Magn Reson 2021;23:1–13.
23. Fahmy AS, Neisius U, Chan RH, et al. Three-dimensional deep convolutional neural networks for automated myocardial scar quantification in hypertrophic cardiomyopathy: a multicenter multivendor study. Radiology 2020;294:52–60.
24. Fahmy AS, Rowin EJ, Chan RH, Manning WJ, Maron MS, Nezafat R. Improved quantification of myocardium scar in late gadolinium enhancement images: deep learning based image fusion approach. J Magn Reson Imaging 2021;54:303–312.
25. Heidenreich JF, Gassenmaier T, Ankenbrand MJ, Bley TA, Wech T. Self-configuring nnU-net pipeline enables fully automatic infarct segmentation in late enhancement MRI after myocardial infarction. Eur J Radiol 2021;141:109817.
26. Kotu LP, Engan K, Eftestøl T, Ørn S, Woie L. Segmentation of scarred and non-scarred myocardium in LG enhanced CMR images using intensity-based textural analysis. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:5698–5701.
27. Kurzendorfer T, Breininger K, Steidl S, Brost A, Forman C, Maier A. Myocardial scar segmentation in LGE-MRI using fractal analysis and ran*dom forest classification*. 2018 24th International Conference on Pattern Recognition (ICPR), 2018;3168–3173.
28. Larroza A, Materka A, López-Lereu MP, Monmeneu JV, Bodí V, Moratal D. Differentiation between acute and chronic myocardial infarction by means of texture analysis of late gadolinium enhancement and cine cardiac magnetic resonance imaging. Eur J Radiol 2017;92:78–83.
29. Larroza A, López-Lereu MP, Monmeneu JV, et al. Texture analysis of cardiac cine magnetic resonance imaging to detect nonviable segments in patients with chronic myocardial infarction. Med Phys 2018;45:1471–1480.
30. Lau F, Hendriks T, Lieman-Sifry J, Sall S, Golden D. ScarGAN: Chained Generative Adversarial Networks to Simulate Pathological Tissue on Cardiovascular MR Scans. In: Stoyanov D, et al., eds. Deep learning in medical image analysis and multimodal learning for clinical decision support. DLMIA 2018, ML-CDS 2018. Lecture Notes in Computer Science, 11045. Cham: Springer; 2018.
31. Mantilla J, Paredes JL, Bellanger JJ, et al. Detection of fibrosis in late gadolinium enhancement cardiac MRI using kernel dictionary learning-based clustering. Comput Cardiol 2015;42:357–360.
32. Merino-Caviedes S, Cordero-Grande L, Perez Rodriguez M, et al. A variational method for scar segmentation with myocardial contour correction in DE-CMR images. 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI) 2016;956–959.
33. Metwally MK, El-Gayar N, Osman NF. Improved technique to detect the infarction in delayed enhancement image using k-mean method. Lect Notes Comput Sci 2010;6112 LNCS:108–119.
34. Moccia S, Banali R, Martini C, et al. Automated scar segmentation from CMR-LGE images using a deep learning approach. Comput Cardiol 2018;2018:1–4.
35. Moccia S, Banali R, Martini C, et al. Development and testing of a deep learning-based strategy for scar segmentation on CMR-LGE images. MAGMA 2019;32:187–195.

36. Moccia S, Cagnoli A, Martini C, et al. A novel approach based on spatio-temporal features and random forest for scar detection using cine cardiac magnetic resonance images. Comput Cardiol 2020;2020:5–8.
37. Morisi R, Donini B, Lanconelli N, et al. Semi-automated scar detection in delayed enhanced cardiac magnetic resonance images. Int J Mod Phys C 2015;26.
38. Rajchl M, Yuan J, White JA, et al. Interactive hierarchical-flow segmentation of scar tissue from late-enhancement cardiac MR images. IEEE Trans Med Imaging 2014;33:159–172.
39. Rukundo O. Evaluation of deep learning-based myocardial infarction quantification using segment CMR software. 2020. arXiv. Available at http://arxiv.org/abs/2012.09070.
40. Wang L, Wong KCL, Zhang H, Liu H, Shi P. Noninvasive computational imaging of cardiac electrophysiology for 3-D infarct. IEEE Trans Biomed Eng 2011; 58:1033–1043.
41. Wang SH, McCann G, Tyukin I. Myocardial infarction detection and quantification based on a convolution neural network with online error correction capabilities. 2020 International Joint Conference on Neural Networks (IJCNN) 2020;1-8.
42. Zabihollahy F, White JA, Ukwatta E. Myocardial scar segmentation from magnetic resonance images using convolutional neural network. Med Imaging 2018 Comput Diagnosis SPIE 2018;10575:106. Available at https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10575/2293518/Myocardial-scar-segmentation-from-magnetic-resonance-images-using-convolutional-neural/10.1117/12.2293518.full.
43. Zabihollahy F, White JA, Ukwatta E. Convolutional neural network-based approach for segmentation of left ventricle myocardial scar from 3D late gadolinium enhancement MR images. Med Phys 2019;46:1740–1751.
44. Zabihollahy F, Rajchl M, White JA, Ukwatta E. Fully automated segmentation of left ventricular scar from 3D late gadolinium enhancement magnetic resonance imaging using a cascaded multi-planar U-Net (CMPU-Net). Med Phys 2020; 47:1645–1655.
45. Zhang Z, Liu C, Ding W, et al. Multi-modality pathology segmentation framework: application to cardiac magnetic resonance images. Lect Notes Comput Sci 2020;12554 LNCS:37–48.
46. Zhang X, Noga M, Punithakumar K. Fully automated deep learning based segmentation of normal, infarcted and edema regions from multiple cardiac MRI sequences. Lect Notes Comput Sci 2020;12554 LNCS:82–91.
47. Zhuang X. Multivariate mixture model for myocardial segmentation combining multi-source images. IEEE Trans Pattern Anal Mach Intell 2019; 41:2933–2946.
48. Mendrik AM, Aylward SR. Beyond the leaderboard: insight and deployment challenges to address research problems. arXiv 2018;1811.03014.
49. Kramer CM. Role of cardiac MR imaging in cardiomyopathies. J Nucl Med 2015; 56:39S–45S.
50. Kochanek KD, Xu J, Arias E. Key findings Data from the National Vital Statistics System How long can we expect to live? 2019. Available at https://www.cdc.gov/nchs/products/databriefs/db395.htm.
51. Raisi-Estabragh Z, Gkontra P, Jaggi A, et al. Repeatability of cardiac magnetic resonance radiomics: a multi-centre multi-vendor test-retest study. Front Cardiovasc Med 2020;7:289.
52. Keenan NG, Captur G, McCann GP, et al. Regional variation in cardiovascular magnetic resonance service delivery across the UK. Heart 2021; 0:1–6.