VILNIUS UNIVERSITY
FACULTY OF MEDICINE

Study Programme: Medicine (English)

Institute of Clinical Medicine, Clinic of the Cardiovascular Diseases

Bagusat Leopold Maximilian Johannes, 6 th Year, 5 Group

Integrated Study Master‘s Thesis

The Role of Artificial Intelligence in Cardiovascular Magnetic Resonance: Comprehensive Literature Review

Supervisor prof. Dr. Sigita Glaveckaitė

Consultant PhD student Žygimantas Abramikas

Head of the Clinic of the Cardiovascular Diseases prof. Dr. Sigita Glaveckaitė

2025

Email: maximilian.bagusat@mf.stud.vu.lt

Abbreviations:

Abbreviation:	Definition:
4CH	Four-Chamber
ACSNet	Anatomical Convolutional Segmentation Network
AI	Artificial Intelligence
ANN	Artificial Neural Network
BA	Balanced Accuracy
CMR	Cardiovascular Magnetic Resonance
CNN	Convolutional Neural Network
CoV	Coefficient of Variation
DL	Deep Learning
ECV	Extracellular Volume
EDV	End-Diastolic Volume
EDVi	End-Diastolic Volume Index
EF	Ejection Fraction
FCN	Fully Convolutional Network
FT	Feature Tracking
GAN	Generative Adversarial Network
GLS	Global Longitudinal Strain
GRS	Global Radial Strain
GCS	Global Circumferential Strain
HCM	Hypertrophic Cardiomyopathy
HfpEF	Heart Failure with Preserved Ejection Fraction
ICC	Intraclass Correlation Coefficient
ICD	Implantable Cardioverter-Defibrillator
LGE	Late Gadolinium Enhancement
LV	Left Ventricle / Left Ventricular
MI	Myocardial Infarction
ML	Machine Learning
MWT	Maximum Wall Thickness
R	Pearson Correlation Coefficient
RV	Right Ventricle / Right Ventricular

SCD	Sudden Cardiac Death
TRL	Technology Readiness Level
VNE	Virtual Native Enhancement
XAI	Explainable Artificial Intelligence

Summary/Abstract:

Objectives: Currently the analysis of cardiovascular magnetic resonance (CMR) images is entirely dependent on the expertise of individual trained professionals and thus is time consuming and introduces observer variation. Artificial intelligence (AI) is becoming more and more relevant when it comes to addressing these limitations, allowing for automation of repetitive processes such as analysis of cardiac morphology, function and myocardial structure. The aim of this review was to assess the role of AI in CMR imaging, with a specific focus on three domains: volumetry and mass, myocardial scar detection and quantification, and myocardial wall thickness measurements. To accomplish this task a systematic review was conducted using the PRISMA technique including high-quality studies published between November 2020 and January 2024.

Results: In total, 15 studies were analyzed. Compared to human expert evaluation, CMR-based AI systems have been shown to significantly reduce intra- and interobserver variabilities, with segmentation Dice scores of ≥ 0.90 for both left and right ventricular volumetry and ICCs up to 0.94 for left ventricular volume measurements. For myocardial wall thickness, AI demonstrated improved test–retest reproducibility with a CoV of 4.3%, compared to 5.7–12.1% for human observers. In the field of tissue characterization, AI solutions allow assessment of myocardial fibrotic changes without contrast media; for instance, Virtual Native Enhancement (VNE) achieved an ICC of 0.94 compared to LGE for scar burden. These advances are valuable in clinical practice for decision-making and reduce required sample sizes in research. AI-based analysis also significantly improves efficiency, reducing analysis time from 10–15 minutes to under 30 seconds in some implementations. Despite advancement of AI based CMR analysis the need for expert oversight and manual review in more complex cases is still needed, as well as generalizability, data standardization, and ethical issues still limits widespread use of AI solutions in CMR image analysis.

Conclusions: Across all three analyzed domains, AI consistently demonstrated improvements in reproducibility, efficiency, and diagnostic consistency when compared to traditional manual human expertise-based methods. Successful integration of AI into clinical CMR workflows will depend in

the future on further validation in larger patient cohorts, cross-center generalizability, and sustained expert supervision.

Key Words: Artificial Intelligence; Cardiovascular Magnetic Resonance Imaging; Wall Thickness Measurement; Cardiac Volumetry; Myocardial Mass; Myocardial Scar Detection; Deep Learning; Segmentation; Clinical Workflow Optimization; Observer Variability.

Introduction:

Cardiovascular magnetic resonance (CMR) imaging is widely acknowledged as a gold-standard modality for non-invasive cardiac assessment, offering highly detailed and accurate visualization of cardiac structures, function, and pathology. Due to its exceptional ability to capture precise myocardial details as well as characterize myocardial structure, CMR has become a cornerstone in the diagnosis, monitoring, and management of a diverse array of cardiovascular diseases (CVDs), including cardiomyopathies, myocardial infarction, congenital heart defects, and other conditions characterized by structural or functional cardiac abnormalities.

However, despite these strengths, the accuracy of CMR interpretation is closely linked to clinician experience and skill level. This dependence introduces observer variability, characterized by discrepancies in measurements and interpretations not only between different clinicians (inter-observer variability) but also within the same clinician over repeated assessments (intra-observer variability). Such variability is most apparent in segmentation tasks, which involve outlining left and right ventricular (LV and RV) volumes, measuring myocardial wall thickness or mass, or identifying myocardial tissue structural changes through late gadolinium enhancement (LGE) imaging. Additionally, traditional manual segmentation techniques require extensive manual contouring, typically taking clinicians several minutes. This time-intensive nature leads to efficiency challenges, specifically with regards to high-volume clinical settings where quick and consistent interpretation is crucial. AI-based methods have demonstrated real time savings in this context. With some of the studies showing that automated segmentation tools can complete accurate volumetric analysis in a fraction of the time needed for manual contouring (1). Given these limitations, there has been growing interest to explore computational solutions that could help improve the accuracy and efficiency of CMR interpretation.

Artificial intelligence (AI) has begun to show great promise in medical imaging and is presenting opportunities to automate complex image analysis tasks, with the potential to improve diagnostic accuracy as well as and streamline clinical workflows. The application of AI-driven segmentation methods has already demonstrated potential in addressing some limitations associated with manual image interpretation. Fully convolutional networks trained on large-scale datasets have shown human-level results in the segmentation of LV and RV boundaries, achieving robust accuracy across a variety of measurements (2). Other proposed tools use architectures incorporating anatomical priors to quantify biventricular function, volume, mass, and ejection fraction within seconds. (3). Innovations like this point toward AI-tools for clinicians becoming more mature and relevant lately.

For example, the automatic quantification of the left ventricle via AI systems has been shown to decrease the processing time whilst keeping close agreement with manual expert analysis, needing only minimal corrections (4). More recent developments include virtual native enhancement (VNE) which facilitates detailed images without the use of contrast (5). The implications and technical details will be discussed later in the results section. In parallel to VNE, another branch of AI enabled systems took on the challenge of automating fibrosis and scar quantification to further bolster the analysis of clinically relevant metrics without falling victim to the variation that is inherent to manual analysis (6). The components that make up the variance in the analysis of CMR images, specifically with regards to the analysis of left ventricular ejection fraction and mass will be visualized in the following Figure.

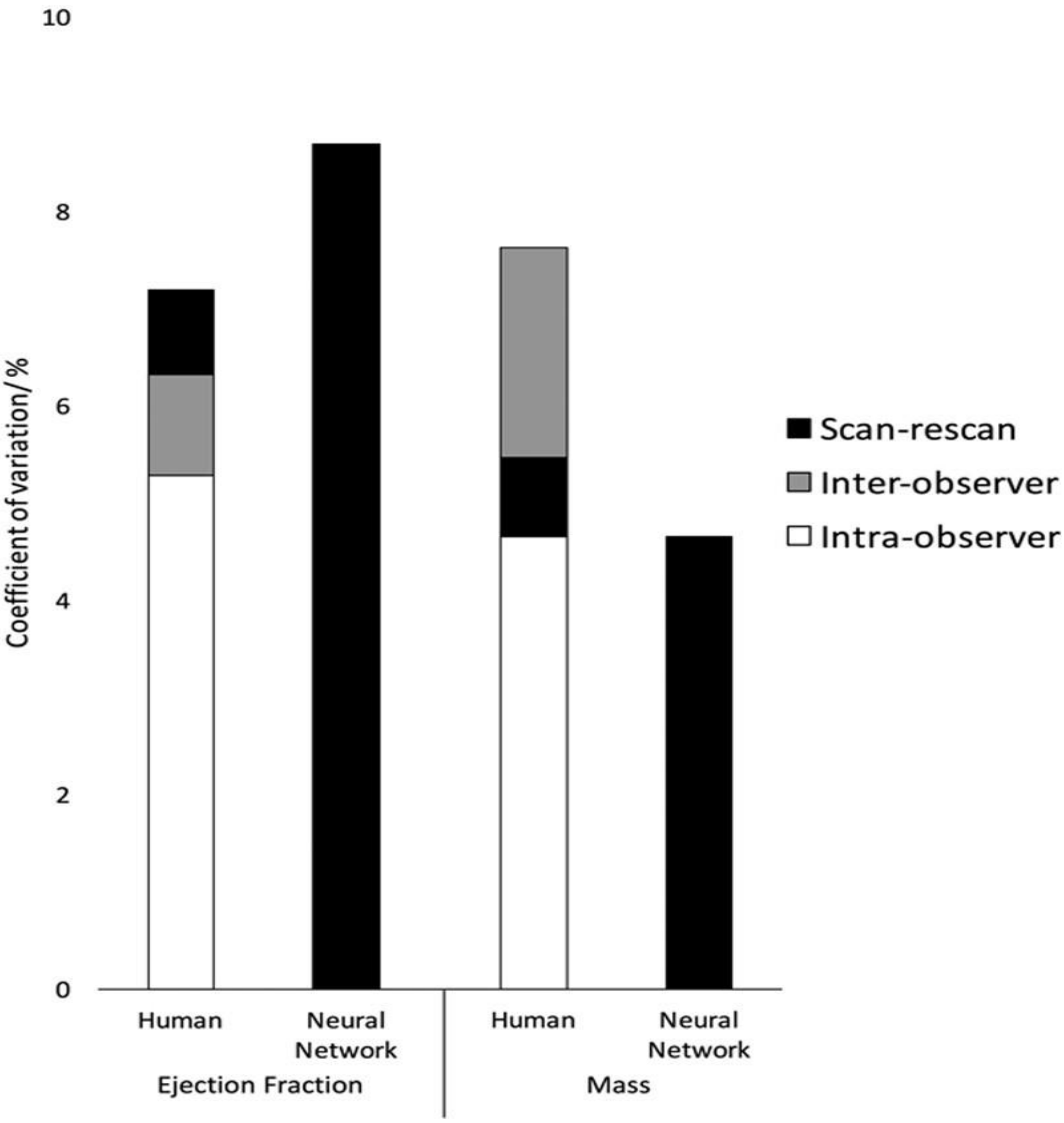

Figure 1: **Contributions of scan–rescan, inter-observer, and intra-observer variability to measurement error in CMR-derived left ventricular ejection fraction and mass. Computational solutions leave scan-rescan variability.** (Reproduced with permission from: Bhuva AN, Bai W, Lau C, Davies RH, Ye Y, Bulluck H, et al. *A multicenter, scan–rescan, human and machine learning CMR study.* JACC Cardiovasc Imaging. 2019 Oct;12(10):1829–40. https://doi.org/10.1016/j.jcmg.2018.08.025. Licensed under CC BY 4.0.) (7).

Figure 1 is a visual representation of the make-up of variation that is encountered in human expert analysis. In human analysis the variance in both mass and ejection fraction is made up of inter-observer, intra-observer and scan-rescan variabilities, with intra-observer variance making up the largest part. As evidenced by the illustration, AI allows for the elimination of variance between observers and, with regards to Mass, leads to an overall lower coefficient of variation. Although the diagram provides a very important notion, it also demonstrates that although AI has the power to significantly improve accuracy, in areas like Ejection fractions, the overall coefficient of variation is still larger than that of humans.

Furthermore, AI-driven segmentation benefits from its ability to consistently interpret extensive annotated datasets, potentially decreasing observer variability seen in manual approaches and improving reproducibility across clinical settings. Beyond the LV, automated techniques have also been helpful in enhancing the accuracy of four-chamber cine imaging, including improved characterization of the RV, which has traditionally posed difficulties due to its complex geometry.

Additionally, AI-based myocardial deformation analysis technique (feature tracking) is increasingly being employed to detect subtle changes in myocardial strain patterns, thus offering potential benefits in prognostication and patient management, particularly in diseases such as hypertrophic cardiomyopathy (HCM) and heart failure. These approaches have also demonstrated potential in high-risk clinical scenarios, such as predicting sudden cardiac death (SCD), guiding patient stratification for implantable cardioverter-defibrillator (ICD) therapy, and informing treatment strategies based on detailed analysis of myocardial fibrosis and structural remodeling (8).

Beyond just segmentation tasks, AI models that have been trained on CMR data have shown increasing potential in spotting subtle imaging markers. Such markers could be used to check disease progression and other outcomes. Features like this are often hard to properly detect by hand. By having another tool that can help in identifying diseases earlier, more informed clinical decisions can be made, and thus better care can be provided for patients. Recent studies have shown that some AI systems are able to with increasing accuracy predict adverse cardiological events by recognizing imaging patterns that can be associated with underlying diseases. Although these advances see to be very promising, there needs to be great care when implementing AI systems, and expert oversight is needed. This becomes clearer when looking at more complex cases, or cases that very close to certain borders. In these situations, human expert evaluation cannot be replaced yet, demonstrating the need for a combination of these two pillars of image analysis. With these questions in mind, this thesis systematically evaluates the performance and clinical viability of AI in comparison to human expert assessment across three key areas of CMR analysis:

1. **Myocardial wall thickness measurement**, with particular attention to diagnostic accuracy and reproducibility in myocardial hypertrophy.
2. **Volumetric and mass quantification**, evaluating automated segmentation approaches in the functional analysis of both left and right ventricles.
3. **Myocardial scar detection and quantification**, examining the reliability of AI methods in detecting and quantifying myocardial fibrosis and infarction

Methodology:

1. Data Sources and Search Strategy

This systematic review was conducted using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) technique. Included are studies and papers that investigate the diagnostic accuracy of AI applications in CMR wall thickness measurements, volumetry and mass as well as myocardial scar detection and quantification. As illustrated in Figure 2, the search strategy for this review was meticulously designed to include relevant and high-quality studies published between November 2020 and January 2024 using the following keywords with MeSH terms. We have used three concepts: artificial intelligence ("artificial intelligence"), diagnostic value ("diagnostic value" OR "diagnosis"), and cardiac MRI ("Cardiovascular magnetic resonance" OR "CMR" OR "cardiac magnetic resonance" OR "late gadolinium enhancement" OR "delayed gadolinium enhancement" OR "LGE" OR "wall thickness" OR "volumes" OR "mass" OR "scar" OR "ejection fraction"). The terms were combined by "OR" in each domain, and then concepts were combined by "AND". Search results were imported into Zotero reference management software (Zotero version 6.0.37).

2. Study selection:

To determine study eligibility, the following inclusion criteria were used:

(1) CMR must have been performed with either 1.5 T or 3 T field strength machine;

(2) AI techniques such as machine learning and convolutional neural networks were considered relevant;

(3) Study should be written in English language;

(4) All non-freely available texts were excluded to ensure that the review could be based on sources that are readily accessible;

(5) Data on AI accuracy assessment against human expert evaluation should be present;

(6) Must revolve around one of the three previously determined topics: wall thickness measurements, volumetric and mass assessments, myocardial scar detection and quantification.

3. Study Quality:

When selecting sources, the Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA) where followed.

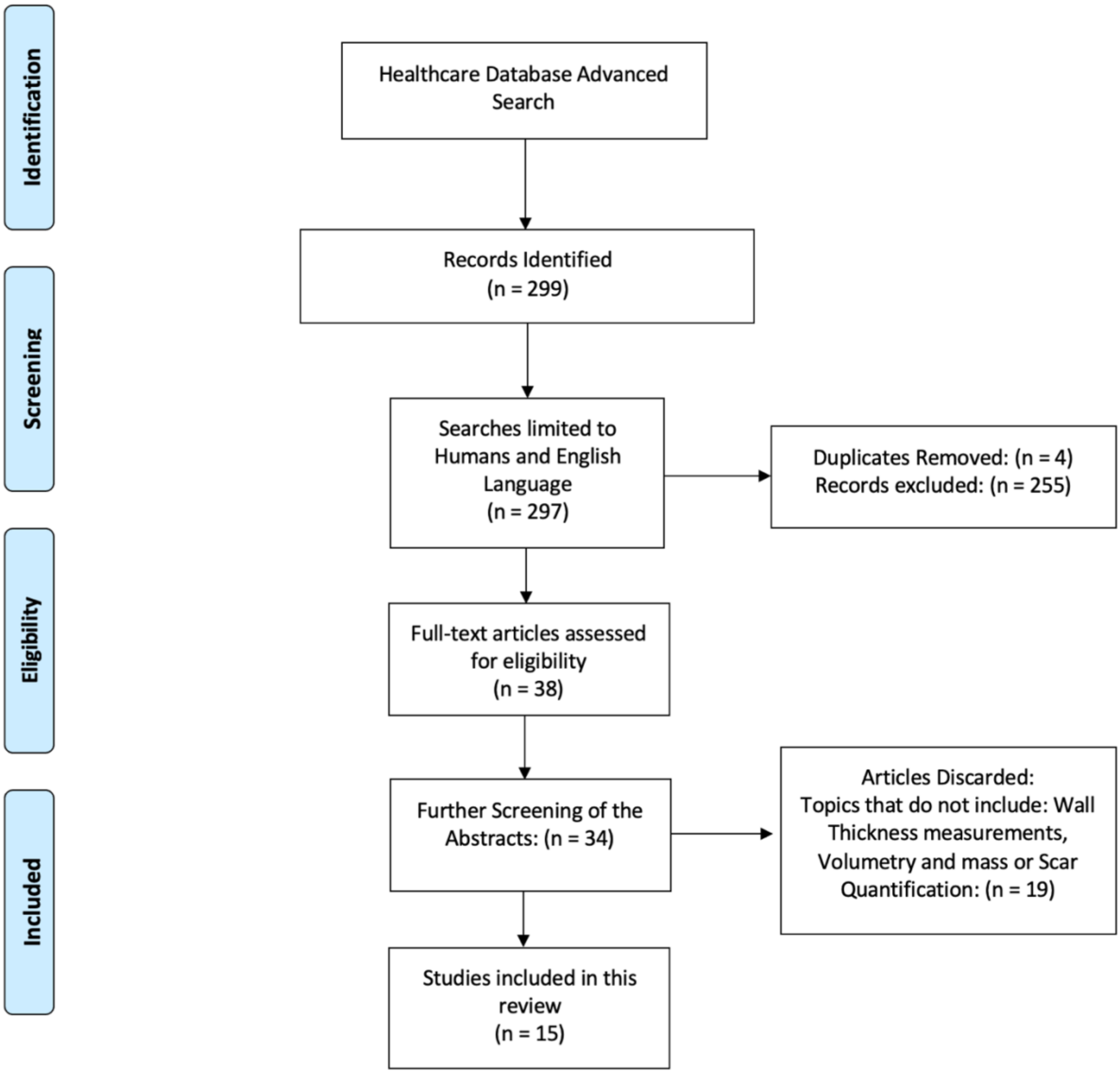

Figure 2: PRISMA search strategy flow chart according to 2009 guidance.

Basic Principles of AI Used for Diagnostic Purposes in CMR:

AI applications in CMR primarily leverage machine learning (ML) and deep learning (DL) techniques, particularly convolutional neural networks (CNNs), which have demonstrated remarkable success in various diagnostic tasks.

1. Machine Learning and Deep Learning in Cardiovascular Magnetic Resonance (CMR):

In recent years, machine learning (ML) has emerged as a game-changer in cardiovascular imaging, including CMR. ML refers to a family of computational techniques that enable systems to recognize patterns and make data-driven predictions without requiring specific, hand-coded instructions.

Among these techniques DL, a specialized branch of ML, has shown tremendous promise, particularly through its use of artificial neural networks (ANNs). These networks, modeled loosely after the human brain, consist of multiple layers that progressively extract complex features from data.

Within the DL category, convolutional neural networks (CNNs) have become the gold standard for medical image analysis. CNNs are especially well-suited for processing visual information, such as CMR images, because of their ability to automatically detect and learn relevant features like edges, textures, and structures. Their application in CMR imaging has significantly advanced the automated segmentation of cardiac chambers, myocardial tissue characterization, and even disease classification.

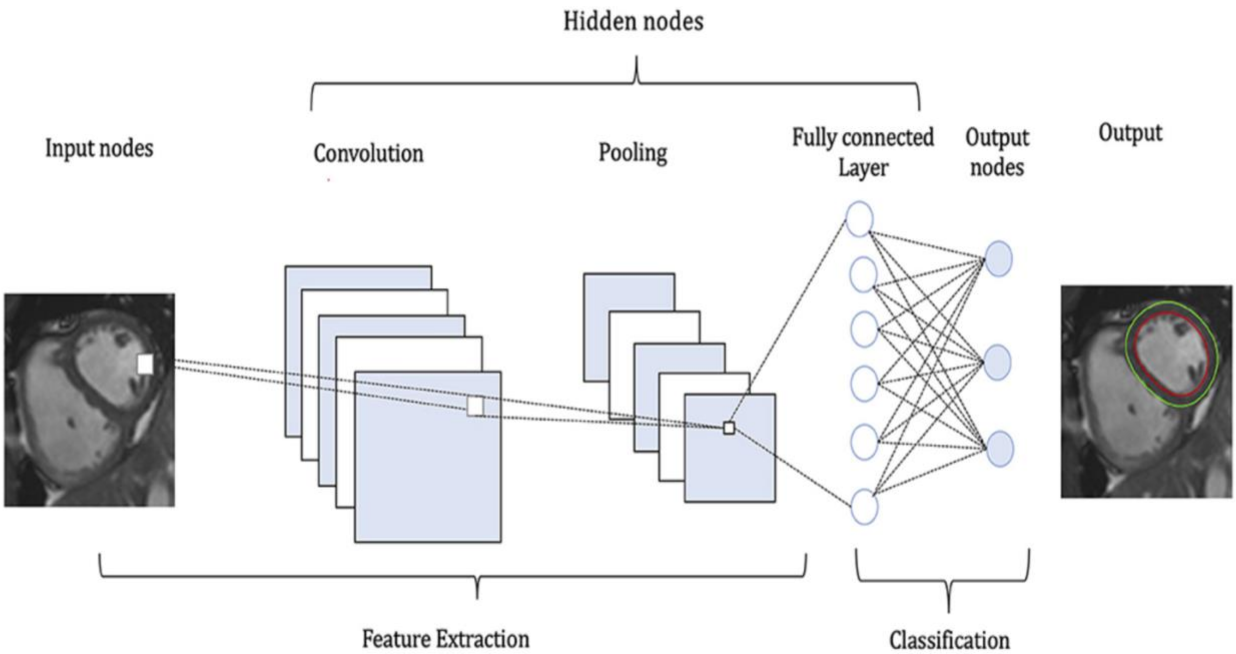

Figure 3: Pipeline of a convolutional neural network (CNN) applied to cardiac magnetic resonance (CMR) image segmentation. (Reproduced with permission from: Fotaki A, et al. Licensed under CC BY 4.0.) (9)

As illustrated in Figure 3, a typical CNN architecture used for CMR segmentation consists of an input layer that receives the original image, followed by multiple hidden layers where feature extraction occurs through convolution and pooling operations. These layers capture increasingly complex patterns within the image. The extracted features are then passed to fully connected layers that perform classification, ultimately generating an output, such as delineated contours of the left ventricular cavity. This automated approach streamlines image interpretation and has demonstrated performance comparable to that of human experts in tasks like cardiac chamber segmentation.

2. AI in Image Acquisition and Preprocessing:

While much attention is given to AI's role in image analysis, it is equally impactful in optimizing the image acquisition process itself. Standard CMR protocols can be lengthy and demanding, often requiring patients to remain still for extended periods. This can be uncomfortable and, in some cases, unfeasible, particularly for critically ill or pediatric patients.

AI-driven methods, particularly those leveraging compressed sensing techniques and reconstruction algorithms, are now being used to:

- shorten scan times by acquiring fewer data points without compromising diagnostic accuracy;
- enhance image resolution, ensuring that even with shorter acquisitions, image quality remains clinically acceptable;
- correct for motion artifacts, such as those caused by breathing or irregular heartbeats, by applying intelligent post-processing techniques.

Taken together, these advancements help create a more comfortable experience for patients and reduce the likelihood of repeat scans. This not only improves overall workflow efficiency but also eases the burden on healthcare systems.

3. AI in Segmentation and Feature Extraction:

Segmenting cardiac structures is one of the most time-consuming and demanding parts of CMR analysis. Clinicians are often required to manually outline areas such as the LV, RV, and myocardium across numerous image slices and phases of the cardiac cycle. This manual approach not only takes considerable time but can also lead to variability between and within observers, where even experienced clinicians might produce slightly different results on the same scan.

To address these challenges, AI-based segmentation tools, particularly those using convolutional neural network (CNN) architectures, are increasingly being used to:

- automatically delineating the LV, RV, and myocardial contours, reducing the need for manual intervention;
- improving reproducibility and standardization across different datasets and institutions;
- enhancing scar quantification in late gadolinium enhancement (LGE) imaging, a critical step in diagnosing myocardial fibrosis and scarring.

AI-based segmentation tools are now capable of achieving human-expert-level accuracy, with the added benefit of performing tasks 10 to 20 times faster than manual methods. In high-volume clinical environments, this time savings translates into faster reporting, reduced workload for clinicians, and more timely patient care.

4. AI the principles of AI-based clinical trials, specifically data/training models:

Another essential principle in the development of AI models for clinical use is the careful handling of training and validation processes. Machine learning algorithms typically require large, high-quality datasets to achieve reliable performance and to avoid overfitting, a phenomenon where models capture noise instead of meaningful patterns. As outlined by Tam et al., most AI studies to date have relied heavily on internal validation techniques, such as k-fold cross-validation, which assess model consistency within the training data but do not guarantee generalizability (10). True external validation, involving testing on an independent patient cohort, remains uncommon yet is crucial for demonstrating a model's real-world applicability. Without robust external validation, reported performance metrics may overestimate a model's clinical utility, highlighting the importance of rigorous study design and transparent reporting in AI-based clinical trials.

This process is summarized in Figure 4, which illustrates the typical workflow for supervised learning, beginning with data collection and model training, followed by internal validation, and concluding with external validation on an independent dataset to ensure generalizability (10).

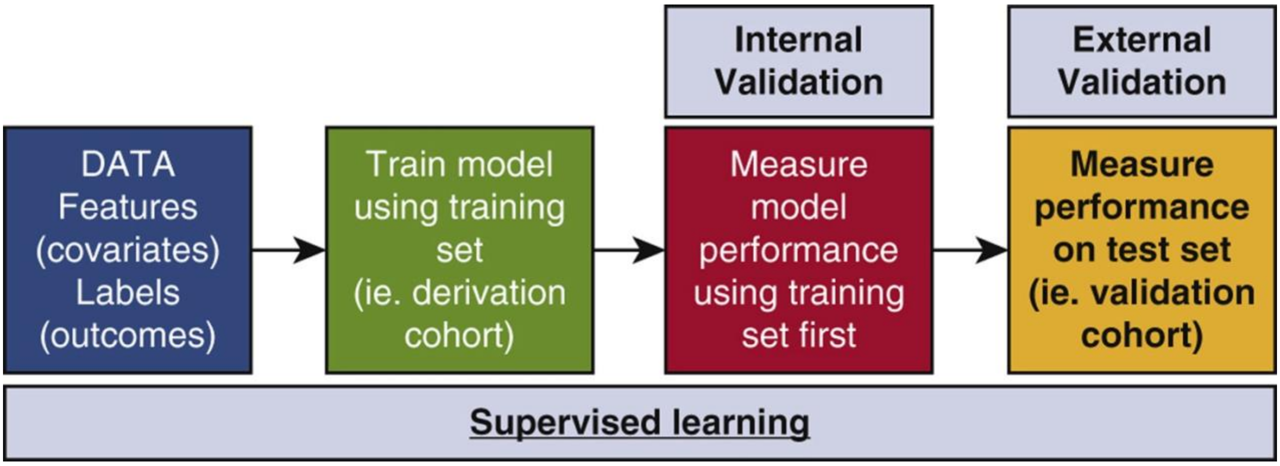

Figure 4: Typical supervised learning workflow, including data preparation, model training, internal validation, and external validation. Reproduced with permission from Tam DY et al. *Artificial intelligence to predict mortality: The rise of the machines*. J Am Coll Cardiol. 2020;75(23):2844–6. License Number: 6016521038349 (10)

Results:

I. Wall Thickness Measurements:

Accurately quantifying the maximum wall thickness (MWT) of the left ventricle is crucial for diagnosing, managing, and risk stratifying patients with hypertrophic cardiomyopathy (HCM).

However, as previously mentioned, this measurement remains susceptible to human variability (8). Addressing this challenge, Augusto et al. conducted a study comparing the reproducibility of an AI-driven machine learning model with assessments made by 11 international expert observers. Their analysis included 60 adult patients with HCM, each undergoing paired CMR scans on the same day, resulting in a dataset of 1,440 individual MWT measurements (8).

The AI system demonstrated superior consistency compared to the human experts. Specifically, the model achieved a test-retest difference of 0.7 mm (SD 0.6), whereas variability among the experts ranged from 1.1 mm (SD 0.9) to 3.7 mm (SD 2.0), depending on the observer (8). The coefficient of variation (CoV) was also lower for the AI model at 4.3%, compared to 5.7% to 12.1% for the human group (8). Figure 5 visually illustrates this difference, showing the AI model's narrower limits of agreement relative to each expert. Furthermore, Bland-Altman analysis reinforced the AI's advantage, with limits of agreement between −2.0 mm and 1.7 mm, while the least consistent expert showed limits exceeding 10 mm (8).

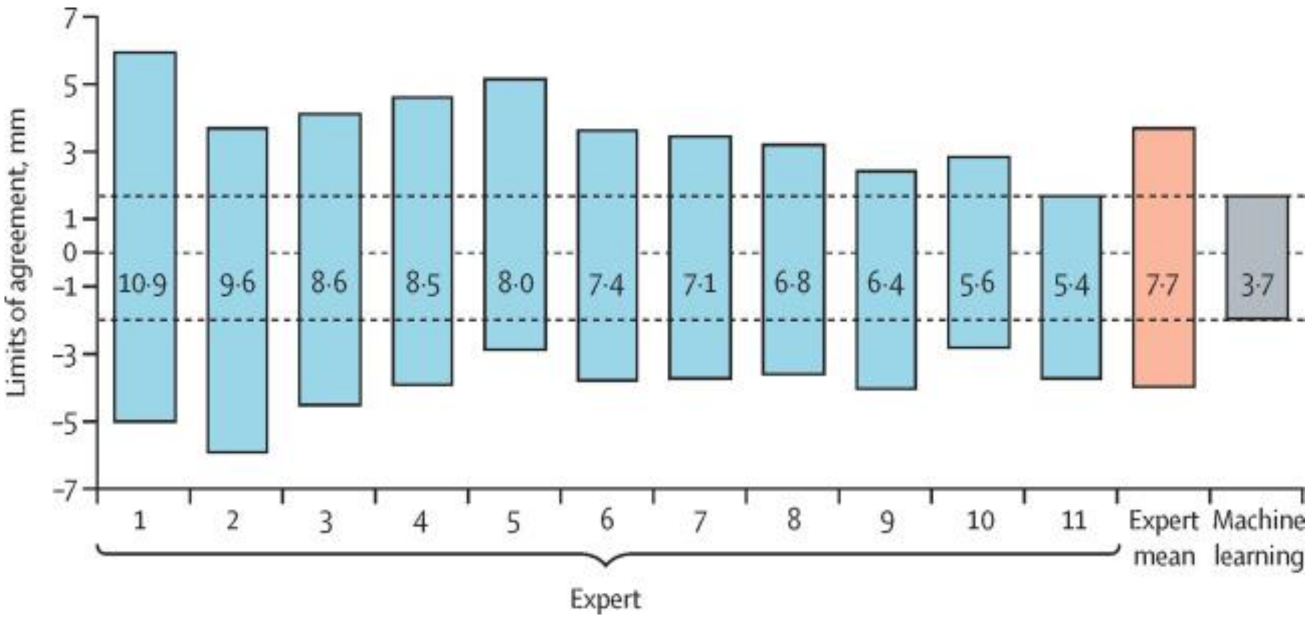

Figure 5: Bland-Altman plots comparing test-retest agreement for maximum wall thickness (MWT) between AI and 11 international experts. (Reproduced with permission from: Augusto JB, Davies RH, Bhuva AN, et al. *Diagnostic performance of machine learning in hypertrophic cardiomyopathy using cardiovascular magnetic resonance*. JACC Cardiovasc Imaging. 2020;13(5):1125–1136. Licensed under CC BY 4.0.) (8)

The diagnostic consistency of AI also outperformed that of the experts. For the HCM threshold of MWT >15 mm, AI reclassified only 8% of patients between test and retest. In contrast, human experts reclassified between 7% and 20% of patients (8). Across the cohort, AI diagnosed 64% of patients with HCM, aligning within the broader and more variable range of 45% to 83% reported by experts

(8). These results highlight the significant inconsistency that remains in manual interpretation. This variability is further illustrated by the variation across all observers (Figure 6), in the example of extreme septal hypertrophy.

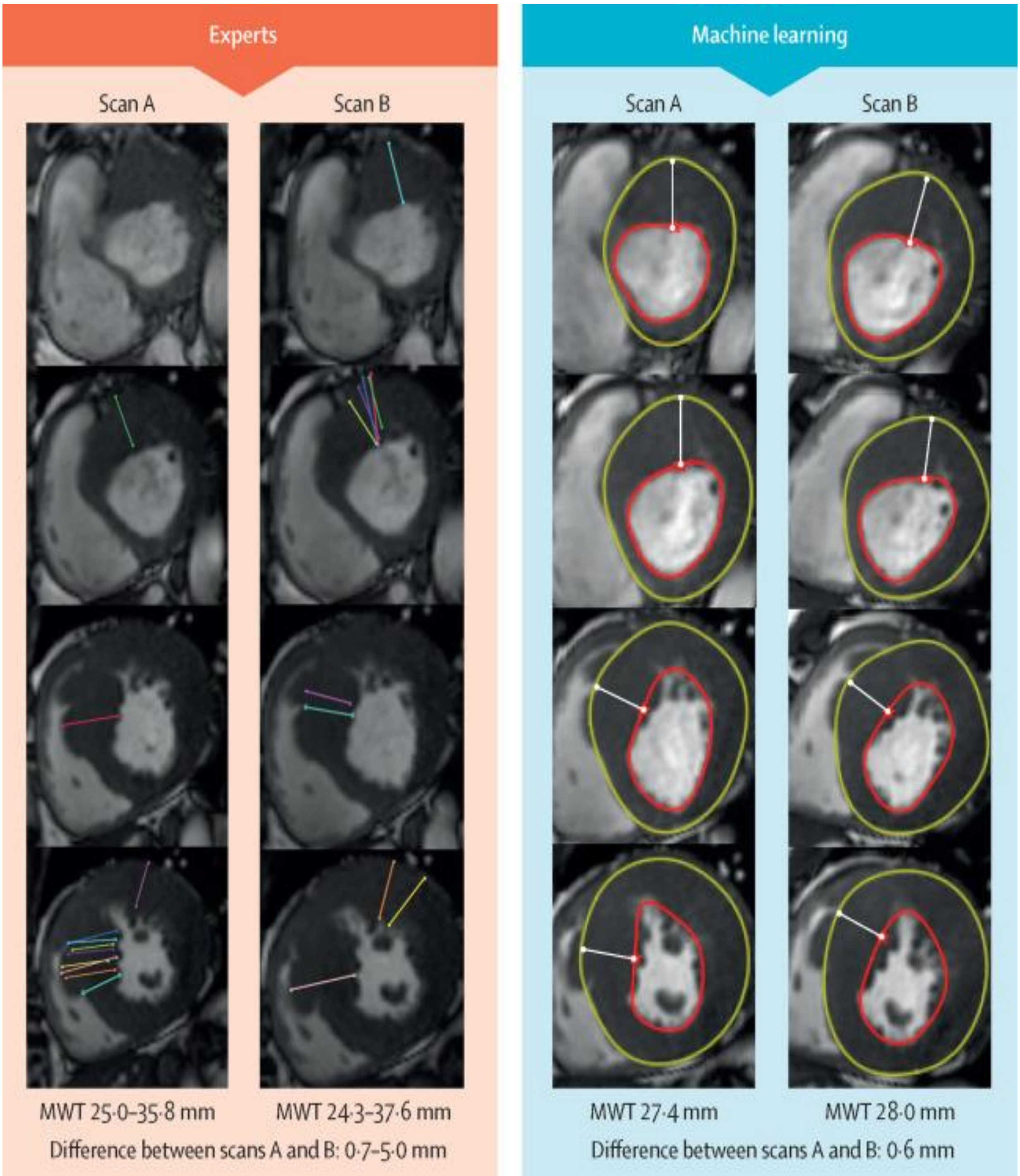

Figure 6: Example of segmentation in extreme septal hypertrophy (Reproduced with permission from: Augusto JB, Davies RH, Bhuva AN, et al. *Diagnostic performance of machine learning in hypertrophic cardiomyopathy using cardiovascular magnetic resonance*. JACC Cardiovasc Imaging. 2020;13(5):1125–1136. Licensed under CC BY 4.0.) (8)

In Figure 6 on the left side, nine expert readers selected different image phases and drew measurements in different locations, leading to a wide range of values and noticeable differences between the two scans. In contrast, the ML algorithm (right side) automatically selected the optimal phase for both scans and produced nearly identical measurements, with less than a 1 mm difference. his example illustrates how AI may help reduce variability and improve the reliability of cardiac imaging, especially in complex cases like hypertrophic cardiomyopathy. AI also showed greater consistency when it came to risk stratification, specifically in ICD decision-making, where it maintained 100% agreement between test and retest scans adhering to the MWT >30 mm threshold. Three human experts altered ICD recommendations on retest, affecting the clinical management of up to four patients (8). The study also reported that the machine learning model demonstrated lower scan–rescan variability compared to the group of 11 experts (8). Furthermore, with regards to HCM Risk-SCD scores, the system was able to produce a minimal change of 0.19% over repeated scans(8). In comparison, expert analysis showed a greater dispersal or results, reaching 0.59% (8). With regards to clinical trials and research, improvements in accuracy and consistency allow for smaller participant numbers, for example being able to detect a change of 2 mm in MWT could require a reduction in sample size by an average factor of 2.3 (8). The observed reduction where in a range from a factor of 1.6 to 4.6 (8). These results are promising for their potential impact on clinical care as well as with a regard to trial efficiency and reducing the related costs. The interclass correlation coefficient (ICC) for human observers was calculated at 0.82 (95% CI 0.69-0.90), with two-thirds of the experts pairs having poor concordance (concordance correlation coefficient <0.90) (8). In one extreme case 2 observers disagreed by 11.4 mm on one single scan (8). The system utilized in this case was based on a convolutional neural network (CNN) that was trained on 1,923 patients with various cardiac conditions, from multiple centers. In order to further bolster their approach, a mathematical approach in the Laplace equation was applied. With this method researchers could establish a dense correspondance between myocardial borders by calculating the maximum distance between them (8).

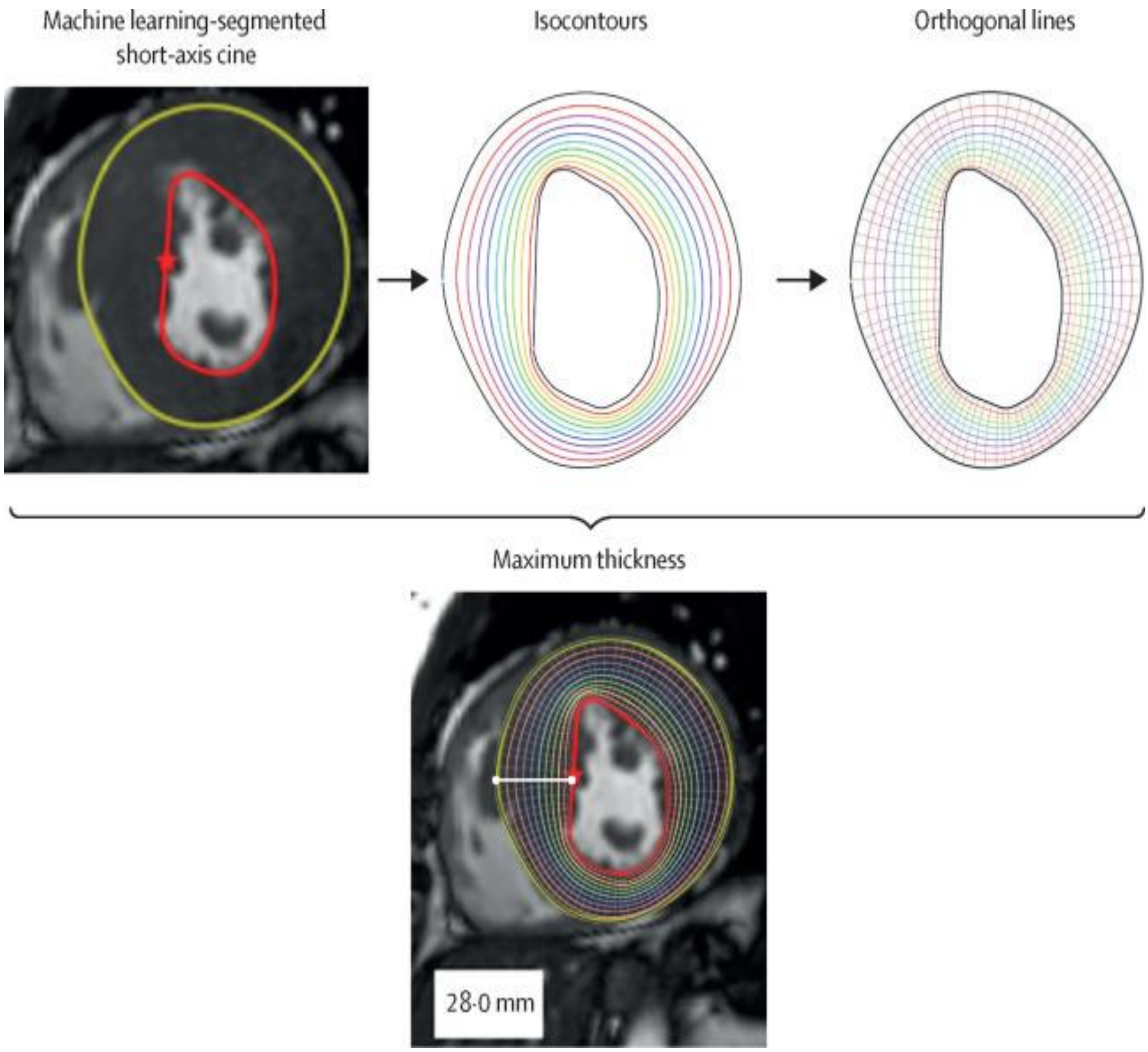

Figure 7: AI-based measurement of left ventricular wall thickness using Laplace's equation. (Reproduced with permission from: Augusto JB, Davies RH, Bhuva AN, et al. *Diagnostic performance of machine learning in hypertrophic cardiomyopathy using cardiovascular magnetic resonance*. JACC Cardiovasc Imaging. 2020;13(5):1125–1136. Licensed under CC BY 4.0.) (8)

As is illustrated in Figure 7, machine learning first outlines the inner and outer walls of the left ventricle at end-diastole. These outlines act as reference points for solving Laplace's equation, which creates a smooth gradient across the heart muscle. From this gradient, the algorithm draws evenly spaced lines that run straight from the inner to the outer wall. Wall thickness is then measured along each line, and the longest one represents the maximum thickness. This approach avoids overlapping or inconsistent measurements and ensures greater accuracy and reproducibility than manual methods.

Human variance was driven by several factors, including differing choices of slice selection, cardiac phase identification, and variability in selecting precise measurement locations. Inconsistent handling of trabeculation and papillary muscle inclusion further contributed to inter-observer differences. By applying standardized rules across all images and cardiac phases, AI was able to avoid a majority of these issues.

In summary, AI based solutions for MWT measurements are on a promising path. Across the bench, results in a range very similar to those of experts where achieved, however there remained some areas in which occasional errors remained. This was a theme that spanned nearly all assessed papers in this portion and specifically underlines the need for expert oversight and manual review in more complex cases, as well as during the development of programs that are AI powered and are meant to improve clinical practice.

II. AI in Volumetric and Mass Assessments:

Artificial intelligence (AI) systems are being used with increasing frequency in the assessment of ventricular volumes and myocardial mass. Some of these systems have begun to tackle human constraints, with a growing body of evidence highlighting improvements across a variety of clinical contexts.

As has been highlighted previously, AI algorithms, specifically Deep Learning architectures like Fully Convolutional Neural Networks (FCNNs), have been shown to robustly agree with manual segmentations performed by trained experts. Bai et al. employed the previously mentioned architecture of an FCN and trained it on a dataset consisting of 4,875 subjects and 93,500 pixelwise annotated images. The results were assessed using technical metrics like Dice coefficients and other clinically relevant metrics like left ventricular end-diastolic volume (LVEDV) and end-systolic volumes (LVESV). Within their large-scale validation study, the employed FCN architecture was able to report Dice similarity coefficients of 0.94 for LV segmentation and 0.90 for right ventricular (RV) segmentation, ranges that are comparable to those of human experts (2).

(a) Absolute difference

	Auto vs Manual (n = 600)	O1 vs O2 (n = 50)	O2 vs O3 (n = 50)	O3 vs O1 (n = 50)
LVSV (mL)	6.1 (5.6)	6.6 (4.1)	5.6 (4.1)	4.2 (3.2)
LVEF (%)	3.2 (2.9)	3.1 (2.1)	3.0 (2.4)	3.8 (1.8)
LVCO (L/min)	0.4 (0.3)	0.4 (0.2)	0.3 (0.2)	0.3 (0.2)
RVSV (mL)	8.1 (6.8)	7.1 (5.5)	5.3 (4.2)	5.4 (4.8)
RVEF (%)	4.3 (3.6)	7.8 (4.4)	3.7 (2.7)	5.7 (3.9)
RVCO (L/min)	0.5 (0.4)	0.4 (0.3)	0.3 (0.2)	0.3 (0.3)

(b) Relative difference

	Auto vs Manual (n = 600)	O1 vs O2 (n = 50)	O2 vs O3 (n = 50)	O3 vs O1 (n = 50)
LVSV (%)	7.0 (5.8)	7.4 (4.1)	6.5 (4.8)	4.8 (3.3)
LVEF (%)	5.4 (4.8)	5.1 (3.7)	4.9 (3.8)	6.6 (3.2)
LVCO (%)	7.0 (5.8)	7.4 (4.1)	6.5 (4.8)	4.8 (3.3)
RVSV (%)	9.6 (8.3)	8.1 (6.9)	6.1 (4.4)	7.1 (8.5)
RVEF (%)	7.5 (6.2)	12.3 (6.6)	6.5 (5.0)	10.7 (7.9)
RVCO (%)	9.6 (8.3)	8.1 (6.9)	6.1 (4.4)	7.1 (8.5)

LVSV: left ventricular stroke volume, LVEF: left ventricular ejection fraction, LVCO: left ventricular cardiac output, RVSV: right ventricular stroke volume, RVEF: right ventricular ejection fraction, RVCO: right ventricular cardiac output.

Table 1: Comparison of AI segmentation variability against human observer (O1, O2, O3) variability across volumetric parameters. (Reproduced with permission from: 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. Licensed under CC BY 4.0.) (2)

The data presented in Table 1 illustrates the variability in volumetric measurements between an automated AI segmentation model and manual expert analysis, as well as among human observers. The "Auto vs Manual" column shows how AI-derived values compare to those obtained by experts across a large dataset of 600 subjects. In contrast, the "O1 vs O2," "O2 vs O3," and "O3 vs O1" columns reflect inter-observer variability between three independent human experts, each reviewing the same 50 cases. Within left ventricular stroke volume (LVSV), ejection fraction (LVEF) and cardiac output (LVCO) the variability achieved was either lower or within that of the experts. The relative difference for right ventricular ejection fraction (RVEF) was seen to be higher within the human observer group. Reaching up to 12.3%, compared to 7.5% for AI versus manual results (2). Results like these further support AI´s potential to reduce subjectivity in cardiac measurements via more consistency. The paper of Hatipoglu et al. assessed a commercially available segmentation tool in a cohort of 300 patients with various cardiac pathologies like hypertrophic cardiomyopathies or ischemic heart disease. The segmentation tool was able to achieve an intraclass correlation coefficient (ICC) of 0.959 for LV EDV and 0.946 for LV ejection fraction (EF) (8). However, reproducibility for RV metrics was somewhat lower, with RVEF yielding an ICC of 0.784, likely because of the RV's complex geometry and anatomical variability (1). Hatipoglu et al. showed that CoV for LVEDV decreased from 9.1% with manual expert analysis to 3.5% with AI-assisted contouring, while CoV for RVEF fell from 9.9% to 7.1% with AI intervention (1). Although undeniably these are improvements, the physicians participating in this reserach underwent a survey on their confidence in the AI-system. While being receptive to new and improved technology, they universally expressed concerns about overall system validity and transparency, issues that will re-emerge later this paper. This before mentioned decrease in the CoV is further visualized in the Bland–Altman plot (Figure 8), which highlights the narrower limits of agreement when using AI for left ventricular volume and ejection fraction calculation.

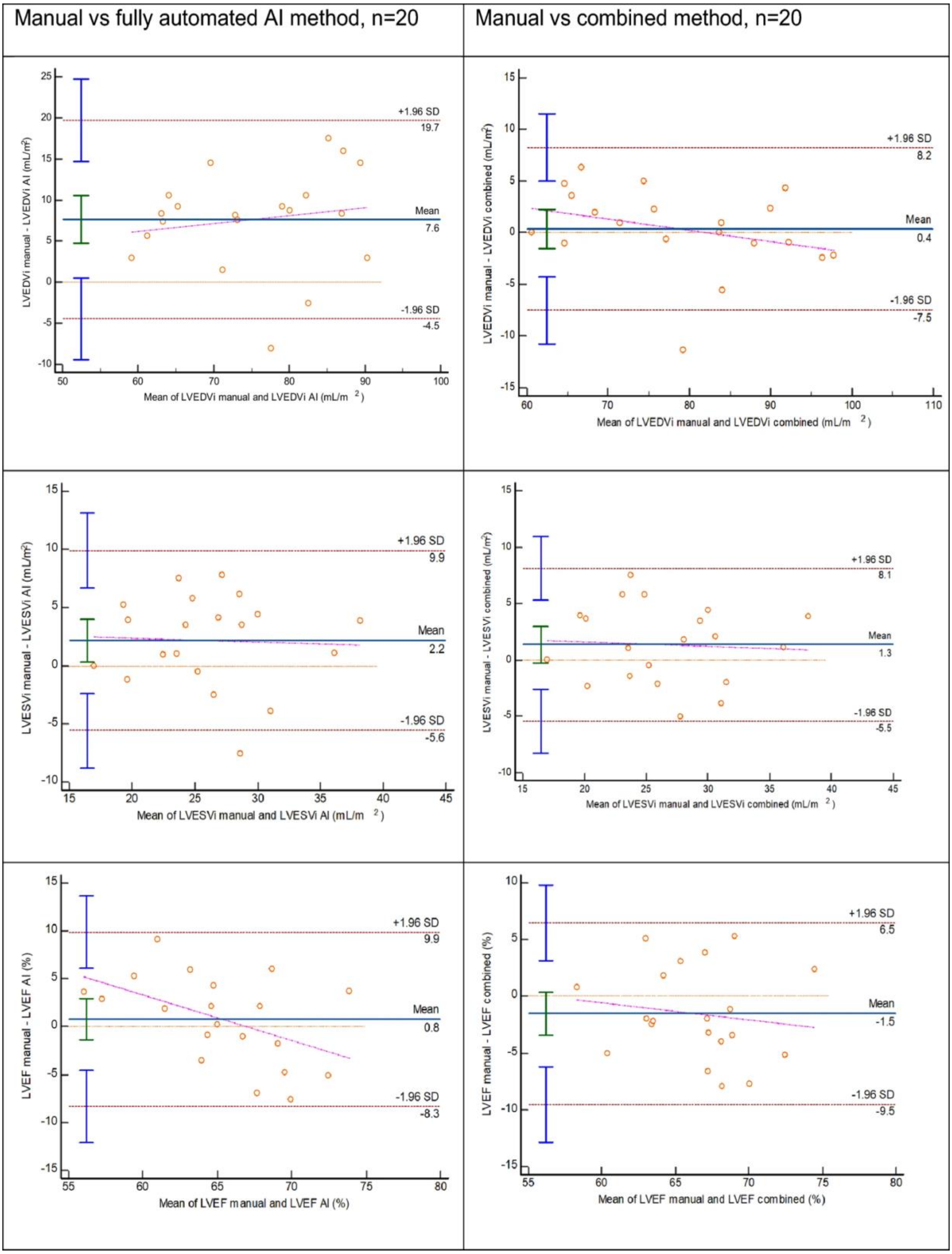

Figure 8: Bland–Altman plot demonstrating improved agreement between manual and AI analysis for left ventricular volumes and ejection Fraction (Adapted from Hatipoglu S., Gatehouse P., Alpendurada F. et al., Int J Cardiovasc Imaging. 2022;38:2413–2424. Licensed under CC BY 4.0.)(1)

Figure 8 presents Bland–Altman plots comparing manual measurements of LV parameters with those derived from a fully automated AI method (left column) and a combined approach that includes manual adjustment of AI-generated contours (right column). In each plot, the blue horizontal line represents the mean difference (bias) between the two methods, while the red dashed lines mark the 95% limits of agreement (±1.96 standard deviations). The pink dotted line illustrates the regression trend, indicating whether the measurement difference varies across the range of values. Each orange dot corresponds to an individual patient comparison. In the top row, the fully automated method tended to underestimate LVEDVi compared to manual analysis, with a bias of 7.6 mL/m², which was reduced to 0.4 mL/m² in the combined method, alongside narrower limits of agreement (1). A similar improvement is seen in the middle row for LVESVi, where the bias decreased from 2.2 mL/m² to 1.3 mL/m² following manual correction (1). The bottom row shows that although the average difference in LVEF was small, the fully automated method exhibited wider variability and a downward trend, which appeared more stable in the combined method. Overall, these observations support that supplementing AI-derived measurements with expert review offers the best results.

The time-intensive nature of manual contouring, requiring on average 10 to 15 minutes per scan, presents significant workflow bottlenecks, especially in high-throughput clinical environments. In contrast, AI models process complete datasets in a fraction of that time. The paper by Böttcher et al. assessed whether an automated volumetric anlysis that is based on a deep learning architecture could reduce segmentation time. With a reduction of segmentation time to a median of 8.4 seconds, they proved that a signifcant increase in efficiency could be achieved (4). Similarly Hatipoglu et al. also supported these findings by reporting that when compared to manual segmentation, AI-based analysis was roughly 42 times faster (1). Such results are promising and could have great impact, especially in centers where large datasets are processed or where the timely anlysis of scans is important.

A great benefit that AI seems to offer is its versatility across a range of pathologies. Whilst accurate results across healthy populations alone are promising, AI based systems have been shown to also be highly accurate in more complex cohorts. For example, Hatipoglu et al. demonstrated consistent performance in segmentation in cases involving dilated cardiomyopathy, hypertrophic cardiomyopathy, ischemic heart disease, and other congenital heart defects (1). Adaptability like this could prove especially helpful in centers that see a greater variety of pathologies, like specialized cardiac care centers. Bai et al. where further able to support this finding by using their fully convolutional network (FCN)-based model on over 1,734 participants in the UK Biobank dataset, showing its capacity to handle population-level imaging studies (2). Being able to scale AI models like this could prove to be highly beneficial when fast anaylsis of large amounts of data is needed.

Despite these clear advantages, several studies have also pointed out recurring limitations that are related to volumetric and mass assessment. Issues related to this are the occasional failure to produce accurate delineation results for endocardial cotours or to pick the innappropriate cardiac phase. This is especially evident in the anylsis of the right ventricle (RV). A possible reasonsn for this could be the greater anatomical variability of the RV. More issues where encountered by Hatipoglu et al., who found that segmentation errors often arose from underfitting of the LV endocardial border or due to inconsistent handling of trabeculations and papillary muscles (1). Bai et al. found that on rare occasions AI could produce biologically implausible measurements, and further underline the need for expert review and oversight (2). With these issues in mind, the proper control mechanisms need to be in place in order for safe and effective adoption of AI supported systems. Also, specifically in mire cimolex cases, the need for expert oversight cannot be overstated.

Another issue that emerged was that of dataset generalizability. Curiale et al. was able to show that models that are trained on a very specific dataset, will perform less effectively when used in local datasets. This could be due to simple discrepancies in scanner technologies, imaging protocols and simply due to different patient demographics (3). Their findings underscore the need for retraining and fine-tuning AI algorithms to ensure that diagnostic performance remains consistently high across all settings.

Overall, AI is not only changing simple imaging tasks but its also proving an effective tool in clinical research and trials. With faster analysis and sufficiently accurate results, the impact on research and clincial trials is not to be overseen. More efficient and accurate segmentation allows for more effective servicing of large-scale studies as it greatly reduces the manual labor required for the analysis of cardiac anatomy and allows for more standardized collection of data. Bai et al. pointed out that their AI system's reproducibility was critical for use in multi-center trials where inter-site variability could otherwise confound volumetric measurements (2).

The literature so far suggests that AI can offer marked improvements over manual segmentation, particularly in terms of speed, reproducibility, and consistency. LV quantification consistently demonstrates higher reliability across studies, whereas RV assessment continues to pose a greater challenge, however with AI still achieving clinically acceptable results. Papers like that of Bhuva et al. highlighted the human limitations by noting that much of the error in expert analysis stems from observer-related factors, further underscoring that automated methods could play an important role in standardization of measurements (7). AI is increasingly regarded as a transformative tool in medical imaging. Deep learning models provide exceptional efficiency and reproducibility, however their application is most effective when combined with human expert oversight, to ensure that occasional segmentation errors can be dealt with. The evidence strongly supports the integration into clinical practice, where it has the potential to streamline numerous processes.

III. AI for Myocardial Scar Detection and Quantification

Artificial intelligence (AI) has shown considerable promise in the automated detection and quantification of myocardial scar and fibrosis on cardiac magnetic resonance (CMR) imaging. Two of the selected studies, one by Popescu et al. and the other by Zhang et al., assess AI's capability in improving the efficiency, accuracy, and through that the reproducibility of myocardial scar assessments.

Popescu et al. used the Anatomical Convolutional Segmentation Network (ACSNet). ACSNet is a DL model that is able to automatically segment LV myocardium, blood pool and scar regions in late gadolinium enhanced (LGE) CMR images. This program achieved segmentation results that had a balanced accuracy (BA) of 96% for the identification of LV region of interest (ROI) (6). Additionally, ACSNet achieved a dice coefficient of 0.93 for the segmentation of the LV and 0.79 for myocardium segmentation (6). When applied to scar regions, ACSNet achieved a Dice coefficient of 0.57 and a BA of 75%, clearly outperforming previous interobservers and many other existing algorithms.

This minimal difference suggests that AI can closely replicate expert performance, while significantly reducing manual workload and variability. Importantly, the model performed reliably across different regions of the LV, including the apex and base areas where traditional methods often struggle due to high variability between observers. (6). This is visually demonstrated in Figure 8, which shows ACSNet's scar segmentation closely matching manual contours across LV regions, also including areas prone to high interobserver disagreement.

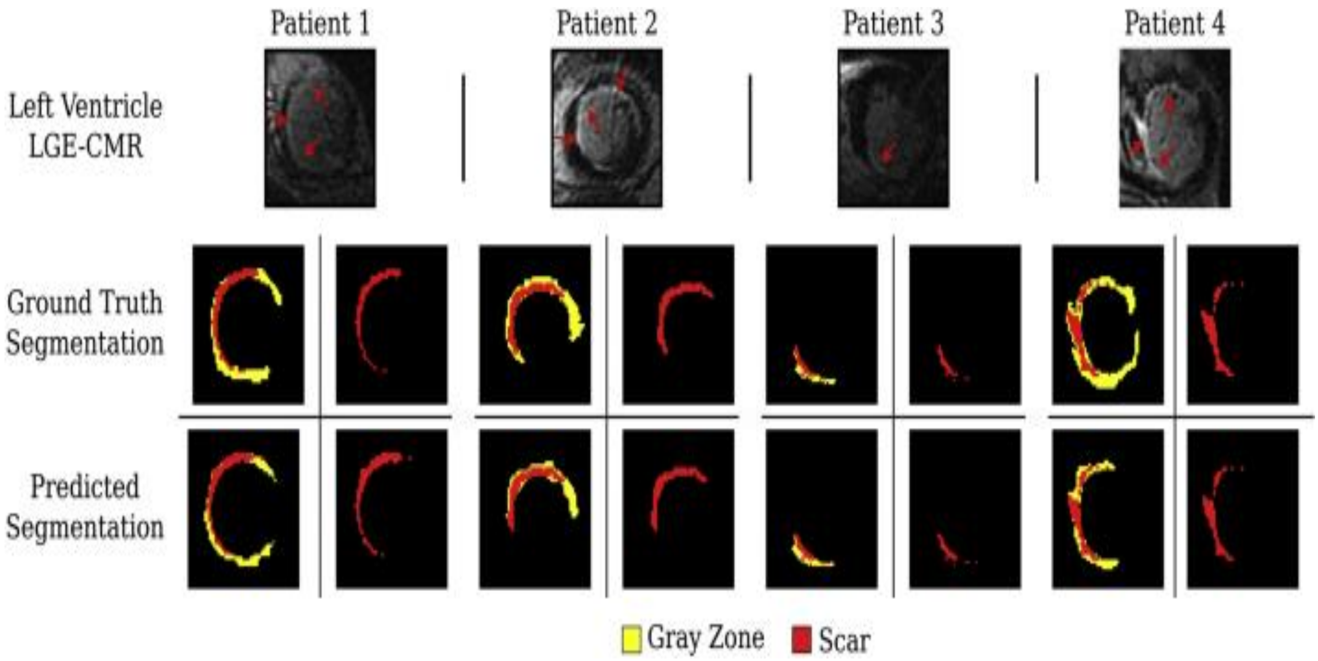

Figure 9: ACSNet-based scar segmentation results overlaid on late gadolinium enhancement (LGE)-CMR images. The figure illustrates the original image in the first row (Left Ventricle LGE-CMR), the manually segmented portion in the middle row (Ground Truth Segmentation)

followed by the ACSNet Predicted Segmentation. (Adapted with permission from: Popescu DM, et al. *Anatomically informed deep learning for myocardial infarction segmentation from late gadolinium enhancement cardiovascular magnetic resonance*. J Cardiovasc Magn Reson. 2022;24(1):14. Licensed under CC BY 4.0.) (6)

As is visualized in the Figure 9, the top row shows the original short-axis CMR scans; the middle row depicts ground truth annotations for scar (red) and gray zone (yellow); and the bottom row displays ACSNet's automated predictions. Patients 1–3 show typical cases with high concordance between predicted and manual segmentations. This can be seen by comparing the Gray Zone and Scar predicted areas to the ground truth row. Patient 4 is used to represent an outlier with reduced gray zone accuracy. Although the ACSNet undermeasured the Gray Zone of patient 4, it was still able to accurately predict scar tissue, demonstrating the robustness of the programm.

In contrast to ACSNet, Zhang et al. focused on developing Virtual Native Enhancement (VNE), a novel AI technology capable of generating LGE-like images without the need for contrast agents. VNE combines LGE-equivalent images from cine CMR and native T1 mapping sequences using a generative adversarial network (GAN) (5). This contrast-free approach offers numerous practical advantages, including reduced scan time, lower cost, and avoidance of contrast-associated risks.

VNE showed a high correlation with traditional LGE-CMR for scar quantification. In the independent test set of 66 patients with prior myocardial infarction (MI), VNE achieved a Pearson correlation coefficient (R) of 0.89 and an intraclass correlation coefficient (ICC) of 0.94 for myocardial scar size, when compared to LGE-CMR analysis (5). These findings indicate a strong correlation, which is further illustrated in Figure 10, where VNE and LGE images of scar distribution are practically identical. This is further shown by the bullseye transmurality plots.

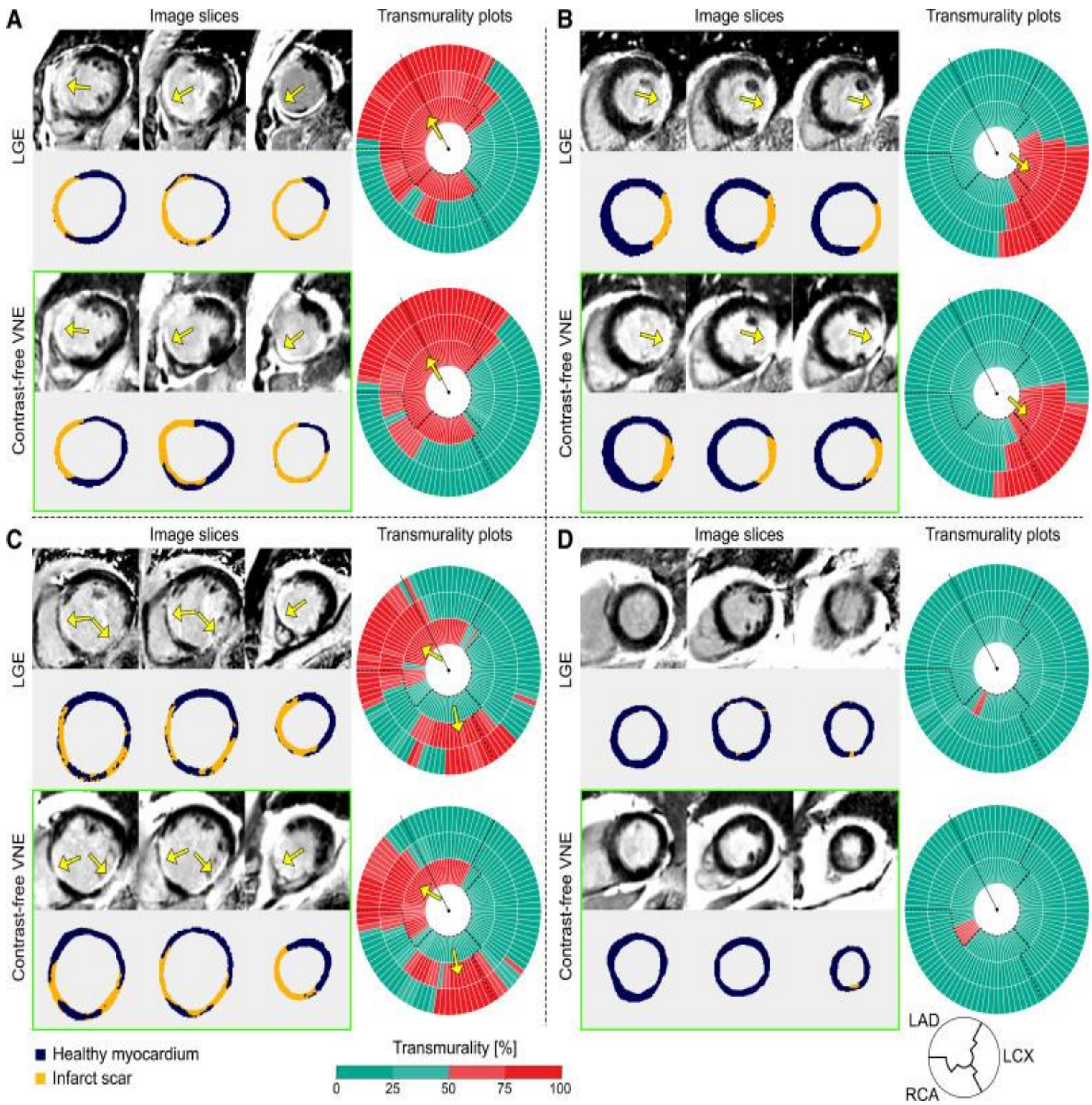

Figure 10: Side-by-side comparison of VNE-generated scar maps and traditional LGE-CMR images, including bullseye plots for scar transmurality analysis. (Reproduced with permission from Zhang Q, Burrage MK, Shanmuganathan M, et al. Artificial Intelligence for Contrast-Free MRI: *Scar Assessment in Myocardial Infarction Using Deep Learning–Based Virtual Native Enhancement.* Circulation. 2022;146(20):1492–1503. (CC BY 4.0.) (5)

The bullseye plots presented in Figure 10 illustrate that the VNE system was able to very strongly match localization of scar tissue when compared to the manual method.

Additionally, transmurality measurements, critical for viability assessments, also demonstrated strong agreement, with R = 0.84 and ICC = 0.90 (5). This performance suggests that VNE could potentially replace LGE imaging in specific clinical scenarios, offering a rapid and contrast-free solution.

Both analyzed studies emphasize the advantages of AI in enhancing image quality and achieving more consistent segmentation. For example, Popescu et al. demonstrated that ACSNet performed better than interexpert segmentation, maintaining reliable accuracy even in anatomically ambiguous regions of the left ventricle (6). Similarly, Zhang et al. noted that VNE images received higher image quality ratings than standard LGE, with all five blinded expert reviewers scoring VNE significantly better than LGE in a multiobserver analysis.

Beyond image quality and accuracy, AI methods demonstrated meaningful clinical implications. ACSNet's scar quantification was integrated with established diagnostic criteria, facilitating direct extraction of clinical features such as scar burden, which plays a key role in sudden cardiac death (SCD) risk stratification and therapeutic planning (6). Similarly, VNE's scar quantification closely matched LGE-derived results on bullseye plots, which depict scar transmurality and guide viability-based revascularization decisions.

Although there where positive outcomes, both studies identified important limitations. ACSNet produced lower Dice scores (0.57) when segmenting dense scar tissue when compared to myocardium (6). This suggests that further refinement is necessary. Whilst not showing any false-positive identifications, it sensitivity of 77% for scar detection and produced several false-negatives. These false-negatives seemed to mainly occur in patients who previously had subendocardial infarctions (5). This emphasizes the need for expert oversight with complementary clinical data.

In conclusion, both ACSNet and VNE represent significant strides toward improving myocardial scar detection and quantification. The reviewed AI systems have been proven to enhance reproducibility whilst also tackling the issue of interobserver variability. Furthermore, approaches like VNE's contrast-free capabilities could make advanced CMR diagnostics more accessible, especially in settings where contrast use is contraindicated.

Discussion:

This systematic review compared the performance of artificial intelligence (AI) and human experts in cardiovascular magnetic resonance (CMR) imaging with a focus on three key areas: myocardial wall thickness measurements, volumetric and mass assessment, and myocardial scar detection and quantification. Across all domains, AI produced results that were in strong agreement with those of trained experts. This approach also offered advantages in speed, reproducibility and overall scalability, which could have great implications, discussed later on.

A sector in which AI had great success in reducing inter-observer variabilities was maximum wall thickness (MWT) measurements, with the tested systems perform particularly good in left ventricular maximum wall thickness (MWT). The clinical significance of accurate MWT measurements cannot be denied, especially when diagnosing hypertrophic cardiomyopathy (HCM) and or risk stratifying patients for sudden cardiac death (SCD). The review was able to show that through AI integration, the variance between measurements in human anlysis, which was sometimes as high as 10 mm within one scan, could be significantly reduced (8).

An important finding was that AI systems achieved narrower limits of agreement and generally more consistent classification of patients who are around a critical clinical threshold lie the 15 mm MWT cut-off for HCM diagnosis and the >30 mm cut-off for SCD risk stratification (8).

Notably, AI managed to avoid diagnostic reclassification errors which are often seen among human observers. These results suggest that AI could help reduce uncertainty in clinical decisions, such as the recommendation for implantable cardioverter-defibrillators (ICD). However, even though AI was able to achieve excellent reproducibility, occasionally it created biologically implausible contours in less than 1% of cases, supporting the need for human control, especially when scan quality is suboptimal (8).

AI also demonstrated significant advancements in volumetry and myocardial mass assessment, outperforming traditional manual workflows in both reproducibility and speed. Multiple studies reported that AI models, particularly those leveraging fully convolutional networks (FCNs), achieved Dice coefficients exceeding 0.90 for left and right ventricular segmentation. This is well within the range of human inter-observer agreement (2,3). Intraclass correlation coefficients (ICCs) were also high for left ventricular (LV) end-diastolic volume index (EDVi) and ejection fraction (EF), suggesting that AI segmentation can reliably contour in clinical and research settings (3) With the issue of observer variabilities continously reemerging, a key strength of AI assisted volumetric assessment is the ability to achieve consistent results acorss scans and thus reduce these variabilites. One of the variables often used to assess AI accuracy is the coefficient of variation (CoV). This value indicates the variance between results, and in the quantification of the left ventricle, AI achieved overall lower CoV values when compared to manual anaylsis. Reductions in variation like this could help improve diagnostic accuracy and the reliability of produced results. This could become particularly relevant when applied to clinical trials and research. In the following figure, the metrics along which the AI-systems where assessed are summarized.

CMR Task	AI CoV (%)	AI Dice	AI ICC	Human CoV (%)	Human Dice	Human ICC	Overall AI vs Human Performance
Wall Thickness (MWT)	4.3 (8)	—	—	5.7–12.1 (8)	—	—	AI superior in reproducibility
Volumetry (LV EDVi)	—	0.94 (2)	0.959 (1)	—	0.92–0.94(2)	0.92–0.95 (1)	Comparable, AI slightly faster
Volumetry (RV EF)	—	0.90 (2)	0.784 (1)	—	0.87-0.89 (2)	0.643 (1)	AI performs well; RV remains challenging
Myocardial Scar Burden	—	0.57 (6)	0.94 (5)	—	0.50–0.55 (6)	0.90 (5)	AI matches or exceeds human variability

Table 2: Comparison of AI and human expert performance in CMR analysis. Metrics include Dice coefficient, intraclass correlation coefficient (ICC), and coefficient of variation (CoV), where higher Dice and ICC or lower CoV indicate better performance. Data adapted from references (1–9).

Table 2 summarizes that in all metrics assessed, AI empowered anaylsis was able to achieve high level of agreement with expert analysis. Intraclass correlation coefficients (ICC) where equally strong, specifically for scar burden and LV volume, with Ai reaching a level of 0.94 (2). An area in which AI was able to achieve more reproducible results when compared to experts was maximunm wall thickness assessment. This is reflected by the lower coefficient of variation (CoV) compared to human analysis: 4.3% vs. 5.7-12.1 (8). The results visualized in this table highlights the ever growing role of AI in CMR imaging.

Moving on from variability, efficiency is another very important factor in the assessment of volumetry and mass. Traditionally, the manual contouring process could take anywhere from 10-15 minutes (4). Some of the reviewed AI solutions where able to routinely process large set of data in under 30 seconds (1,4). This gain in efficiency holds not only the potential to reduce the workload of radiologists, but also to improve turnaround times and allows for the saved time to be spent on more cases in which human expertise is necessary. Despite these benefits, some issues remain. Mainly when looking at the segmentation of the right ventricle, it seems to be more error prone due to its complex anatomy and the presence of trabeculated myocardial borders which can sometimes skew the results. Other issues that emerged where the selection of incorrect phases during cardiac cycle analysis as well as the under segmentation of the endocardial border. These issues further support the need for the incorporation of safeguards and other validation mechanisms to enable the best possible outcomes.

In addition to segmenting cardiac structures, AI has also been applied to functional analysis, most notably in the assessment of myocardial strain using feature tracking (FT). FT is a post-processing technique that measures how the heart muscle deforms during the cardiac cycle, using routine cine-CMR images without the need for additional sequences (11). This enables evaluation of strain parameters such as global longitudinal strain (GLS), radial strain (GRS), and circumferential strain (GCS), offering insight into myocardial function.

Several recent studies have evaluated the performance of AI-assisted FT-CMR for strain analysis across different cardiac conditions. Gröschel et al. showed that AI-generated strain values closely matched expert manual measurements in healthy individuals, but accuracy was reduced in patients with left ventricular hypertrophy, particularly in basal and lateral segments (12). Pryds et al. found that FT-CMR correlated well with speckle tracking echocardiography for GLS, although systematic differences in strain values were observed, suggesting the two methods are not interchangeable (13). While global strain values were generally reliable, localized tracking errors, especially near the papillary muscles, remained a challenge. Despite these limitations, AI-enhanced FT-CMR appears to offer reproducible and efficient functional assessment, particularly useful in cases where strain abnormalities have diagnostic or prognostic value like HCM and Heart-failure with preserved ejection fraction (HfpEF).

When applied to myocardial scar detection and quantification, AI solutions showed equally promising results. Popescu et al.'s ACSNet model and Zhang et al.'s Virtual Native Enhancement (VNE) technique both demonstrated high concordance with expert manual assessments of scar burden and transmurality, key metrics used in the management of ischemic heart disease and viability-based decision-making for revascularization. VNE works by combining data from cine CMR images and from native T1 maps to create a LGE-equivalent image (5). This, if consistently accurate would introduce a novel, contrast-free approach to cardiac imaging. This technique could benefit patient populations where gadolinium administration is contraindicated, for example those with chronic kidney disease. VNE was able to achieve a Pearson correlation coefficient of 0.89 and an ICC of 0.94 for scar size compared to traditional LGE-CMR (2). The levels of agreement for transmurality metrics where also similar (5). Meanwhile, ACSNet was able to provide highly accurate segmentation and reduced variability across the apical and basal regions of the myocardium. These specific regions are known to be prone to observer discrepancies.

From a clinical perspective, AI´s ability to streamline scar quantification processes while also maintaining diagnostic accuracy could prove transformative. AI based scar quantification could directly inform clinical risk stratification tools such as SCD risk scores and viability assessment frameworks. Despite many positive advancements, several limitations remain. Popescu et al.'s ACSNet model for example reported lower Dice scores when segmenting dense scar tissue when compared to myocardium. Similarly in Zhang et al.'s assessment, their VNE model also demonstrated decreased sensitivity when detecting small subendocardial infarctions. Faults like these both underline the need for proepr validation of results.

Overall the broader implication for AI integration remain complicated, however what has been showed is that being able to employ AI tools to standardize CMR interpretation could help enhance diagnostic consistency and so reduce variability. Both those factors could greatly influence multi-center studies, an area in which the afore mentioned benefits of AI could hlelp researchers reach adequate statistical power with smaller samplem sizes. Such gains in turn could contribute to substantially lowering costs that arer associated with studies that aim to detect only small changes in cardiac structure or function.

Nevertheless, the adoption of AI in clinical CMR practice introduces a range of ethical and regulatory challenges. One major concern of the implementatiomn of AI-systems is the "black box" phenomenon, which limits interpretability and may erode clinician trust due to a lack of understanding of the implemented systems. This lack of transparency can complicate the communication between physician and patient, especially when AI-derived insights inform high-stakes decisions. An additional risk is that AI models trained on homogenous datasets could introduce biase and thereby affect the diagnostic performance in certain patient groups. Addressing concerns like these will require both improved and more diverse datasets for training and the implementation of fairness metrics during development.

Due to these many factors that influence the successful integration of AI systems, regulatory agencies like the FDA and EMA are emphasizing the need for external validation as well as close surveillance to ensure proper patient safety and diagnostic relaibility. Due to the nature of this being an emerging field these frameworks are still evolving and will hopefully play an important role in balancing benefits with ethical considerations.

In parallel with developments in segmentation and classification-focused AI models, a promising new direction is emerging in CMR imaging: radiomics. Although the primary focus of this thesis has been on AI approaches for segmentation and quantification, radiomics represents an emerging complementary field. It involves the high-throughput extraction of numerous imaging features, such as shape, texture, and intensity, that can be correlated with clinical outcomes, disease phenotypes, or histopathological findings using statistical or machine learning techniques. Unlike conventional AI segmentation algorithms, which primarily focus on delineating anatomical structures, radiomics attempts to reveal underlying tissue heterogeneity that may not be visually representable. Within the CMR domain, radiomics has shown promise in areas such as fibrosis detection, myocardial tissue characterization, and outcome prediction. For instance, Nakamori et al. demonstrated that radiomic texture features derived from late gadolinium enhancement (LGE) images were not only capable of distinguishing between ischemic and non-ischemic cardiomyopathy but also showed superior performance compared to traditional assessment (14). More importantly, in a cohort of patients with recent-onset dilated cardiomyopathy, these features provided enhanced discrimination between noncollagenous extracellular matrix expansion which is often associated with myocardial inflammation, and mild to moderate collagenous fibrosis. Differentiations like this are clinically significant, as they may guide early therapeutic strategies aimed at targeting inflammation before irreversible fibrotic remodeling occurs. Building on these findings, Neisius et al. demonstrated that radiomic analysis of native T1 mapping can enhance diagnostic precision by detecting subtle textural differences in myocardial tissue (15). Notably, their model successfully distinguished hypertensive heart disease from hypertrophic cardiomyopathy, even when conventional T1 values appeared within normal ranges. This suggests that radiomics may uncover microstructural fibrosis patterns missed by traditional analysis, offering promise for earlier and more accurate disease differentiation. This manual limitation is clearly visualized in Figure 11, where native T1 and ECV values significantly overlap between non-collagenous and collagenous myocardial expansion groups, highlighting the challenge of differentiating early fibrotic changes using conventional methods.

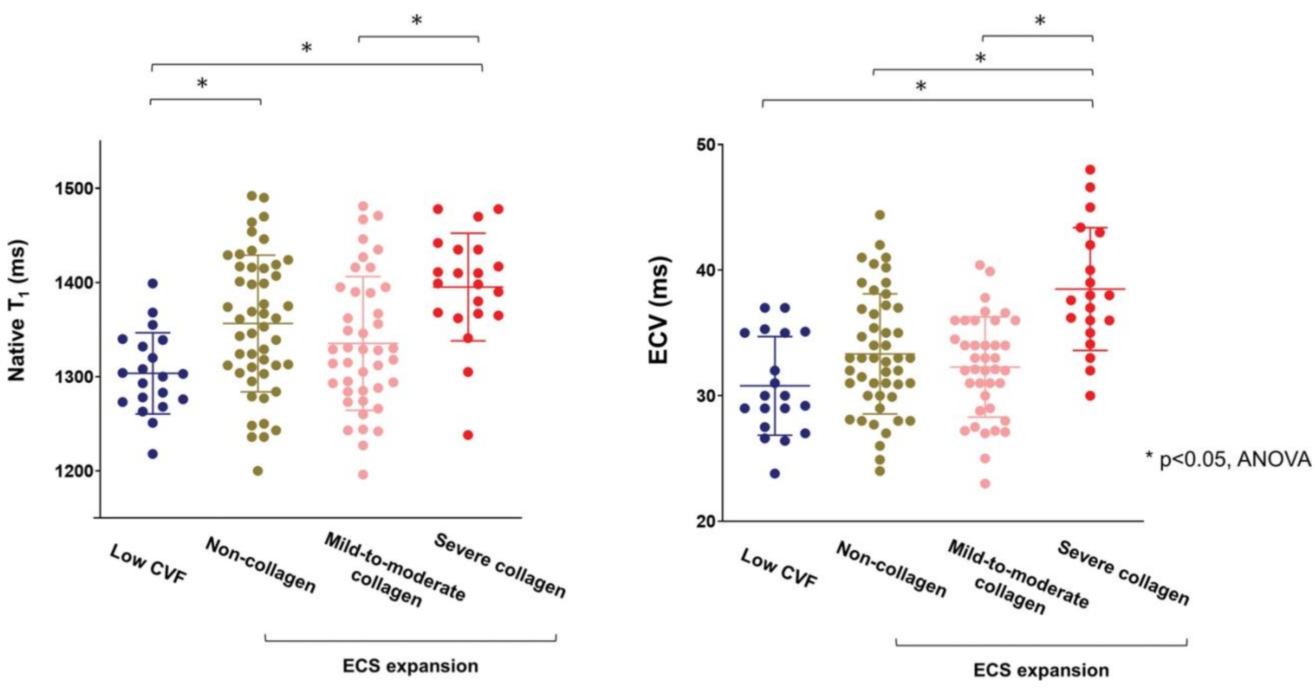

Figure 11: Comparisons of native T₁ and ECV among 4 different histopathological phenotypes (*Adapted with permission from Nakamori S, Amyar A, Fahmy AS, et al. Cardiovascular magnetic*

resonance radiomics to identify components of the extracellular matrix in dilated cardiomyopathy. Radiology. 2024;310(1):96–107. © Radiological Society of North America (License Nr.: 6014750007676*) (14)*

Figure 11 illustrates the limitations of conventional tissue markers, native T1 and extracellular volume (ECV), in distinguishing between histopathological subtypes of myocardial extracellular expansion. While both metrics increase with the severity of fibrosis, there is substantial overlap between non-collagenous and early collagenous expansion, making precise differentiation difficult. This overlap highlights a key challenge in conventional CMR tissue characterization and underscores the potential value of radiomics. By extracting higher-order image features beyond average signal intensity, radiomic analysis may improve discrimination between inflammation-related and fibrotic myocardial changes. This could be especially valuable in conditions like early dilated cardiomyopathy, where clinical management hinges on accurate tissue phenotyping.

From a methodological perspective, radiomics often involves handcrafted feature extraction followed by machine learning classifiers. This contrasts with the deep learning models discussed earlier in this thesis, which typically learn features directly from the image data. As outlined by Mayerhoefer et al., radiomics features can be grouped into various classes (16).

Features like these are inherently sensitive to image acquisition and reconstruction parameters, and without standardization this could limit reproducibility across sites and institutions. Raisi-Estabragh et al. also underlined the need for robust feature selection in order to avoid overfitting and to ensure appropriate interpretation (17). This distinction suggests that radiomics offers complimentary value, especially if further stratified with genomic data in multimodal approaches. Similarly to simple segmentation tasks, some issues currently remain in radiomics that limit the current clinical utility. Many studies are retrospective, rely on small datasets and lack any external validation. Furthermore, the absence of standardized protocols for image acquisition, feature extraction, and statistical modeling negatively impact reproducibility. Mayerhoefer et al. underscore the sensitivity of radiomic features to technical factors such as reconstruction algorithms, and segmentation methods, which can distort feature values and limit comparison of data across different studies (16). Raisi-Estabragh et al. also share these concerns in the context of CMR, noting that radiomics models often suffer from lack of generalizability due to the high variance in imaging protocols (17). They emphasize that without rigorous validation and standardization, clinical adoption will remain limited. These issues are critical to address if radiomics is to progress beyond academic research into routine clinical care.

As these field evolve, combining technological development with clinical guidance is going to be essential for successful implementation. A recent multisociety position paper by Mastrodicasa et al.

created a structured framework for assessing the clinical maturity of AI technologies in cardiovascular imaging using the Technology Readiness Level (TRL) model. The created scale ranges from early experimental concepts (TRL1) to full clinical deployment (TRL9) (18). This scale provides a valuable perspective on the advancement of specific AI-based systems in cardiac imaging. Additionally their anaylsis showed that some tasks like automated biventricular segmentation are already nearing or have reached full implementation, while the majority of others remain in earlier stages of clincal readiness. These findings point to the value of a cautious, yet deliberate approach to integrating AI tools, tailored to the current maturity of the technology. Figure 12 provides a concise visual summary of AI readiness across cardiac MRI and CT domains, helping contextualize which tools are closest to clinical translation and which still require substantial validation.

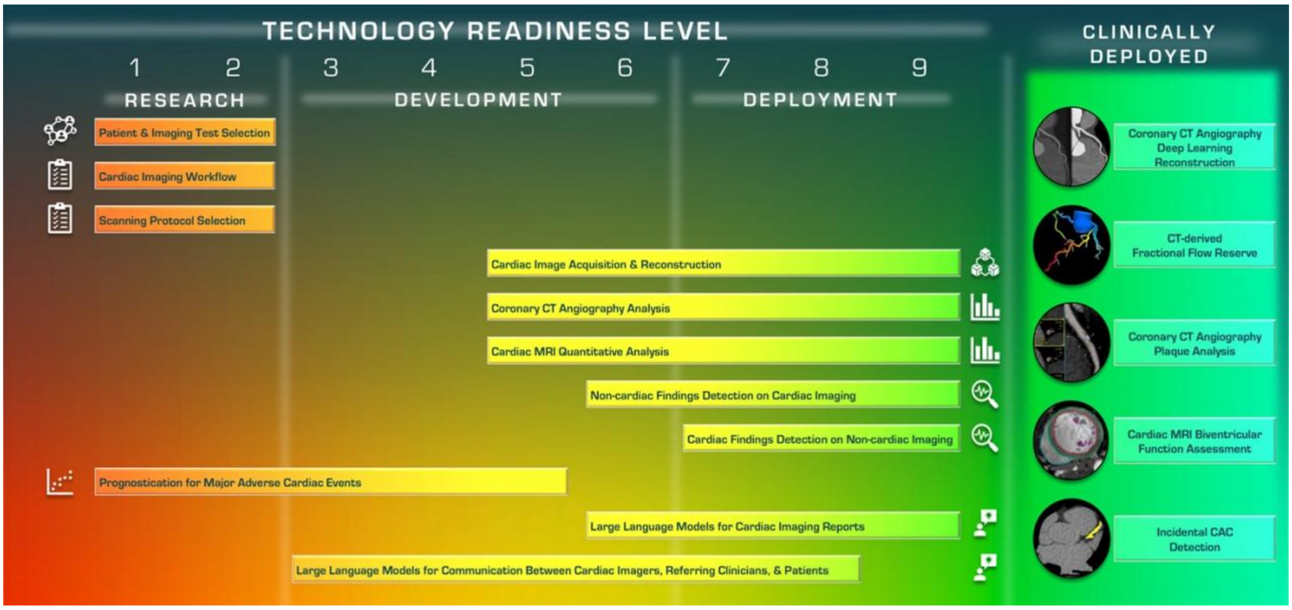

Figure 12: Technology readiness levels (TRLs) from 1 (early-stage research) to 9 (clinical implementation) across AI applications in cardiac MRI and CT. (Reproduced with permission from Mastrodicasa D, van Assen M, Huisman M, et al. *Use of AI in Cardiac CT and MRI: A Scientific Multisociety Statement*. Radiology. 2025;314(1):e240516. © RSNA, used under CC BY 4.0.) (18)

Looking to the future, further research should prioritize the development of AI models with improved generalizability across different scanners, imaging protocols, and patient populations. Although current evidence shows that AI systems perform best when fine-tuned to local datasets, the push for externally validated, multi-institutional models is essential for wider clinical adoption. Additionally, as regulatory bodies increasingly scrutinize AI-driven diagnostics, ensuring model transparency and explainability will be crucial to building clinician trust and securing formal approval pathways.

Finally, the most plausible way of integrating AI in CMR imaging is the adoption of a system where AI handles routine, high-volume tasks such as volumetry, scar quantification and wall thickness measurements, while human experts focus on oversight, the interpretation of complex cases and clinical decision-making. By effectively combining AI systems with human oversight, cardiac imaging could benefit greatly, as well as improving patient outcomes and optimizing overall healthcare delivery and research productivity.

Conclusion:

This thesis systematically examined the role of artificial intelligence (AI) in cardiovascular magnetic resonance (CMR) imaging, focusing on myocardial wall thickness, volumetric and mass assessments, and myocardial scar detection. Across these areas, AI consistently demonstrated improvements in reproducibility, efficiency, and diagnostic consistency when compared to traditional methods. In the context of hypertrophic cardiomyopathy, AI significantly reduced inter-observer variability in wall thickness measurements, supporting more reliable classification of patients around key clinical thresholds, such as sudden cardiac death risk. Volumetric and mass quantification also benefited from AI-driven segmentation, achieving strong concordance with expert values while markedly improving processing times. Although right ventricular analysis remains more technically complex, AI models continue to evolve and show promise in narrowing this performance gap.

In myocardial scar detection, contrast-free techniques such as Virtual Native Enhancement have shown high agreement with conventional late gadolinium enhancement, potentially offering safer alternatives for patients with contraindications to contrast agents. In addition to these segmentation-based applications, radiomics has emerged as a complementary AI method, enabling the extraction of high-dimensional imaging features linked to fibrosis, functional impairment, and clinical outcomes. Early evidence suggests radiomics may improve risk stratification and disease phenotyping, particularly when integrated with conventional imaging or clinical data. However, its clinical utility remains limited by methodological variability and the need for prospective, standardized validation. Overall, while AI and radiomics hold substantial potential to enhance CMR interpretation, their successful integration into clinical workflows will depend on further validation, cross-center generalizability, and sustained expert supervision.

Practical recommendations:

This review highlights several practical considerations for the successful integration of AI into clinical CMR workflows. With the technology continuing to demonstrate great potential, its implementation in clinical settings must still be approached with care, precision, and with the patient as the most important factor. A hybrid model in which AI supports but does not replace human expertise may offer the most clinically effective approach. This aligns with current expert consensus, like the European Society of Cardiology, which emphasizes that AI's role is to enhance clinical judgment rather than substituting it (21). This would ensure both improved precision and patient safety. This is especially critical in tasks such as left ventricular wall thickness measurement, cardiac volumetry, and myocardial scar detection, where even small discrepancies can impact diagnostic and therapeutic decisions. While AI models have consistently shown high levels of reproducibility and efficiency, oversight by trained cardiologists and radiologists remains essential to protect against errors. This is particularly true in borderline cases or anatomical anomalies. Human review also helps maintain clinical accountability and provides personalized patient care.

A second key recommendation concerns the standardization of AI use across healthcare institutions. Currently, differences in scanner hardware and imaging protocols can affect how AI models perform. To improve generalizability and reliability, hospitals should prioritize models that have been validated in multi-center studies. To move from research to routine care, AI algorithms must undergo rigorous clinical validation and demonstrate added value over existing workflows, particularly in terms of efficiency, reproducibility, and diagnostic accuracy

There are multicenter studies ongoing, which aim to assess exactly this. An example of this would be Winther et al. ν-Net Study, in which a DL algorithm that is to assess biventricular mass and function, using multicenter CMR data. One issue faced by the researchers was that the openly accessible datasets were very limited, with one of the used sets only containing 61 cases with freely accessible contours (19). Although the amount of information to train the algorithm on was limited, the authors conclude that the assessed neural network is ready to be employed on a grander scale. Collaborative initiatives developing shared imaging datasets may help support continuous refinement of AI tools and improve their robustness across a wider range of clinical environments.

Interpretability is another pillar of trustworthy AI integration. Tools that fall under the umbrella of explainable AI (XAI), such as saliency maps and attention-based visualization, can enhance model transparency and build clinician confidence in AI-generated outputs. When a model can visually demonstrate which image features contributed to a particular diagnosis or segmentation, clinicians are better positioned to validate or challenge those results. One study demonstrated this by using radiomic features extracted from LGE-CMR to successfully distinguish cardiac sarcoidosis from post-COVID myocardial inflammation (23). These conditions often appear similar on visual inspection and by accurately being able to tell them apart demonstrates how new and adapted, AI-supported systems could help generate insights that might not be immediately visible to clinicians (23). As regulatory bodies like the FDA and EMA move toward stricter guidelines for AI in healthcare, incorporating XAI features will also be essential for compliance.

Alongside AI, radiomics is emerging as a complementary technology with the potential to add a new layer of diagnostic and prognostic value in CMR imaging. Unlike AI-based segmentation models, radiomics involves the extraction of quantitative imaging features such as texture, shape, and intensity from standard CMR sequences, which can then be correlated with histological or clinical outcomes. Radiomics has already demonstrated utility in predicting myocardial fibrosis, stratifying risk in chronic coronary syndromes, and identifying subtle myocardial abnormalities that might otherwise be overlooked. For example, radiomics analysis of non-contrast cine-CMR images has been shown to accurately differentiate infarcted from viable myocardium (22). An advancement like this could potentially serve as a non-invasive alternative to the traditional contrast-based methods. However, its clinical integration will require further standardization in acquisition protocols and feature extraction, as well as external validation in large, multi-center cohorts. As such, radiomics is best viewed as a complementary tool to conventional AI, particularly in scenarios demanding deeper tissue characterization or complex risk stratification.

Quality control protocols must also be established and/or improved. While AI reduces inter-observer variability, it is not immune to segmentation errors or performance degradation in lower-quality scans. Studies such as Hatipoglu et al. (1) have shown that even the most advanced models sometimes misidentify structures or underfit endocardial contours. Besides poor scan quality, another paper from Jathanna et al. found that whilst the assessed systems can reliably segment cardiac scar tissue, there remains the issue of significant heterogeneity in model performance and evaluation standards (25). Institutions should therefore create systematic checkpoints where AI outputs are reviewed and approved by human experts, particularly in high-risk scenarios. The integration of flagging systems for low-confidence predictions could further support safer and more reliable analysis.

Equally important is clinician education. For AI to be used safely and effectively, the physicians who rely on its outputs must understand its strengths and limitations. Radiologists, cardiologists, and imaging technologists should receive targeted training in the capabilities, interpretability, and common pitfalls of the AI tools they use. This is particularly crucial in areas where AI performance remains inconsistent, such as segmentation of the anatomically complex right ventricle, as noted in studies by Bai et al. (2) and Hatipoglu et al. (1). Fostering a culture of digital literacy among clinicians will be essential to ensure that AI tools are used critically rather than passively. Building familiarity and confidence with these technologies may help practitioners engage more thoughtfully with AI-generated outputs.

Beyond any techincal oversight, the neccesary infrastructure needs to be put in place as well. This includes more than just providing adequate computational power and storage. One of the most important factors is the seamless integration into existing clinical paltforms. A possible starting point woul be to employ AI first in areas where the gain in efficiency will have the biggest impact, for example emergency departments, high-throguhput imaging centers and surgical pre-assessment units. This is consistent with findings that AI systems can significantly streamline cardiovascular imaging workflows and reduce the number of manual tasks required in segmentation, reporting, and interpretation, leading to more efficient and standardized care (20).

A study that demonstrated this efficiency gain was the paper by Böttcher et al. in which the employed AI-system was able to reduce segmentation times from several minutes to seconds, allowing for quicker diagnosis treatment decisions. In settings like acute care, time savings like this directly influence patient outcomes. Another factor that could directly impact patient outcomes was demonstrated by Jam et al. who have shown that besides imaging, machine learning algorithms can be applied for predicting mortality after cardiac surgery and outperform the traditional scoring systems (26).

Another area that deserves attention is the clinical adoption of contrast-free imaging methods, like Virtual Native Enhancement (VNE) developed by Zhang et al. (5). Innovations like these not only reduce scan times and the costs associated with it, but also minimizes the amount of gadolinium-baased contrast agents that patients are exposed to. This could play a particularly important role in patients with chronic kidney disease or of there are other contraindications for the use of contrast. Another paper that echoed these findings was Cau et al., who emphasize that emerging non-comtrast AI models have the potential to reduce the use of contrast agent whilst maintaining diagnostic accuracy (24). This has the potential to positively and directly impact patient safety and improve overall accesivility.

Besides just focussing on the technological advancements and the integration of new and imrpoved systems, it is essential for healthcare providers to stay up to date with ethical and regulatory considerations in order to provide the best outcomes.Whilst greater efficiency and objectivity are compelling factors, it is important to recognize and identify risks like algorithmic bias, lack of transparency (Black-Box phenomenon) and data privacy concerns. Developers and clinicians need to work together to create AI framewros that clearly prioritze fairness, accountability and most important of all, patient autonomy. Algorithms need to be trained and tested on diverse datasets to avoid biases in performance which could affect underrepresemnted groups.

Regulatory agencies are required to also evolve their oversight frameworks to be able to adequately accommodate new AI-driven tools. Just like with traditional medical devices, requirements for external validation, post-market surveillance and mandatory reporting of adverse events needs to be rigorously implemented.

Finally, interdisciplinary collaboration will be essential. Creating a connection between clinicians, data scientists, engineers, and policymakers will help ensure that AI technologies are developed in ways that genuinely address clinical needs wihtout disregarding the regulatory and ethical aspects. Academic institutions and professional societies should support research initiatives and promote the co-design of AI systems. Another important factor is engaging healthcare professionals early in the development process in order create AI tools that are not only technologically advanced but also clinically intuitive and relevant.

In summary, while integrating AI into cardiovascular magnetic resonance (CMR) imaging holds significant promise for enhancing diagnostic consistency, reducing variability, and improving efficiency, being able to realize these benefits will require more than just technological readiness. A collaborative approach that prioritizes ethical responsibility and that respects clinical needs will be essential to ensure best possible patient care.

References

1. Hatipoglu S, Mohiaddin RH, Gatehouse P, Alpendurada F, Baksi AJ, Izgi C, et al. Performance of artificial intelligence for biventricular cardiovascular magnetic resonance volumetric analysis in the clinical setting. The International Journal of Cardiovascular Imaging. 2022 Jun 29;38(11):2413–24.
2. Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. Journal of Cardiovascular Magnetic Resonance. 2018 Sep 14;20(1).
3. Curiale AH, Calandrelli E, Lucca Dellazoppa, Trevisan M, Bonifacio JP. Automatic Quantification of Volumes and Biventricular Function in Cardiac Resonance. Validation of a New Artificial Intelligence Approach. Revista Argentina de Cardiología [Internet]. 2021 Oct 1;89(4):337–41.
4. Böttcher B, Beller E, Busse A, Cantré D, Yücel S, Öner A, et al. Fully automated quantification of left ventricular volumes and function in cardiac MRI: clinical evaluation of a deep learning-based algorithm. The International Journal of Cardiovascular Imaging. 2020 Jul 16;36(11):2239–47.

5. Zhang Q, Burrage MK, Shanmuganathan M, Gonzales RA, Lukaschuk E, Thomas KE, et al. Artificial Intelligence for Contrast-Free MRI: Scar Assessment in Myocardial Infarction Using Deep Learning–Based Virtual Native Enhancement. Circulation. 2022 Nov 15;146(20):1492–503.
6. Popescu DM, Abramson HG, Yu R, Lai C, Shade JK, Wu KC, et al. Anatomically informed deep learning on contrast-enhanced cardiac magnetic resonance imaging for scar segmentation and clinical feature extraction. Cardiovascular Digital Health Journal. 2022 Feb 1;3(1):2–13.
7. Bhuva AN, Bai W, Lau C, Davies RH, Ye Y, Bulluck H, et al. A Multicenter, Scan-Rescan, Human and Machine Learning CMR Study to Test Generalizability and Precision in Imaging Biomarker Analysis. Circulation: Cardiovascular Imaging. 2019 Oct;12(10).
8. Augusto JB, Davies RH, Bhuva AN, Knott KD, Seraphim A, Alfarih M, et al. Diagnosis and risk stratification in hypertrophic cardiomyopathy using machine learning wall thickness measurement: a comparison with human test-retest performance. The Lancet Digital Health. 2021 Jan;3(1):e20–8.
9. Fotaki A, Puyol-Antón E, Chiribiri A, Botnar R, Pushparajah K, Prieto C. Artificial Intelligence in Cardiac MRI: Is Clinical Adoption Forthcoming? Frontiers in Cardiovascular Medicine. 2022 Jan 10;8.
10. Tam DY, Chung D, Fremes SE. Commentary: Artificial intelligence to predict mortality: The rise of the machines? The Journal of Thoracic and Cardiovascular Surgery. 2020 Aug;163(6).
11. Chudgar PD, Burkule NJ, Kamat NV, Rege GM, Jantre MN. Myocardial Strain Imaging Using Feature Tracking Method of Cardiac MRI: Our Initial Experience of This Novel Parameter as an Additional Diagnostic Tool. Indian journal of radiology and imaging - new series/Indian journal of radiology and imaging/Indian Journal of Radiology & Imaging [Internet]. 2022 Aug 30 [cited 2025 Apr 18];32(04):479–87. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC9705126/
12. Gröschel J, Kuhnt J, Viezzer D, Hadler T, Hormes S, Barckow P, et al. Comparison of manual and artificial intelligence based quantification of myocardial strain by feature tracking—a cardiovascular MR study in health and disease. European Radiology [Internet]. 2023 Aug 18 [cited 2025 Mar 6];34(2):1003–15. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC10853310/
13. Kasper Pryds, Larsen AH, Hansen MS, Yoon A, Rasmus Stilling Tougaard, Hansson NH, et al. Myocardial strain assessed by feature tracking cardiac magnetic resonance in patients with a variety of cardiovascular diseases – A comparison with echocardiography. Scientific

Reports [Internet]. 2019 Aug 5;9(1). Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6683180/

14. Shiro Nakamori, Amine Amyar, Fahmy AS, Ngo LH, Ishida M, Nakamura S, et al. Cardiovascular Magnetic Resonance Radiomics to Identify Components of the Extracellular Matrix in Dilated Cardiomyopathy. Circulation. 2024 May 29;150(1):7–18.
15. Neisius U, El-Rewaidy H, Nakamori S, Rodriguez J, Manning WJ, Nezafat R. Radiomic Analysis of Myocardial Native T1 Imaging Discriminates Between Hypertensive Heart Disease and Hypertrophic Cardiomyopathy. JACC: Cardiovascular Imaging. 2019 Oct;12(10):1946–54.
16. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to Radiomics. Journal of Nuclear Medicine. 2020 Feb 14;61(4):488–95.
17. Raisi-Estabragh Z, Izquierdo C, Campello VM, Martin-Isla C, Jaggi A, Harvey NC, et al. Cardiac magnetic resonance radiomics: basic principles and clinical perspectives. European Heart Journal - Cardiovascular Imaging. 2020 Mar 6;21(4):349–56.
18. Domenico Mastrodicasa, Marly van Assen, Huisman M, Leiner T, Williamson EE, Nicol ED, et al. Use of AI in Cardiac CT and MRI: A Scientific Statement from the ESCR, EuSoMII, NASCI, SCCT, SCMR, SIIM, and RSNA. Radiology. 2025 Jan 1;314(1).
19. Winther HB, Hundt C, Schmidt B, Czerner C, Bauersachs J, Wacker F, et al. ν-net. JACC: Cardiovascular Imaging. 2018 Jul;11(7):1036–8.

Supporting Literature:

20. Leiner T, Rueckert D, Suinesiaputra A, Baeßler B, Nezafat R, Išgum I, et al. Machine learning in cardiovascular magnetic resonance: basic concepts and applications. Journal of Cardiovascular Magnetic Resonance. 2019 Oct 7;21(1)
21. Miranda' I, Luz ' JM, Pereira' AR, Augusto' JB. Artificial intelligence in cardiovascular imaging algorithms – what is used in clinical routine? [Internet]. www.escardio.org. 2024. Available from: https://www.escardio.org/Councils/Council-for-Cardiology-Practice-(CCP)/Cardiopractice/artificial-intelligence-in-cardiovascular-imaging-algorithms-what-is-used-in-c
22. Avard E, Shiri I, Ghasem Hajianfar, Abdollahi H, Kiara Rezaei Kalantari, Houshmand G, et al. Non-contrast Cine Cardiac Magnetic Resonance image radiomics features and machine learning algorithms for myocardial infarction detection. Computers in Biology and Medicine. 2022 Feb 1;141(141):105145–5.

23. Mushari NA, Soultanidis G, Duff L, Trivieri MG, Fayad ZA, Robson P, et al. An assessment of PET and CMR radiomic features for the detection of cardiac sarcoidosis. Frontiers in Nuclear Medicine. 2024 Jan 16;4.
24. Cau R, Pisu F, Suri JS, Mannelli L, Scaglione M, Masala S, et al. Artificial Intelligence Applications in Cardiovascular Magnetic Resonance Imaging: Are We on the Path to Avoiding the Administration of Contrast Media? Diagnostics [Internet]. 2023 Jan 1;13(12):2061. Available from: https://www.mdpi.com/2075-4418/13/12/2061
25. Jathanna N, Podlasek A, Sokol A, Auer D, Chen X, Jamil-Copley S. Diagnostic utility of artificial intelligence for left ventricular scar identification using cardiac magnetic resonance imaging—A systematic review. Cardiovascular Digital Health Journal. 2021 Dec;2(6):S21–9.