

CURRENT AND FUTURE ROLE OF ARTIFICIAL INTELLIGENCE IN CARDIAC IMAGING

EDITED BY: Steffen Erhard Petersen, Karim Lekadir, Alistair A. Young and
Tim Leiner

PUBLISHED IN: Frontiers in Cardiovascular Medicine





frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88966-058-2

DOI 10.3389/978-2-88966-058-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

CURRENT AND FUTURE ROLE OF ARTIFICIAL INTELLIGENCE IN CARDIAC IMAGING

Topic Editors:

Steffen Erhard Petersen, Queen Mary University of London, United Kingdom

Karim Lekadir, University of Barcelona, Spain

Alistair A. Young, King's College London, United Kingdom

Tim Leiner, University Medical Center Utrecht, Netherlands

Citation: Petersen, S. E., Lekadir, K., Young, A. A., Leiner, T., eds. (2020). Current and Future Role of Artificial Intelligence in Cardiac Imaging. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-058-2

Table of Contents

04	<i>Editorial: Current and Future Role of Artificial Intelligence in Cardiac Imaging</i>
	Karim Lekadir, Tim Leiner, Alistair A. Young and Steffen E. Petersen
11	<i>Machine Learning for Assessment of Coronary Artery Disease in Cardiac CT: A Survey</i>
	Nils Hampe, Jelmer M. Wolterink, Sanne G. M. van Velzen, Tim Leiner and Ivana Išgum
19	<i>Machine Learning Approaches for Myocardial Motion and Deformation Analysis</i>
	Nicolas Duchateau, Andrew P. King and Mathieu De Craene
28	<i>Artificial Intelligence for Cardiac Imaging-Genetics Research</i>
	Antonio de Marvao, Timothy J. W. Dawes and Declan P. O'Regan
38	<i>Image-Based Cardiac Diagnosis With Machine Learning: A Review</i>
	Carlos Martin-Isla, Victor M. Campello, Cristian Izquierdo, Zahra Raisi-Estabragh, Bettina Baeßler, Steffen E. Petersen and Karim Lekadir
57	<i>From Compressed-Sensing to Artificial Intelligence-Based Cardiac MRI Reconstruction</i>
	Aurélien Bustin, Niccolo Fuin, René M. Botnar and Claudia Prieto
76	<i>Deep Learning for Cardiac Image Segmentation: A Review</i>
	Chen Chen, Chen Qin, Huaqi Qiu, Giacomo Tarroni, Jinming Duan, Wenjia Bai and Daniel Rueckert
109	<i>AI in Cardiac Imaging: A UK-Based Perspective on Addressing the Ethical, Social, and Political Challenges</i>
	Matthew E. Fenech and Olly Buston
117	<i>Mini Review: Deep Learning for Atrial Segmentation From Late Gadolinium-Enhanced MRIs</i>
	Kevin Jamart, Zhaohan Xiong, Gonzalo D. Maso Talou, Martin K. Stiles and Jichao Zhao
129	<i>Artificial Intelligence in Cardiac Imaging With Statistical Atlases of Cardiac Anatomy</i>
	Kathleen Gilbert, Charlene Mauger, Alistair A. Young and Avan Suinesiaputra



Editorial: Current and Future Role of Artificial Intelligence in Cardiac Imaging

Karim Lekadir^{1*}, Tim Leiner², Alistair A. Young³ and Steffen E. Petersen^{4,5}

¹ Universitat de Barcelona, Artificial Intelligence in Medicine Lab (BCN-AIM), Departament de Matemàtiques and Informàtica, Barcelona, Spain, ² Department of Radiology, Utrecht University Medical Centre, Utrecht, Netherlands, ³ School of Biomedical Engineering & Imaging Sciences, King's College London, London, United Kingdom, ⁴ Barts Heart Centre, Barts Health NHS Trust, London, United Kingdom, ⁵ NIHR Barts Biomedical Research Centre, William Harvey Research Institute, Queen Mary University of London, London, United Kingdom

Keywords: artificial intelligence, cardiac imaging modalities, big data, cardiac image analysis, cardiovascular personalized medicine, AI adoption and translation

Editorial on the Research Topic

Current and Future Role of Artificial Intelligence in Cardiac Imaging

INTRODUCTION

Cardiovascular disease is currently the most common cause of morbidity and mortality worldwide (1) and thus remains an important focus for both biomedical and technological research. In the age of personalized medicine, cardiac imaging is expected to play an important role to enable more accurate and advanced quantification of structural and functional changes due to cardiovascular disorders. However, despite advances in cardiac imaging modalities, such as echocardiography, cardiovascular magnetic resonance, cardiac computed tomography, and nuclear cardiology, the heart remains a challenging anatomical organ to image and assess compared to other organ systems. The main challenges faced by cardiac imaging include the perpetual cardiac and respiratory motions, the complex geometry of the ventricles, atria and arteries, the oblique orientation of the heart with respect to the body, and the small size of some of the cardiac structures, such as the coronary arteries, trabeculae and papillary muscles, as well as the large variability in imaging conditions and protocols (including non-contrasted and contrast-enhanced cardiac imaging sequences).

Consequently, advanced tools are needed to optimize the use of cardiac imaging and to support clinicians throughout the whole value-chain of cardiovascular practice, including improved image acquisition, automated cardiac quantification, cardiac tissue characterization, imaging biomarker discovery, and clinical decision support. In this context, artificial intelligence (AI), including machine learning and computer vision, has emerged as one of the most promising topics over the last 5 years. Combined with the exponential increase in computing power, AI provides unprecedented opportunities to leverage the available collections of cardiac imaging data for developing more robust cardiac image analysis algorithms, to uncover currently unknown clinical knowledge on cardiac health and disease, and to build novel software tools that will impact clinical cardiology. This area is expected to benefit from the current efforts to provide access to large-scale and high-quality image data for the scientific community. In the US, for example, existing studies such as the Multi-Ethnic Study of Atherosclerosis and the Framingham Heart Study have for a long time compiled thousands of cardiac images. More recently, in the United Kingdom, the UK Biobank has been acquiring cardiac MRI images from tens of thousands of individuals (2).

OPEN ACCESS

Edited and reviewed by:

Hendrik Tevaearai Stahel,
Bern University Hospital, Switzerland

*Correspondence:

Karim Lekadir
karim.lekadir@ub.edu

Specialty section:

This article was submitted to
Cardiovascular Imaging,
a section of the journal
Frontiers in Cardiovascular Medicine

Received: 20 May 2020

Accepted: 30 June 2020

Published: 07 August 2020

Citation:

Lekadir K, Leiner T, Young AA and
Petersen SE (2020) Editorial: Current
and Future Role of Artificial Intelligence
in Cardiac Imaging.
Front. Cardiovasc. Med. 7:137.
doi: 10.3389/fcvm.2020.00137

In Europe, the euCanSHare project funded by the European Commission is developing a data sharing and analytics platform to facilitate access to large-scale cardiac imaging and non-imaging data from multiple centers (www.eucanshare.eu). Similar initiatives and large cohorts are expected to emerge across the globe in the years to come, which will further enhance the potential of AI in cardiac imaging and clinical cardiology.

To promote and guide further research and developments in AI for cardiac imaging, several experts and leading institutions in the field have recently published position and review papers that outline the initial achievements, discuss the current challenges and identify future perspectives. For example, Dey et al. summarized the most promising AI methods for cardiac imaging by distinguishing between the use of classical AI and advanced approaches (3). Al'Aref et al. reviewed some clinical applications of AI in cardiac imaging by considering each cardiac imaging modality separately (4). Litjens et al. focused on the sole application of deep learning methods for cardiac image analysis (5). Finally, Petersen et al. outlined the current challenges and emerging opportunities (6), emphasizing the importance of addressing non-technical aspects of AI in cardiac imaging such as patient acceptance, data protection and AI regulation.

While these papers provided an overall presentation and promotion of the field, the goal of this special issue entitled “current and future role of AI in cardiac imaging” is to compile more detailed and focused reviews covering the whole value-chain of AI in cardiac imaging. Specifically, we have invited active experts across the globe to submit in-depth reviews on several key areas of AI in cardiac imaging, including (1) cardiac image reconstruction, (2) cardiac image segmentation, (3) cardiac shape and motion analysis, (4) computer-aided diagnosis, (5) imaging-genetics integration, and (6) socio-ethical impact and regulations. This comprehensive special issue, totaling nearly 1,000 references from the field, will constitute an unprecedented resource for researchers, both novice and experienced, to study in detail the methods, applications, strategies, datasets, tools, hypotheses, limitations and opportunities that are of direct relevance to each aspect of AI in cardiac imaging. Importantly, for each paper, we requested the authors to provide descriptions and discussions for both AI and clinical audiences, to enhance the democratization and promotion of AI in cardiac imaging, and thus future collaborations and developments in the field.

SPECIAL ISSUE CONTENT

This special issue covers six specific areas and application domains of AI in cardiac imaging, as shown in **Figure 1** and presented as follows:

Cardiac Image Reconstruction

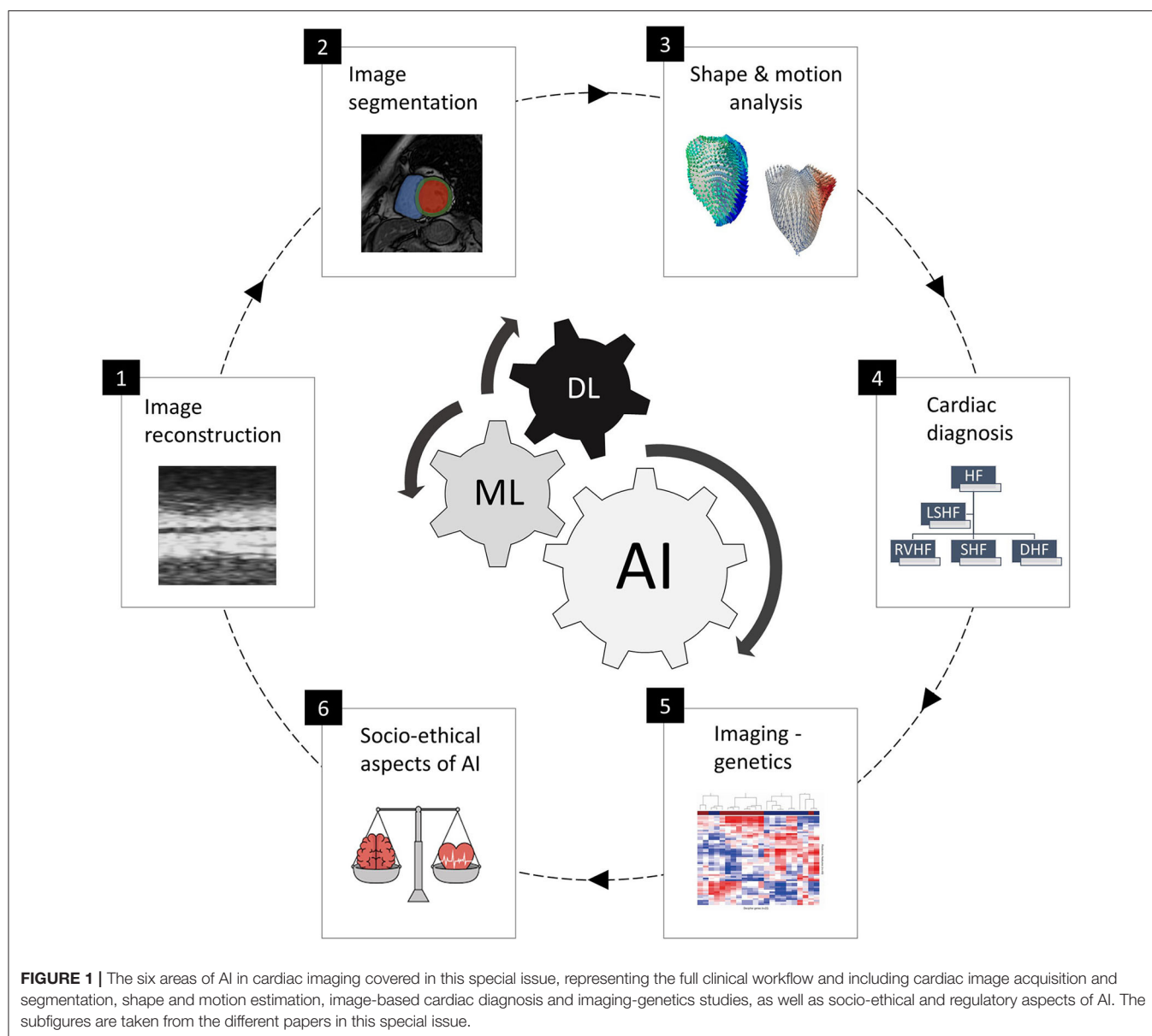
The first paper of this special issue focuses on the very first step of the cardiac imaging workflow, i.e., enhancing cardiac image acquisition using AI. Fast and portable cardiac imaging such as echocardiography inherently suffers from low image quality, while high resolution cardiac imaging such as CMR requires long acquisition times to address the cardiac and respiratory motions, as well as the need to highlight the

different types of cardiac structures, tissues and vessels. For a long time, enhancing and accelerating image acquisition for modalities such as CMR was targeted by developing new CMR imaging/physics sequences and techniques, such as efficient pulse sequences, motion compensation techniques, multiple radio-frequency receiver coils for parallel imaging and compressed sensing. In this special issue, Bustin et al. thoroughly surveyed emerging AI techniques for accelerating and enhancing CMR image reconstruction. Concretely, AI provides a unique opportunity to perform CMR acquisition using under-sampling strategies that acquire less image data than needed, followed by learning-based estimation of the sparse domain from existing data. The authors reviewed first the initial learning-based techniques based on dictionaries of transforms (from low to high resolution domains) learned from the acquired under-sampled data itself. Subsequently, they focused their attention on recent deep learning-based approaches, which learn the reconstruction transforms from low-resolution to high-resolution images offline based on training data. They surveyed in detail the many advances over the last 2 years, describing recent applications to both 2D dynamic cardiac imaging and 3D whole-heart CMR imaging. As deep learning based CMR reconstruction is novel as well as popular, they concluded their review by discussing the future avenues to improve real-world validation, including assessment of reconstruction quality and generalization.

Cardiac Image Segmentation

In the next review papers, this special issue addressed the next stage of cardiac image analysis, namely cardiac image segmentation, which is by far the most covered topic by AI researchers in cardiac imaging. Deep learning approaches have been particularly popular as they have shown to generate highly accurate results when trained on large manually segmented data. At the same time, clinicians have been highly receptive to such automated black-box methods as they accelerate their clinical work without interfering in the decision making. In this special issue, Chen et al. put together a comprehensive review of deep learning techniques for cardiac image segmentation, totaling over 100 papers describing applications to various imaging modalities (echocardiography, MRI, and cardiac CT) and to the main cardiac structures (ventricle, atria, and vessels). They concluded that there is no universally optimal deep learning implementation for cardiac image segmentation, and suggested that algorithms need to be customized and optimized for each application depending on the imaging modality, protocol (e.g. contrast vs. non-contrast), and cardiac structure (left vs. right ventricle). For the immediate future, they discussed the importance of developing segmentation methods that can generalize well across various imaging modalities, scanners, and pathologies. Finally, in order to encourage reproducible research, the authors provided a summary of public datasets for training and testing new deep learning models, as well as public code repositories that include recently developed techniques.

In a more focused review, Jamart et al. surveyed one of the most challenging cardiac image segmentation applications, namely atrial segmentation from late enhanced cardiac MRI. This contrast enhanced imaging sequence is the technique of



choice in clinical practice to quantify fibrosis and assess atrial fibrillation. However, the task is complicated by the geometrical complexity and small size of the atrial chambers, which are also constrained by thin walls. Moreover, the anatomical boundaries on the late enhanced MRI often lack clear contrast, which can further mislead the segmentation algorithms. As a result, until the advent of deep learning, very few techniques had been attempted for automated segmentation of the atria. In this review, the authors summarized the recent deep learning developments in this field, including multi-stage and multi-scale conventional neural networks to address the image class imbalance that is inherent to atrial segmentation (the atrial cavity represents only a small fraction of the image volume), as well as to provide contextual cues to better differentiate the atrial boundaries from the surrounding structures. The best reported performance in

the survey reached a 93.2% segmentation accuracy, which is highly promising given the complexity of the task. Arguably, the most important future work in this domain is the automated detection of the fibrosis using advanced AI approaches, which will represent an important Research Topic in the years to come.

Cardiac Shape and Motion Analysis

Image segmentation of the cardiac boundaries as described above is a pre-requisite in clinical practice to estimate standard clinical indices such as chamber volumes and ejection fraction for cardiac assessment. However, existing research has shown that more detailed information about cardiac geometry and regional motion is expected to improve future clinical assessment of normal and abnormal cardiac (dys)function. In this special issue, Gilbert et al. presented a review of the so-called statistical atlases

of cardiac anatomy, which are widely used to model complex shape and function variability across subpopulations, and thus to extract new descriptors of cardiac geometry of relevance to cardiac health and disease. In this paper, the authors firstly describe the existing and newer methods for building clinically useful statistical cardiac atlases from annotated cardiac contours. Subsequently, the paper discussed that, powered by supervised or unsupervised machine learning algorithms, statistical cardiac shape analysis can be used to automatically identify and quantify abnormal shape deviations, and to provide morphometric indices that are optimally associated with clinical factors.

In another review, Duchateau et al. described in great depth the literature on cardiac motion quantification and analysis based on machine learning. With these techniques, the idea is to learn new advanced representations and patterns of myocardial motion and deformation (displacement, velocity, deformation, torsion, strain) from representative samples of cardiac images, such as echocardiography or MRI. In particular, two families of approaches were reviewed, namely (1) traditional techniques that apply machine learning onto explicit features of myocardial motion/deformation (displacement fields calculated from the images), and (2) more recent approaches based on neural networks applied directly to the image data to extract and analyze new spatiotemporal signatures from local image patches around the myocardium. In this paper, the authors described the entire workflow of steps and methods required to derive physiologically meaningful and clinically useful analyses of cardiac motion. Finally, they discussed the next steps toward clinical adoption of machine learning based cardiac motion quantification, including community benchmarking, standardization initiatives, and clinical interpretability of the extracted spatiotemporal signatures for abnormality localization.

Computer-Aided Diagnosis

Generally, the main AI developments in cardiac imaging have mostly addressed cardiac image analysis tasks before clinical decision making, with the aim to enhance cardiac image acquisition, facilitate cardiac image segmentation, and estimate advanced indices of cardiac shape and function. However, AI is also expected to impact clinical decision making in the future, such as to enable earlier and more precise diagnosis, as well as treatment planning and response estimation. To illustrate this, this special issue includes two review papers centered on cardiac diagnosis. First, Martin-Isla et al. surveyed in detail the area of image-based cardiac diagnosis using machine learning. This included a step-by-step description of the techniques for building and validating new AI models of cardiac diagnosis. The authors also described emerging techniques for more precise diagnosis, including radiomics (omics for radiology) (7). The authors then reviewed more than 100 papers on AI- and image-driven diagnosis of coronary heart disease, cardiomyopathy, heart failure, and valve disease. Interestingly, the survey showed that some complex cardiac diseases are yet to be extensively investigated by the AI community, such as atrial fibrillation. Finally, the paper discussed current obstacles that limit the applicability of AI-driven diagnosis in cardiac imaging, in particular the lack of interpretability, which must be addressed

in the years to come to enable clinicians to understand and trust the AI-generated diagnoses and decisions.

In a second review paper on AI for cardiac diagnosis, Hampe et al. focused on the assessment of coronary artery disease from non-invasive CT using machine learning. In clinical practice, catheter-guided X-ray angiography and intravascular ultrasound provide detailed information on coronary stenosis and plaque composition, but they are limited by their invasive nature. CT imaging provides a promising alternative but offers reduced contrast between the atherosclerotic plaque constituents, and thus AI is expected to play a role for the extraction of detailed and clinically useful information from the non-invasive images. This review surveyed classical and modern machine learning methods to estimate coronary stenosis, to discriminate between calcified, non-calcified and mixed plaques in CT, and to characterize fibrous and lipidic plaque constituents. Furthermore, the paper described recent methods, in particular deep learning based, for predicting fractional flow reserve directly from CT, by learning the relationship between CT features and fractional flow reserve based on a training sample of corresponding CT and invasive imaging. In their discussion, the authors noted an important limitation of current models, which are based on small training samples, as images with manually characterized plaques are more difficult to obtain than, for example, manual annotations of cardiac boundaries. Thus, this field of AI in cardiac imaging is expected to further develop for future clinical use as additional and larger datasets become available in the years to come.

AI Integration With Non-imaging Data

To realize the promise of precision medicine in cardiology, cardiac imaging is a central piece of the puzzle. However, non-imaging data play an important role, in particular—omics and health data, as they allow to build multi-scale AI models that integrate patient-specific biomolecular, phenotypic, environmental, and clinical information. Such integrated AI models are expected to lead to improved diagnosis and treatment selection, as well as to better clinical outcomes. This special issue included an interesting review by de Marvao et al. of AI-driven integrated cardiac imaging-genetics studies, which aim to characterize the complex interplay between cardiac imaging phenotypes, environmental and genetic factors. Several concrete examples of AI-empowered imaging-genetics studies were provided, such as to enable more stratified diagnosis of heart failure, predict treatment response in cardiomyopathic patients, identify genetic variants or proteomic signatures of high-risk atherosclerotic plaques, or predict positive cardiac remodeling after cardiac resynchronization therapy. While the use of AI in cardiovascular imaging-genetics is shown to have great potential, the review noted that the challenges of AI in genetics and imaging separately are amplified by combining these very large data. Thus, further research is expected to address more ambitious whole-genome and high-resolution whole-heart imaging studies, and to derive multi-scale AI solutions for clinical practice integrating imaging, biological and clinical data.

Ethical, Social, and Political Issues

While the review papers described above dealt with technical and clinical aspects of AI in cardiac imaging, this special issue

concludes with a paper by Fenech and Buston that reviews the ethical, social, and political issues that are being investigated to facilitate future acceptance and deployment of the AI solutions in cardiac imaging. These include clarifying the impact of AI solutions on the roles of cardiologists, radiologists, and other doctors, future liability of clinicians vs. AI manufacturers, as well as on the altered relationships between healthcare professionals, patients, their relatives, and administrators. Data sharing and privacy issues were also reviewed in the paper, focusing on the challenges to manage patient informed consent for AI solutions that remain difficult to understand (and trust) by the general public and clinicians alike. From a social point of view, initial studies reviewed in this paper suggest that there is a concern that AI solutions may remain biased and in fact exacerbate health inequalities. Furthermore, the authors discussed the need to include patients and citizens in the AI development process, to take into close consideration their requirements, expectations, and behaviors. Other important issues, such as algorithmic transparency, fairness, and regulation, were also discussed at length. For addressing these key issues and to optimize adoption by clinicians, patients, and regulators, the paper emphasized the importance of developing principles and translating them into policies in the years to come. In the general context of AI, imaging and cardiology in particular are expected to play an important role, as exemplified by the fact that they have been the healthcare

domains with the greatest number of FDA approvals for novel data-driven technologies in the recent years.

FUTURE PERSPECTIVES

This special issue reviewed AI developments and opportunities for each task of the cardiac imaging clinical workflow, surveying in detail cardiac imaging acquisition, segmentation and quantification, clinical decision support and precision cardiovascular medicine through integration with genomics data, in addition to ethical, social, and regulatory aspects. A close look at the statistics from **Figures 2, 3**, gathered from the publications reviewed in this special issue, shows a continuous increase in the research output in AI for cardiac imaging over the last 5 years. Interestingly, while cardiac image segmentation has received the most attention in the field as shown in **Figure 3**, due to the urgent need to accelerate the contouring process, it is also the only cardiac imaging task for which there is a decrease in the number of publications between 2018 and 2019 (**Figure 2**). This may be explained by the advances made possible by deep learning in the field, combined with an increasing interest from the community to invest in other AI applications such as cardiac image reconstruction or computer-aided diagnosis.

While this special issue described six main AI applications in cardiac imaging separately, we believe the ultimate aim should

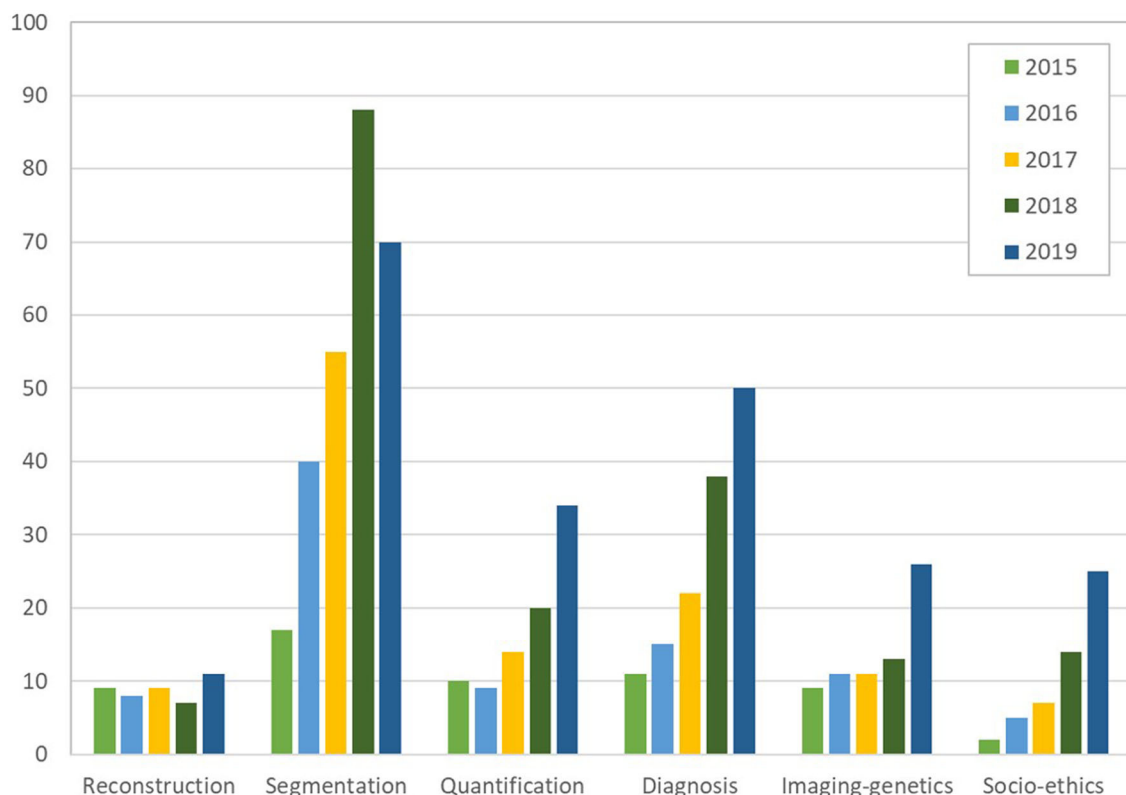


FIGURE 2 | Number of papers reviewed in the six categories of cardiac imaging tasks in the last 5 years (2015 to 2019), showing an increase in the research output continuously and for all tasks, except for segmentation which decreased between 2018 and 2019 (over 550 papers reviewed in total). The subfigures are taken from the different papers in this special issue.

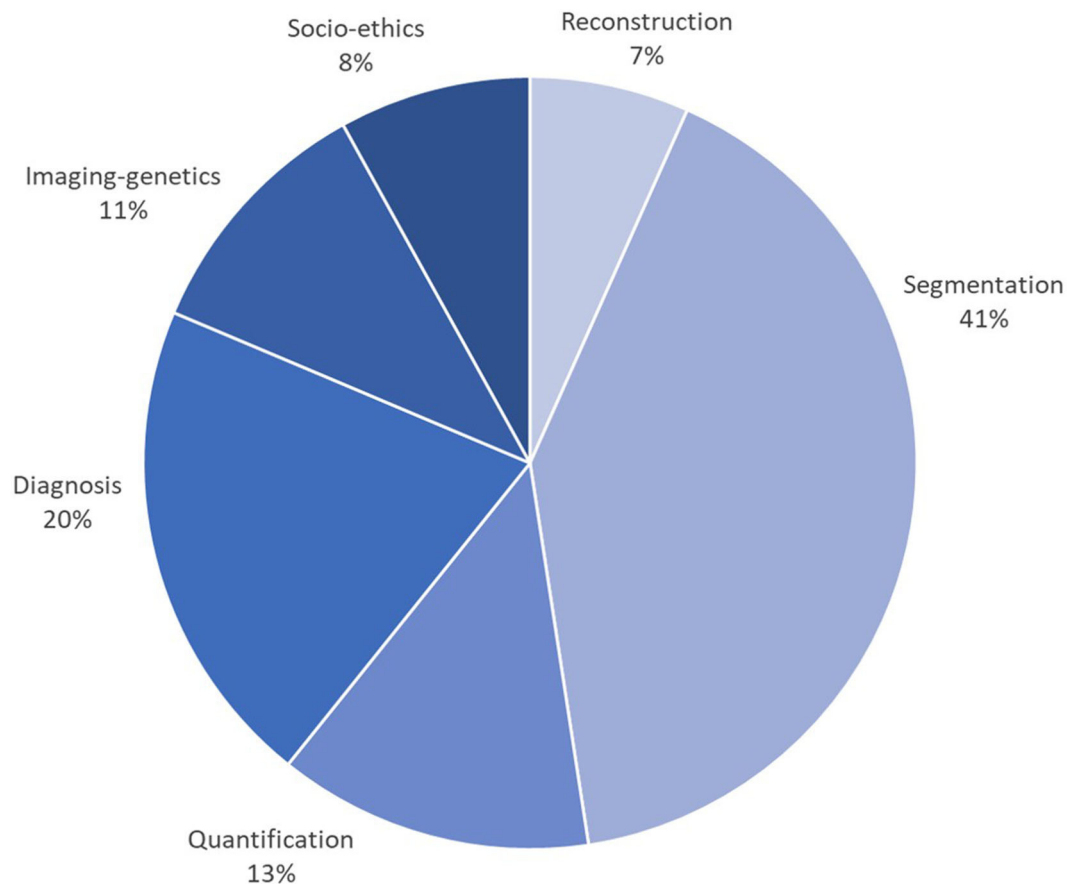


FIGURE 3 | Percentage of papers surveyed per cardiac imaging task, showing that approx. 40% of efforts have been dedicated to segmentation, followed by diagnosis (20%), then shape and motion quantification (13%). The subfigures are taken from the different papers in this special issue.

be to integrate these separate tasks into one single smooth, efficient and user-friendly pipeline for clinical cardiologists. As an example, this multi-task integration is the scope of the UK-based research project “SmartHeart: Next-generation cardiovascular healthcare via integrated image acquisition, reconstruction, analysis and learning” funded by UK’s Engineering and Physical Sciences Research Council.

Additionally, while the special issue covered all important steps of cardiac imaging, it is worth mentioning other research tasks, at their very initial stages of development, which have not yet been covered in this special issue. For example, “automated quality control” is expected to enhance the cardiac image analysis workflow as large volumes of research and clinical data become available, and as the demand for AI-driven automation and robustness will increase in clinical practice. Here, it is worth listing a few preliminary works, such as automated quality control of CMR images using a deep learning approach to identify suboptimal image contrast or heart coverage (8). Other works have instead focused on quality control of the final image segmentation results using classical AI (9) or neural networks (10). Another area that may benefit from AI is “image-based computational cardiology,” which builds patient-specific digital models of the heart to simulate treatment response. While this

area has been traditionally addressed using pure physiological and mechanistic models, researchers are now investigating the integration of machine learning to improve the accuracy and speed of the personalized simulated outputs (11). Furthermore, as larger datasets become available, it is expected that predictive models of disease progression will be developed and validated, including by integrating imaging with non-imaging predictors (e.g., socio-demographic, biomarker, lifestyle and genomic data). Finally, while this special issue is dominated by AI applications in echocardiology, CMR and cardiac CT, there have also been machine learning applications in “nuclear cardiology” (12), and these are expected to increase in the years to come as larger nuclear medicine datasets (both PET and SPECT) become available.

To conclude this editorial, we wish to emphasize the need in the next years for more concerted efforts dedicated to enhancing the technical and clinical advances described in this special issue, but also to address non-technical and non-clinical aspects of AI in cardiac imaging. Importantly, there is a need for community-defined standards and guidelines for validating and adopting future AI solutions, including metrics and procedures to evaluate performance, bias and errors, clinical effectiveness, degree of interpretability, and even

cost-effectiveness. Benchmarking datasets and tools are also required to enable transparent and comparative analysis of the AI solutions across research institutions and players. For example, a test-retest reference dataset was recently compiled to assess reproducibility of machine learning CMR studies (13), while an international challenge on multi-center and multi-vendor cardiac imaging segmentation was organized to test generalizability across scanners (Siemens, Philips, General Electric and Canon)¹. Finally, ethical and regulatory aspects will need to be established in a multi-stakeholder collaboration between experts in AI, bioethics and cardiac imaging, but also with the involvement of patient associations, private companies, and public authorities.

¹<https://www.ub.edu/mnms/>

REFERENCES

1. Timmis A, Townsend N, Gale CP, Torbica A, Lettino M, Petersen SE, et al. European Society of Cardiology: cardiovascular disease statistics 2019. *Eur Heart J*. (2020) 41:12–85. doi: 10.1093/eurheartj/ehz859
2. Petersen SE, Matthews PM, Francis JM, Robson MD, Zemrak F, Boubertakh R, et al. UK Biobank's cardiovascular magnetic resonance protocol. *J Cardiovasc Magn Reson*. (2015) 18:8. doi: 10.1186/s12968-016-0227-4
3. Dey D, Slomka PJ, Leeson P, Comaniciu D, Shrestha S, Sengupta PP, et al. (2019). Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. *J Am Coll Cardiol*. 73:1317–35. doi: 10.1016/j.jacc.2018.12.054
4. Al' Aref SJ, Anchouche K, Singh G, Slomka PJ, Kolli KK, et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *Eur Heart J*. (2019) 40:1975–86. doi: 10.1093/eurheartj/ehy404
5. Litjens G, Ciompi F, Wolterink JM, de Vos BD, Leiner T, Teuwen J, et al. State-of-the-art deep learning in cardiovascular image analysis. *JACC Cardiovasc Imaging*. (2019) 12:1549–65. doi: 10.1016/j.jcmg.2019.06.009
6. Petersen SE, Abdulkareem M, Leiner T. Artificial intelligence will transform cardiac imaging—opportunities and challenges. *Front Cardiovasc Med*. (2019) 6:133. doi: 10.3389/fcvm.2019.00133
7. 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. *Eur Heart J Cardiovasc Imaging*. (2020) 21:349–56. doi: 10.1093/ehjci/jeaa028
8. Zhang L, Pereañez M, Piechnik SK, Neubauer S, Petersen SE, Frangi AF. Image quality assessment for population cardiac magnetic resonance imaging. In: *Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics*. Cham: Springer (2019). p. 299–321. doi: 10.1007/978-3-030-13969-8_15
9. Albà X, Lekadir K, Pereañez M, Medrano-Gracia P, Young AA, Frangi AF. Automatic initialization and quality control of large-scale cardiac MRI segmentations. *Med Image Anal*. (2018) 43:129–41. doi: 10.1016/j.media.2017.10.001
10. Robinson R, Valindria VV, Bai W, Oktay O, Kainz B, Suzuki H, et al. Automated quality control in image segmentation: application to the UK Biobank cardiovascular magnetic resonance imaging study. *J Cardiovasc Magn Reson*. (2019) 21:18. doi: 10.1186/s12968-019-0523-x
11. Corral-Acero J, Margara F, Marciniak M, Rodero C, Loncaric F, Feng Y, et al. The 'Digital Twin' to enable the vision of precision cardiology. *Eur Heart J*. (2020). doi: 10.1093/eurheartj/ehaa159
12. Betancur J, Commandeur F, Motlagh M, Sharir T, Einstein AJ, Bokhari S, et al. Deep learning for prediction of obstructive disease from fast myocardial perfusion SPECT: a multicenter study. *JACC Cardiovasc Imaging*. (2018) 11:1654–63. doi: 10.1016/j.jcmg.2018.01.020
13. 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. *Circ Cardiovasc Imaging*. (2019) 12:e009214. doi: 10.1161/CIRCIMAGING.119.009214

AUTHOR CONTRIBUTIONS

All co-authors discussed the structure and content of the special issue and paper. KL wrote the first draft. TL, AY, and SP revised the manuscript.

FUNDING

The work of KL was partly funded by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 825903 (euCanSHare project), and by the Ramon y Cajal Programme of the the Spanish Ministry of Economy and Competitiveness under grant no. RYC-2015-17183.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Lekadir, Leiner, Young and Petersen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Machine Learning for Assessment of Coronary Artery Disease in Cardiac CT: A Survey

Nils Hampe^{1,2,3*}, Jelmer M. Wolterink^{1,2,3}, Sanne G. M. van Velzen³, Tim Leiner⁴ and Ivana Išgum^{1,2,3,5}

¹ Department of Biomedical Engineering and Physics, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands, ² Amsterdam Cardiovascular Sciences, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands, ³ Image Sciences Institute, University Medical Center Utrecht, Utrecht, Netherlands, ⁴ Department of Radiology, University Medical Center Utrecht, Utrecht, Netherlands, ⁵ Department of Radiology and Nuclear Medicine, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands

OPEN ACCESS

Edited by:

Fabrizio Ricci,
G. d'Annunzio University of Chieti and
Pescara, Italy

Reviewed by:

Maria A. Zuluaga,
Institut Eurécom, France
John Hoe,
MediRad Associates Ltd, Singapore

*Correspondence:

Nils Hampe
n.hampe@amsterdamumc.nl

Specialty section:

This article was submitted to
Cardiovascular Imaging,
a section of the journal
Frontiers in Cardiovascular Medicine

Received: 30 September 2019

Accepted: 12 November 2019

Published: 26 November 2019

Citation:

Hampe N, Wolterink JM,
van Velzen SGM, Leiner T and Išgum I
(2019) Machine Learning for
Assessment of Coronary Artery
Disease in Cardiac CT: A Survey.
Front. Cardiovasc. Med. 6:172.
doi: 10.3389/fcvm.2019.00172

Cardiac computed tomography (CT) allows rapid visualization of the heart and coronary arteries with high spatial resolution. However, analysis of cardiac CT scans for manifestation of coronary artery disease is time-consuming and challenging. Machine learning (ML) approaches have the potential to address these challenges with high accuracy and consistent performance. In this mini review, we present a survey of the literature on ML-based analysis of coronary artery disease in cardiac CT. We summarize ML methods for detection and characterization of atherosclerotic plaque as well as anatomically and functionally significant coronary artery stenosis.

Keywords: machine learning, coronary artery disease, atherosclerotic plaque, coronary artery stenosis, cardiac CT

1. INTRODUCTION

Diagnosis and monitoring of coronary artery disease (CAD) is increasingly based on non-invasive imaging with computed tomography (CT), allowing excellent visualization of the coronary arteries with high spatial resolution. Cardiac CT exams consist of hundreds of slices and the number of cardiac CT studies has been steadily increasing (1). This has led to an increased workload for medical professionals, which in combination with shortages of trained cardiac imagers (2) might lead to cardiac CT underuse in the clinic. Machine learning (ML) could offer a way to address these challenges and facilitate automatic cardiac CT analysis with consistent and accurate results. Furthermore, ML algorithms might enable an increased range of secondary diagnoses.

This survey provides an overview of ML algorithms for detection, characterization, and quantification of CAD in cardiac CT. We searched PubMed for articles related to ML-based assessment of CAD in cardiac CT published within the last 10 years (search strategy in **Supplementary Materials**) which led to inclusion of 59 studies. The structure of this survey is as follows. We provide a brief primer on ML in section 2. Applications of ML for automatic detection and characterization of atherosclerotic plaque are summarized in section 3. Studies focusing on ML for anatomical and functional evaluation of luminal stenosis are summarized in section 4. Finally, section 5 provides a discussion of outstanding challenges for transfer of ML algorithms into the clinic.

2. MACHINE LEARNING

Machine learning comes in many flavors, but most applications in cardiac CT use supervised learning. In supervised learning, a model is optimized to provide the correct labels as defined by the reference standard during *training*, and predict a label to new and unseen samples during *testing*.

Each sample can be described based on characteristics or *features*. Among the simplest ML algorithms are k-nearest neighbor (kNN) classifiers, which look for training samples with similar feature values to a test sample, and assign the test sample to the majority class among these training samples. Linear classifier (LC) models like support vector machines (SVM) aim to find a linear combination of features to separate samples in different classes. Alternatively, samples can be separated by thresholding feature values along a single axis. This is unlikely to lead to highly accurate classifiers, but by consecutively applying thresholds, a decision tree model can be built for more accurate classification.

ML performance can often be improved by combining predictions of multiple models. Ensembles (E) combine predictions of multiple simultaneously executed models, e.g., by averaging predictions of decision trees in a random forest (RF). In boosting (BO), models are applied consecutively and each model is trained to correct errors of its predecessors. Finally, artificial neural networks (ANNs) transform samples into targets through layers of trainable neurons, which are loosely based on biological neurons. While ANNs have been around since the 1950s, it has recently become possible to train networks that have many layers, i.e., *deep learning*. The success of deep learning in medical image analysis has been to a large extent due to the inclusion of trainable image filters in so-called convolutional neural networks (CNNs), which can be trained to extract valuable features from raw image data (3). For a more in-depth introduction to ML and deep learning, please refer to Jordan and Mitchell (4).

3. ATHEROSCLEROTIC PLAQUE DETECTION, CHARACTERIZATION, AND QUANTIFICATION

CT offers a non-invasive alternative to e.g., catheter-guided X-ray angiography, optical coherence tomography, and intravascular ultrasound (IVUS) for atherosclerotic plaque visualization. Characterization and quantification of plaque in CT provide insight in different stages of CAD (5). In this section, we survey analysis of methods for calcified plaque (section 3.1) and non-calcified and mixed plaque (section 3.2). Reviewed papers are listed in **Table 1**.

3.1. Calcified Plaque

Coronary artery calcification (CAC) quantification or *scoring* is typically performed in dedicated non-contrast-enhanced, ECG-triggered, calcium scoring CT images (CSCT). Using dedicated software, an expert identifies voxels with a density over 130 Hounsfield units (HU) in the coronary arteries. Identified CAC is then quantified according to its volume, density, or a

TABLE 1 | Publications related to analysis of (A) calcified and (B) non-calcified and mixed atherosclerotic plaque.

(A) Calcified plaque	CSCT	Chest CT	CCTA	Training	Testing	Classifier
de Vos et al. (6)	✓	✓		1,554	1,036	CNN
Cano-Espinosa et al. (7)		✓		4,973	1,000	CNN
Lessmann et al. (8)		✓		1,181	506	CNN
Yang et al. (9)	✓		✓	32	40	SVM
Wolterink et al. (10)			✓	150	100	CNN
Wolterink et al. (11)	✓			384	570	RF
Shahzad et al. (12)	✓			209	157	kNN
Işgum et al. (13)		✓		337	231	kNN, SVM
Sánchez et al. (14)	✓			200	76	kNN
Liu et al. (15)	✓			*	31	SVM
Kurkure et al. (16)	✓			100	105	SVM
Brunner et al. (17)	✓			*	30	SVM
(B) Non-calcified plaque	Detect	Characterize		Training	Testing	Classifier
Kolossváry et al. (18)		✓		*	7	LC
Masuda et al. (19)		✓		*	78	BO
Zreik et al. (20)	✓	✓		98	65	CNN
Zhao et al. (21)	✓	✓		*	18	SVM
Jawaid et al. (22)	✓			*	32	SVM
Wei et al. (23)	✓			*	83	LC
Yamak et al. (24)		✓		-	3	E
Kelm et al. (25)	✓	✓		*	229	RF
Zuluaga et al. (26)	✓			1/13 [†]	14/2	SVM

Check marks in (A) indicate detection (Detect) or characterization (Characterize) of plaque, check marks in (B) indicate analysis in dedicated non-contrast-enhanced calcium scoring CT (CSCT), chest CT (Chest CT) or coronary CT angiography (CCTA) images. The number of patients included for method development (Training) and evaluation (Testing) are listed, * indicates cross-validation and - indicates training on non-patient data. The classifier with which the primary result was obtained is indicated (Classifier, see section 2 for abbreviations).

[†]A total of 15 scans was divided into training sets ranging from 1 to 13 and respective test sets comprised of the remaining scans.

combination of both (27). CAC cannot only be quantified in CSCT, but also in other kinds of CT images visualizing the heart, such as cardiac CT angiography (CCTA) and non-gated chest CT. Calcium scoring is not considered a difficult task for trained clinicians, but it is time-consuming when performed in large numbers of images. Hence, automatic ML-based methods have been proposed.

ML-based calcium scoring methods proposed prior to the advent of deep learning have focused on identification of CAC lesions among a large set of samples, i.e., groups of connected voxels above 130 HU. Samples are described with features such as size, shape, appearance and location to distinguish CAC from other candidate lesions such as calcifications in the aorta. Location features are of particular importance, as recognized by Liu et al. (15), Kurkure et al. (16), and Brunner et al. (17) who proposed a heart coordinate system. Similarly, Sánchez et al. (14) described candidate locations relative to anatomical landmarks.

Işgum et al. (13) used multi-atlas registration to estimate the location of the coronary artery tree, while Shahzad et al. (12) and Wolterink et al. (11) estimated the location of three major coronary arteries for per-vessel calcium scoring. Yang et al. (9) extracted coronary artery centerlines in CCTA images and propagated these to CSCT images of the same patients to provide location features.

Deep learning-based methods have typically classified individual voxels instead of candidate lesions. Due to the extreme imbalance between numbers of CAC and background voxels in CT images, Wolterink et al. (10) proposed to use two CNNs, where one CNN identified candidate voxels in CCTA and the second CNN further discriminated among identified candidates. Similarly, Lessmann et al. (8) used two CNNs to identify calcified voxels in chest CT. Cano-Espinosa et al. (7) and de Vos et al. (6) avoid voxel-based classification altogether by directly regressing calcium scores in chest CT, enabling automatic scoring in less than a second.

Automatic CAC scoring methods have been validated in large data sets (28) and in other types of CT scans in which the heart is routinely visualized, such as attenuation correction images for PET-CT (29) and CT images acquired for radiotherapy treatment planning (30–32). Wolterink et al. presented a public data set with reference standard for standardized evaluation of CAC scoring in CSCT (33).

3.2. Non-calcified Plaque

Non-calcified plaque is typically lipid-rich and vulnerable to rupture, causing acute coronary syndrome (34). ML-based analysis methods in CCTA have been developed for detection or localization of non-calcified plaque, as well as characterization of lipid and fibrous plaque components.

Coronary artery localization by means of centerline extraction is a typical preprocessing step for ML-based plaque analysis. Traditionally, many automatic centerline extraction methods have been based on minimum cost paths between proximal and distal artery points (35, 36). ML has been used to verify automatic centerline extraction results with an RF (25) or CNN (37). Alternatively, centerlines can be iteratively extracted based on a single seed point. Wolterink et al. (38) showed how such a tracker can be guided by a 3D CNN that locally detects the artery orientation.

Coronary artery centerlines can be used to reconstruct CCTA volumes into images that allow better plaque visualization and identification. Zhao et al. (21), Jawaid et al. (22), Wei et al. (23), and Zuluaga et al. (26) used cross-sectional images along the coronary artery centerline to extract features describing the vessel wall shape and texture. In Jawaid et al. (22) and Wei et al. (23), these features were used in an SVM or linear classifier to determine whether the image contained non-calcified plaque. Similarly, Zuluaga et al. (26) used such features to train an SVM classifying lesion segments as either healthy or diseased, i.e., containing non-calcified or calcified plaque. Zhao et al. (21) trained an SVM to classify cross-sectional images as healthy or containing non-calcified, calcified, or mixed plaque. For the same task, Zreik et al. (20) trained a recurrent CNN that did not depend on hand-crafted feature extraction.

Kelm et al. (25) used an RF classifier to classify whether non-calcified or calcified plaque was present along a coronary artery centerline segment.

Characterization of individual components in non-calcified plaque is a challenging task due to low-contrast boundaries between plaque components (39). Yamak et al. (24) exploited additional attenuation data provided by dual-energy CT to characterize plaque in manually determined regions of interest in axial slices. To validate their model in patient scans, manual CCTA annotations by an expert were used. However, obtaining reliable manual reference annotations for non-calcified plaque in CCTA is challenging. Kolossváry et al. (18) determined the reference standard in CCTA through registration of histology images to *ex-vivo* CCTA scans. Features were extracted for each cross-sectional image and lesions were classified into advanced or early stage atherosclerosis using a linear classifier. Alternatively, Masuda et al. (19) used an *in-vivo* IVUS-based reference standard to train a boosting classifier with histogram-based features distinguishing fibrous from lipid plaque in CCTA.

4. CORONARY STENOSIS DETECTION AND CHARACTERIZATION

Non-invasive assessment of CAD-induced stenotic lesions in CT prior to invasive treatment may prevent unnecessary costs and complications (40). Therefore, CT images have long been used to assess the *anatomical* significance of lesions by a local measurement of luminal narrowing. However, determination of the *functional* significance of a lesion by taking physiology into account can better stratify patients in need of treatment (41). In this section, we review ML algorithms for the detection and quantification of anatomically (section 4.1) and functionally (section 4.2) significant stenosis. Reviewed papers are listed in **Table 2**.

4.1. Anatomical Significance

Identification of anatomically significant stenotic lesions in CCTA, i.e., those lesions causing a luminal narrowing of at least 50%, allows a first assessment of the severity of stenosis in patients with symptoms of CAD. While this assessment is often based on visual estimation by a clinician, this is a difficult task (56) with substantial inter-observer variability (57). ML-based automatic approaches could reduce this variability.

Stenosis detection typically requires a local measurement of the lumen diameter and an estimation of the healthy lumen diameter. These estimates can be based on automatically extracted centerlines (section 3.2). Many centerline extraction methods also estimate the luminal radius at each centerline point, assuming a circular coronary artery profile (25, 38). However, circular artery profiles are not a realistic assumption for diseased vessel segments. Automatically extracted centerlines can also be used as an initialization for more detailed lumen segmentation. Huang et al. (44) used centerlines to obtain a reformatted image in which the lumen was segmented using a 3D CNN. Lee et al. (42) use centerlines to

TABLE 2 | Publications related to (A) anatomically and (B) functionally significant stenosis detection.

	Structure		Patients		
	Artery	Myocardium	Training	Testing	Classifier
(A) Anatomical significance					
Lee et al. (42)	✓		412	136	CNN
Freiman et al. (43)	✓		*	90	CNN
Zreik et al. (20)	✓		98	65	CNN
Huang et al. (44)	✓		45	7	CNN
Kang et al. (45)	✓		*	42	SVM
Xiong et al. (46)		✓	*	140	BO
Mukhopadhyay et al. (47)		✓	*	27	ANN
Kelm et al. (25)	✓		*	229	RF
Zuluaga et al. (48)	✓		9	9	SVM
(B) Functional significance					
Kumamaru et al. (49)	✓		*	131	CNN
Wang et al. (50)	✓		8	63	ANN
Hae et al. (51)		✓	932	279	BO
Dey et al. (52)	✓		*	254	BO
Zreik et al. (53)		✓	*	166	SVM
Han et al. (54)		✓	*	252	BO
Itu et al. (55)	✓		-	87	ANN

Check marks indicate arterial (Artery) or myocardial (Myocardium) analysis. The number of patients included for method development (Training) and evaluation (Testing) are listed, * indicates cross-validation and - indicates training on non-patient data. The classifier with which the primary result was obtained is indicated (Classifier, see section 2 for abbreviations).

obtain a tube-shaped prior that is deformed to segment the coronary lumen.

Lumen segmentation is often considered a preprocessing step for stenosis detection, but it has been shown that stenosis degree can also be directly determined based on image data. Zuluaga et al. (48) detected stenosis and artery bifurcations with an SVM based on features obtained from concentric circles in cross-sectional images. Similarly, Kang et al. (45) used geometrical and plaque features in an SVM to detect obstructive lesions (> 50% narrowing) and non-obstructive lesions (25–50% narrowing). Zreik et al. (20) used a recurrent CNN to detect anatomically significant stenosis along the centerline. Freiman et al. (43) detected stenosis of at least intermediate severity (> 40% narrowing) using deep sparse autoencoders, a variation on CNNs.

Coronary stenoses are located in the arteries, but may restrict blood flow to myocardial segments. Mukhopadhyay et al. (47) used an ML approach to identify myocardial segments (58) affected by coronary stenosis. Hand-crafted feature vectors describing the endocardial surface shape were combined using a bag-of-words approach and classified with an ANN to identify affected segments. Xiong et al. (46) performed analysis of the full myocardium to detect existence of at least one anatomically significant stenosis. Instead of the shape of the endocardial

surface, features in this approach described the attenuation and wall thickness of myocardial segments.

4.2. Functional Significance

The sensitivity of CCTA-based anatomical stenosis evaluation for detection of functionally significant stenosis is high when evaluated visually, but its specificity is moderate (41). The current reference standard for determination of functional significance of a stenosis is given by its fractional flow reserve (FFR), i.e., the ratio of flow distal of the stenosis to the flow proximal of the stenosis. FFR is measured invasively by inserting a special catheter in the coronary artery under hyperemic conditions. FFR below 0.80 indicates need for intervention (59). Treatment based on invasive FFR measurements can improve patient outcomes (59), but measurement of FFR is still relatively uncommon, which is due to associated cost and risk, as well as lack of vasodilator drugs (60).

FFR estimation based on CCTA scans (FFR_{CT}) could provide reproducible physical measurement without the drawbacks of invasive procedures. FFR_{CT} has traditionally been based on computational fluid dynamics (CFD) (61, 62), i.e., numerical simulation of blood flow in a coronary tree model extracted from CCTA using lumen segmentation methods (section 4.1). These methods are accurate (63) but computationally expensive due to their iterative nature. This precludes their deployment on local workstations, and instead CFD simulations are typically performed on off-site dedicated systems. ML could be used to significantly speed up estimation of FFR_{CT} .

Itu et al. (55) proposed an ANN model to predict an FFR value for each segment in the coronary artery tree, given local features based on the segment's geometry and global features based on the most severe stenoses. To train this model, a large data set of 12,000 synthetic coronary artery trees was generated and a reference standard was obtained through conventional CFD simulation. By only performing CFD simulations once in a training phase, the time required to perform FFR_{CT} was reduced by two orders of magnitude. The diagnostic value of this method has been demonstrated thoroughly (64–76). Yu et al. (77) further demonstrated additional prognostic value of CT morphological index for the method proposed by Itu et al. (55). Wang et al. (50) proposed to use a recurrent ANN that can model long-range dependencies between segments.

Both conventional CFD-based FFR_{CT} and the methods proposed in Wang et al. (50) and Itu et al. (55) are based only on the geometry of the coronary artery tree model, and are thus susceptible to errors by the segmentation method used to obtain this model (78). Instead, Dey et al. (52) proposed to combine geometric features with semi-automatically obtained plaque and attenuation gradient measurements to identify arteries with functionally significant stenosis. Other methods skip explicit coronary artery centerline extraction and lumen segmentation altogether. Kumamaru et al. (49) trained a CNN to extract a map showing the contrast-enhanced territories in CCTA and used this map in a classifier to predict the minimum FFR value in a patient. Alternatively, analysis can be moved from the cause—the coronary arteries—to the effect, i.e., the myocardium. Han et al. (54) subdivided the separated endocardium and epicardium

into the American Heart Association (AHA) 17 segments (58), with 3 features per segment characterizing perfusion and wall thickness. However, the trained boosting classifier showed only moderate accuracy for patientwise prediction of abnormal FFR values. For the same purpose, Zreik et al. (53) trained an SVM based on features from myocardial regions extracted from CCTA. Clinical evaluation of this method yielded improved diagnostic accuracy of FFR_{CT} over visual evaluation of stenosis (79). Hae et al. (51) increased accuracy of FFR-prediction by including the tissue volume subtended to a stenotic lesion in analysis. However, determination of lesion position required additional analysis including artery tree segmentation.

5. DISCUSSION

We have presented a survey of applications of ML for detection, characterization and quantification of atherosclerotic plaque and stenosis in cardiac CT. We found that while ML has been a mainstay of cardiac image analysis for years, the recent emergence of deep learning has accelerated progress in the field. Machine learning has the potential to unburden clinicians from time-consuming tasks and change diagnostic procedures, thereby reducing healthcare costs. Moreover, low-cost ML-based analysis could be added to screening studies as a secondary goal. In this survey, we have focused on ML for CAD analysis. For a broader scope the reader is referred to Al'Aref et al. (80), Litjens et al. (81), Nicol et al. (82), Petersen et al. (83), and Singh et al. (84).

We have reviewed plaque and stenosis analysis methods in separate sections, but formation of plaque and stenosis is naturally related and many papers have proposed simultaneous analysis [e.g., (48, 53)]. Moreover, (semi-)automatic identification of plaque or stenosis is often only an intermediate step for prediction of cardiovascular events. Motwani et al. (85) used stenosis scores and plaque characteristics to develop a model for 5 years all-cause mortality prediction. Similarly, Johnson et al. (86) showed that an ML model taking into account per segment coronary artery characteristics can outperform hand-crafted models for prediction of adverse cardiac events. Van Rosendael et al. (87) developed a model for all-cause mortality prediction in combination with future myocardial infarction based only on hand-crafted features derived from CCTA scans. Furthermore, some methods directly predict presence of CAD from medical images, i.e., chest CT (88) or non-contrast-enhanced cardiac CT (89). While these approaches only require one label per patient and large data sets are thus not expensive to obtain, the interpretability of predictions may be limited. Interpretability might constitute an opportunity, not only to improve reliability but also as it might increase

medical knowledge by quantifying the diagnostic relevance of underlying phenomena.

The readiness of automatic analysis methods for clinical implementation depends on the complexity of the task, but also on other factors. ML algorithms require large training sets, and tasks with abundant data may be easier to automate. For example, obtaining a ground truth for e.g., non-calcified plaque characterization is very challenging. Therefore, data sets are generally small and ML algorithms remain at an early developmental stage. In contrast, large data sets are available for the development of ML-based CAC scoring methods, which has led to highly accurate results in both dedicated cardiac CT images (11) and other CT images visualizing the heart (8, 32). Similarly, ML-based FFR_{CT} development is aided by the availability of large data sets with CFD-derived reference values. An important remaining step toward clinical application of FFR_{CT} lies in performance evaluation specifically for subjects around the FFR threshold of 0.8, which were shown to be most challenging (90). Furthermore, a recent study showed that not all CCTA exams are suitable for FFR_{CT} analysis (78).

Many challenges in the adoption of machine learning methods in the clinic are not exclusive to CAD detection in cardiac CT. For example, ML algorithms could show unexpected behavior, motivating research into ML interpretability and explainability (91). Furthermore, it is important to point out that ML algorithms are often trained and evaluated on single center studies with high risk for selective biases, and under exclusion of low quality scans.

Despite these challenges, current rapid development allows for justifiable hope that the importance of ML algorithms in cardiac CT will not cease to increase in near future, with benefits for clinicians and patients alike.

AUTHOR CONTRIBUTIONS

NH and SV drafted the manuscript, which was critically revised and edited by JW, TL, and II. Authors agree to be accountable for all aspects of the work.

FUNDING

This study has been financially supported by Pie Medical Imaging BV.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcvm.2019.00172/full#supplementary-material>

REFERENCES

- Levin DC, Parker L, Halpern EJ, Rao VM. Coronary CT angiography: reversal of earlier utilization trends. *J Am Coll Radiol*. (2019) 16:147–55. doi: 10.1016/j.jacr.2018.07.022
- Dreisbach JG, Nicol ED, Roobottom CA, Padley S, Roditi G. Challenges in delivering computed tomography coronary angiography as the first-line test for stable chest pain. *Heart*. (2018) 104:921–7. doi: 10.1136/heartjnl-2017-311846
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005
- Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science*. (2015) 349:255–60. doi: 10.1126/science.aaa8415

5. Insull W. The pathology of atherosclerosis: plaque development and plaque responses to medical treatment. *Am J Med.* (2009) 122(1, Suppl.):S3–14. doi: 10.1016/j.amjmed.2008.10.013
6. de Vos BD, Wolterink JM, Leiner T, Jong PAD, Lessmann N, Išgum I. Direct automatic coronary calcium scoring in cardiac and chest CT. *IEEE Trans. Med. Imaging.* (2019) 38:2127–38. doi: 10.1109/TMI.2019.2899534
7. Cano-Espinosa C, González G, Washko GR, Cazorla M, Estépar RSJ. Automated agatston score computation in non-ECG gated CT scans using deep learning. In: *Proceedings of SPIE—the International Society for Optical Engineering*. vol. 10574 (2018). Available online at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6095680/>
8. Lessmann N, Ginneken BV, Zreik M, Jong PAD, Vos BDD, Viergever MA, et al. Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions. *IEEE Trans Med Imaging.* (2018) 37:615–25. doi: 10.1109/TMI.2017.2769839
9. Yang G, Chen Y, Ning X, Sun Q, Shu H, Coatrieux JL. Automatic coronary calcium scoring using noncontrast and contrast CT images. *Med Phys.* (2016) 43:2174. doi: 10.1118/1.4945045
10. Wolterink JM, Leiner T, de Vos BD, van Hamersvelt RW, Viergever MA, Išgum I. Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Med Image Anal.* (2016) 34:123–36. doi: 10.1016/j.media.2016.04.004
11. Wolterink JM, Leiner T, Takx RAP, Viergever MA, Išgum I. Automatic coronary calcium scoring in non-contrast-enhanced ECG-triggered cardiac CT with ambiguity detection. *IEEE Trans Med Imaging.* (2015) 34:1867–78. doi: 10.1109/TMI.2015.2412651
12. Shahzad R, van Walsum T, Schaap M, Rossi A, Klein S, Weustink AC, et al. Vessel specific coronary artery calcium scoring: an automatic system. *Acad Radiol.* (2013) 20:1–9. doi: 10.1016/j.acra.2012.07.018
13. Išgum I, Prokop M, Niemeijer M, Viergever MA, van Ginneken B. Automatic coronary calcium scoring in low-dose chest computed tomography. *IEEE Trans Med Imaging.* (2012) 31:2322–34. doi: 10.1109/TMI.2012.2216889
14. Sánchez CI, Niemeijer M, Išgum I, Dumitrescu A, Suttrop-Schulten MSA, Abramoff MD, et al. Contextual computer-aided detection: improving bright lesion detection in retinal images and coronary calcification identification in CT scans. *Med Image Anal.* (2012) 16:50–62. doi: 10.1016/j.media.2011.05.004
15. Liu Q, Qian Z, Marvasti I, Rinehart S, Voros S, Metaxas DN. Lesion-specific coronary artery calcium quantification for predicting cardiac event with multiple instance support vector machines. In: *Medical Image Computing and Computer-Assisted Intervention: MICCAI*, Vol. 13. Beijing (2010). p. 484–92.
16. Kurkure U, Chittajallu DR, Brunner G, Le YH, Kakadiaris IA. A supervised classification-based method for coronary calcium detection in non-contrast CT. *Int J Cardiovasc Imaging.* (2010) 26:817–28. doi: 10.1007/s10554-010-9607-2
17. Brunner G, Chittajallu DR, Kurkure U, Kakadiaris IA. Toward the automatic detection of coronary artery calcification in non-contrast computed tomography data. *Int J Cardiovasc Imaging.* (2010) 26:829–38. doi: 10.1007/s10554-010-9608-1
18. Kolossváry M, Karády J, Kikuchi Y, Ivanov A, Schlett CL, Lu MT, et al. Radiomics versus visual and histogram-based assessment to identify atheromatous lesions at coronary CT angiography: an *ex vivo* study. *Radiology.* (2019) 293:190407. doi: 10.1148/radiol.2019190407
19. Masuda T, Nakaura T, Funama Y, Okimoto T, Sato T, Higaki T, et al. Machine-learning integration of CT histogram analysis to evaluate the composition of atherosclerotic plaques: validation with IB-IVUS. *J Cardiovasc Comput Tomogr.* (2019) 13:163–9. doi: 10.1016/j.jcct.2018.10.018
20. Zreik M, van Hamersvelt RW, Wolterink JM, Leiner T, Viergever MA, Išgum I. A recurrent CNN for automatic detection and classification of coronary artery plaque and stenosis in coronary CT angiography. *IEEE Trans Med Imaging.* (2019) 38:1588–98. doi: 10.1109/TMI.2018.2883807
21. Zhao F, Wu B, Chen F, Cao X, Yi H, Hou Y, et al. An automatic multi-class coronary atherosclerosis plaque detection and classification framework. *Med Biol Eng Comput.* (2019) 57:245–57. doi: 10.1007/s11517-018-1880-6
22. Jawaid MM, Riaz A, Rajani R, Reyes-Aldasoro CC, Slabaugh G. Framework for detection and localization of coronary non-calcified plaques in cardiac CTA using mean radial profiles. *Comput Biol Med.* (2017) 89:84–95. doi: 10.1016/j.combiomed.2017.07.021
23. Wei J, Zhou C, Chan HP, Chughtai A, Agarwal P, Kuriakose J, et al. Computerized detection of noncalcified plaques in coronary CT angiography: evaluation of topological soft gradient prescreening method and luminal analysis. *Med Phys.* (2014) 41:081901. doi: 10.1118/1.4885958
24. Yamak D, Panse P, Pavlicek W, Boltz T, Akay M. Non-calcified coronary atherosclerotic plaque characterization by dual energy computed tomography. *IEEE J Biomed Health Informat.* (2014) 18:939–45. doi: 10.1109/JBHI.2013.2295534
25. Kelm BM, Mittal S, Zheng Y, Tsybmal A, Bernhardt D, Vega-Higuera F, et al. Detection, grading and classification of coronary stenoses in computed tomography angiography. In: *Medical Image Computing and Computer-Assisted Intervention: MICCAI*, Vol. 14. Toronto, ON (2011). p. 25–32.
26. Zuluaga MA, Hush D, Delgado Leyton EJJ, Hernández Hoyos M, Orkisz M. Learning from only positive and unlabeled data to detect lesions in vascular CT images. In: *Medical Image Computing and Computer-Assisted Intervention: MICCAI*, Vol. 14. Toronto, ON (2011). p. 9–16.
27. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M, Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol.* (1990) 15:827–32. doi: 10.1016/0735-1097(90)90282-T
28. Takx RAP, de Jong PA, Leiner T, Oudkerk M, de Koning HJ, Mol CP, et al. Automated coronary artery calcification scoring in non-gated chest CT: agreement and reliability. *PLoS ONE.* (2014) 9:e91239. doi: 10.1371/journal.pone.0091239
29. Išgum I, de Vos BD, Wolterink JM, Dey D, Berman DS, Rubeaux M, et al. Automatic determination of cardiovascular risk by CT attenuation correction maps in Rb-82 PET/CT. *J Nucl Cardiol.* (2018) 25:2133–42. doi: 10.1007/s12350-017-0866-3
30. Emaus MJ, Išgum I, van Velzen SGM, van den Bongard HJGD, Gernaat SAM, Lessmann N, et al. Bragatston study protocol: a multicentre cohort study on automated quantification of cardiovascular calcifications on radiotherapy planning CT scans for cardiovascular risk prediction in patients with breast cancer. *BMJ Open.* (2019) 9:e028752. doi: 10.1136/bmjopen-2018-028752
31. Gernaat SAM, Išgum I, de Vos BD, Takx RAP, Young-Afat DA, Rijnberg N, et al. Automatic coronary artery calcium scoring on radiotherapy planning CT scans of breast cancer patients: reproducibility and association with traditional cardiovascular risk factors. *PLoS ONE.* (2016) 11:e0167925. doi: 10.1371/journal.pone.0167925
32. Gernaat SAM, van Velzen SGM, Koh V, Emaus MJ, Išgum I, Lessmann N, et al. Automatic quantification of calcifications in the coronary arteries and thoracic aorta on radiotherapy planning CT scans of Western and Asian breast cancer patients. *Radiother Oncol.* (2018) 127:487–92. doi: 10.1016/j.radonc.2018.04.011
33. Wolterink JM, Leiner T, de Vos BD, Coatrieux JL, Kelm BM, Kondo S, et al. An evaluation of automatic coronary artery calcium scoring methods with cardiac CT using the orCaScore framework. *Med Phys.* (2016) 43:2361. doi: 10.1118/1.4945696
34. Virmani R, Burke AP, Farb A, Kolodgie FD. Pathology of the vulnerable plaque. *J Am Coll Cardiol.* (2006) 47(Suppl. 8):C13–18. doi: 10.1016/j.jacc.2005.10.065
35. Lesage D, Angelini ED, Bloch I, Funka-Lea G. A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes. *Med Image Anal.* (2009) 13:819–45. doi: 10.1016/j.media.2009.07.011
36. Schaap M, Metz CT, van Walsum T, van der Giessen AG, Weustink AC, Mollet NR, et al. Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *Med Image Anal.* (2009) 13:701–14. doi: 10.1016/j.media.2009.06.003
37. Gülsün MA, Funka-Lea G, Sharma P, Rapaka S, Zheng Y. Coronary centerline extraction via optimal flow paths and CNN path pruning. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, editors. *Medical Image Computing and Computer-Assisted Intervention: MICCAI*. Lecture Notes in Computer Science. Athens: Springer International Publishing (2016). p. 317–25.
38. Wolterink JM, van Hamersvelt RW, Viergever MA, Leiner T, Išgum I. Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier. *Med Image Anal.* (2019) 51:46–60. doi: 10.1016/j.media.2018.10.005

39. Kristanto W, van Ooijen PMA, Jansen-van der Weide MC, Vliegenthart R, Oudkerk M. A meta analysis and hierarchical classification of HU-based atherosclerotic plaque characterization criteria. *PLoS ONE*. (2013) 8:e73460. doi: 10.1371/journal.pone.0073460
40. Pijls NHJ, Tanaka N, Fearon WF. Functional assessment of coronary stenoses: can we live without it? *Eur Heart J*. (2013) 34:1335–44. doi: 10.1093/eurheartj/ehs436
41. Meijboom WB, Van Mieghem CAG, van Pelt N, Weustink A, Pugliese F, Mollet NR, et al. Comprehensive assessment of coronary artery stenoses: computed tomography coronary angiography versus conventional coronary angiography and correlation with fractional flow reserve in patients with stable angina. *J Am Coll Cardiol*. (2008) 52:636–43. doi: 10.1016/j.jacc.2008.05.024
42. Lee MCH, Petersen K, Pawlowski N, Glocker B, Schaap M. TETRIS: template transformer networks for image segmentation with shape priors. *IEEE Trans Med Imaging*. (2019) 38:2596–606. doi: 10.1109/TMI.2019.2905990
43. Freiman M, Manjeshwar R, Goshen L. Unsupervised abnormality detection through mixed structure regularization (MSR) in deep sparse autoencoders. *Med Phys*. (2019) 46:2223–31. doi: 10.1002/mp.13464
44. Huang W, Huang L, Lin Z, Huang S, Chi Y, Zhou J, et al. Coronary artery segmentation by deep learning neural networks on computed tomographic coronary angiographic images. In: *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, Vol. 2018. Honolulu, HI (2018). p. 608–11.
45. Kang D, Dey D, Slomka PJ, Arsanjani R, Nakazato R, Ko H, et al. Structured learning algorithm for detection of nonobstructive and obstructive coronary plaque lesions from computed tomography angiography. *J Med Imaging*. (2015) 2:014003. doi: 10.1117/1.JMI.2.1.014003
46. Xiong G, Kola D, Heo R, Elmore K, Cho I, Min JK. Myocardial perfusion analysis in cardiac computed tomography angiographic images at rest. *Med Image Anal*. (2015) 24:77–89. doi: 10.1016/j.media.2015.05.010
47. Mukhopadhyay A, Qian Z, Bhandarkar SM, Liu T, Rinehart S, Voros S. Morphological analysis of the left ventricular endocardial surface and its clinical implications. In: *Medical Image Computing and Computer-Assisted Intervention: MICCAI*, Vol. 15. Nice (2012). p. 502–10.
48. Zuluaga MA, Magnin IE, Hernández Hoyos M, Delgado Leyton EJF, Lozano F, Orkisz M. Automatic detection of abnormal vascular cross-sections based on density level detection and support vector machines. *Int J Comput Assist Radiol Surg*. (2011) 6:163–74. doi: 10.1007/s11548-010-0494-8
49. Kumamaru KK, Fujimoto S, Otsuka Y, Kawasaki T, Kawaguchi Y, Kato E, et al. Diagnostic accuracy of 3D deep-learning-based fully automated estimation of patient-level minimum fractional flow reserve from coronary computed tomography angiography. *Eur Heart J Cardiovasc Imaging*. (2019). doi: 10.1093/ehjci/jez160. [Epub ahead of print].
50. Wang ZQ, Zhou YJ, Zhao YX, Shi DM, Liu YY, Liu W, et al. Diagnostic accuracy of a deep learning approach to calculate FFR from coronary CT angiography. *J Geriatr Cardiol*. (2019) 16:42–8. doi: 10.11909/j.issn.1671-5411.2019.01.010
51. Hae H, Kang SJ, Kim WJ, Choi SY, Lee JG, Bae Y, et al. Machine learning assessment of myocardial ischemia using angiography: development and retrospective validation. *PLoS Med*. (2018) 15:e1002693. doi: 10.1371/journal.pmed.1002693
52. Dey D, Gaur S, Ovrehus KA, Slomka PJ, Betancur J, Goeller M, et al. Integrated prediction of lesion-specific ischemia from quantitative coronary CT Angiography using machine learning: a multicenter study. *Eur Radiol*. (2018) 28:2655–64. doi: 10.1007/s00330-017-5223-z
53. Zreik M, Lessmann N, van Hamersvelt RW, Wolterink JM, Voskuil M, Viergever MA, et al. Deep learning analysis of the myocardium in coronary CT angiography for identification of patients with functionally significant coronary artery stenosis. *Med Image Anal*. (2018) 44:72–85. doi: 10.1016/j.media.2017.11.008
54. Han D, Lee JH, Rizvi A, Gransar H, Baskaran L, Schulman-Marcus J, et al. Incremental role of resting myocardial computed tomography perfusion for predicting physiologically significant coronary artery disease: a machine learning approach. *J Nucl Cardiol*. (2018) 25:223–33. doi: 10.1007/s12350-017-0834-y
55. Itu L, Rapaka S, Passerini T, Georgescu B, Schwemmer C, Schoebinger M, et al. A machine-learning approach for computation of fractional flow reserve from coronary computed tomography. *J Appl Physiol*. (2016) 121:42–52. doi: 10.1152/japplphysiol.00752.2015
56. Pugliese F, Hunink MGM, Gruszczynska K, Alberghina F, Malagó R, van Pelt N, et al. Learning curve for coronary CT angiography: what constitutes sufficient training? *Radiology*. (2009) 251:359–68. doi: 10.1148/radiol.2512080384
57. Arbab-Zadeh A, Hoe J. Quantification of coronary arterial stenoses by multidetector CT angiography in comparison with conventional angiography methods, caveats, and implications. *JACC Cardiovasc Imaging*. (2011) 4:191–202. doi: 10.1016/j.jcmg.2010.10.011
58. Cerqueira MD, Weissman NJ, Dilsizian V, Jacobs AK, Kaul S, Laskey WK, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. A statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Circulation*. (2002) 105:539–42. doi: 10.1161/hc0402.102975
59. Tonino PAL, De Bruyne B, Pijls NHJ, Siebert U, Ikeno F, van't Veer M, et al. Fractional flow reserve versus angiography for guiding percutaneous coronary intervention. *N Engl J Med*. (2009) 360:213–24. doi: 10.1056/NEJMoa0807611
60. Petraco R, Park JJ, Sen S, Nijjer SS, Malik IS, Echavarría-Pinto M, et al. Hybrid iFR-FFR decision-making strategy: implications for enhancing universal adoption of physiology-guided coronary revascularisation. *EuroIntervention*. (2013) 8:1157–65. doi: 10.4244/EIJV8I10A179
61. Taylor CA, Fonte TA, Min JK. Computational fluid dynamics applied to cardiac computed tomography for noninvasive quantification of fractional flow reserve: scientific basis. *J Am Coll Cardiol*. (2013) 61:2233–41. doi: 10.1016/j.jacc.2012.11.083
62. Tesche C, Vliegenthart R, Duguay TM, De Cecco CN, Albrecht MH, De Santis D, et al. Coronary computed tomographic angiography-derived fractional flow reserve for therapeutic decision making. *Am J Cardiol*. (2017) 120:2121–7. doi: 10.1016/j.amjcard.2017.08.034
63. Nørgaard BL, Leipsic J, Gaur S, Seneviratne S, Ko BS, Ito H, et al. Diagnostic performance of noninvasive fractional flow reserve derived from coronary computed tomography angiography in suspected coronary artery disease: the NXT trial (Analysis of Coronary Blood Flow Using CT Angiography: next steps). *J Am Coll Cardiol*. (2014) 63:1145–55. doi: 10.1016/j.jacc.2013.11.043
64. Baumann S, Renker M, Schoepf UJ, De Cecco CN, Coenen A, De Geer J, et al. Gender differences in the diagnostic performance of machine learning coronary CT angiography-derived fractional flow reserve - results from the MACHINE registry. *Eur J Radiol*. (2019) 119:108657. doi: 10.1016/j.ejrad.2019.108657
65. Coenen A, Kim YH, Kruk M, Tesche C, De Geer J, Kurata A, et al. Diagnostic accuracy of a machine-learning approach to coronary computed tomographic angiography-based fractional flow reserve: result from the MACHINE consortium. *Circulat Cardiovasc Imaging*. (2018) 11:e007217. doi: 10.1161/CIRCIMAGING.117.007217
66. Duguay TM, Tesche C, Vliegenthart R, De Cecco CN, Lin H, Albrecht MH, et al. Coronary computed tomographic angiography-derived fractional flow reserve based on machine learning for risk stratification of non-culprit coronary narrowings in patients with acute coronary syndrome. *Am J Cardiol*. (2017) 120:1260–6. doi: 10.1016/j.amjcard.2017.07.008
67. Hu X, Yang M, Han L, Du Y. Diagnostic performance of machine-learning-based computed fractional flow reserve (FFR) derived from coronary computed tomography angiography for the assessment of myocardial ischemia verified by invasive FFR. *Int J Cardiovasc Imaging*. (2018) 34:1987–996. doi: 10.1007/s10554-018-1419-9
68. Kurata A, Fukuyama N, Hirai K, Kawaguchi N, Tanabe Y, Okayama H, et al. On-site computed tomography-derived fractional flow reserve using a machine-learning algorithm - clinical effectiveness in a retrospective multicenter cohort. *Circulat J*. (2019) 83:1563–71. doi: 10.1253/circj.CJ-19-0163
69. Nous FMA, Coenen A, Boersma E, Kim YH, Kruk MBP, Tesche C, et al. Comparison of the diagnostic performance of coronary computed tomography angiography-derived fractional flow reserve in patients with versus without diabetes mellitus (from the MACHINE consortium). *Am J Cardiol*. (2019) 123:537–43. doi: 10.1016/j.amjcard.2018.11.024

70. Tang CX, Wang YN, Zhou F, Schoepf UJ, Assen Mv, Stroud RE, et al. Diagnostic performance of fractional flow reserve derived from coronary CT angiography for detection of lesion-specific ischemia: a multi-center study and meta-analysis. *Eur J Radiol.* (2019) 116:90–7. doi: 10.1016/j.ejrad.2019.04.011
71. Tesche C, De Cecco CN, Albrecht MH, Duguay TM, Bayer RR, Litwin SE, et al. Coronary CT angiography-derived fractional flow reserve. *Radiology.* (2017) 285:17–33. doi: 10.1148/radiol.2017162641
72. Tesche C, Otani K, De Cecco CN, Coenen A, De Geer J, Kruk M, et al. Influence of coronary calcium on diagnostic performance of machine learning CT-FFR: results from MACHINE registry. *JACC Cardiovasc Imaging.* (2019). doi: 10.1016/j.jcmg.2019.06.027
73. von Knebel Doeberitz PL, De Cecco CN, Schoepf UJ, Duguay TM, Albrecht MH, van Assen M, et al. Coronary CT angiography-derived plaque quantification with artificial intelligence CT fractional flow reserve for the identification of lesion-specific ischemia. *Eur Radiol.* (2019) 29:2378–87. doi: 10.1007/s00330-018-5834-z
74. Wardziak L, Kruk M, Pleban W, Demkow M, Ruzyllo W, Dzielinska Z, et al. Coronary CTA enhanced with CTA based FFR analysis provides higher diagnostic value than invasive coronary angiography in patients with intermediate coronary stenosis. *J Cardiovasc Comput Tomogr.* (2019) 13:62–7. doi: 10.1016/j.jcct.2018.10.004
75. Yu M, Lu Z, Shen C, Yan J, Wang Y, Lu B, et al. The best predictor of ischemic coronary stenosis: subtended myocardial volume, machine learning-based FFRCT, or high-risk plaque features? *Eur. Radiol.* (2019) 29:3647–57. doi: 10.1007/s00330-019-06139-2
76. Zhou F, Tang CX, Schoepf UJ, Tesche C, Rollins JD, Liu H, et al. Machine learning using CT-FFR predicts proximal atherosclerotic plaque formation associated with LAD myocardial bridging. *JACC Cardiovasc Imaging.* (2019) 12:1591–3. doi: 10.1016/j.jcmg.2019.01.018
77. Yu M, Lu Z, Li W, Wei M, Yan J, Zhang J. CT morphological index provides incremental value to machine learning based CT-FFR for predicting hemodynamically significant coronary stenosis. *Int J Cardiol.* (2018) 265:256–61. doi: 10.1016/j.ijcard.2018.01.075
78. Pontone G, Weir-McCall JR, Baggiano A, Del Torto A, Fusini L, Guglielmo M, et al. Determinants of rejection rate for coronary CT angiography fractional flow reserve analysis. *Radiology.* (2019) 292:597–605. doi: 10.1148/radiol.2019182673
79. van Hamersvelt RW, Zreik M, Voskuil M, Viergever MA, Išgum I, Leiner T. Deep learning analysis of left ventricular myocardium in CT angiographic intermediate-degree coronary stenosis improves the diagnostic accuracy for identification of functionally significant stenosis. *Eur Radiol.* (2019) 29:2350–9. doi: 10.1007/s00330-018-5822-3
80. Al'Aref SJ, Anchouche K, Singh G, Slomka PJ, Kolli KK, Kumar A, et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *Eur Heart J.* (2019) 40:1975–86. doi: 10.1093/eurheartj/ehy404
81. Litjens G, Ciompi F, Wolterink JM, de Vos BD, Leiner T, Teuwen J, et al. State-of-the-art deep learning in cardiovascular image analysis. *JACC Cardiovasc Imaging.* (2019) 12(8 Pt 1):1549–65. doi: 10.1016/j.jcmg.2019.06.009
82. Nicol ED, Norgaard BL, Blanke P, Ahmadi A, Weir-McCall J, Horvat PM, et al. The future of cardiovascular computed tomography: advanced analytics and clinical insights. *JACC.* (2019) 12:1058–72. doi: 10.1016/j.jcmg.2018.11.037
83. Petersen SE, Abdulkareem M, Leiner T. Artificial intelligence will transform cardiac imaging-opportunities and challenges. *Front Cardiovasc Med.* (2019) 6:133. doi: 10.3389/fcvm.2019.00133
84. Singh G, Al'Aref SJ, Van Assen M, Kim TS, van Rosendael A, Kolli KK, et al. Machine learning in cardiac CT: basic concepts and contemporary data. *J Cardiovasc Comput Tomogr.* (2018) 12:192–201. doi: 10.1016/j.jcct.2018.04.010
85. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J.* (2017) 38:500–7. doi: 10.1093/eurheartj/ehw188
86. Johnson KM, Johnson HE, Zhao Y, Dowe DA, Staib LH. Scoring of coronary artery disease characteristics on coronary CT angiograms by using machine learning. *Radiology.* (2019) 292:354–62. doi: 10.1148/radiol.2019182061
87. Van Rosendael AR, Maliakal G, Kolli KK, Beecy A, Al'Aref SJ, Dwivedi A, et al. Maximization of the usage of coronary CTA derived plaque information using a machine learning based algorithm to improve risk stratification; insights from the CONFIRM registry. *J Cardiovasc Comput Tomogr.* (2018) 12:204–9. doi: 10.1016/j.jcct.2018.04.011
88. Dormer JD, Halicek M, Ma L, Reilly CM, Schreiber E, Fei B. Convolutional neural networks for the detection of diseased hearts using CT images and left atrium patches. In: *Proceedings of SPIE—the International Society for Optical Engineering*, Vol. 10575. San Diego, CA (2018). doi: 10.1117/12.2293548
89. Mannil M, von Spiczak J, Manka R, Alkadhi H. Texture analysis and machine learning for detecting myocardial infarction in noncontrast low-dose computed tomography: unveiling the invisible. *Investigat Radiol.* (2018) 53:338–43. doi: 10.1097/RLL.0000000000000448
90. Cook CM, Petraco R, Shun-Shin MJ, Ahmad Y, Nijjer S, Al-Lamee R, et al. Diagnostic accuracy of computed tomography-derived fractional flow reserve: a systematic review. *JAMA Cardiol.* (2017) 2:803–10. doi: 10.1001/jamacardio.2017.1314
91. Samek W, Montavon G, Vedaldi A, Hansen LK, Müller KR, editors. Explainable AI: interpreting, explaining and visualizing deep learning. In: *Lecture Notes in Artificial Intelligence, Lecture Notes Computer Science of the Art Surveys*. Springer International Publishing (2019). Available online at: <https://www.springer.com/gp/book/9783030289539>

Conflict of Interest: II and TL received institutional research projects by Dutch Technology Foundation cofunded by Pie Medical Imaging and Philips Healthcare (P15-26), the Netherlands Organisation for Health Research and Development with participation of Pie Medical Imaging (104003009). II received institutional research project by Dutch Technology Foundation cofunded by Pie Medical Imaging (12726). II and TL are cofounders and shareholders of Quantib-U BV.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Hampe, Wolterink, van Velzen, Leiner and Išgum. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Machine Learning Approaches for Myocardial Motion and Deformation Analysis

Nicolas Duchateau^{1*}, Andrew P. King² and Mathieu De Craene³

¹ CREATIS, CNRS UMR 5220, INSERM U1206, Université, Lyon, France, ² School of Biomedical Engineering and Imaging Sciences, King's College London, London, United Kingdom, ³ Philips Research Paris, Suresnes, France

OPEN ACCESS

Edited by:

Steffen Erhard Petersen,
Queen Mary University of London,
United Kingdom

Reviewed by:

Antonio De Marvao,
Imperial College London,
United Kingdom
Wenjia Bai,
Imperial College London,
United Kingdom
Alireza Sojoudi,
University of Calgary, Canada

*Correspondence:

Nicolas Duchateau
nicolas.duchateau@creatis.insa-lyon.fr

Specialty section:

This article was submitted to
Cardiovascular Imaging,
a section of the journal
Frontiers in Cardiovascular Medicine

Received: 14 October 2019

Accepted: 12 December 2019

Published: 09 January 2020

Citation:

Duchateau N, King AP and De
Craene M (2020) Machine Learning
Approaches for Myocardial Motion
and Deformation Analysis.
Front. Cardiovasc. Med. 6:190.
doi: 10.3389/fcvm.2019.00190

Information about myocardial motion and deformation is key to differentiate normal and abnormal conditions. With the advent of approaches relying on data rather than pre-conceived models, machine learning could either improve the robustness of motion quantification or reveal patterns of motion and deformation (rather than single parameters) that differentiate pathologies. We review machine learning strategies for extracting motion-related descriptors and analyzing such features among populations, keeping in mind constraints specific to the cardiac application.

Keywords: machine learning, computer-aided diagnosis, myocardial motion, myocardial strain, cardiac imaging

1. INTRODUCTION

1.1. Myocardial Motion and Deformation Analysis: What For?

Pump efficiency can discriminate failing from healthy hearts, as quantified by volume and ejection fraction. Clinicians are well aware of the limitations of these simple measurements to face the complexity of heart disease, and recommend finer markers of cardiac mechanical dysfunction (1). Myocardial motion (displacement or velocity) and deformation (strain or strain rate) are richer descriptors of (ab)normal cardiac function (2, 3). They can provide characteristic spatiotemporal signatures for disease at each location of the myocardium and each instant of the cardiac cycle. They are often projected onto anatomically-relevant directions to facilitate interpretations (4). Interestingly, they can be estimated from routine modalities such as echocardiography and magnetic resonance (MR) (5), and have therefore been thoroughly investigated for a wide range of applications.

1.2. Machine Learning for Myocardial Motion and Deformation Analysis: What For?

Machine learning builds upon models whose optimal parameters are learnt from a set of samples representative of the studied population. This data-driven approach is more flexible than traditional methods (e.g., variational), as demonstrated for myocardial segmentation (6, 7), and has strong potential for the analysis of complex descriptors such as myocardial motion and deformation. In essence, machine learning seeks to learn data representations (either explicit or hidden) for better solving a supervised problem or for characterizing the data distribution. This often involves dimensionality reduction to facilitate the analysis of high-dimensional descriptors, and requires navigating between the low-dimensional/latent space and high-dimensional/original space for better interpretation.

1.3. Which Data Approach for Learning?

Over the years, researchers have gained detailed knowledge of the complexity of cardiac mechanics, and proposed physiologically-relevant motion and deformation descriptors, from global strain in a single anatomical direction to richer representations such as 3D+t vector or tensor fields. Most approaches decompose the analysis into two steps (**Figure 1A**): the extraction of motion/deformation descriptors from image sequences, followed by their analysis over a population of interest. Machine learning can address both parts, and we discuss these topics separately (sections 2 and 3). Deep neural networks (8) may address the two parts in **Figure 1A**, but also enable the analysis of population data directly from the image sequences by looking for image features not necessarily interpretable or visualizable, but optimal to answer the clinical question of interest (**Figure 1B**). We specifically comment on this strategy, which is more recent and preliminary, in section 4.7.

2. MOTION AND DEFORMATION ESTIMATION

Traditionally, myocardial motion fields have been estimated from images using standard image registration techniques such as optical flow (9), free-form deformation (10), or block matching (11). Naturally, this depends on the algorithm ability to catch motion-related structures, which strongly varies with the imaging modality. Tags and speckles can directly be tracked within the myocardium in tagged MR and 2D/3D echography (within the limits of tag fading, speckles temporal consistency, and out-of-plane motion), contrary to cine MR where algorithms tend to approximate motion from endocardial/epicardial contour tracking. A dedicated review (5) details the standards for spatial and temporal resolution and the influence of imaging parameters on the estimation of myocardial deformation.

Approaches based on neural networks challenge the variational formulation of motion estimation, as shown on

video image sequences with the FlowNet2 convolutional neural network (CNN) architecture (12) that focuses on optical flow. Similar approaches have been applied to cardiac imaging (13, 14), but raise several methodological questions. First, the generalization ability of the trained networks to estimate a wide range of deformations *at multiple scales* still needs to be verified. This is critical for specific disease traits of lower prevalence. Furthermore, robustness to a variety of routine clinical imaging conditions (different image qualities, fields of view, devices, etc.) needs to be established. Second, supervised CNN-based motion estimators such as FlowNet2 do not embed any regularization, and are therefore sensitive to imaging noise if it differs from the training database. This not the case for unsupervised approaches like (13), which use an intensity-based loss, combined with a regularization term as in classical image registration. Finally, motion features can boost segmentation performances (15–17), as looking at several frames improves the manual segmentation of physicians. Further details are given in a review dedicated to deep learning for motion estimation in medical imaging (18).

Statistical models learnt from data can act as regularizers for tracking algorithms. (19) used dictionary learning as a sparse basis for cardiac motion fields to feed the regularization. Within deep learning, auto-encoders can encode spatial transformations into a low-dimensional space and provide powerful projection and reconstruction operators to connect with the tracking in the original image space (20).

Additional constraints specific to the cardiac application can provide more plausible registration outputs, such as invertibility (the myocardium does not fold) and incompressibility, as investigated for the diffeomorphic LogDemons (21) and free-form-deformation algorithms (22). Temporal consistency has been enforced through 4D representations of motion (23, 24), for multiple pairwise transformations simultaneously (25), or for intra/inter-subject mappings (26). Motion and deformation estimation with machine learning should also consider these aspects for better consistency and robustness.

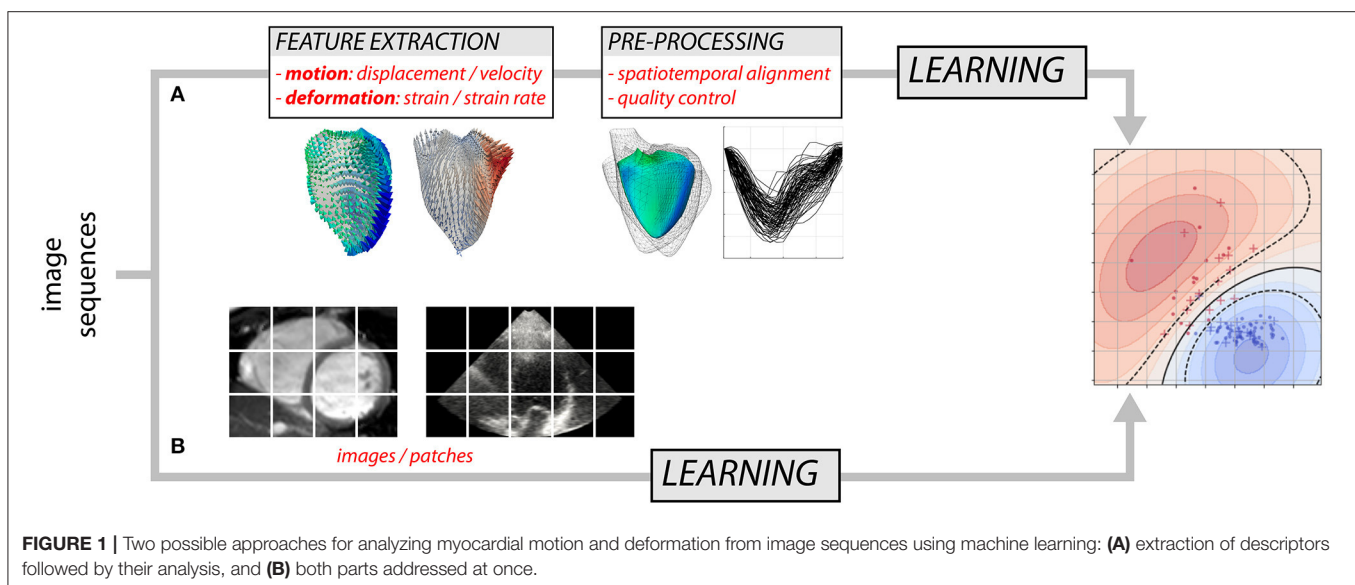


FIGURE 1 | Two possible approaches for analyzing myocardial motion and deformation from image sequences using machine learning: **(A)** extraction of descriptors followed by their analysis, and **(B)** both parts addressed at once.

3. MOTION AND DEFORMATION ANALYSIS

3.1. Before the Analysis: Data Normalization

Cardiac image data often need to be normalized in terms of anatomy, frame rate or cycle phases, before any statistical or machine learning analysis.

Image sequences can be registered using a 4D transformation model based on e.g., free-form deformation (10) or demons (26). This approach quantifies the spatiotemporal differences between the image sequences, analyzed statistically afterwards through deformation-based morphometry methods.

Motion or deformation descriptors (or any other data) from a given individual can also be transported to a reference template (generally, a central case at end-diastole). This involves local reorientation of the motion/deformation fields (27, 28), adjusted to the addressed clinical question (29). Temporal differences between sequences can also be normalized by resampling before the motion extraction [e.g., piece-wise linear interpolation (30)]. Recent approaches transport the whole subject-specific trajectory instead of the descriptors of interest, with specific computational considerations (31, 32). Automatically estimating multiple templates across the sequence may also be well adapted to the cardiac circular/periodic dynamics (33).

In both strategies, existing data correspondences facilitate the normalization. Spatial alignment can rely on anatomical landmarks (apex, valve ring, etc.) or point-to-point correspondences obtained from model-based tracking of the anatomy. Temporal alignment can use physiologically-relevant instants, such as the maximum contraction (10) or QRS and valve events (28).

3.2. Learning From Motion and Deformation Data

Machine learning can benefit a wide range of clinical problems. Unsupervised approaches learn a data representation that uncovers useful insights into the data distribution, but without explicit reference to a particular clinical question. Clustering and dimensionality reduction techniques fall into this category. Supervised approaches train a model for a specific task, and labels/annotations are provided as supervision. For example, diagnosing disease may involve binary labels for supervision (disease/healthy) and the task would be to predict these labels from the motion data. The type of labels determines the task addressed by the model: categorical labels mean classification, whereas discrete or continuous labels imply regression. Supervised approaches also involve learning a (lower dimension) representation of the data that facilitates the classification/regression, but this representation can be formed in an unsupervised or supervised way, as described below.

3.2.1. Unsupervised Learning

Unsupervised motion and deformation analysis shares objectives with statistical atlases, regarding how to characterize variability across a population. Pioneering works directly applied a principal component analysis (PCA) on myocardial displacements at each spatiotemporal location (34) over a healthy population, later

extended through the estimation of local abnormalities in the myocardial velocities of a given subject compared to a reference population (28, 35). However, these analyses consider each spatial location and temporal instant independently from the others. The statistical analysis can also consider the motion patterns over the entire cardiac cycle as high-dimensional objects, as simply demonstrated through a PCA on temporal strain traces concatenated over the heart segments (36, 37). This approach reminds earlier work on Active Appearance Motion Models (38), which statistically analyzed both displacement and image intensity information over the entire cardiac cycle.

More advanced strategies estimate a low-dimensional space that encodes the high-dimensional myocardial motion/deformation data and navigate through this space, although this requires specific care. Myocardial shapes across a population can be considered as originating from one or several references under the action of a transformation such as a diffeomorphic warping. In this case, the space of myocardial shapes is related to the (known) non-linear high-dimensional space of diffeomorphic transformations. This space is a manifold, and known tools exist to perform statistics on such transformations and therefore on myocardial shapes while preserving this data structure (39, 40). Myocardial motion/deformation patterns may also be considered as originating from a non-linear high-dimensional manifold, but in this case the manifold is unknown. Machine learning allows estimating this space from data, and can overcome the limitations of linear techniques such as PCA that ignore this known structure. A general framework (41) groups the vast variety of existing manifold learning techniques. A graph is built across high-dimensional samples to approximate the manifold, and diagonalization, and dimensionality reduction processes provide a low-dimensional space that encodes the data. Techniques generally differ on how input samples are related within the graph, either locally (e.g., distance between neighbors, or local structure variations expressed in the graph Laplacian) or globally (e.g., geodesic distance). These techniques improve the statistical analysis of myocardial motion and deformation patterns. They can represent the continuum of disease from normality while preserving the data structure (42). The unsupervised representation of populations is particularly interesting when existing labels are not fully trusted, as in heart failure with preserved ejection fraction (43, 44) or when a supervised formulation of the clinical problem is uncertain, such as outcome from cardiac resynchronization therapy (45).

Nonetheless, these techniques normally lack explicit mappings between the high-dimensional and low-dimensional spaces, which are typically approximated using out-of-sample reconstruction/regression (46) and are therefore inexact. Deep learning auto-encoders explicitly address this by simultaneously learning how to encode and decode high-dimensional data with a limited number of parameters while minimizing the reconstruction error. However, this also requires constraining the distribution of samples in the latent space so that a statistical analysis can still be performed on it afterwards, as in variational auto-encoders (47). These techniques are promising for the analysis of myocardial motion and deformation and start

being used in cardiac imaging for segmentation (48, 49) or segmentation-based biomarkers (50).

3.2.2. Supervised Learning

As noted above, designing a supervised learning model traditionally consists of two steps (**Figure 1A**). First, the input data are transformed to a new representation that facilitates the task performance. Second, a classification or regression model is trained to predict the label given the new representation. More recent techniques such as deep learning combine these two steps: the representation is learnt and optimized during the model training (**Figure 1B**). Below, we first summarize works using supervised learning in the traditional way and then we briefly review more recent deep learning approaches.

The new data representation can be estimated using knowledge of the labels (supervised way) or without such knowledge (unsupervised). In other words, although the final classification or regression model is supervised, the transformation to a new representation can be unsupervised. Examples include the dimensionality reduction methods reviewed in section 3.2.1, such as PCA (51–53) or non-linear manifold learning (53, 54). The use of hand-crafted features such as volumes/diameters/strains (55) and radius/thickness (56, 57) also falls into this category, although one could argue that knowledge of the task was also used to design these features. A supervised approach was taken in Dawes et al. (58), in which supervised PCA was used to find the principal components of displacement data related to survival.

Classification or regression come once the new representation is obtained. Many classification algorithms have been used, including support vector machines (SVM) (55, 59), random forests (55), variants of dictionary learning (59–61) and ridge logistic regression (57). Regression applications rely on svm (62) and multiscale kernel regression (54).

Recent research has increasingly focused on deep learning for both classification and regression from dynamic imaging data. In these approaches, the activations of intermediate network layers can stand as a transformed representation formed in a supervised way. Inputs to these models are commonly dynamic image intensity data, but segmentation data has also been used (63). For classification, variants of auto-encoders have been a common architecture choice. An auto-encoder is a deep learning-based dimensionality reduction technique, and classification can be performed in the low-dimensional latent space learnt without supervision (53), or in a supervised way by including classification accuracy into the loss function (48, 63, 64). Auto-encoders are attractive as they allow examining the classification features in the original image space, leading to more interpretable analyses. CNNs have also been proposed for classification (65), and a challenge on automated diagnosis was recently organized (7). Regression tasks such as estimating volume and/or ejection fraction may also involve CNNs (66), as tested on the recent Kaggle Challenge data¹. Variational auto-encoders have also been used to perform regression in the latent space (50).

A wide set of classification applications involved myocardial motion or deformation, including identifying abnormal wall motion (59, 61), predicting therapy response (67) and survival (58, 64), and diagnosing myocardial infarct (16, 60, 65, 68) or pathology (7, 48, 57, 63). Regression applications aimed at localizing myocardial infarct (54), grading myocardial motion defects (62), and estimating volumes (66).

Detecting some form of abnormality is a common theme for supervised learning applications, for which two main strategies exist. In the first one, the transformed representation only involves healthy subjects: the distribution of samples in the low-dimensional space therefore represents healthy variations, and subsequent subjects who fall away from the healthy distribution are considered abnormal, as investigated on myocardial velocities (28, 35) and shapes (69). The other strategy learns a low-dimensional representation from both healthy and pathological subjects, where supervised classification can be applied afterwards (70).

4. SPECIFICITIES OF THE CLINICAL CONTEXT

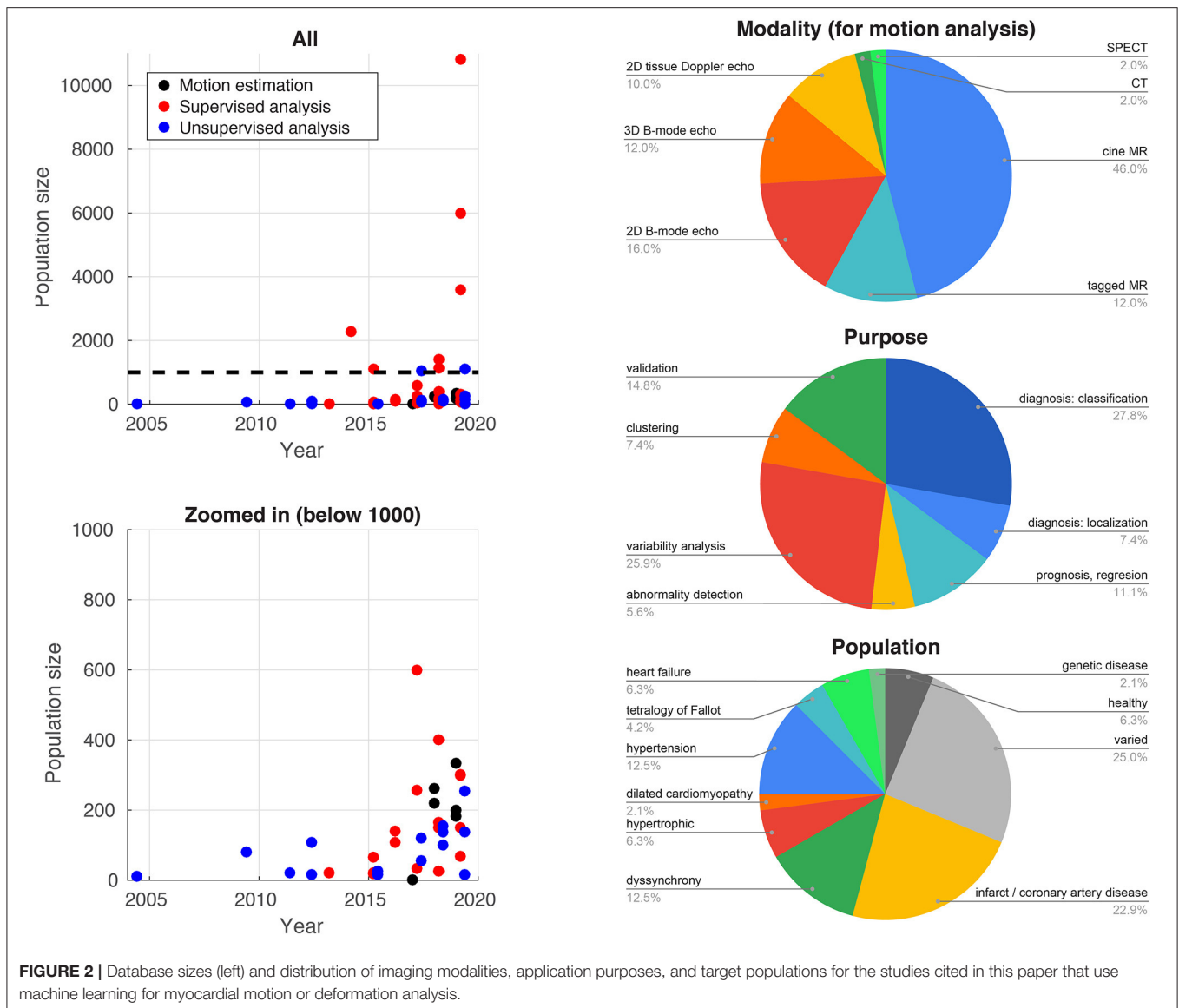
4.1. Physiological Consistency

Learning algorithms utilize a low-dimensional representation of the high-dimensional motion/deformation data, where the population variability is either rendered through diagonalization according to inter-subject distances, or correlated to labels of interest. Transforming to and from this representation involves interpolation between samples. Regularizing the low-dimensional space ensures smoother interpolation and generates new samples that are physiologically plausible (49, 71). In both of these works, the low-dimensional space produced by the encoding part of a CNN was regularized to map smoothly to a set of input shapes, labeled images, or slice locations. This notion of joint projection from the image and label space is also inherently present in more classical manifold learning techniques such as partial least squares. Similar notions need to be extended to motion fields, whilst mapping similar pathological conditions to close locations in the latent space.

4.2. Spatiotemporal Analysis

Most learning techniques consider high-dimensional inputs as high-dimensional column vectors or a set of patches, and disregard the spatiotemporal characteristics of motion and deformation. Few works explicitly addressed this issue for the statistical analysis of populations. A bilinear statistical model was used on cardiac shapes (72) to distinguish inter-subject variations from individual heart dynamics. (73, 74) explicitly addressed the problem through spatiotemporal tensor decomposition. Duchateau et al. (75) tuned up the contributions of the spatial, temporal, and magnitude dimensions to analyze changes in deformation patterns through registration. Jia et al. (31) and Guigui et al. (32) transported temporal trajectories without explicitly extracting motion or deformation descriptors beforehand. These strategies, limited to variability analyses, pave the ground for better considering spatiotemporal aspects with machine learning.

¹ Available online at: <https://www.kaggle.com/c/second-annual-data-science-bowl/data>



4.3. Interpretability

Many tasks may benefit from somehow “interpretable” learnt models, i.e., a user should have ways to inspect the input data characteristics that led to the output prediction or representation. The recent trend toward more complex learning models (such as deep learning) has raised the interest for this property, since these models are generally harder to interpret than simpler ones. One approach consists in defining a simpler model that is “locally similar” to the global complex model (i.e., it has similar performance for similar inputs) (76). For deep learning based approaches, “saliency maps” can be produced, which show which parts of the input data were important in producing the output. Alternatively, regression or autoencoders can be used to reconstruct cases from the low-dimensional latent space and examine features in the original-high dimensional space, with clear benefits for interpretability as demonstrated in Clough et al. (48), Puyol-Anton et al. (53), Biffi et al. (63), and Bello et al. (64).

4.4. Database Size and Heterogeneity

Traditionally, difficulties in accessing and reliably annotating databases of medical images have led to smaller databases in medical imaging compared to computer vision applications. Recent initiatives such as the UK Biobank project² (77) now provide large-scale annotated imaging databases, fuelling a rise in more data-intensive methods such as deep learning. **Figure 2** illustrates this high increase over recent years for the studies reported in this paper. The impact of these large databases is high: reporting reference ranges for cardiac functional biomarkers is now possible with much greater confidence (78, 79), in addition to detecting effects otherwise hidden with smaller databases, as shown for genome data (77). Data heterogeneity is also crucial when choosing or curating a database for a specific task, i.e., the database should include sufficient subjects to cover a

² Available online at: <https://www.ukbiobank.ac.uk/>

range of values for the output label and guarantee the model generalizability. More pathology-focused databases such as those in the Cardiac Atlas Project³ (80) have an important role to play in this respect.

4.5. Validation and Standardization Initiatives

As analyzing the tracking output is sensitive to processing errors, in particular for multi-centric data, tracking algorithms should be benchmarked to prevent bias due to different manufacturers or settings/practices. To ensure reproducibility of clinical decision-making from these data, standardization initiatives arose from academic, clinical, and industrial actors of cardiac imaging. Strain estimation was compared across vendors for synthetic and real images (81). Outputs were consistent regarding the differentiation between pathological and healthy regions, and the identification of ambiguous zones. However, statistically significant differences among vendors were reported, including differences around 15% for the biggest scars. These differences call for benchmarks on more realistic datasets (both regarding geometry and image quality), obtained e.g., from simulation frameworks that mix image formation and biomechanical models with real images (82).

Complementary standardization of imaging are also investigated through deep learning, for the control of e.g., the full coverage of the ventricles (83), the view/plane (84, 85), and the image quality in general (78) or due to motion-related artifacts (86).

4.6. Multiple Modalities/Descriptors

Most studies only consider a single type of motion or deformation descriptor at once from a single acquisition and a single modality, unlike clinical reasoning, which repeats acquisitions in the same or different modality and uses different types of measurements and descriptors. Recent works addressed these limitations within the framework of manifold learning. (30) enforced the complementarity of multimodal acquisitions (tagged MR and 3D echocardiography) using canonical correlation analysis and partial least squares methods. (87) used a similar strategy to better relate myocardial shape and deformation descriptors. Puyol-Anton et al. (70) investigated multi-view linear discriminant analysis for classification purposes. Finally, the more generic framework of multiple kernel learning allows reducing the dimensionality and examining the weights attributed to each descriptor. It was applied to supervised (67) and unsupervised (43–45, 88) problems, to investigate multiple descriptors among which motion-based ones, which could come from different modalities or different views of a single modality.

4.7. Complexity of the Models and Data Descriptors

Machine learning relies on models whose complexities should be adjusted to the question being answered. Researchers should keep in mind that such models only provide an approximation

of reality, and try to minimize this error (e.g., by refining the model, adding more data or relevant descriptors, or estimating uncertainties). We strongly recommend to start with simple data descriptors and models, and carefully benchmark the retained methods against simpler models or even standard statistics.

Deep learning approaches allow circumventing the design of hand-crafted features (**Figure 1B**), and therefore go beyond a substantial limitation of standard machine learning. They mainly have been used for supervised problems and avoiding segmentation. The ACDC challenge (7) included a diagnosis challenge not necessarily requiring segmentation, although all participants opted for segmentation-based diagnosis. Regression-based estimation of cardiac parameters directly from images was proposed in (66, 89, 90), and may also strengthen the segmentation-based estimation of such parameters (91). However, as already pointed out, this direct strategy may also limit interpretability, and therefore transfer to clinical practice.

5. CONCLUSION

Machine learning offers wide possibilities to automate processing, and notably extract and analyze myocardial motion and deformation. Driven by advances in cardiac segmentation and large databases collection, there is potential for substantially improving the characterization of the cardiac function and impacting clinical practice. Changes cover the automation of time-consuming and user-dependent tasks such as feature extraction, higher performance on supervised problems such as (earlier) diagnosis, prognosis, and risk stratification, and new unsupervised data representations for knowledge discovery such as clustering or phenotyping. Nonetheless, motion and deformation are rich but complex high-dimensional data. Efforts need to be continued to reduce uncertainties, approximations, and crucial misinterpretations along the analysis pipeline, from careful problem definition, compliance with the mathematical and physiological data properties, algorithms benchmarking/validation/testing, and health actors' education.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

ND was partially supported by the LABEX PRIMES of Université de Lyon (ANR-11-LABX-0063) and the MIC-MAC JCJC project (ANR-19-CE45-0005). AK was partially supported by the Wellcome EPSRC Centre for Medical Engineering at Kings College London (WT 203148/Z/16/Z) and the UK EPSRC (EP/P001009/1 and EP/R005516/1). MD was partially supported by the European Union Horizon 2020 Programme for Research and Innovation (CardioFunXion, H2020-14-MSCA-ITN-642676).

³ Available online at: <https://www.cardiacatlas.org/>

ACKNOWLEDGMENTS

The articles discussed in this review were selected by querying PubMed over the last 10 years with the terms (myocardial [OR] cardiac) [AND] learning [AND] (motion [OR] deformation), complemented by the authors' knowledge, and examining the

publication profile of the authors of the already selected articles. Papers not using spatial or temporal motion or deformation patterns but single measurements such as peak values or timings, and papers addressing cardiac respiratory motion were removed from this selection, although we acknowledge their importance for the more complete analysis of cardiac function.

REFERENCES

- Cikes M, Solomon S. Beyond ejection fraction: an integrative approach for assessment of cardiac structure and function in heart failure. *Eur Heart J*. (2016) 37:1642–50. doi: 10.1093/eurheartj/ehv510
- Bijnens B, Cikes M, Claus P, Sutherland G. Velocity and deformation imaging for the assessment of myocardial dysfunction. *Eur J Echocardiogr*. (2009) 10:216–26. doi: 10.1093/ejehocardi/jen323
- Bijnens B, Cikes M, Butakoff C, Sitges M, Crispi F. Myocardial motion and deformation: what does it tell us and how does it relate to function? *Fetal Diagn Ther*. (2012) 32:5–16. doi: 10.1159/000335649
- D'hooge J, Heimdal A, Jamal F, Kukulski T, Bijnens B, Rademakers F, et al. Regional strain and strain rate measurements by cardiac ultrasound: principles, implementation and limitations. *Eur J Echocardiogr*. (2000) 1:154–70. doi: 10.1053/euje.2000.0031
- Amzulescu M, De Craene M, Langet H, Pasquet A, Vancraeynest D, Pouleur A, et al. Myocardial strain imaging: review of general principles, validation, and sources of discrepancies. *Eur Heart J Cardiovasc Imaging*. (2019) 20:605–19. doi: 10.1093/ehjci/jez041
- Petitjean C, Dacher J. A review of segmentation methods in short axis cardiac MR images. *Med Image Anal*. (2011) 15:169–84. doi: 10.1016/j.media.2010.12.004
- Bernard O, Lalande A, Zotti C, Cervenansky F, Yang X, Heng P, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE Trans Med Imaging*. (2018) 37:2514–25. doi: 10.1109/TMI.2018.2837502
- Litjens G, Ciompi F, Wolterink J, de Vos B, Leiner T, Teuwen J, et al. State-of-the-art deep learning in cardiovascular image analysis. *JACC Cardiovasc Imaging*. (2019) 12:1549–65. doi: 10.1016/j.jcmg.2019.06.009
- McLeod K, Prakosa A, Mansi T, Sermesant M, Pennec X. An incompressible log-domain demons algorithm for tracking heart tissue. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Toronto, ON (2011). p. 55–67. doi: 10.1007/978-3-642-28326-0-6
- Perperidis D, Mohiaddin R, Rueckert D. Spatio-temporal free-form registration of cardiac MR image sequences. *Med Image Anal*. (2005) 9:441–56. doi: 10.1016/j.media.2005.05.004
- Perrot V, Garcia D. Back to basics in ultrasound velocimetry: tracking speckles by using a standard PIV algorithm. In: *2018 IEEE International Ultrasonics Symposium (IUS)*. Kobe (2018) p. 206–12. doi: 10.1109/ULTSYM.2018.8579665
- Ilg E, Mayer N, Saikia T, Keuper M, Dosovitskiy A, Brox T. FlowNet 2.0: evolution of optical flow estimation with deep networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI (2017) p. 2462–70. doi: 10.1109/CVPR.2017.179
- Morales MA, Izquierdo-Garcia D, Aganj I, Kalpathy-Cramer J, Rosen BR, Catana C. Implementation and validation of a three-dimensional cardiac motion estimation network. *Radiol Artif Intell*. (2019) 1:e180080. doi: 10.1148/ryai.2019180080
- Østvik A, Smistad E, Espeland T, Rye Berg E, Lovstakken L. Automatic myocardial strain imaging in echocardiography using deep learning. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA 2018, ML-CDS 2018. Lecture Notes in Computer Science*. Cham: Springer (2018). p. 309–16. doi: 10.1007/978-3-030-00889-5-35
- Qin C, Bai W, Schlemper J, Petersen S, Piechnik S, Neubauer S, et al. Joint learning of motion estimation and segmentation for cardiac MR image sequences. *First MICCAI Workshop, DART 2019, and First International Workshop, MIL3D 2019, Shenzhen, Held in Conjunction with MICCAI 2019. Shenzhen* (2018) 11071:472–80. doi: 10.1007/978-3-030-00934-2-53
- Xu C, Xu L, Gao Z, Zhao S, Zhang H, Zhang Y, et al. Direct delineation of myocardial infarction without contrast agents using a joint motion feature learning architecture. *Med Image Anal*. (2018) 50:82–94. doi: 10.1016/j.media.2018.09.001
- Yan W, Wang Y, van der Geest R, Tao Q. Cine MRI analysis by deep learning of optical flow: adding the temporal dimension. *Comput Biol Med*. (2019) 111:103356. doi: 10.1016/j.combiomed.2019.103356
- Haskins G, Kruger U, Yan P. Deep learning in medical image registration: a survey. *arXiv preprint arXiv:190302026* (2019).
- Ouzir N, Basarab A, Liebgott H, Harbaoui B, Tourneret JY. Motion estimation in echocardiography using sparse representation and dictionary learning. *IEEE Trans Image Proc*. (2017) 27:64–77. doi: 10.1109/TIP.2017.2753406
- Krebs J, Delingette H, Mailhé B, Ayache N, Mansi T. Learning a probabilistic model for diffeomorphic registration. *IEEE Trans Med Imaging*. (2019) 38:2165–76. doi: 10.1109/TMI.2019.2897112
- Mansi T, Pennec X, Sermesant M, Delingette H, Ayache N. iLogDemons: a demons-based registration algorithm for tracking incompressible elastic biological tissues. *Int J Comput Vis*. (2011) 92:92–111. doi: 10.1007/s11263-010-0405-z
- Heyde B, Alessandrini M, Hermans J, Barbosa D, Claus P, D'hooge J. Anatomical image registration using volume conservation to assess cardiac deformation from 3D ultrasound recordings. *IEEE Trans Med Imaging*. (2015) 35:501–11. doi: 10.1109/TMI.2015.2479556
- Shi W, Jantsch M, Aljabar P, Pizarro L, Bai W, Wang H, et al. Temporal sparse free-form deformations. *Med Image Anal*. (2013) 17:779–89. doi: 10.1016/j.media.2013.04.010
- De Craene M, Piella G, Camara O, Duchateau N, Silva E, Doltra A, et al. Temporal diffeomorphic free-form deformation: application to motion and strain estimation from 3D echocardiography. *Med Image Anal*. (2012) 16:427–50. doi: 10.1016/j.media.2011.10.006
- De Craene M, Camara O, Bijnens BH, Frangi AF. Large diffeomorphic FFD registration for motion and strain quantification from 3D-US sequences. *Functional Imaging and Modeling of the Heart. FIMH 2009. Lecture Notes in Computer Science*. Berlin; Heidelberg: Springer (2009) p. 437–46. doi: 10.1007/978-3-642-01932-6-47
- Peyrat J, Delingette H, Sermesant M, Xu C, Ayache N. Registration of 4D cardiac CT sequences under trajectory constraints with multichannel diffeomorphic demons. *IEEE Trans Med Imaging*. (2010) 29:1351–68. doi: 10.1109/TMI.2009.2038908
- Rao A, Chandrashekar R, Sanchez-Ortiz G, Mohiaddin R, Aljabar P, Hajnal J, et al. Spatial transformation of motion and deformation fields using nonrigid registration. *IEEE Trans Med Imaging*. (2004) 23:1065–76. doi: 10.1109/TMI.2004.828681
- Duchateau N, De Craene M, Piella G, Silva E, Doltra A, Sitges M, et al. A spatiotemporal statistical atlas of motion for the quantification of abnormalities in myocardial tissue velocities. *Med Image Anal*. (2011) 15:316–28. doi: 10.1016/j.media.2010.12.006
- Duchateau N, De Craene M, Pennec X, Merino B, Sitges M, Bijnens B. Which reorientation framework for the atlas-based comparison of motion from cardiac image sequences? *Spatio-temporal Image Analysis for Longitudinal and Time-Series Image Data. STIA 2012. Lecture Notes in Computer Science*. Berlin; Heidelberg: Springer (2012) 7570:25–37. doi: 10.1007/978-3-642-33555-6-3
- Puyol-Anton E, Sinclair M, Gerber B, Amzulescu M, Langet H, De Craene M, et al. A multimodal spatiotemporal cardiac motion atlas

- from MR and ultrasound data. *Med Image Anal.* (2017) 40:96–110. doi: 10.1016/j.media.2017.06.002
31. Jia S, Duchateau N, Mocerri P, Sermesant M, Pennec X. Parallel transport of surface deformations from pole ladder to symmetrical extension. *Shape in Medical Imaging. ShapeMI 2018. Lecture Notes in Computer Science.* Cham:Springer (2018). p. 116–24. doi: 10.1007/978-3-030-04747-4_11
 32. Guigui N, Jia S, Sermesant M, Pennec X. Symmetric algorithmic components for shape analysis with diffeomorphisms. *Processinds GSI LNCS.* (2019). p. 759–68. doi: 10.1007/978-3-030-26980-7-79
 33. Rohe M, Sermesant M, Pennec X. Low-dimensional representation of cardiac motion using Barycentric Subspaces: a new group-wise paradigm for estimation, analysis, and reconstruction. *Med Image Anal.* (2018) 45:1–12. doi: 10.1016/j.media.2017.12.008
 34. Rougon N, Petitjean C, Preteux F. Building and using a statistical 3D motion atlas for analyzing myocardial contraction in MRI. In: *Proceedings SPIE Medical Imaging.* San Diego, CA (2004) 5370:253–64. doi: 10.1117/12.535609
 35. De Craene M, Duchateau N, Tobon-Gomez C, Ghafaryasl B, Piella G, Rhode K, et al. SPM to the heart: mapping of 4D continuous velocities for motion abnormality quantification. In: *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI).* IEEE: Barcelona. (2012) p. 454–7. doi: 10.1109/ISBI.2012.6235582
 36. Tabassian M, Alessandrini M, Herbots L, Mirea O, Pagourelas E, Jasaityte R, et al. Machine learning of the spatio-temporal characteristics of echocardiographic deformation curves for infarct classification. *Int J Cardiovasc Imaging.* (2017) 33:1159–67. doi: 10.1007/s10554-017-1108-0
 37. Tabassian M, Sunderji I, Erdei T, Sanchez-Martinez S, Degiovanni A, Marino P, et al. Diagnosis of heart failure with preserved ejection fraction: machine learning of spatiotemporal variations in left ventricular deformation. *J Am Soc Echocardiogr.* (2018) 31:1272–84.e9. doi: 10.1016/j.echo.2018.07.013
 38. Bosch J, Mitchell S, Lelieveldt B, Nijland F, Kamp O, Sonka M, et al. Automatic segmentation of echocardiographic sequences by active appearance motion models. *IEEE Trans Med Imaging.* (2002) 21:1374–83. doi: 10.1109/TMI.2002.806427
 39. Arsigny V, Commowick O, Pennec X, Ayache N. A log-Euclidean framework for statistics on diffeomorphisms. *Med Image Comput Comput Assist Interv.* (2006) 9(Pt 1):924–31. doi: 10.1007/11866565-113
 40. Pennec X. Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *J Math Imaging Vis.* (2006) 25:127. doi: 10.1007/s10851-006-6228-4
 41. Yan S, Xu D, Zhang B, Zhang H, Yang Q, Lin S. Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Trans Pattern Anal Mach Intell.* (2007) 29:40–51. doi: 10.1109/TPAMI.2007.250598
 42. Duchateau N, De Craene M, Piella G, Frangi A. Constrained manifold learning for the characterization of pathological deviations from normality. *Med Image Anal.* (2012) 16:1532–49. doi: 10.1016/j.media.2012.07.003
 43. Sanchez-Martinez S, Duchateau N, Erdei T, Fraser A, Bijnsens B, Piella G. Characterization of myocardial motion patterns by unsupervised multiple kernel learning. *Med Image Anal.* (2017) 35:70–82. doi: 10.1016/j.media.2016.06.007
 44. Sanchez-Martinez S, Duchateau N, Erdei T, Kunszt G, Aakhus S, Degiovanni A, et al. Machine learning analysis of left ventricular function to characterize heart failure with preserved ejection fraction. *Circ Cardiovasc Imaging.* (2018) 11:e007138. doi: 10.1161/CIRCIMAGING.117.007138
 45. Cikes M, Sanchez-Martinez S, Claggett B, Duchateau N, Piella G, Butakoff C, et al. Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. *Eur J Heart Fail.* (2019) 21:74–85. doi: 10.1002/ehf.1333
 46. Bengio Y, Paiement J, Vincent P, Delalleau O, Le Roux N, Ouimet M. Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering. *Advances in Neural Information Processing Systems.* Vancouver, BC (2003) p. 177–84.
 47. Kingma D, Welling M. Auto-encoding variational bayes. In: *Proceedings ICLR.* Banff, AB (2014).
 48. Clough J, Oksuz I, Puyol-Anton E, Ruijsink B, King A, Schnabel J. Global and local interpretability for cardiac MRI classification. In: *Proceedings Medical Image Computing and Computer Assisted Intervention Lecture Notes in Computer Science.* Shenzhen (2019). p. 656–64. doi: 10.1007/978-3-030-32251-9-72
 49. Painchaud N, Skandarani Y, Judge T, Bernard O, Lalande A, Jodoin P. Cardiac MRI segmentation with strong anatomical guarantees. In: *Proceedings Medical Image Computing and Computer Assisted Intervention Lecture Notes in Computer Science.* Shenzhen (2019). p. 632–40. doi: 10.1007/978-3-030-32245-8-70
 50. Puyol-Anton E, Ruijsink B, Clough J, Oksuz I, Rueckert D, Razavi R, et al. Assessing the impact of blood pressure on cardiac function using interpretable biomarkers and variational autoencoders. In: *Proceedings Statistical Atlases and Computational Modelling of the Heart-Medical Image Computing and Computer Assisted Intervention Lecture Notes in Computer Science.* Shenzhen (2019).
 51. Bai W, Shi W, de Marvao A, Dawes T, O'Regan D, Cook S, et al. A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of healthy subjects and an analysis of shape and motion. *Med Image Anal.* (2015) 26:133–45. doi: 10.1016/j.media.2015.08.009
 52. Piras P, Teresi L, Puddu PE, Torromeo C, Young AA, Suinesiaputra A, et al. Morphologically normalized left ventricular motion indicators from MRI feature tracking characterize myocardial infarction. *Sci Rep.* (2017) 7:12259. doi: 10.1038/s41598-017-12539-5
 53. Puyol-Anton E, Ruijsink B, Langet H, De Craene M, Piro, Schnabel J, et al. Learning associations between clinical information and motion-based descriptors using a large scale MR-derived cardiac motion atlas. In: *Proceedings Statistical Atlases and Computational Modelling of the Heart-Medical Image Computing and Computer Assisted Intervention Lecture Notes in Computer Science.* Shenzhen (2019). p. 94–102. doi: 10.1007/978-3-030-12029-0-11
 54. Duchateau N, De Craene M, Allain P, Saloux E, Sermesant M. Infarct localization from myocardial deformation: prediction and uncertainty quantification by regression from a low-dimensional space. *IEEE Trans Med Imaging.* (2016) 35:2340–52. doi: 10.1109/TMI.2016.2562181
 55. Narula S, Shameer K, Omar A, Dudley J, Sengupta P. Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography. *J Am Coll Cardiol.* (2016) 68:2287–95. doi: 10.1016/j.jacc.2016.08.062
 56. Bleton H, Margeta J, Lombaert H, Delingette H, Ayache N. Myocardial infarct localization using neighbourhood approximation forests. In: *Proceedings Statistical Atlases and Computational Modelling of the Heart-Medical Image Computing and Computer Assisted Intervention Lecture Notes in Computer Science.* Munich (2015). p. 108–16. doi: 10.1007/978-3-319-28712-6-12
 57. Zheng Q, Delingette H, Ayache N. Explainable cardiac pathology classification on cine MRI with motion characterization by semi-supervised learning of apparent flow. *Med Image Anal.* (2019) 56:80–95. doi: 10.1016/j.media.2019.06.001
 58. Dawes T, de Marvao A, Shi W, Fletcher T, Watson G, Wharton J, et al. Machine learning of three-dimensional right ventricular motion enables outcome prediction in pulmonary hypertension: a cardiac MR imaging study. *Radiology.* (2017) 283:381–90. doi: 10.1148/radiol.2016161315
 59. Mantilla J, Garreau M, Bellanger J, Paredes J. Machine learning techniques for LV wall motion classification based on spatio-temporal profiles from cardiac cine MRI. In: *ProICMLA '13 Proceedings of the 2013 12th International Conference on Machine Learning and Applications.* Washington, DC (2013). p. 167–72. doi: 10.1109/ICMLA.2013.36
 60. Peressutti D, Bai W, Shi W, Tobon-Gomez C, Jackson T, Sohal M, et al. Towards left ventricular scar localisation using local motion descriptors. In: *Proceedings Statistical Atlases and Computational Modelling of the Heart-Medical Image Computing and Computer Assisted Intervention Lecture Notes in Computer Science.* Athens (2016). p. 30–9. doi: 10.1007/978-3-319-28712-6-4
 61. Mantilla J, Paredes J, Bellanger J, Donal E, Leclercq C, Medina R, et al. Classification of LV wall motion in cardiac MRI using kernel Dictionary Learning with a parametric approach. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* Milan (2015) p. 7292–5. doi: 10.1109/EMBC.2015.7320075
 62. Marin T, Kalayeh M, Pretorius P, Wernick M, Yang Y, Brankov J. Numerical observer for cardiac motion assessment using machine learning. In: *Proceedings SPIE Medical Imaging.* Lake Buena Vista, FL (2011). p. 79660G. doi: 10.1117/12.878186

63. Biffi C, Oktay O, Tarroni G, Bai W, de Marvao A, Doumou G, et al. Learning interpretable anatomical features through deep generative models: application to cardiac remodeling. In: *Proceedings Modelling of the Heart-Medical Image Computing and Computer Assisted Intervention Lecture Notes in Computer Science*. Granada (2018). p. 464–71. doi: 10.1007/978-3-030-00934-2-52
64. Bello GA, Dawes TJW, Duan J, Biffi C, de Marvao A, Howard LS, et al. Deep-learning cardiac motion analysis for human survival prediction. *Nat Mach Intell*. (2019) 1:95–104. doi: 10.1038/s42256-019-0019-2
65. Kusunose K, Abe T, Haga A, Fukuda D, Yamada H, Harada M, et al. A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. *JACC Cardiovasc Imaging*. (2019). doi: 10.1016/j.jcmg.2019.02.024. [Epub ahead of print].
66. Luo G, Dong S, Wang K, Zuo W, Cao S, Zhang H. Multi-views fusion CNN for left ventricular volumes estimation on cardiac MR images. *IEEE Trans Biomed Eng*. (2018) 65:1924–34. doi: 10.1109/TBME.2017.2762762
67. Peressutti D, Sinclair M, Bai W, Jackson T, Ruijsink J, Nordsletten D, et al. A framework for combining a motion atlas with non-motion information to learn clinically useful biomarkers: application to cardiac resynchronisation therapy response prediction. *Med Image Anal*. (2017) 35:669–84. doi: 10.1016/j.media.2016.10.002
68. Suinesiaputra A, Ablin P, Alba X, Alessandrini M, Allen J, Bai W, et al. Statistical shape modeling of the left ventricle: myocardial infarct classification challenge. *IEEE J Biomed Health Inform*. (2018) 22:503–15. doi: 10.1109/JBHI.2017.2652449
69. Zhang X, Cowan B, Bluemke D, Finn J, Fonseca C, Kadish A, et al. Atlas-based quantification of cardiac remodeling due to myocardial infarction. *PLoS ONE*. (2014) 9:e110243. doi: 10.1371/journal.pone.0110243
70. Puyol-Anton E, Ruijsink B, Gerber B, Amzulescu MS, Langet H, De Craene M, et al. Regional multi-view learning for cardiac motion analysis: application to identification of dilated cardiomyopathy patients. *IEEE Trans Biomed Eng*. (2019) 66:956–66. doi: 10.1109/TBME.2018.2865669
71. Oktay O, Ferrante E, Kamnitsas K, Heinrich M, Bai W, Caballero J, et al. Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. *IEEE Trans Med Imaging*. (2017) 37:384–95. doi: 10.1109/TMI.2017.2743464
72. Hoogendoorn C, Sukno F, Ordas S, Frangi A. Bilinear models for spatio-temporal point distribution analysis: application to extrapolation of left ventricular, biventricular and whole heart cardiac dynamics. *Int J Comput Vis*. (2009) 85:237–52. doi: 10.1007/s11263-009-0212-6
73. McLeod K, Sermesant M, Beerbaum P, Pennec X. Spatio-temporal tensor decomposition of a polyaffine motion model for a better analysis of pathological left ventricular dynamics. *IEEE Trans Med Imaging*. (2015) 34:1562–75. doi: 10.1109/TMI.2015.2405579
74. McLeod K, Tondel K, Calvet L, Sermesant M, Pennec X. Cardiac motion evolution model for analysis of functional changes using tensor decomposition and cross-sectional data. *IEEE Trans Biomed Eng*. (2018) 65:2769–80. doi: 10.1109/TBME.2018.2816519
75. Duchateau N, Giraldeau G, Gabrielli L, Fernandez-Armenta J, Penela D, Evertz R, et al. Quantification of local changes in myocardial motion by diffeomorphic registration via currents: application to paced hypertrophic obstructive cardiomyopathy in 2D echocardiographic sequences. *Med Image Anal*. (2015) 19:203–19. doi: 10.1016/j.media.2014.10.005
76. Ribeiro M, Singh S, Guestrin C. “Why Should I Trust You?” explaining the predictions of any classifier. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*. San Diego: CA (2016) p. 135–44. doi: 10.18653/v1/N16-3020
77. Littlejohns T, Sudlow C, Allen N, Collins R. UK Biobank: opportunities for cardiovascular research. *Eur Heart J*. (2019) 40:1158–66. doi: 10.1093/eurheartj/ehx254
78. Ruijsink B, Puyol-Anton E, Oksuz I, Sinclair M, Bai W, Schnabel J, et al. Fully automated, quality-controlled cardiac analysis from CMR: validation and large-scale application to characterize cardiac function. *JACC Cardiovasc Imaging*. (2019). doi: 10.1016/j.jcmg.2019.05.030. [Epub ahead of print].
79. Attar R, Pereanez M, Gooya A, Alba X, Zhang L, de Vila M, et al. Quantitative CMR population imaging on 20,000 subjects of the UK Biobank imaging study: LV/RV quantification pipeline and its evaluation. *Med Image Anal*. (2019) 56:26–42. doi: 10.1016/j.media.2019.05.006
80. Fonseca C, Backhaus M, Bluemke D, Britten R, Chung J, Cowan B, et al. The Cardiac Atlas Project - An imaging database for computational modeling and statistical atlases of the heart. *Bioinformatics*. (2011) 27:2288–95. doi: 10.1093/bioinformatics/btr360
81. D’hooge J, Barbosa D, Gao H, Claus P, Prater D, Hamilton J, et al. Two-dimensional speckle tracking echocardiography: standardization efforts based on synthetic ultrasound data. *Eur Heart J Cardiovasc Imaging*. (2015) 17:693–701. doi: 10.1093/ehjci/jev197
82. Alessandrini M, Heyde B, Queirós S, Cygan S, Zontak M, Somphone O, et al. Detailed evaluation of five 3D speckle tracking algorithms using synthetic echocardiographic recordings. *IEEE Trans Med Imaging*. (2016) 35:1915–26. doi: 10.1109/TMI.2016.2537848
83. Zhang L, Gooya A, Pereanez M, Dong B, Piechnik S, Neubauer S, et al. Automatic assessment of full left ventricular coverage in cardiac cine magnetic resonance imaging with fisher discriminative 3D CNN. *IEEE Trans Biomed Eng*. (2018) 66:1975–86. doi: 10.1109/TBME.2018.281952
84. Margeta J, Criminisi A, Cabrera Lozoya R, Lee D, Ayache N. Fine-tuned convolutional neural nets for cardiac MRI acquisition plane recognition. *Comput Methods Biomech Biomed Engin*. (2017) 5:339–49. doi: 10.1080/21681163.2015.1061448
85. Østvik A, Smistad E, Aase S, Haugen B, Lovstakken L. Real-time standard view classification in transthoracic echocardiography using convolutional neural networks. *Ultrasound Med Biol*. (2019) 45:374–84. doi: 10.1016/j.ultrasmedbio.2018.07.024
86. Oksuz I, Ruijsink B, Puyol-Anton E, Clough J, Cruz G, Bustin A, et al. Automatic CNN-based detection of cardiac MR motion artefacts using k-space data augmentation and curriculum learning. *Med Image Anal*. (2019) 55:136–47. doi: 10.1016/j.media.2019.04.009
87. Di Folco M, Clarysse P, Mocerri P, Duchateau N. Learning interactions between cardiac shape and deformation: application to pulmonary hypertension. In: *Proceedings Statistical Atlases and Computational Modelling of the Heart-Medical Image Computing and Computer Assisted Intervention Lecture Notes in Computer Science*. Shenzhen (2019).
88. Nogueira M, De Craene M, Sanchez-Martinez S, Chowdhury D, Bijmens B, Piella G. Analysis of nonstandardized stress echocardiography sequences using multiview dimensionality reduction. *Med Image Anal*. (2019) 60:101594. doi: 10.1016/j.media.2019.101594
89. Zhen X, Wang Z, Islam A, Bhaduri M, Chan I, Li S. Multi-scale deep networks and regression forests for direct bi-ventricular volume estimation. *Med Image Anal*. (2016) 30:120–9. doi: 10.1016/j.media.2015.07.003
90. Xue W, Brahm G, Pandey S, Leung S, Li S. Full left ventricle quantification via deep multitask relationships learning. *Med Image Anal*. (2018) 43:54–65. doi: 10.1016/j.media.2017.09.005
91. Luo G, Dong S, Wang W, Wang K, Cao S, Tam C, et al. Commensal correlation network between segmentation and direct area estimation for bi-ventricle quantification. *Med Image Anal*. (2020) 59:101591. doi: 10.1016/j.media.2019.101591

Conflict of Interest: MD was employed by Philips Research Paris. ND and AK have research publications and/or projects with researchers and engineers from private companies, but this did not influence the contents of this review.

Copyright © 2020 Duchateau, King and De Craene. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Artificial Intelligence for Cardiac Imaging-Genetics Research

Antonio de Marvao[†], Timothy J. W. Dawes[†] and Declan P. O'Regan^{*}

MRC London Institute of Medical Sciences, Imperial College London, London, United Kingdom

OPEN ACCESS

Edited by:

Steffen Erhard Petersen,
Queen Mary University of London,
United Kingdom

Reviewed by:

Alexander Teumer,
University of Greifswald, Germany
Avan Suinesiaputra,
The University of Auckland,
New Zealand

*Correspondence:

Declan P. O'Regan
declan.oregan@imperial.ac.uk

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Cardiovascular Imaging,
a section of the journal
Frontiers in Cardiovascular Medicine

Received: 30 September 2019

Accepted: 27 December 2019

Published: 21 January 2020

Citation:

de Marvao A, Dawes TJW and
O'Regan DP (2020) Artificial
Intelligence for Cardiac
Imaging-Genetics Research.
Front. Cardiovasc. Med. 6:195.
doi: 10.3389/fcvm.2019.00195

Cardiovascular conditions remain the leading cause of mortality and morbidity worldwide, with genotype being a significant influence on disease risk. Cardiac imaging-genetics aims to identify and characterize the genetic variants that influence functional, physiological, and anatomical phenotypes derived from cardiovascular imaging. High-throughput DNA sequencing and genotyping have greatly accelerated genetic discovery, making variant interpretation one of the key challenges in contemporary clinical genetics. Heterogeneous, low-fidelity phenotyping and difficulties integrating and then analyzing large-scale genetic, imaging and clinical datasets using traditional statistical approaches have impeded process. Artificial intelligence (AI) methods, such as deep learning, are particularly suited to tackle the challenges of scalability and high dimensionality of data and show promise in the field of cardiac imaging-genetics. Here we review the current state of AI as applied to imaging-genetics research and discuss outstanding methodological challenges, as the field moves from pilot studies to mainstream applications, from one dimensional global descriptors to high-resolution models of whole-organ shape and function, from univariate to multivariate analysis and from candidate gene to genome-wide approaches. Finally, we consider the future directions and prospects of AI imaging-genetics for ultimately helping understand the genetic and environmental underpinnings of cardiovascular health and disease.

Keywords: artificial intelligence, machine learning, deep learning, genetics, genomics, imaging-genetics, cardiovascular imaging, cardiology

INTRODUCTION

Cardiovascular conditions remain the leading cause of mortality and morbidity worldwide (1), with genetic factors playing a significant role in conferring risk for disease (2). High-throughput DNA sequencing and genotyping technologies, such as whole-genome sequencing and high-resolution array genotyping, have developed at an extraordinary pace since the first draft of the human genome was published in 2001 at a cost of \$0.5-1 billion (3). Continuous improvements have so far outpaced Moore's law, with the sequencing cost per genome currently estimated to be \$1,000 (4), enabling cost-effective sequencing of millions of humans. At the same time, technological advances in physics, engineering, and computing have enabled a step-change improvement in cardiovascular imaging, facilitating the shift from one dimensional, low-fidelity descriptors of the cardiovascular system to high-resolution multi-parametric phenotyping. These capabilities are not limited to research settings but are increasingly available in clinical echocardiography, nuclear imaging, computerized tomography (CT), and cardiovascular magnetic resonance (CMR) practice. An unprecedented volume of clinical data is also becoming available, from smartphone-linked wearable sensors (5) to the numerous variables included in the electronic health records of entire

populations (6). However, the volume, heterogeneity, complexity, and speed of accumulation of these datasets now make human-driven analysis impractical. Artificial intelligence (AI) methods such as machine learning (ML), are particularly suited to tackling the challenges of “Big Data” and have shown great promise in addressing complex classification, clustering, and predictive modeling tasks in cardiovascular research. Cardiac imaging-genetics refers to the integrated research methods that aim to identify and characterize the genetic variants that influence functional, physiological, and anatomical phenotypes derived from cardiovascular imaging.

In the same way that basic statistical literacy has become a routine aspect of clinical practice, a basic understanding of AI’s strengths, applications, and limitations is becoming essential for practicing researchers and clinicians. Here we introduce common AI principles, review applications in imaging-genetics research, and discuss future directions and prospects in this field.

IMAGING-GENETICS: FROM SINGLE GENE HYPOTHESIS-TESTING TO GENOME-WIDE HYPOTHESES GENERATION

Imaging-genetics aims to dissect and characterize the complex interplay between imaging-derived phenotypes and environmental and genetic factors. Many principles and approaches originated from neuroimaging research, where the first attempts at integrating multi-parametric phenotypes, obtained from structural and functional brain MRI, with genetic data were carried out (7). To help manage the computational and statistical challenges inherent to the use of “Big Data” squared (high-dimensional imaging \times high-dimensional genetic data), interrogations were limited to pre-defined regions of interest in the brain and candidate genes or SNPs, based on *a priori* assumptions about the biology of disease (8). Similar, “hypothesis-led” designs underpinned candidate gene and linkage studies that established causal relationships between rare genetic variants and rare conditions, such as those that first identified the role of myosin heavy-chain beta in hypertrophic cardiomyopathy (HCM) (9) and of titin in dilated cardiomyopathy (DCM) (10).

The increased affordability of DNA sequencing and genotyping resulted in genetic information becoming available in large numbers of subjects. This has contributed to shift the focus to genetic discovery and the study of common, complex disease traits. These traits are not characterized by a single gene mutation leading to a large change on the phenotype but attributable to the cumulative effects of many loci. Although the effect sizes of individual loci are relatively modest, composite effects can significantly alter the probability of developing disease (11). The “common disease—common variant” hypothesis underpins genome wide association studies (GWAS), where subjects are genotyped for hundreds of thousands of common variants. For example, a study into the genetic determinants of hypertension in over 1 million subjects, identified 901 loci that were associated with systolic blood pressure (SBP) and these

explained 5.7% of the variance observed (12). Even though these single nucleotide polymorphisms (SNPs) explain only a small proportion of phenotypic variance they provide relevant, hypothesis-generating biological or therapeutic insights. The rapid development of complementary high-throughput technologies, able to characterize the transcriptome, epigenome, proteome, and metabolome now enables us to search for molecular evidence of gene causality and to understand the mechanisms and pathways involved in health and disease (13). These large biological multi-omics data sets and their computational analysis are conceptually similar to the more established study of genomics and examples of such work are included in this review.

IMAGING-GENETICS: FROM ONE-DIMENSIONAL PHENOTYPING TO MULTIPARAMETRIC IMAGING

Several biological and technical reasons have been proposed to explain the “missing heritability” of complex cardiovascular traits. However, a common factor limiting many genotype-phenotype studies was that the ability to characterize phenotypes rapidly and accurately, significantly lagged behind our ability to describe the human genotype (14). Phenotyping was characterized by imprecise quantification, sparsity of measurements, high intra- and inter- observer variability, low signal to noise ratios, reliance on geometric assumptions, and adequate body habitus, poor standardization of measurement techniques and the tendency to discretize continuous phenotypes (15). Commonly, the complexity of the cardiovascular system was distilled into a small number of continuous one-dimensional variables [e.g. volumetric assessment of the left ventricle (16)] or, convenient dichotomies, such as responders vs. non-responders (17), leading to a loss of statistical power (18).

The imaging community responded to calls for more accurate and precise, high-dimensional phenotyping (19, 20) with the roll out of developments in echocardiography (e.g., tissue doppler, speckle-tracking, and 3D imaging), CMR (e.g., tissue characterization, 4D flow, 3D imaging, diffusion tensor imaging, spectroscopy, and real-time scanning), CT (e.g., improved spatial and temporal resolution, radiation dose reduction techniques, functional assessment of coronary artery flow using FFR-CT, and coronary plaque characterization), and nuclear cardiology (e.g., improvements in radiopharmaceuticals and hardware resulting in increased accuracy and reduced radiation exposure). In parallel, computational approaches have become increasingly integral to the clinical interpretation of these much larger datasets (21–23) and several have obtained FDA approval (24).

IMAGING-GENETICS: A “BIG DATA” SQUARED PROBLEM

Leveraging these deeper phenotypes is an attractive proposition but the joint analysis of high-dimensional imaging and genetic data poses major computational and theoretical challenges. An early example of a neuroimaging GWAS investigated the

association between 448,293 SNPs and 31,622 CMR voxels in a cohort of 740 subjects (25). This study highlighted difficulties correcting for multiple testing (1.4×10^{10} tests were performed) and the need for unprecedented computational power (300 parallel cores).

Simultaneously assessing the statistical significance of several hundred thousand tests vastly increases the number of anticipated type I errors. If the probability of incorrectly rejecting the null hypothesis in one test with a pre-set α of 0.05 is 5%, then under the same conditions, the probability of incorrectly rejecting the null hypothesis at least once if 100 tests are performed is 99.4%. Therefore, an adjustment for the number of tests being carried out is required. The simplest approach for adjustment for multiple testing is the Bonferroni correction, where the pre-set α is recalculated as α/m , where m represents the number of independent tests being performed. However, this method is overly conservative when m is large, leading instead to many false negatives. An alternative, extensively-validated method is the Benjamini–Hochberg Procedure (26). Using this approach, instead of controlling for the chance of any false positives, an acceptable maximum fixed percentage of false discoveries (the expected proportion of rejected hypotheses that are false positives) is set.

A further consideration in the statistical analysis of high-dimensional cardiac phenotypes is that a clinically significant signal will not originate from a single voxel but across many voxels in extended, anatomically coherent areas. Indeed, approaches such as threshold-free cluster enhancement (TFCE), which were developed in neuroimaging (27), have recently applied in cardiovascular research (28). Using such methods, both signal size and contiguity with surrounding signal patterns contribute to inference statistics.

ARTIFICIAL INTELLIGENCE

Artificial intelligence, machine learning, and deep learning are terms that are interlinked, have some overlap but are often incorrectly used interchangeably. AI refers to the overarching field of computer science focused on simulating human cognitive processes. As a subset of AI, machine learning refers to the family of algorithms that share a capacity to perform tasks like classification, regression, or clustering based on patterns or rules iteratively learnt directly from the data without using explicit instructions. ML algorithms can be further subdivided into supervised, unsupervised, and reinforcement learning.

Supervised learning is the most common form of traditional ML and involves the training of models on pairs of input and expected outputs (“labeled” data) and then their deployment to make predictions in previously unseen data. It includes such approaches as nearest neighbor, support vector machines, random forests and naïve Bayes classifiers. Unsupervised learning algorithms are used to address clustering or dimensionality reduction problems by detecting patterns and structures within the data without any prior knowledge or constraints. In other words, the model organizes “unlabeled” data into groupings that share common, previously undefined

characteristics. Examples including k-means clustering, t-distributed stochastic neighbor embedding (t-SNE), and association rule learning algorithms. The use of reinforcement learning algorithms (e.g., deep Q networks), common in robotics and gaming applications (29) has now also been trialed in the navigation of 3D datasets for anatomical landmark detection (30).

Deep learning (DL) is a specific ML method inspired by the way that the human brain processes data and draws conclusions. To achieve this, DL applications use a layered structure of algorithms, called an artificial neural network that imitates the biological neural network of the human brain. The word “deep” in “deep learning” refers to the number of layers through which the data is transformed. The most common DL models are convolutional neural networks (CNN), which are extremely efficient at extracting features and often superior to traditional ML in larger, more complex datasets such as medical imaging and genomics (31, 32). However, feature and process interpretability is more amenable in classical ML as even simple DL networks can operate as “black-boxes.” While the computational and time requirements of DL are much higher during training, subsequent inference is extremely fast and DL approaches can be used to accelerate supervised, unsupervised, and reinforcement learning. Indeed, while traditional ML is carried out using central processing units (CPUs), DL was only made possible thanks to the development of graphics processing units (GPUs), which have a massively parallel architecture consisting of thousands of cores and were designed to handle vast numbers of tasks simultaneously.

During the training stage of supervised learning algorithms, the labeled data is divided into training, validation, and testing subsets to reduce overfitting and estimate how well the models generalize. No standard methodologies exist to determine optimum proportions allocated to each set. The training set usually includes a large proportion of the available data and is used for the development of the model. The validation set is used to estimate overall model performance during development and fine-tune the algorithm’s hyperparameters (e.g., the number of network layers which could not be learnt). Dividing data into training and validation subsets can be done randomly at the onset of the process or by using a cross-validation approach. This involves dividing the entire dataset into folds of equal size and then training the algorithms in all the folds except one that is left out for validation. The process is repeated until all folds have been used as a validation set and the overall performance of the model is calculated as the average across all validation sets. Finally, an independent (ideally external) test set should be used to assess the model’s generalizability.

Despite ML’s vast potential and significant performance breakthroughs in fields such as speech recognition, natural language processing, and computer vision, these approaches are not without limitations and vulnerabilities. Some of these are shared with classical statistical approaches (33) while others are entirely novel (34). A significant potential pitfall of ML models derives from the presence of unrecognized confounders that can be present in both the training and testing sets, if they originated from the same dataset. This could result in overfitting

of the model to the training data, achieving an artificially inflated performance with poor generalization to other data sets in subsequent studies. The gold-standard approach to address this issue is to obtain a validation dataset acquired by an independent group under real-world conditions. Another possible cause of unsatisfactory generalization of an AI system is if the training data is not an accurate representation of the wider population. For example, an AI model trained on a healthy cohort may not generalize well to a general population that includes extreme disease phenotypes, and a system trained on images from a specific CMR scanner might not perform well when labeling images acquired under different technical conditions. Domain adaptation or transfer learning are fields of AI research that aim to address these challenges.

AI algorithms can also be oversensitive to changes in the input data and therefore vulnerable to unintentional or harmful interference. This was clearly demonstrated in experiments involving “adversarial examples” or inputs that lead the model to make a classification error. For example, the introduction of an imperceptible perturbation in a picture of a benign skin mole resulted in the misclassification as a malignant mole, with 100% confidence (35). The general application of AI has also been hindered by the “black-box” nature of several methodologies. Indeed, full clinical acceptability is only likely if it is possible to explore and scrutinize the predictive features and if the outputs are clinically interpretable.

At a more fundamental level, “Big Data” studies are often no more than observational research. As in classical statistics, observational AI studies cannot test causality and should therefore be considered hypothesis-generating that require further testing. A recent systematic review and meta-analysis of 82 studies applying DL methods to medical imaging found that although the diagnostic performance of DL methods was often reported as equivalent to human experts, few studies tested human vs. DL performance on the same sample and then went on to externally validate their findings (36). Furthermore, apart from a handful of exceptions (37), the effect of AI in routine clinical practice has been rarely tested in the setting of randomized controlled trials. Indeed, it has not been systematically demonstrated that the roll out of AI into clinical practice leads to an improvement in the quality of care, increased efficiency or improved patient outcomes (38). These studies will be required before this technology can be routinely used to help guide clinical care.

Table 1 provides an introduction to some of the technical and methodological aspects that should be considered in AI research.

Nevertheless, the use of machine learning methods in cardiovascular research has grown exponentially over recent years, with an ever increasing set of uses and applications. Traditional supervised ML methods have been applied successfully to classification tasks in extremely diverse input data, ranging from discrimination between sequences underlying *Cis*-regulatory elements from random genome sequences (39), separation of human induced pluripotent stem cell-derived cardiomyocytes of distinct genetic cardiac diseases (CPVT, LQT, HCM) (40) to numerous applications in medical imaging analysis. Examples of this include automated

TABLE 1 | Considerations in the use of machine learning in imaging-genetics research.

Selection of AI approach based on clinical question and data characteristics	Supervised methods suited to classification and prediction tasks involving “labeled” data: e.g., image segmentation or survival prediction. Unsupervised methods useful to identify structures and patterns in unlabeled data: e.g., association and clustering. Reinforcement learning algorithms interact with the environment by producing actions that get rewarded or penalized, while identifying the optimal path to address the problem. DL can be used to accelerate supervised, unsupervised or reinforcement learning but is better suited to larger, more unstructured datasets. Classical ML is more likely to work better in smaller training datasets.
Algorithm selection	Are there “off-the-shelf” algorithms tailored to identical problems or validated in similar data? Transparency, understandability and performance are all important features. Try to avoid “black box” approaches where it is not possible to scrutinize the features that inform the classification or explain the outputs in high-stakes decision-making.
Data pre-processing	Several steps are likely to be required in the preparation of data including anonymization, quality control, data normalization and standardization, addressing how to handle missing data points and outliers, imputation of missing values, etc. Is the training data an accurate representation of the wider data/population (e.g., all expected variation present, same technical characteristics)?
Feature selection	A subset of relevant features (variables or predictors) is selected from high dimensional data allowing for a more succinct representation of the dataset.
Data allocation	Evaluate the available data and plan the proportions of data being allocated into the training, testing, and validation datasets. Other approaches include cross validation, stratified cross validation, leave-one-out, and bootstrapping.
Hardware considerations	Based on the volume of data and methodological approaches are CPU clusters, GPUs, or cloud computing better suited?
Evaluation of model performance	Receiver operating characteristic (ROC) curves with accuracy measured by the area under the ROC curve (AUC), C-statistics, negative predictive value, positive predictive values, sensitivity, and specificity, Hosmer–Lemeshow test for goodness of fit, precision, recall, f-measure. Imaging segmentation accuracy (comparison between human expert labels and automated labels) reported as Dice metric, mean contour distance, and Hausdorff distance. If the accuracy is perfect, have too many predictors been included for the sample size or are there confounding biases hidden in the data that may result in the model overfitting the data? Compare performance against standard statistical approaches (i.e., multivariate regression). If several algorithms are tested report on them all and not just on the best performance.
Publication and transparency	Make code and anonymized sample of data publicly available (e.g., GitHub, Docker containers, R packages, or Code Ocean repositories). Encourage independent scrutiny of the algorithm.
Generalization and replication results	Algorithms should be validated by independent researchers on external cohorts and satisfy the requirements of medical devices and software regulatory frameworks.

quality control during CMR acquisition (41), high-resolution CMR study of cardiac remodeling in hypertension (42) and aortic stenosis (43), and echocardiographic differentiation of restrictive cardiomyopathy from constrictive pericarditis (44). Unsupervised ML analysis have provided new unbiased insights into cardiovascular pathologies such as by establishing subsets of patients likely to benefit from cardiac resynchronization therapy (45) and by agnostic identification of echocardiography derived patterns in patients with heart failure with preserved ejection fraction and controls (46). Traditional ML has also been used for prediction of outcomes such as hospital readmission due to heart failure (47), survival in pulmonary hypertension (48), and population-based cardiovascular risk prediction (49).

More recently, there has been a greater interest in DL approaches, which have been used with great promise in ever larger-scale classification tasks. Applications include the analysis of CMRs (50), echocardiograms (51), and electrocardiograms (52), identification of the manufacturer of a pacemaker from a chest radiograph (53), aortic pressure waveform analysis during coronary angiography (54); automated categorization of HCM and healthy CMRs (55) and detection of atrial fibrillation using smartwatches (56). DL has also been successfully used to address complex survival prediction tasks in pulmonary hypertension (57) and heart transplantation (58).

The analysis of ever larger and complex genome-scale biological datasets is also particularly suited to ML approaches. One of the strengths of these approaches comes from the ability to discover unknown structures in the data and to derive predictive models without requiring *a priori* assumptions about, frequently poorly understood, underlying biological mechanisms (59). The field is large, diverse and fast moving with new opportunities for AI to synthesize data and optimize the prediction of key functional biological features appearing all the time. Applications of traditional ML have ranged from the prediction of quantitative (growth) phenotypes from genetic data (60), to the identification of proteomic biomarkers of disease (61), to the prediction of metabolomes from gene expression (62). As in cardiology research, there has been growing interest in applying DL to the field of functional genomics. Such approaches have been used to predict sequence specificities of DNA- and RNA-binding proteins (31, 63), transcriptional enhancers (64) and splicing patterns (65) and to identify the functional effects of non-coding variants (66, 67). A more in depth discussion of the applications of ML and DL to genomics and other multi-omics data can be found elsewhere (68–71).

ARTIFICIAL INTELLIGENCE IN CARDIOVASCULAR IMAGING-GENETICS

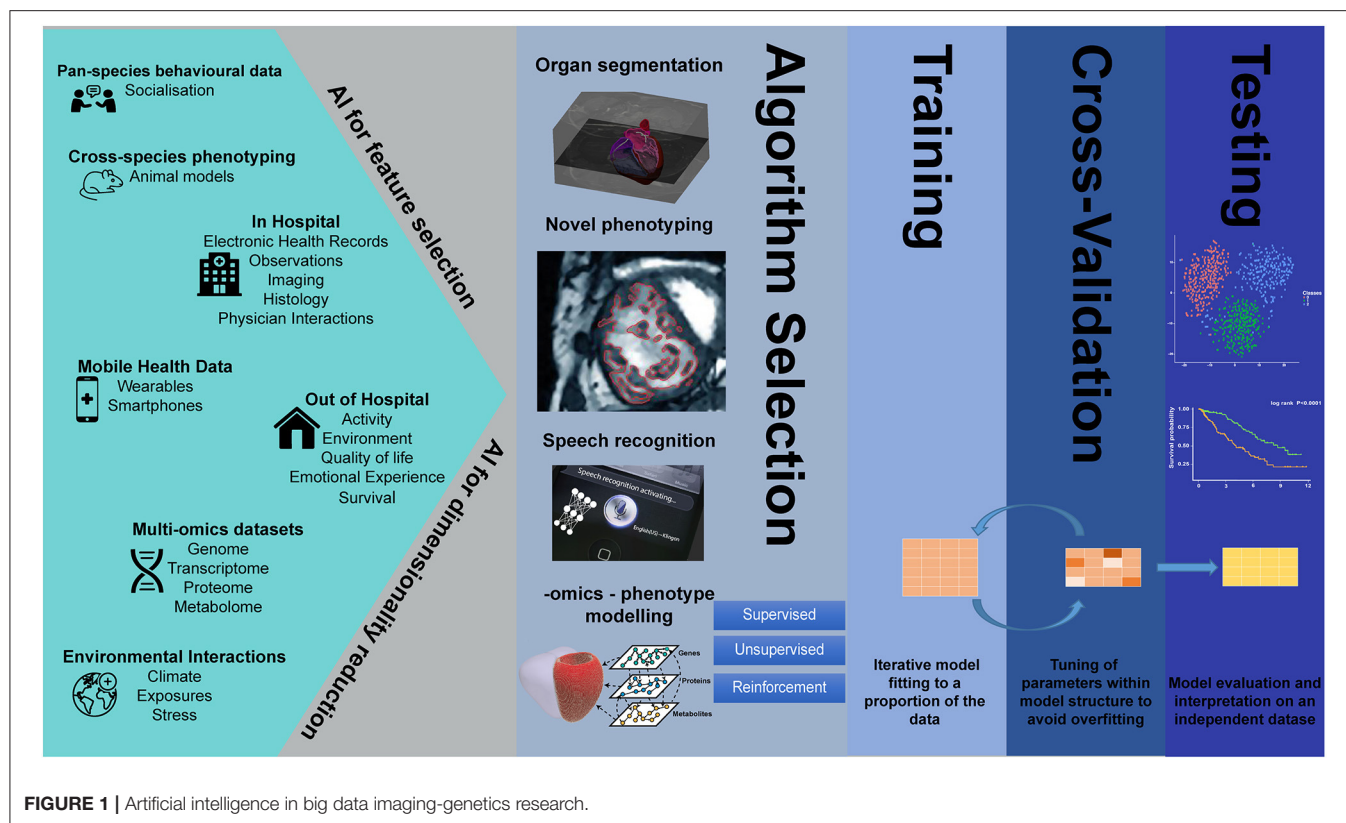
Despite the parallel successes of AI in the fields of genetics and imaging analysis, integrated imaging-genetics research is still an emerging field. However, several studies have already demonstrated the usefulness of AI tools in the analysis of large biological, imaging, and environmental data, in such tasks as dimensionality reduction and feature selection, speech recognition, clustering, image segmentation, natural

language processing, variable classification, and outcome prediction (Figure 1).

To predict which dilated cardiomyopathy patients responded to immunoglobulin G substitution (IA/IgG) therapy, as assessed by echocardiography, two supervised ML approaches, a random forest analysis and a support vector machine algorithm, were used independently on gene expression data derived from 48 endomyocardial biopsies (72). The overlapping set of 4 genes that was identified by both ML approaches was superior to clinical parameters in discriminating between responders and non-responders to therapy. The prediction performance was further improved by adding data on the negative inotropic activity (NIA) of antibodies. A support vector machine classifier, also proved to be extremely helpful in identifying specific proteomic signatures that accurately discriminated between patients with heart failure with reduced ejection fraction (HFrEF) and controls in the absence (73) or presence of chronic kidney disease (74). ML pipelines also often use feature selection to more efficiently process high dimensional phenotypes, distinguishing the most informative features from those that are redundant. For example, an information gain method was used to identify speckle-tracking features able to differentiate athlete's heart from HCM. The combination of three different supervised machine learning algorithms (support-vector machine, random forest, and neural network) trained on this sparser data was then shown to be better at distinguishing the two types of remodeling (ML model sensitivity = 87%; specificity = 82%) than conventional echocardiographic parameters (best parameter was e' —sensitivity = 84%; specificity = 74%) (75).

ML approaches have also been successfully used in the identification of new, useful structures in data. One such study, using a hypothesis-free unsupervised clustering approach, revealed four distinct proteomic signatures with differing clinical risk and survival in patients with pulmonary arterial hypertension (76). ML has similarly been able to identify new sub-phenotypes in heart failure with preserved ejection fraction, classifying subjects into three subgroups associated with distinct clinical, biomarker, hemodynamic, and structural groups with markedly different outcomes (77). Okser et al. used a naïve Bayes classifier in a longitudinal imaging-genetics study of 1,027 young adults to identify a predictive relationship between genotypic variation and early signs of atherosclerosis, as assessed by carotid artery intima-media thickness, which could not be explained by conventional cardiovascular risk factors (78).

Classification problems, such as pixel-wise classification of CMR images, are also particularly suited to supervised classical ML (79, 80) and deep learning approaches (81). These high-resolution representations of whole-heart shape and function can encode multiple phenotypes, such as wall thickness or strain, at each of thousands of points in the model (82). Such high-fidelity models were used in a study aiming to clarify the physiological role of titin-truncating variants (TTNtv), known to be a common cause of DCM but surprisingly also present in ~1% of the general population (83). Mass univariate analyses, adjusted for multiple clinical variables and multiple testing, were carried out at over 40,000 points of a statistical parametric map of 1,409 healthy volunteers. This identified an association between TTNtv



positive status and eccentric remodeling, indicating a previously unproven physiological effect of these variants in subjects without DCM. A similar phenotyping approach was used by Attard et al. in 312 patients to elucidate the physiological mechanisms that underpinned reported association between certain metabolites and survival in patients with pulmonary hypertension (84). Univariate regression models including clinical, hemodynamic, and metabolic data were fitted at each vertices of a 3D cardiac mesh. These showed coherent associations between 6 metabolites and right ventricular adaptation to pulmonary hypertension as well as showing that wall stress was an independent predictor of all-cause mortality.

ML algorithms have also shown promise in predicting outcomes, such as imaging surrogates of disease or response to treatment, from complex sets of clinical and genetic variables. For example, to predict the presence or absence of coronary plaques on CT coronary angiography, a gradient boosting classifier was trained on a proteomic assay and identified two distinct protein signatures (85). A subset of these was found to outperform generally available clinical characteristics in the prediction of patients with high risk plaques ($AUC = 0.79$ vs. $AUC = 0.65$), while a distinct set outperformed clinical variables in predicting absence of coronary disease ($AUC = 0.85$ vs. $AUC = 0.70$). In another study, a combination of random forest and neural network methods were used first to identify the most informative subset of clinical and genomic data and then to predict coronary artery calcium (86). Interestingly, the model trained on SNP data only was highly predictive ($AUC = 0.85$), and better than models trained on clinical data ($AUC = 0.61$) and on a combination

of genomic and clinical data ($AUC = 0.83$). Further validation experiments in patients with less severe coronary artery calcium showed poor predictive accuracy suggesting that the models' predictive value is limited to a range of (high) coronary calcium or that the models do not generalize well in the broader population. Schmitz et al. investigated the performance of 15 different supervised machine learning algorithms in predicting positive cardiac remodeling in patients that underwent cardiac resynchronization therapy (CRT) from clinical and genomic data (87). Several of the approaches demonstrated clear overfitting (accuracy $\sim 100\%$), while the algorithm that was identified as the most useful had a fair performance (accuracy = 83%) in addition to high transparency (predictive features easily identified).

Novel deep learning methods are also starting to make an impact in the imaging-genetics field by enabling unprecedented high-throughput image analysis. For example, DL methods have been able to achieve fully automated analysis of CMRs with a performance that is similar to human experts (88) and permitted the rapid segmentation of 17,000 CMRs that were then used in a GWAS (89). This identified multiple genetic loci and several candidate genes associated with LV remodeling, and enabled the computing of a polygenic risk score (PRS) that was predictive of heart failure in a validation sample of nearly 230,000 subjects (odds ratio 1.41, 95% CI 1.26 – 1.58, for the top quintile vs. the bottom quintile of the LV end-systolic volume).

While the use of AI in cardiovascular imaging-genetics has great potential, the limitations and challenges of AI in genetics (90) and imaging (91) are further amplified by combining these very large data. To date, no methodological approaches have

been able to include whole-genome and high-resolution whole-heart phenotypes, without requiring extensive dimensionality reduction, filtering and/or feature selection, possibly introducing errors or biases to the input data. Even when this challenge is dealt with, multiple testing correction will continue to be problematic, with the potential for false positive findings likely to only be reliably addressed with replication studies. In AI imaging-genetics, no single method is universally applicable, and the choice of whether and how to use ML or DL approaches will remain task, researcher and population specific, creating difficulties in the pooling of data and meta-analyses. It should not be forgotten that conventional analysis remains valid and has advantages when data are scarce or if the aim is to assess statistical significance, which is currently difficult using deep learning methods. Issues related to the lack of interpretability (“black box”) of some ML algorithms are less of an issue in imaging analysis, where accuracy of analysis can be visually verified, but very relevant to integrated imaging-genetics analysis or risk prediction, where identifying and explaining the features driving the algorithm’s output can be virtually impossible. The tendency to over-fit models to training datasets risks reduction in the performance of the model when applied to new populations. These problems are likely to be exacerbated if new test datasets include subjects with differing genetic or physiological backgrounds, data were acquired using different technical conditions (e.g., different scanners or different genotyping batches) or if the quality of data acquired in the research setting significantly differs from real world data sets. Finally, issues regarding privacy, ownership, and consent over vast amounts of genetic and imaging data and legal and ethical considerations for clinicians using integrated imaging-genetics algorithms will become an ever more relevant topic of debate.

Although the application of AI to imaging genetics-research is still new, these promising methods and findings warrant further extensive validation in independent populations. Fully integrated, end-to-end, imaging-genetics DL approaches are theoretically extremely attractive but as yet untested. To confidently implement AI methods in research and clinical practice, challenges regarding standardization of data acquisition and algorithm development and reporting still need to be overcome. Initiatives such as adapting the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) recommendations (92) to machine learning research [TRIPOD-ML (93)] are very much welcome. Ultimately, the additive value of AI-driven decision making may

require robust multi-center studies and randomized controlled trials (94, 95).

FUTURE PERSPECTIVES

The development of body imaging, the elucidation of inheritance and genetics and the application of statistics to medicine were some of the most important medical developments of the past millennium (96). AI now provides an unrivaled ability to integrate these three aspects in imaging-genetics studies of unprecedented scale and complexity. The increasing variety and capabilities of ML tools at the disposal of researchers provide a powerful platform to agnostically revisit classical definitions of disease, to more accurately predict outcomes and to vastly improve our understanding of the genetic and environmental underpinnings of cardiovascular health and pathology. ML approaches will play an increasing role in every field of cardiovascular research, from genomic discovery and deep phenotyping, to mechanistic studies and drug development. Concerted efforts to improve AI study design, reporting, and collaborative validation will greatly contribute to deliver on the great promise of AI and ultimately improve patient care.

AUTHOR CONTRIBUTIONS

AM, TD, and DO’R contributed to the content and writing of this manuscript.

FUNDING

AM, TD, and DO’R research was supported by the British Heart Foundation (RG/19/6/34387, NH/17/1/32725, and RE/13/4/30184); the National Institute for Health Research Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London; and the Medical Research Council, UK. AM acknowledges additional support from the Academy of Medical Sciences (SGL015/1006) and a Mason Medical Research Trust grant.

ACKNOWLEDGMENTS

The authors would like to thank Drs. Wenjia Bai and Carlo Biffi (Department of Computing, Imperial College London, London, UK) for their critical review of this article.

REFERENCES

- Ritchie H, Roser M. *Causes of Death*. Available online at: <https://ourworldindata.org/causes-of-death> (accessed September 25, 2019).
- Kathiresan S, Srivastava D. Genetics of human cardiovascular disease. *Cell*. (2012) 148:1242–57. doi: 10.1016/j.cell.2012.03.001
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. (2001) 409:860–921. doi: 10.1038/35057062
- Wetterstrand KA. *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*. Available online at: www.genome.gov/sequencingcostsdata (accessed September 25, 2019).
- Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. *Sci Transl Med*. (2015) 7:283rv3. doi: 10.1126/scitranslmed.aaa3487
- Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol*. (2019) 20:e262–73. doi: 10.1016/S1470-2045(19)30149-4
- Bogdan R, Salmeron BJ, Carey CE, Agrawal A, Calhoun VD, Garavan H, et al. Imaging genetics and genomics in psychiatry.

- a critical review of progress and potential. *Biol Psychiatry*. (2017) 82:165–75. doi: 10.1016/j.biopsych.2016.12.030
8. Heinz A, Goldman D. Genotype effects on neurodegeneration and neuroadaptation in monoaminergic neurotransmitter systems. *Neurochem Int*. (2000) 37:425–32. doi: 10.1016/S0197-0186(00)00057-7
 9. Geisterfer-Lowrance AA, Kass S, Tanigawa G, Vosberg HP, McKenna W, Seidman CE, et al. A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation. *Cell*. (1990) 62:999–1006. doi: 10.1016/0092-8674(90)90274-I
 10. Herman DS, Lam L, Taylor MR, Wang L, Teekakirikul P, Christodoulou D, et al. Truncations of titin causing dilated cardiomyopathy. *N Engl J Med*. (2012) 366:619–28. doi: 10.1056/NEJMoa1110186
 11. Marian AJ, Belmont J. Strategic approaches to unraveling genetic causes of cardiovascular diseases. *Circ Res*. (2011) 108:1252–69. doi: 10.1161/CIRCRESAHA.110.236067
 12. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet*. (2018) 50:1412–25. doi: 10.1038/s41588-018-0205-x
 13. Leon-Mimila P, Wang J, Huertas-Vazquez A. Relevance of multi-omics studies in cardiovascular diseases. *Front Cardiovasc Med*. (2019) 6:91. doi: 10.3389/fcvm.2019.00091
 14. Houle D, Govindaraju DR, Omholt S. Phenomics: the next challenge. *Nat Rev Genet*. (2010) 11:855–66. doi: 10.1038/nrg2897
 15. Plomin R, Haworth CM, Davis OS. Common disorders are quantitative traits. *Nat Rev Genet*. (2009) 10:872–8. doi: 10.1038/nrg2670
 16. Vasani RS, Glazer NL, Felix JE, Lieb W, Wild PS, Felix SB, et al. Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *JAMA*. (2009) 302:168–78. doi: 10.1001/jama.2009.978-a
 17. Jøkerst JV, Cauwenberghs N, Kuznetsova T, Haddad F, Sweeney T, Hou J, et al. Circulating biomarkers to identify responders in cardiac cell therapy. *Sci Rep*. (2017) 7:4419. doi: 10.1038/s41598-017-04801-7
 18. Stringer S, Wray NR, Kahn RS, Derks EM. Underestimated effect sizes in GWAS: fundamental limitations of single SNP analysis for dichotomous phenotypes. *PLoS ONE*. (2011) 6:e27964. doi: 10.1371/journal.pone.0027964
 19. Bilder RM, Sabb FW, Cannon TD, London ED, Jentsch JD, Parker DS, et al. Phenomics: the systematic study of phenotypes on a genome-wide scale. *Neuroscience*. (2009) 164:30–42. doi: 10.1016/j.neuroscience.2009.01.027
 20. Schork NJ. Genetics of complex disease: approaches, problems, and solutions. *Am J Respir Crit Care Med*. (1997) 156(4 Pt 2):S103–9. doi: 10.1164/ajrccm.156.4.12-tac-5
 21. Suinesiaputra A, Sanghvi MM, Aung N, Paiva JM, Zemrak F, Fung K, et al. Fully-automated left ventricular mass and volume MRI analysis in the UK Biobank population cohort: evaluation of initial results. *Int J Cardiovasc Imag*. (2018) 34:281–91. doi: 10.1007/s10554-017-1225-9
 22. Knackstedt C, Bekkers SCAM, Schummers G, Schreckenber M, Muraru D, Badano LP, et al. Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain the FAST-EFs multicenter study. *J Am Coll Cardiol*. (2015) 66:1456–66. doi: 10.1016/j.jacc.2015.07.052
 23. Taylor CA, Fonte TA, Min JK. Computational fluid dynamics applied to cardiac computed tomography for noninvasive quantification of fractional flow reserve scientific basis. *J Am Coll Cardiol*. (2013) 61:2233–41. doi: 10.1016/j.jacc.2012.11.083
 24. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. (2019) 25:44–56. doi: 10.1038/s41591-018-0300-7
 25. Stein JL, Hua X, Lee S, Ho AJ, Leow AD, Toga AW, et al. Voxelwise genome-wide association study (vGWAS). *Neuroimage*. (2010) 53:1160–74. doi: 10.1016/j.neuroimage.2010.02.032
 26. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J R Stat Soc B*. (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
 27. Smith SM, Nichols TE. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*. (2009) 44:83–98. doi: 10.1016/j.neuroimage.2008.03.061
 28. Biffi C, de Marvao A, Attard MI, Dawes TJW, Whiffin N, Bai W, et al. Three-dimensional cardiovascular imaging-genetics: a mass univariate framework. *Bioinformatics*. (2018) 34:97–103. doi: 10.1093/bioinformatics/btx552
 29. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, et al. Mastering the game of go without human knowledge. *Nature*. (2017) 550:354. doi: 10.1038/nature24270
 30. Alansary A, Oktay O, Li Y, Folgoc LL, Hou B, Vaillant G, et al. Evaluating reinforcement learning agents for anatomical landmark detection. *Med Image Anal*. (2019) 53:156–64. doi: 10.1016/j.media.2019.02.007
 31. Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat Biotechnol*. (2015) 33:831–8. doi: 10.1038/nbt.3300
 32. Bernard O, Lalande A, Zotti C, Cervenansky F, Yang X, Heng PA, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE Trans Med Imaging*. (2018) 37:2514–25. doi: 10.1109/TMI.2018.2837502
 33. Nieuwenhuis S, Forstmann BU, Wagenmakers EJ. Erroneous analyses of interactions in neuroscience: a problem of significance. *Nat Neurosci*. (2011) 14:1105–7. doi: 10.1038/nn.2886
 34. Chen JH, Asch SM. Machine learning and prediction in medicine - beyond the peak of inflated expectations. *N Engl J Med*. (2017) 376:2507–9. doi: 10.1056/NEJMp1702071
 35. Finlayson SG, Bowers JD, Ito J, Zittrain JL, Beam AL, Kohane IS. Adversarial attacks on medical machine learning. *Science*. (2019) 363:1287–9. doi: 10.1126/science.aaw4399
 36. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digital Health*. (2019) 1:e271–e97. doi: 10.1016/S2589-7500(19)30123-2
 37. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. *BMJ Open Respir Res*. (2017) 4:e000234. doi: 10.1136/bmjresp-2017-000234
 38. Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. *Nat Rev Cardiol*. (2016) 13:350–9. doi: 10.1038/nrcardio.2016.42
 39. Lee D, Kapoor A, Safi A, Song L, Halushka MK, Crawford GE, et al. Human cardiac cis-regulatory elements, their cognate transcription factors, and regulatory DNA sequence variants. *Genome Res*. (2018) 28:1577–88. doi: 10.1101/gr.234633.118
 40. Juhola M, Joutsijoki H, Penttinen K, Aalto-Setälä K. Detection of genetic cardiac diseases by Ca(2+) transient profiles using machine learning methods. *Sci Rep*. (2018) 8:9355. doi: 10.1038/s41598-018-27695-5
 41. Tarroni G, Oktay O, Bai W, Schuh A, Suzuki H, Passerat-Palmbach J, et al. Learning-based quality control for cardiac MR images. *IEEE Trans Med Imaging*. (2019) 38:1127–38. doi: 10.1109/TMI.2018.2878509
 42. de Marvao A, Dawes TJ, Shi W, Durighel G, Rueckert D, Cook SA, et al. Precursors of hypertensive heart phenotype develop in healthy adults: a high-resolution 3D MRI study. *JACC Cardiovasc Imaging*. (2015) 8:1260–9. doi: 10.1016/j.jcmg.2015.08.007
 43. Bhuvana AN, Treibel TA, De Marvao A, Biffi C, Dawes TJW, Doumou G, et al. Sex and regional differences in myocardial plasticity in aortic stenosis are revealed by 3D model machine learning. *Eur Heart J Cardiovasc Imaging*. (2019) jez166. doi: 10.1093/ehjci/jez166
 44. Sengupta PP, Huang YM, Bansal M, Ashrafi A, Fisher M, Shameer K, et al. Cognitive machine-learning algorithm for cardiac imaging: a pilot study for differentiating constrictive pericarditis from restrictive cardiomyopathy. *Circ Cardiovasc Imaging*. (2016) 9:e004330. doi: 10.1161/CIRCIMAGING.115.004330
 45. Cikes M, Sanchez-Martinez S, Claggett B, Duchateau N, Piella G, Butakoff C, et al. Machine learning-based phenotyping in heart failure to identify responders to cardiac resynchronization therapy. *Eur J Heart Fail*. (2019) 21:74–85. doi: 10.1002/ehf.1333
 46. Sanchez-Martinez S, Duchateau N, Erdei T, Kunszt G, Aakhus S, Degiovanni A, et al. Machine learning analysis of left ventricular function to characterize heart failure with preserved ejection fraction. *Circ Cardiovasc Imaging*. (2018) 11:e007138. doi: 10.1161/CIRCIMAGING.117.007138

47. Mortazavi BJ, Downing NS, Bucholz EM, Dharmarajan K, Manhapra A, Li SX, et al. Analysis of machine learning techniques for heart failure readmissions. *Circ Cardiovasc Qual Outcomes*. (2016) 9:629–40. doi: 10.1161/CIRCOUTCOMES.116.003039
48. Dawes TJW, de Marvao A, Shi W, Fletcher T, Watson GMJ, Wharton J, et al. Machine learning of three-dimensional right ventricular motion enables outcome prediction in pulmonary hypertension: a cardiac MR imaging study. *Radiology*. (2017) 283:381–90. doi: 10.1148/radiol.2016161315
49. Weng SF, Reys J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS ONE*. (2017) 12:e0174944. doi: 10.1371/journal.pone.0174944
50. 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) 20:65. doi: 10.1186/s12968-018-0471-x
51. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, et al. Fully automated echocardiogram interpretation in clinical practice. *Circulation*. (2018) 138:1623–35. doi: 10.1161/CIRCULATIONAHA.118.034338
52. Hannun AY, Rajpurkar P, Haghighpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. (2019) 25:65–9. doi: 10.1038/s41591-018-0268-3
53. Howard JP, Fisher L, Shun-Shin MJ, Keene D, Arnold AD, Ahmad Y, et al. Cardiac rhythm device identification using neural networks. *JACC Clin Electrophysiol*. (2019) 5:576–86. doi: 10.1016/j.jacep.2019.02.003
54. Howard JP, Cook CM, van de Hoef TP, Meuwissen M, de Waard GA, van Lavieren MA, et al. Artificial intelligence for aortic pressure waveform analysis during coronary angiography. *Mach Learn Patient Safety*. (2019) 12:2093–101. doi: 10.1016/j.jcin.2019.06.036
55. Biffi C, Oktay O, Tarroni G, Bai W, De Marvao A, Doumou G, et al. (eds). *Learning Interpretable Anatomical Features Through Deep Generative Models: Application to Cardiac Remodeling*. Cham: Springer International Publishing (2018).
56. Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, et al. Passive detection of atrial fibrillation using a commercially available smartwatch. *JAMA Cardiol*. (2018) 3:409–16. doi: 10.1001/jamacardio.2018.0136
57. Bello GA, Dawes TJW, Duan J, Biffi C, de Marvao A, Howard L, et al. Deep learning cardiac motion analysis for human survival prediction. *Nat Mach Intell*. (2019) 1:95–104. doi: 10.1038/s42256-019-0019-2
58. Medved D, Ohlsson M, Hoglund P, Andersson B, Nuges P, Nilsson J. Improving prediction of heart transplantation outcome using deep learning techniques. *Sci Rep*. (2018) 8:3613. doi: 10.1038/s41598-018-21417-7
59. Angermueller C, Parnamaa T, Parts L, Stegle O. Deep learning for computational biology. *Mol Syst Biol*. (2016) 12:878. doi: 10.15252/msb.20156651
60. Martens K, Hallin J, Warringer J, Liti G, Parts L. Predicting quantitative traits from genome and phenome with near perfect accuracy. *Nat Commun*. (2016) 7:11512. doi: 10.1038/ncomms11512
61. Swan AL, Mobasheri A, Allaway D, Liddell S, Bacardit J. Application of machine learning to proteomics data: classification and biomarker identification in postgenomics biology. *OMICS*. (2013) 17:595–610. doi: 10.1089/omi.2013.0017
62. Zelezniak A, Vowinkel J, Capuano F, Messner CB, Demichev V, Polowsky N, et al. Machine learning predicts the yeast metabolome from the quantitative proteome of kinase knockouts. *Cell Syst*. (2018) 7:269–83 e6. doi: 10.1016/j.cels.2018.08.001
63. Zeng H, Edwards MD, Liu G, Gifford DK. Convolutional neural network architectures for predicting DNA-protein binding. *Bioinformatics*. (2016) 32:i121–i7. doi: 10.1093/bioinformatics/btw255
64. Liu F, Li H, Ren C, Bo X, Shu W. PEDLA: predicting enhancers with a deep learning-based algorithmic framework. *Sci Rep*. (2016) 6:28517. doi: 10.1038/srep28517
65. Leung MK, Xiong HY, Lee LJ, Frey BJ. Deep learning of the tissue-regulated splicing code. *Bioinformatics*. (2014) 30:i121–9. doi: 10.1093/bioinformatics/btu277
66. Kelley DR, Snoek J, Rinn JL. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res*. (2016) 26:990–9. doi: 10.1101/gr.200535.115
67. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods*. (2015) 12:931–4. doi: 10.1038/nmeth.3547
68. Eraslan G, Avsec Z, Gagneur J, Theis FJ. Deep learning: new computational modelling techniques for genomics. *Nat Rev Genet*. (2019) 20:389–403. doi: 10.1038/s41576-019-0122-6
69. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet*. (2015) 16:321–32. doi: 10.1038/nrg3920
70. Min S, Lee B, Yoon S. Deep learning in bioinformatics. *Brief Bioinform*. (2017) 18:851–69. doi: 10.1093/bib/bbw068
71. Zou J, Huss M, Abid A, Mohammadi P, Torkamani A, Telenti A. A primer on deep learning in genomics. *Nat Genet*. (2019) 51:12–8. doi: 10.1038/s41588-018-0295-5
72. Ameling S, Herda LR, Hammer E, Steil L, Teumer A, Trimpert C, et al. Myocardial gene expression profiles and cardiodepressant autoantibodies predict response of patients with dilated cardiomyopathy to immunoadsorption therapy. *Eur Heart J*. (2013) 34:666–75. doi: 10.1093/eurheartj/ehs330
73. Rossing K, Bosselmann HS, Gustafsson F, Zhang ZY, Gu YM, Kuznetsova T, et al. Urinary proteomics pilot study for biomarker discovery and diagnosis in heart failure with reduced ejection fraction. *PLoS ONE*. (2016) 11:e0157167. doi: 10.1371/journal.pone.0157167
74. Farmakis D, Koeck T, Mullen W, Parisiis J, Gogas BD, Nikolaou M, et al. Urine proteome analysis in heart failure with reduced ejection fraction complicated by chronic kidney disease: feasibility, and clinical and pathogenetic correlates. *Eur J Heart Fail*. (2016) 18:822–9. doi: 10.1002/ehf.544
75. Narula S, Shameer K, Salem Omar AM, Dudley JT, Sengupta PP. Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography. *J Am Coll Cardiol*. (2016) 68:2287–95. doi: 10.1016/j.jacc.2016.08.062
76. Sweatt AJ, Hedlin HK, Balasubramanian V, Hsi A, Blum LK, Robinson WH, et al. Discovery of distinct immune phenotypes using machine learning in pulmonary arterial hypertension. *Circ Res*. (2019) 124:904–19. doi: 10.1161/CIRCRESAHA.118.313911
77. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiadu M, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation*. (2015) 131:269–79. doi: 10.1161/CIRCULATIONAHA.114.010637
78. Okser S, Lehtimäki T, Elo LL, Mononen N, Peltonen N, Kahonen M, et al. Genetic variants and their interactions in the prediction of increased pre-clinical carotid atherosclerosis: the cardiovascular risk in young Finns study. *PLoS Genet*. (2010) 6:e1001146. doi: 10.1371/journal.pgen.1001146
79. Shi W, Lombaert H, Bai W, Ledig C, Zhuang X, Marvao A, et al. Multi-atlas spectral PatchMatch: application to cardiac image segmentation. *Med Image Comput Assist Interv*. (2014) 17(Pt 1):348–55. doi: 10.1007/978-3-319-10404-1_44
80. Bai W, Shi W, de Marvao A, Dawes TJ, O'Regan DP, Cook SA, et al. A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of healthy subjects and an analysis of shape and motion. *Med Image Anal*. (2015) 26:133–45. doi: 10.1016/j.media.2015.08.009
81. Duan J, Bello G, Schlemper J, Bai W, Dawes TJW, Biffi C, et al. Automatic 3D Bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach. *IEEE Trans Med Imaging*. (2019) 38:2151–64. doi: 10.1109/TMI.2019.2894322
82. de Marvao A, Dawes TJ, Shi W, Minas C, Keenan NG, Diamond T, et al. Population-based studies of myocardial hypertrophy: high resolution cardiovascular magnetic resonance atlases improve statistical power. *J Cardiovasc Magnetic Resonance*. (2014) 16:16. doi: 10.1186/1532-429X-16-16
83. Schafer S, de Marvao A, Adami E, Fiedler LR, Ng B, Khin E, et al. Titin-truncating variants affect heart function in disease cohorts and the general population. *Nat Genet*. (2017) 49:46–53. doi: 10.1038/ng.3719
84. Attard MI, Dawes TJW, de Marvao A, Biffi C, Shi W, Wharton J, et al. Metabolic pathways associated with right ventricular adaptation to pulmonary hypertension: 3D analysis of cardiac magnetic resonance imaging. *Eur Heart J Cardiovasc Imaging*. (2019) 20:668–76. doi: 10.1093/ehjci/jej175

85. Bom MJ, Levin E, Driessen RS, Danad I, Van Kuijk CC, van Rossum AC, et al. Predictive value of targeted proteomics for coronary plaque morphology in patients with suspected coronary artery disease. *EBioMedicine*. (2019) 39:109–17. doi: 10.1016/j.ebiom.2018.12.033
86. Oguz C, Sen SK, Davis AR, Fu YP, O'Donnell CJ, Gibbons GH. Genotype-driven identification of a molecular network predictive of advanced coronary calcium in ClinSeq(R) and Framingham Heart Study cohorts. *BMC Syst Biol*. (2017) 11:99. doi: 10.1186/s12918-017-0474-5
87. Schmitz B, De Maria R, Gatsios D, Chrysanthakopoulou T, Landolina M, Gasparini M, et al. Identification of genetic markers for treatment success in heart failure patients: insight from cardiac resynchronization therapy. *Circ Cardiovasc Genet*. (2014) 7:760–70. doi: 10.1161/CIRCGENETICS.113.000384
88. Bhuvana 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. *Circ Cardiovasc Imaging*. (2019) 12:e009214. doi: 10.1161/CIRCIMAGING.119.009759
89. Aung N, Vargas JD, Yang C, Cabrera CP, Warren HR, Fung K, et al. Genome-wide analysis of left ventricular image-derived phenotypes identifies fourteen loci associated with cardiac morphogenesis and heart failure development. *Circulation*. (2019) 140:1318–30. doi: 10.1161/CIRCULATIONAHA.119.041161
90. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. (2017) 169:1177–86. doi: 10.1016/j.cell.2017.05.038
91. Petersen SE, Abdulkareem M, Leiner T. Artificial intelligence will transform cardiac imaging-opportunities and challenges. *Front Cardiovasc Med*. (2019) 6:133. doi: 10.3389/fcvm.2019.00133
92. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group. *Circulation*. (2015) 131:211–9. doi: 10.1161/CIRCULATIONAHA.114.014508
93. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet*. (2019) 393:1577–9. doi: 10.1016/S0140-6736(19)30037-6
94. Dawes TJW, Bello G, O'Regan DP. *Multicentre Study of Machine Learning to Predict Survival in Pulmonary Hypertension*. OSF. Available online at: doi: 10.17605/OSF.IO/BG6T9 (accessed September 25, 2019).
95. Lin H, Li R, Liu Z, Chen J, Yang Y, Chen H, et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomized controlled trial. *EClinicalMedicine*. (2019) 9:52–9. doi: 10.1016/j.eclinm.2019.03.001
96. Looking back on the millennium in medicine. *N Engl J Med*. (2000) 342:42–9. doi: 10.1056/NEJM200001063420108

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 de Marvao, Dawes and O'Regan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Image-Based Cardiac Diagnosis With Machine Learning: A Review

Carlos Martin-Isla^{1*}, Victor M. Campello¹, Cristian Izquierdo¹, Zahra Raisi-Estabragh^{2,3}, Bettina Baeßler⁴, Steffen E. Petersen^{2,3} and Karim Lekadir¹

¹ Departament de Matemàtiques & Informàtica, Universitat de Barcelona, Barcelona, Spain, ² Barts Heart Centre, Barts Health NHS Trust, London, United Kingdom, ³ William Harvey Research Institute, Queen Mary University of London, London, United Kingdom, ⁴ Department of Diagnostic & Interventional Radiology, University Hospital Zurich, Zurich, Switzerland

Cardiac imaging plays an important role in the diagnosis of cardiovascular disease (CVD). Until now, its role has been limited to visual and quantitative assessment of cardiac structure and function. However, with the advent of big data and machine learning, new opportunities are emerging to build artificial intelligence tools that will directly assist the clinician in the diagnosis of CVDs. This paper presents a thorough review of recent works in this field and provide the reader with a detailed presentation of the machine learning methods that can be further exploited to enable more automated, precise and early diagnosis of most CVDs.

Keywords: cardiovascular disease, automated diagnosis, cardiac imaging, artificial intelligence, machine learning, deep learning, radiomics

OPEN ACCESS

Edited by:

Sebastian Kelle,
Deutsches Herzzentrum Berlin,
Germany

Reviewed by:

Joao Bicho Augusto,
Barts Heart Centre, United Kingdom
John Hoe,
MediRad Associates Ltd, Singapore

*Correspondence:

Carlos Martin-Isla
carlos.martinisla@ub.edu

Specialty section:

This article was submitted to
Cardiovascular Imaging,
a section of the journal
Frontiers in Cardiovascular Medicine

Received: 01 November 2019

Accepted: 06 January 2020

Published: 24 January 2020

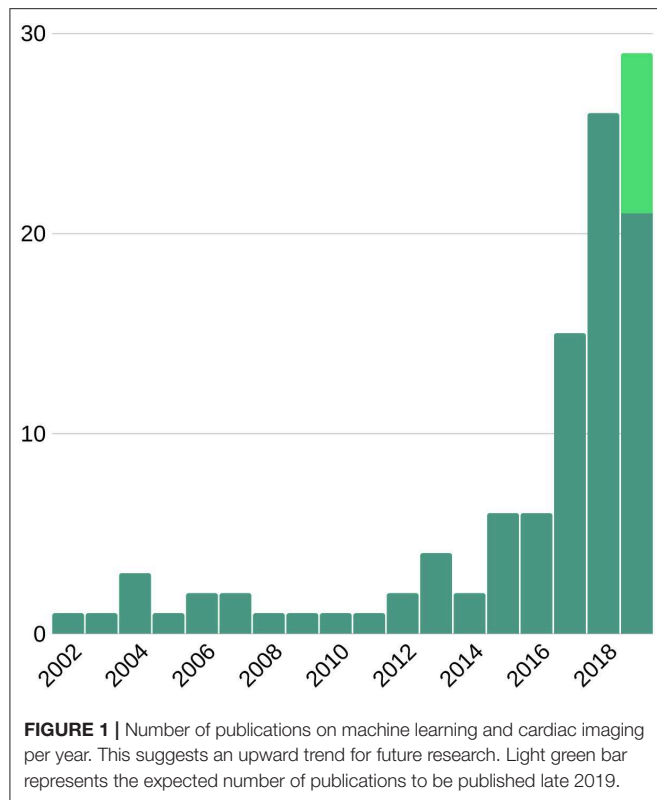
Citation:

Martin-Isla C, Campello VM,
Izquierdo C, Raisi-Estabragh Z,
Baeßler B, Petersen SE and Lekadir K
(2020) Image-Based Cardiac
Diagnosis With Machine Learning: A
Review. *Front. Cardiovasc. Med.* 7:1.
doi: 10.3389/fcvm.2020.00001

1. INTRODUCTION

Despite significant advances in diagnosis and treatment, cardiovascular disease (CVD) remains the most common cause of morbidity and mortality worldwide, accounting for approximately one third of annual deaths (1, 2). Early and accurate diagnosis is key to improving CVD outcomes. Cardiovascular imaging has a pivotal role in diagnostic decision making. Current image analysis techniques are mostly reliant on qualitative visual assessment of images and crude quantitative measures of cardiac structure and function. In order to optimize the diagnostic value of cardiac imaging, there is need for more advanced image analysis techniques that allow deeper quantification of imaging phenotypes. In recent years, the development of big data and availability of high computational power have driven exponential advancement of artificial intelligence (AI) technologies in medical imaging (**Figure 1**). Machine learning (ML) approaches to image-based diagnosis rely on algorithms/models that *learn* from past clinical examples through identification of hidden and complex imaging patterns. Existing work already demonstrates the incremental value of image-based cardiovascular diagnosis with ML for a number of important conditions such as coronary artery disease (CAD) and heart failure (HF). The superior diagnostic performance of AI image analysis has the potential to substantially alleviate the burden of cardiovascular disease through facilitation of faster and more accurate diagnostic decision making.

In this paper we describe the main ML techniques and the procedures required to successfully design, implement, and validate new ML tools for image-based diagnosis. We also present a comprehensive review of existing literature pertaining to applications of ML for image-based diagnosis of CVD.



2. OVERVIEW OF PIPELINE FOR IMAGE-BASED MACHINE LEARNING DIAGNOSIS

The overall pipeline to build ML tools for image-based cardiac diagnosis is schematically described in the following section, as well as in **Figure 2**. In short, it requires (1) input imaging datasets from which suitable imaging predictors can be extracted, (2) accurate output diagnosis labels, and (3) a suitable ML technique that is typically chosen and optimized depending

Abbreviations: Machine learning abbreviations: AI, Artificial Intelligence; AUC, Area Under Curve; ANN, Artificial Neural Networks; BN, Bayesian Network; CNN, Convolutional Neural Network; CL, Clustering; DL, Deep Learning; DT, Decision Tree; GA, Genetic Algorithm; GAN, Generative Adversarial Network; GBRT, Gradient Boosting Trees; kNN, k-Nearest Neighbors; LDA, Linear Discriminant Analysis; LR, Logistic Regression; ML, Machine Learning; PCA, Principal Component Analysis; PLS, Partial Least Squares; RF, Random Forest; ROC, Receiver Operating Characteristic Curve; Se, Sensitivity; Sp, Specificity; SVM, Support Vector Machine; VAE, Variational Autoencoder.

Cardiac imaging and clinical abbreviations: ARV, Abnormal Right Ventricle; ASD, Atrial Septal Defect; CAC, Coronary Artery Calcium; CAD, Coronary Artery Disease; CMR, Cardiac Magnetic Resonance; CT, Computed Tomography; CTA, Computed Tomography Angiography; CVD, Cardiovascular Disease; DCM, Dilated Cardiomyopathy; ECG, Electrocardiography; echo, Echocardiography; HCM, Hypertrophic Cardiomyopathy; HF, Heart Failure; HFpEF, Heart Failure with preserved Ejection Fraction; HHD, Hypertensive Heart Disease; ICA, Invasive Coronary Angiography; IR, Iterative Reconstruction; LV, Left Ventricle; MACE, Major Adverse Cardiovascular Event; MI, Myocardial Infarction; MR, Mitral Regurgitation; MRI, Magnetic Resonance Imaging; MYO, Myocarditis; NRS, Napkin Ring Sign; PET, Positron Emission Tomography; ROI, Region Of Interest; RV, Right Ventricle; SPECT, Single Positron Emission Computed Tomography.

on the application to predict the cardiac diagnosis (output) based on the imaging predictors (input). Additional non-imaging predictors (e.g., electrocardiogram data, genetic data, sex, or age) are often integrated into the ML model and typically improve model performance.

In this section, we will first discuss the input and output variables in more detail, before introducing common used ML techniques and their applications.

2.1. Data, Input and Output Variables

2.1.1. Sources of Cardiovascular Imaging Data

Robust ML models are reliant on the availability of sufficient and accurate data. Thus, data preparation is an important pre-requisite to derive that perform well on internal and external validation. Within cardiac imaging, there is increasing availability of quality sources of organized big data through various biobanks, bioresources, and registries. Available cohorts can be classified into population-based and clinical cohorts. Population cohorts such as the UK Biobank follow the health status of a representative sample of individuals from the general population and thus are particularly useful for risk stratification. In contrast, clinical cohorts, such as the Barts BioResource or the European cardiovascular magnetic resonance (EuroCMR) registry, are composed of clinical imaging from patients and therefore more suitable for building diagnostic tools. These datasets are an invaluable resource for the development and validation of ML diagnostic models (see **Table 1** for examples of additional cardiac imaging datasets).

2.1.2. Input Variables

Before an ML model can be built for image-based diagnosis estimation, it is necessary to suitably define the imaging inputs. Imaging inputs may be the raw imaging data (i.e., pixel intensities), conventional cardiac indices (and other transformed quantitative image parameters) or radiomics features extracted from the image. See **Figures 3** and **4** for additional information about input variables.

2.1.2.1. Conventional imaging indices

Conventional imaging indices include measures commonly used in routine clinical image analysis such as ventricular volumes in end diastole/systole and ventricular ejection fractions.

Estimation of these clinical indices requires prior contouring of the endocardial and epicardial boundaries of the relevant cardiac chambers. Deep learning approaches have been used to develop automated/semi-automated contouring tools for more efficient and reproducible segmentation of cardiac chambers.

Since manual delineation of these boundaries is tedious and subject to errors, many automatic or semi-automatic tools have been developed (see **Table 2** for examples of existing tools). Note that recently, many deep learning (DL) based approaches have been published for accurate and robust segmentation of the cardiac boundaries with promising results, however this is beyond the scope of this review [more details on this, as well as a basic introduction to ML, in cardiac magnetic resonance imaging (MRI) can be found in recent work by (3)].

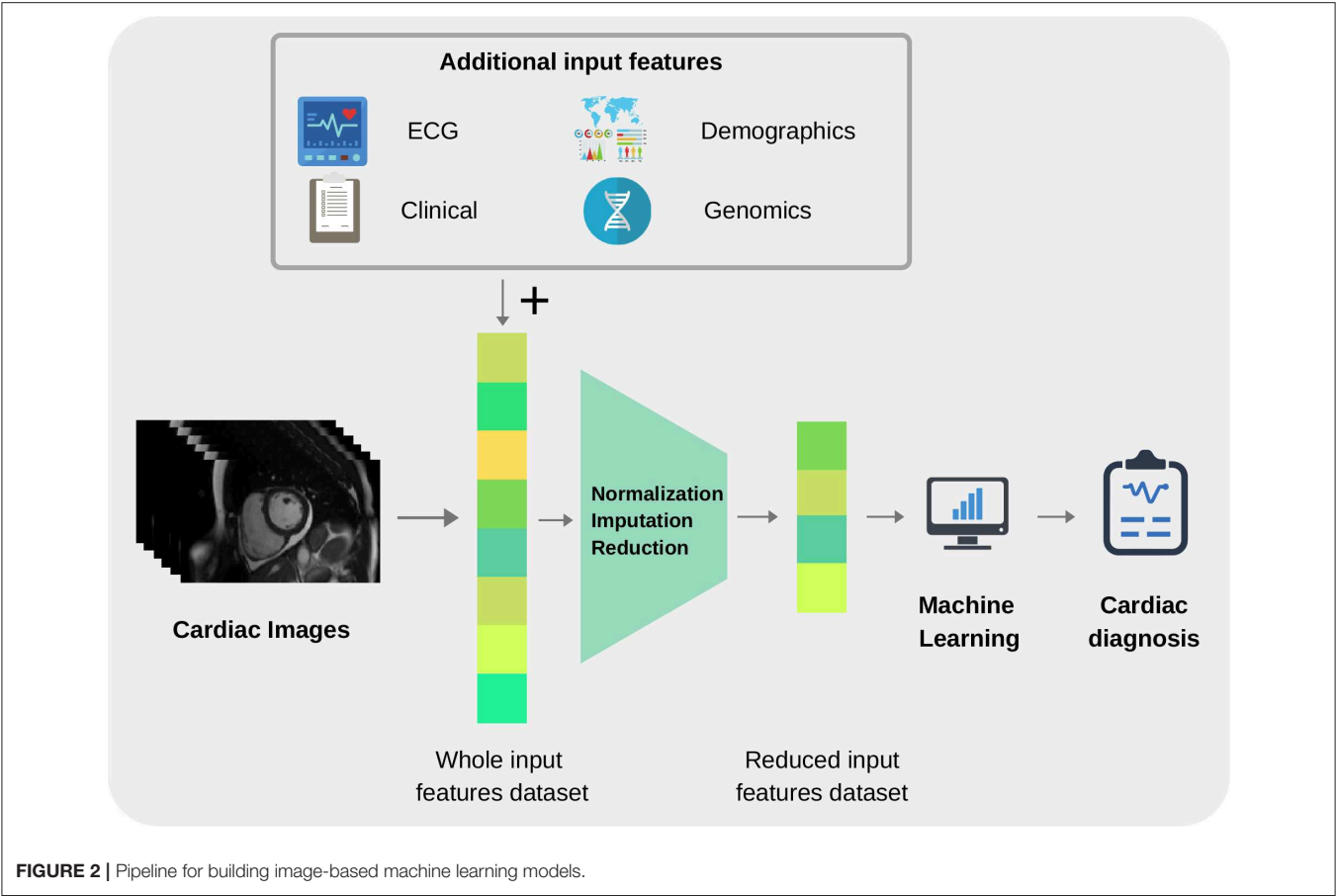


TABLE 1 | Selection of cardiac imaging datasets available.

Name	Country	Modality	Size	Year
Framingham Heart Study	USA	Echo/MRI/CT	>5,000	1948
Study of Health in Pomerania	DE	MRI	>8,000	1997
The Tromso Study	NO	Echo	3,287	1999
Multi-Ethnic Study of Atherosclerosis	USA	MRI	2,450	2000
UK Biobank	UK	MRI	20,000	2006
HUNT Study	NO	Echo	1,296	2006
Defibrillators to Reduce Risk by MRI Evaluation	USA	MRI	450	2007
Barts BioResource	UK	Echo / MRI	>10,000	2007
European CMR Registry	EU	MRI	>27,000	2007
NEO Study	NL	MRI	1,205	2008
SunnyBrook Cardiac Data	USA	MRI	45	2009
Registry of Fast Myocardial Perfusion Imaging with next gen SPECT	USA	SPECT	>20,400	2009
The German National Cohort	DE	MRI	20,000	2011
Maastricht Study	NL	Echo / CT	3,451	2012
Canadian Alliance for Healthy Hearts and Minds	CA	MRI	9,700	2013
Challenge on Endocardial 3D Ultrasound Segmentation	FR	Echo	45	2014
Hamburg City Health Study	DE	MRI	>45,000	2016
Automated Cardiac Diagnosis Challenge Dataset	FR	MRI	150	2017
Cardiac Acquisitions for Multi-structure Ultrasound Segmentation	FR	Echo	500	2019

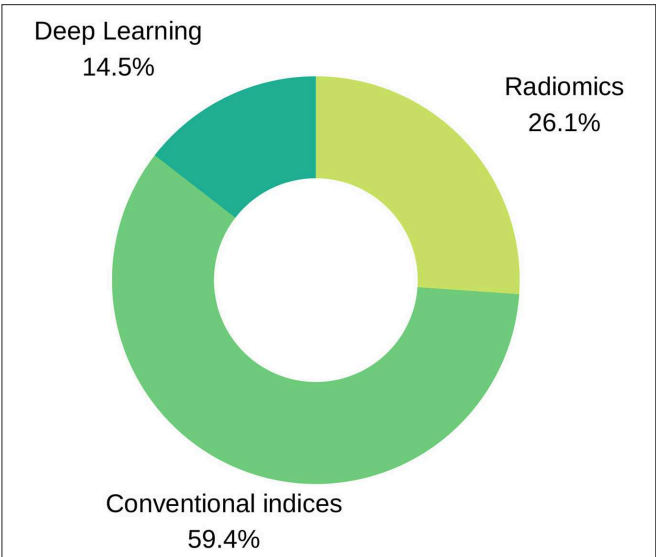


FIGURE 3 | Input variables type distribution in reviewed literature. As seen in the pie chart, conventional indices are the predominant features for training ML models, followed by radiomics and deep learning techniques.

TABLE 2 | Selection of cardiac structural and functional analysis softwares.

Name	Producer	Modality
CMRtools	Cardiovascular Imaging Solutions	MRI
suiteHEART	NeoSoft	MRI
CVI42	Circle Cardiovascular Imaging	MRI/CT
Medis Suite	Medis	MRI/CT
iNtuition	Terarecon	MRI/CT
Segment	Medviso	MRI/CT/SPECT
syngo.via	Siemens	MRI/CT/SPECT
IntelliSpace Portal	Philips	MRI/CT/echo
VevoLAB	Visualsonics	Echo
QLAB	Philips	Echo
TOMTEC	Philips	Echo

ejection fraction, RV and LV volume end-systole and end-diastole, myocardial mass, as well as the patient's height and weight. In Chen et al. (5), the authors integrated a set of 32 variables from clinical data, including ejection fraction, blood pressure, sex, age, as well as other conventional risk factors, to diagnose dilated cardiomyopathy (DCM). Juarez-Orozco et al. (6) merged ejection fractions at rest and stress with a pool of clinical parameters to predict ischemia and adverse cardiovascular events using ML.

Regarding motion, strain and single intensity analysis, in Mantilla et al. (7), global spatio-temporal image features are extracted to feed a support vector machine (SVM) classifier for LV wall motion assessment. Pairwise single intensity and variance regional differences in SPECT perfusion studies mimics the clinical procedure of qualitatively comparing stress and rest images in Bagher-Ebadian et al. (8). Contractility differences and multiscale wall motion assessment are performed by means of apparent flow in Moreno et al. (9) and Zheng et al. (10) where each feature describes an oriented velocity at a given position along the cardiac ROI.

2.1.2.2. Radiomics features

Radiomics analysis is the process of converting digital images to minable data. Analysis of the data through application of various statistical and mathematical processes allows quantification of various shape and textural characteristics of the image, referred to as radiomics features (Table 3). Radiomics analysis quantifies more advanced and complex characteristics of the cardiac chambers than is visually perceptible. Similarly to clinical imaging indices, radiomics requires the delineation of the cardiac structures before the features can be extracted.

Introduced in 2012 (11, 12), the radiomics paradigm was, for a long time, mostly exploited in oncology (13). Recently, a number of works have shown the promise of radiomics combined with ML for image-aided diagnosis of CVD. For instance, Cetin et al. (14) demonstrated that about 10 radiomics features integrated within an ML model are sufficient to discriminate between several major CVDs. More recently, researchers at Harvard University, Neisius et al. (15) have built an ML model with 6 radiomic features calculated from T1 mapping

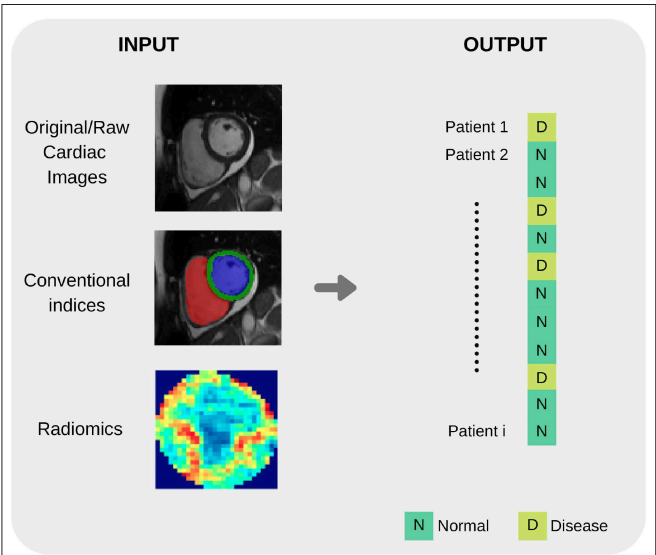


FIGURE 4 | Summary of common input and output variables for image-based diagnosis ML algorithms. Different cardiac imaging input features such as raw data, conventional indices extracted from a ROI or radiomics (delineation of cardiac anatomy is required for the last two cases) and desired output. Both structures shape the most basic requirement for a ML cardiac imaging application, data.

Some recent works will be listed to illustrate the use of conventional imaging indices as inputs for ML-based diagnosis models. In Khened et al. (4), an artificial neural network (ANN) was built to automatically diagnose several cardiac diseases such as hypertrophic cardiomyopathy (HCM), myocardial infarction (MI) and abnormal RV (ARV), by using as input LV and RV

TABLE 3 | Radiomics features overview.

Type	Description	Examples
Shape features	Describe geometric characteristics of the cardiac structures	Volume, surface area, sphericity, diameters, axis, surface to volume ratio, flatness
Intensity (First order)	Statistics on the intensity distributions within the region of interest (ROI)	Mean intensity, range, skewness (asymmetry) and entropy
Texture GLCM (Second order)	Quantifies the spatial relationship of the pixels in the ROI	Contrast, correlation
Texture GLSZM (Higher order)	Quantifies the number of connected voxels that share the same intensity level	Gray level non-uniformity, zone entropy
Texture GLRLM (Higher order)	Quantifies the gray level runs in the ROI	Run entropy, long run emphasis and short run emphasis
Texture NGTDM (Higher order)	Quantifies the difference between a gray value and the average gray value of its neighbors within a predefined distance	Busyness, strength
Texture GLDM (Higher order)	Quantifies the gray level dependencies in the ROI	Dependence non-uniformity, dependence entropy and dependence variance
Fractal dimension	Determines the ratio of change in detail to the change in scale	

sequences to differentiate between hypertensive heart disease (HHD) and HCM.

2.1.2.3. Raw imaging data

Whole raw images may also be used as the input for the ML model, without any pre-processing or calculation of hand-crafted input imaging features. About 10% of published reports rely on this type of modeling. In this case, the optimal features for predicting the cardiac diagnoses are self-learned automatically by the ML techniques based on the training sample, as opposed to *a priori* definition by the AI scientist.

For illustration, it is worth mentioning the work by Betancur et al. (16), an end-to-end DL model, estimating per-vessel CAD probability without any assumed subdivision of the input coronary territories from imaging data. The authors in Wolterink et al. (17) built a coronary artery calcification (CAC) detector, also based on DL trained on raw CT images. A similar DL model directly built from raw echo images was demonstrated in Lu et al. (18) to identify dilated cardiomyopathy cases. Also from raw echo images, the authors in Kusunose et al. (19) built a DL model for automatic detection of regional wall motion abnormalities.

2.1.3. Output

ML algorithms may be developed using supervised or unsupervised learning methods. Supervised learning requires

accurately labeled training examples. In the simplest form, the output is a binary variable which takes a value of 1 for a diseased individual and 0 for a control healthy subject. To obtain a robust ML model, it is recommended to use a balanced training sample, comprising a similar number of healthy and diseased subjects. Note that the binary classification can be easily extended to the multi-class case if several diseases or stages of disease are to be included in the ML model. Thus, supervised learning algorithms link the input variables to labeled outputs. Unsupervised learning is the training of algorithms without definition of the output. Through this technique, the ML algorithm groups the sample through recognition of inherent patterns within the data. In general, supervised learning outperforms unsupervised learning and so is the preferred method in situations where the ground truth is known. However, unsupervised learning has unique value for discovery of novel disease sub-types and patient stratification e.g., different pheno-groups of hypertensive heart disease or CAD.

2.2. Machine Learning Techniques

ML, refers to the use of computer algorithms that have the capacity to learn to perform given tasks from example data without the need for explicitly programmed instructions, i.e., image-based cardiac diagnosis in our case. This field of AI uses advanced statistical techniques to extract predictive or discriminatory patterns from the training data in order to perform the most accurate predictions on new data. We present the most commonly used ML techniques in the field of cardiac imaging and diagnosis for a non-expert audience and discuss their benefits and drawbacks (see **Table 4** and **Figure 5** for additional information). A list of diagnostic applications for each method will be provided as examples.

2.2.1. Logistic Regression

A Logistic Regression (LR) model is used to estimate the probability of a given output based on input variables in a continuous fashion, in contrast with a binary classifier. Final probabilities add up to one, so one obtains a stratification into all possible outcomes and the odds for each one. One property of this model is that a slight change in the input value may disproportionately impact the final probability prediction, as can be seen in **Figure 6A**. Additionally, the input vector dimension (number of predictor variables) must be kept low, as this can lead to costly model training processes and risks overfitting of the model to the training dataset with resultant poor generalisability of the model. Thus, when dealing with a large number of input variables, dimensionality reduction algorithms, such as principal component analysis (PCA) or linear discriminant analysis (LDA), are applied to reduce the number of predictors to those that are most informative. LR is a valuable model to be selected when different sources of data must be integrated in a binary classification task and low complexity is required.

In the literature, several works have applied LRs for their particular application. For example, Zheng et al. (10) applied a sequence of four LRs to classify patients according to cardiac pathologies by using shape features extracted from cine MRI

TABLE 4 | Overview of machine learning techniques.

Technique	Description	Advantages	Disadvantages
Logistic Regression	Extension of linear regression that outputs a binary classification	Simple and explainable; Does not require empirical parameter tuning nor input feature normalization	Not suitable for non-linear problems; Prone to overfitting
Support Vector Machine	Finds the optimal boundary between classes	Can handle different types of non-linear class separations; Does not require large training samples	Requires hyperparameter tuning and non-linear kernel selection; Not suitable for very large datasets
Random Forest	Generates a set of hierarchical decision queries over the input and output data	Automatically defines feature importances; Does not require input feature normalization	Prone to overfitting; Requires definition of depth and number of trees
Artificial Neural Network	Models complex classification tasks by propagating input data through a network of non-linear transformations	Generalizes well when trained on large samples	Difficult to interpret; Requires prior selection of a network design (e.g., depth of the network); Requires large training data
Convolutional Neural Network	ANNs adjusted for the processing and classification of image data	Flexible design depending on the applications; Can learn the optimal features directly from the images	Same limitations as ANNs
Clustering	Finds subgroups within the input feature space in an unsupervised manner	Useful to discover subgroups when the groups labels are not unknown a priori; Simple and fast	Sensitive to initialization and scale; Difficult to estimate the number of subgroups

per segment. Thus they obtained a simple and easily interpretable model with only three input features per classifier. In another example, Arsanjani et al. (20) used a combination of classifiers improved with a LR to diagnose obstructive CAD using SPECT images. Finally, a LR was also applied by Baessler et al. (21) to diagnose acute or chronic heart failure-like myocarditis.

2.2.2. Support Vector Machine (SVM)

Support vector machines (SVMs) are supervised ML models whereby the optimal linear or non-linear boundary segregating the data into two or more classes is identified, as can be seen in **Figure 6B**. Prior to application of SVMs, the function which will be used for segregating the data should be selected, the so called kernel function. The most used kernels are the linear

function or the Gaussian function. The remaining parameters of the SVM model are chosen empirically by training a set of models and keeping the settings as for the model with the lowest error. Since this model is insensitive to non-discriminating dimensions, a dimension reduction could be applied to the input variables to ease the training and obtain a better generalization as for linear regression. One major drawback of SVM is that it becomes memory expensive when large amounts of data are processed. SVM is a good choice to identify non-linearity and sparsity in the input data : different kernels can be used to fit different distributions.

Amongst all ML methods presented in this review, SVM is one of the most frequently used techniques and some works find this model to obtain the best performance. For example, Conforti and Guido (22) presented a comparison of SVM models with different kernels (polynomial, Gaussian and Laplacian functions), the original 105 features and a feature selection of 25 as input for the early diagnosis of myocardial infarction. Similarly, Arsanjani et al. (23) and Ciecholewski (24) found that a SVM model outperformed previous algorithms used in the task of CAD identification by using data extracted from SPECT images. In the first example, a second degree polynomial was used as kernel while in the second, a Gaussian function showed better performance. A SVM was also the best model when predicting acute coronary syndrome for 228 patients using histological, ECG and echo qualitative features, as shown by Berikol et al. (25). As a final example, Borkar and Annadate (26) obtained a very good accuracy for discrimination of DCM and atrial septal defect (ASD) patients using radiomics features and a SVM using a Gaussian kernel function.

2.2.3. Random Forest (RF)

This popular technique consists of a combination of decision trees (DTs) trained on different random samples of the training set, as can be seen in **Figure 6C**. Each DT is a set of rules based on the input features values optimized for accurately classifying all elements of the training set. DTs are nonlinear models and tend to have high variance. If the DT is grown very deep it can pick up irregularities in the training dataset and consequently problems with overfitting may be encountered. This problem is counteracted in a RF through training on different samples of the training dataset. In this way the variance is reduced as the number of DT used, lowering therefore the generalization error and becoming a powerful technique. The final prediction is obtained by selecting the mode (for classification problems) or the mean (for regression problems) of all predictions. Two parameters must be selected for these models: the number of DTs and the depth level for each DT (i.e., the number of decisions). However, one must bear in mind that whilst discriminatory power on training dataset is increased as DT increase in depth, this is often at the expense of losing generalization power. RFs are chosen in order to transform the problem into a set of hierarchical queries represented as DTs. However, RFs are not very resistant to noise.

In the literature, RF or DT have been used frequently and were selected as the best performing model in some works. For

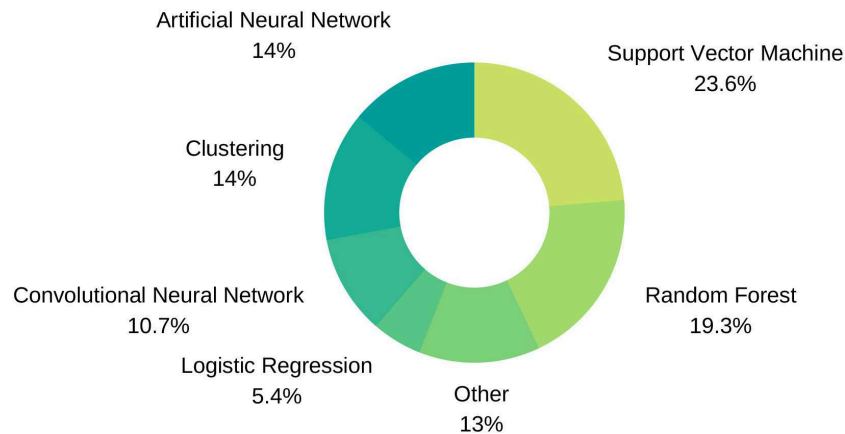


FIGURE 5 | Machine Learning technique distribution.

example, Moreno et al. (9) compared SVM and RF models in binary classification tasks with 2,964 input features for different cardiac pathologies, such as HF or HCM, using optical flow features in cardiac MRI, where the latter model obtained the best performance in most cases. In this case, each DT in the RF model had two depth levels for fast predictions in clinical practice. In another example, Wong et al. (27) a RF outperformed a SVM for infarction detection by means of regional intensity analysis and motion modeling. As a final example, a RF was also used by Baessler et al. (28) to find the most discriminative features in texture analysis for T1-weighted cardiac MRI for HCM and normal patients classification.

2.2.4. Cluster Analysis

Cluster analysis relates to the set of techniques that group together subjects in the form of data points according to similarity or proximity in the parametric space given by quantitative data extracted from input variables (image parameters and/or clinical information), as can be seen in **Figure 6D**. This technique is very useful for patient stratification, since patients with apparently similar pathology, according to existing image analysis techniques, may fall into previously unrecognized subsets which may inform understanding of disease pathophysiology and inform more effective targeted therapies. Some clustering techniques require definition of outcomes, which means that lay on the unsupervised learning ML group. However, in classification tasks a very common supervised clustering strategy is k-nearest neighbors (kNN) clustering, where k is the number of neighbor subjects to look at when finding subgroups. In this case, surrounding diagnosed subjects will determine the outcome for a new patient. Most of the reviewed literature in clustering uses kNN (29, 30).

Additional studies report the use of different cluster analysis for classification and/or discovery of cardiac pheno-groups. For example, Bruse et al. (31) used hierarchical clustering techniques to subdivide 60 patients into three groups, a healthy cohort and two associated with congenital heart disease by using shape features from cardiac MRI. Wojnarski et al. (32) also used a

cluster analysis technique to group bicuspid aortic valve patients using CT data to find three phenotypes, and a RF was applied later to identify biomarker differences for these phenotypes using echo and clinical data.

2.2.5. Artificial Neural Network (ANN)

ANNs are motivated by the structure and interactions of biological neural networks. These models propagate input data in a hierarchical fashion through internal nodes in different layers. Each input line has a corresponding weight that must be estimated and iteratively adjusted during the training process. The ANN adapts until the weights giving optimal model performance are identified (**Figure 6E**). A nonlinear function is applied in each node to the contribution from incoming connections for obtaining its value/activation (net input function). Weight optimization provides the model with great adaptability to complex boundaries separating classes because of the high non-linear combinations of features involved in such models. Moreover, the connections between layers in an ANN can be used to design different networks depending on the application. Some caveats are the lack of an underlying theory for deciding the amount of layers or nodes in each layer, that depends on each problem and the amount of training data, as well as the trend for these models to adapt to the training set due to the large difference between number of parameters/weights of the model and training samples. ANNs are the best choice when large amount of data is available.

In the literature, these techniques have been applied frequently. For example, Tsai et al. (33) used ANNs for detection of HCM and DCM patients using features extracted from echo. And more recently, two works by Nakajima et al. (34, 35), with the same SPECT dataset with 1,001 cases, used ANNs to assess CAD using features extracted from stress and rest images with good accuracy.

2.2.6. Convolutional Neural Network (CNN)

CNNs are an extension of ANNs in which the value of a node in a given layer is affected by the spatial surrounding

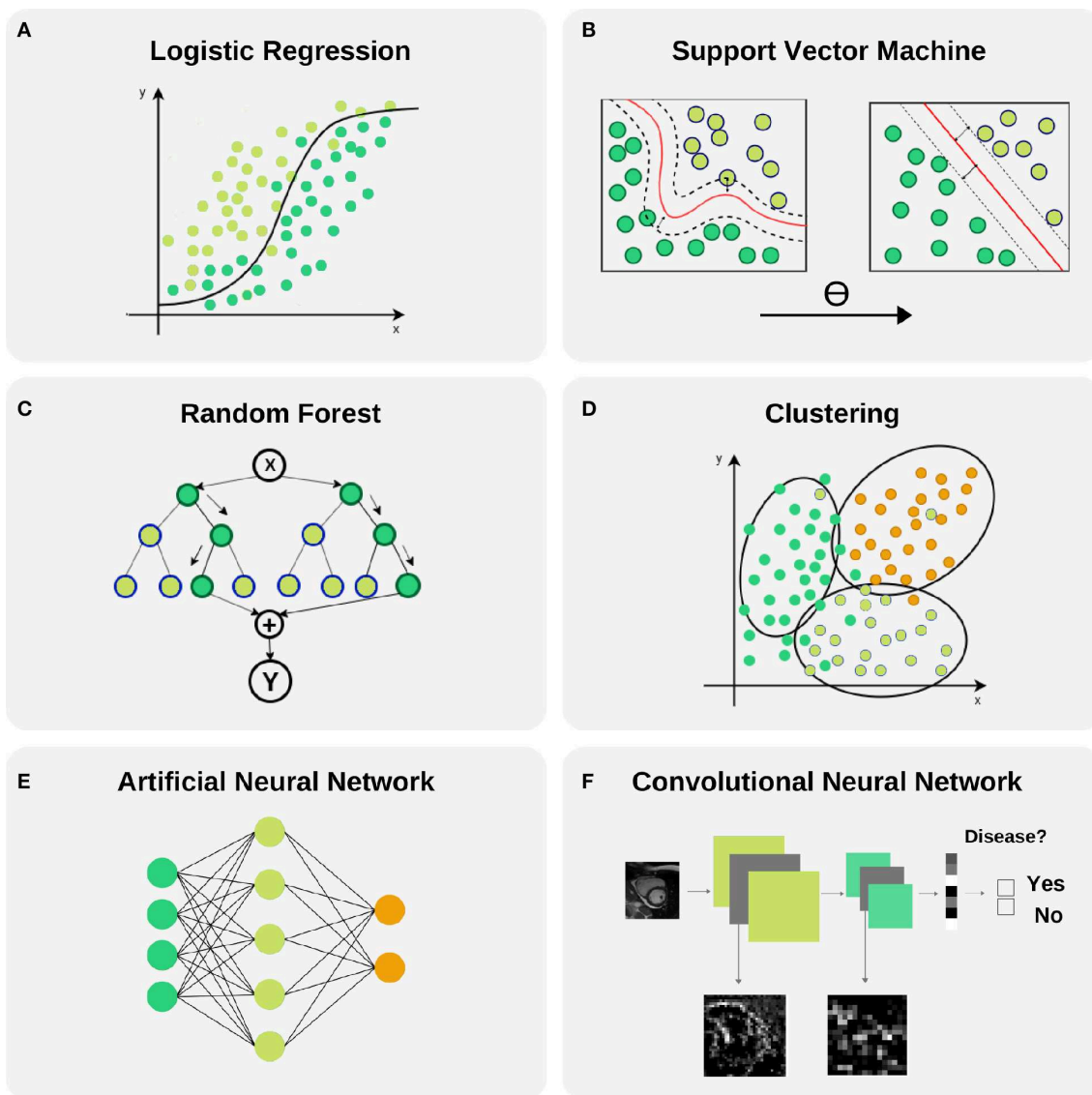


FIGURE 6 | Selected machine learning techniques. **(A)** Logistic Regression is used to model the probability of a binary outcome. In the figure, Y axis represents the probability while X axis is the continuous input variable. Notice that small changes in X produce large variations of the final probability Y, mainly in the central part of the plot where the uncertainty of the model is larger. This model can be extended to a multi-class problems. **(B)** Support Vector Machine models are able to transform a non-linear boundary to a linear one using the kernel trick. During the training process, the distance between classes to the final selected boundary is maximized. **(C)** Random Forest is a technique that combines Decision Trees for reducing the uncertainty in the final prediction. It is based in a recursive binary splitting strategy where upper nodes are intended to be the most discriminative ones and subsequent branching is applied to less relevant variables. **(D)** Clustering is a technique with capability to find subgroups (clusters) along data. There are different cluster techniques, some need a prior number of clusters (kMeans), some of them can be used with output information (kNN), and others are fully unsupervised (meanShift). **(E)** Artificial neural networks are able to model complex non-linear relations between input variables and outcomes by propagating structured data (green nodes—input variables), e.g., radiomics, through hidden layers (blue nodes) to obtain an output (orange nodes). **(F)** Convolutional neural networks are the backbone of Deep Learning applications. They comprise input and output layers separated by multiple hidden layers. Their ability to hierarchically propagate imaging information and extract data-driven features implies automatic detection of relevant cardiac imaging biomarkers within the intermediate layers.

of a node in the previous layer through an operation called convolutional product. These models are specially designed for image processing, where spatial information for the nodes (pixels) is essential for the final prediction. The advantages and disadvantages are shared with ANNs. The main difference that make these models very popular nowadays is that images are provided as input without any feature extraction. These models

are able to extract their own meaningful features for the final prediction, as illustrated in **Figure 6F**. Additional models exist for compressing images to a lower dimensional representation space such as the Variational Autoencoder (VAE) and Generative Adversarial Networks (GANs) where additional analysis can be carried out more easily (e.g., clustering or classification with a SVM model).

TABLE 5 | Common normalization techniques.

Technique	Description	Advantages
Mean/variance normalization	Centering to zero mean and unit variance	Avoid high variance features dominance
Range scaling	Mapping to a given interval	Robustness to small variances, preserve zero entries
Robust scaling	Mapping interval with interquartile information	Robustness to outliers
Image normalization	Brightness/contrast correction	Avoid variability in pixel intensity distribution

A balanced approach should be taken to defining the layers of a CNN; whilst a deeper network loses information from the original image with each new layer, a network with few layers could have problems extracting meaningful features for the final prediction. CNNs are widely used for analysis of images and their application to cardiac imaging is reported in numerous studies. Wolterink et al. (17) presented a framework where two cascading CNNs were able to detect CAC using cardiac computed tomography angiography (CTA) images. Their models had 8–13 convolutional layers that reduced 200×200 features (pixel intensities) to only 32. Zhang et al. (36) used a 13-layer CNN to diagnose HCM, cardiac amyloidosis and pulmonary artery hypertension from echo images of size 224×224 , that were reduced to 4,096 features. Madani et al. (37) used a CNN model to predict left ventricular hypertrophy from echo images of size 120×160 .

2.2.7. Additional Steps

2.2.7.1. Normalization

Due to the diverse nature of different information sources in cardiac medicine, a normalization step is often required prior to model crafting. In general, learning algorithms benefit from standardization of the data set, e.g., some algorithms as SVM will improve cardiovascular predictions if all numerical features are zero centered and have a variance of the same magnitude order. Furthermore, some non-linear transformations can prepare the selected features to create a model more robust to outliers. Some of the most common techniques are mentioned in **Table 5**.

For illustration, Wong et al. (27) shows that feature normalization has a positive impact in the ML model performance. Moreover, categorical variables should be encoded using *Integer encoding*, that consist in referencing each possible categorical value with an integer, or *One-Hot encoding*, that considers each possible categorical value as a new binary variable.

2.2.7.2. Dimensionality reduction and feature selection

Frequently, after extracting features from different sources such as demographic and clinical data, conventional indices and imaging parameters, one ends up with thousands of values defining a single patient. This information is later utilized during the training process of ML models, but the combination of a large number of input parameters with a limited number of samples (as usually happens in the medical field) can make the optimization problem expensive and may limit the generalization

ability of our model. Thus, a dimensionality reduction algorithm is usually applied to the input data, such as principal component analysis (PCA) or linear discriminant analysis (LDA). Another proactive approach is feature selection. Such method will add sequentially the most discriminative features for the particular model instance being trained and dismiss redundant and non-informative ones.

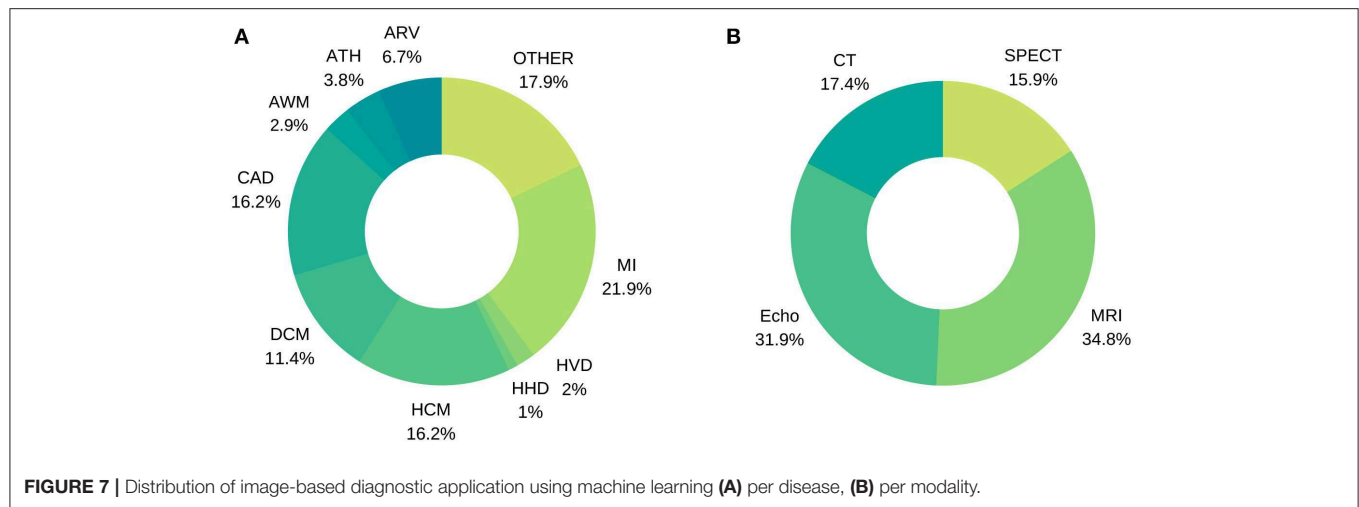
For example, Tabassian et al. (29) aimed to analyze deformation curves of the LV in echocardiographic records of 120 patients. The strain curves obtained were reduced by means of PCA and the result was used to train a strain kNN model. The resultant accuracy was 0.87, significantly higher than the clinician's results, 0.7. Cetin et al. (38) identified HHD from healthy controls in 200 subjects with SVM and sequential forward feature selection. The predictive power of selected radiomics (AUC = 0.76) was substantially improved compared to conventional indices (AUC = 0.62).

2.2.8. Validation

In order to prove the validity of ML applied to cardiac imaging, results must be analyzed from two perspectives: statistical validity, considering the reproducibility with different cohorts and correctness of statistical values obtained (i.e., metrics), and intra-validity, regarding the clinical and real implications of the algorithms on a daily basis (i.e., clinical effectiveness). This is a pairwise co-existence; none of the ML cardiac imaging algorithms will be applied in clinical routine if there is no agreement from both sides. The following sub-sections will describe how the metrics and the clinical effectiveness are considered.

A cohort is sorted in a very specific manner for ML purposes. For the validity of the algorithms, a whole cardiac imaging data set should be split into 3 different subgroups, called *training set*, *validation set*, and *testing set*, respectively. These groups are often selected in such way that subgroups share demographic distributions such as age or sex, in order to represent a real world scenario. Of course, a balanced distribution of control and pathologic subjects is also required. Once the ML model is trained and tested, different metrics are obtained to evaluate its performance.

Accuracy measures the percentage of the algorithm classifying the input data correctly. It is a simple measure used in multiple scientific scenarios if there is no class imbalance (i.e., one class represented by a higher number of individuals compared with the rest). One of the drawbacks of using accuracy as the metric is that there is a knowledge loss when measuring False Positive and False Negative observations. Therefore, *Specificity* (Sp) and *Sensitivity* (Se) are widely used for measuring the performance of the algorithm, this time taking into consideration a possible class imbalance. In order to assess the performance of an algorithm and to understand where there might be a miss-classification issue, a table report called *Confusion Matrix* is used. This specific table layout is typically used to describe the performance of a supervised learning model. Each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class (or vice versa). This way, a computer scientist can have a wider



overview of the parameters that may be changed or which classes are down-performing the algorithm. From sensitivity, specificity and the confusion matrix we can extract a performance plot representation called the receiver operating (ROC) curve. It is created by plotting the true positive rate (TP rate) against the false positive rate (FP rate) at various threshold settings. In ML, the true-positive rate is also known as sensitivity, recall or probability of detection. ROC analysis is related in a direct and natural way to cost/benefit analysis of diagnostic decision making. The area under the ROC curve (AUC) is another metric used to measure algorithms' performance.

It is noticeable that AUC can be derived from decision boundaries obtained by ML models despite the fact that it is trained with discrete outputs. When a trained model is asked to make a prediction, a probability can be computed and used to generate a ROC analysis.

3. DIAGNOSTIC APPLICATIONS—A REVIEW OF LITERATURE

We conducted an organized, pre-defined literature search of two electronic databases (Google Scholar, Scopus). We included studies using a well-defined ML technique for cardiac image analysis using echocardiography, cardiac magnetic resonance, cardiac computed tomography, or single photon emission computed tomography (SPECT). Our search strategy comprised a series of title and whole text searches with search terms combined using Boolean operators. Search results were filtered by subject area, limiting to entries from Cardiology, Computer Science and Engineering fields. We review in detail various achievements in the diagnosis of a wide range of cardiac diseases using image-based ML methods. Statistics about the conducted literature review can be seen in **Figure 7**.

3.1. Myocardial Infarction

Accurate and timely identification of MI helps in guidance of treatment strategies and reduction in the time taken for further tests. While MI diagnostic assessment using imaging is prone to inter- and intra-observer variability and requires significant

amount of time of experts, ML methods offer opportunities to simplify, speed up and quantify the diagnostic process in combination with conventional assessment. For example, Nakada et al. (39) demonstrated that MI diagnosis can be achieved in echo using quantitative motion features, avoiding the inter-observer human variability, as input for an ANN reaching an accuracy of 0.95. Later, Unguru et al. (40) validated these results in mice models by inducing MI in healthy specimens with a prediction accuracy of 0.91, comparing several ML techniques. The same level of accuracy was obtained in the first texture analysis work, by Agani et al. (41), with only 17 subjects and a clustering approach. This echocardiographic research was later extended with a full pool of texture features and 160 subjects by Sudarshan et al. (42). In this work, DT, ANN and SVM models were benchmarked, with the best accuracy obtained using ANN: 0.94 (Se = 0.91, Sp = 0.97). Vidya et al. (43) also performed an intensive texture analysis for 800 subjects, achieving an accuracy of 0.99 using a SVM. In their study, different pre-processing techniques were used to enhance the cardiac images.

Cardiac MRI has particular value in identification of MI. Since 2017, 13 studies were found integrating input variables from this imaging modality. Baeßler et al. (44) used late gadolinium enhancement MRI as a standard reference for non-enhanced MRI discrimination between chronic and subacute MI. Radiomic features in combination with a LR gave an AUC of 0.92 in a cohort of 180 patients. Similarly, segment viability can be detected on cine MRI using also radiomics, as suggested by Larroza et al. (45). This classification between viable, non-viable and remote segments yielded an AUC of 0.84. However, we believe that these encouraging results should be validated with a bigger cohort, and a well-balanced segment viability distribution. Recently, Zhang et al. (46) tried to detect MI from non-enhanced MRI images. 212 patients with chronic MI and 87 healthy control patients were used to train a three-stage DL pipeline. The per-segment AUC for detecting chronic MI was 0.94 (Sp = 0.99, Se = 0.9)

Two consecutive state-of-the-art texture analysis studies were conducted in cardiac CT: Mannil et al. (47) and Mannil et al.

TABLE 6 | Selected studies using image-based ML analysis for the diagnosis of Myocardial Infarction.

Publication	Modality	Biomarker	ML technique	Diagnostic	Sample size	Performance
(44)	MRI	Radiomics	LR	MI	180	ACC = 0.92
(46)	MRI	Conventional	ANN	MI	299	AUC = 0.94
(45)	MRI	Radiomics	SVM	MI	50	AUC = 0.84
(9)	MRI	Conventional	SVM/RF	MI/HCM	45	ACC = 0.94
(94)	MRI	Conventional	DT/CL/SVM	MI	200	ACC = 0.95
(95)	MRI	Conventional	PLS	MI	200	ACC = 0.98
(22)	Echo	Qualitative	SVM	MI	242	ACC = 0.97
(39)	Echo	Conventional	ANN	MI/AP	91	ACC = 0.95
(40)	Echo	Conventional	BN/DT/CL/SVM	MI	42	ACC = 0.91
(42)	Echo	Radiomics	DT/ANN/SVM	MI	160	ACC = 0.94
(41)	Echo	Radiomics	CL	MI	17	ACC = 0.91
(43)	Echo	Radiomics	SVM	MI	800	ACC = 0.99
(29)	Echo	Conventional	CL	MI	120	ACC = 0.87
(47)	CT	Radiomics	RF/CL/ANN	MI	87	ACC = 0.78
(48)	CT	Radiomics	DT	MI	30	ACC = 0.97
(27)	CT	Conventional	SVM/RF	MI	170	ACC = 0.85
(96)	SPECT	Conventional	BN	MI/CAD	728	ACC = 0.78

(48). The former underlines ML ability for detecting MI on non-contrast low radiation dose CT images on the basis of features invisible to the radiologists' eye, obtaining an AUC of 0.78. The latter study evaluates the impact of automatic classification methods using different iterative reconstruction (IR) strengths for contrast-enhancement images, reporting an accuracy of 0.94 (IR 3) and 0.97 (IR 5) for the ML model, while three independent readers achieved 0.73 (IR 5) on average. A summary of MI studies can be found in **Table 6**.

3.2. Cardiomyopathies

Cardiomyopathy is a broad term describing various heart muscle disorders, a first level of subclassification is into ischaemic and non-ischaemic cardiomyopathies. This heterogeneous group of disorders have many causes, signs and symptoms, and require different treatments. The challenge of distinguishing different cardiomyopathies is illustrated by the fact that many of them can be associated with diverse manifestations. Each disease entity is associated with a typical imaging phenotypes. Whilst in routine image analysis, it is not always possible to discriminate individual cardiomyopathies, this may be improved with the more granular and quantitative approach to image analysis in ML models. These premises makes ML-based imaging diagnosis a perfect tool for computer aided analysis of heterogeneous cardiomyopathies. For example, Gopalakrishnan et al. (49) used a set of conventional indices from a pediatric cardiac MRI cohort of 83 subjects to characterize five different cardiomyopathies. In this study, a DT (AUC = 0.79) was compared with other ML methods (AUC = 0.73–0.77). Physiological vs. pathological patterns of HCM remodeling were characterized by Narula et al. (50) using an ensemble of models with conventional indices from 2D echo as input (Se = 0.96, Sp = 0.77).

In 2017, a relevant challenge was organized by Bernard et al. (51). The Automated Cardiac Diagnosis Challenge (ACDC)

aimed to evaluate the performance of different automatic methods for the classification of 150 subjects into 5 categories (healthy, HCM, DCM, ARV and MI) as provided by clinical experts. Several approaches were proposed for this problem. Khened et al. (4) and Wolterink et al. (52) used a set of conventional indices extracted from their own automatic delineations as input for a RF to obtain an accuracy of 0.96 and 0.86 on the test set, respectively. Isensee et al. (53) also used a RF and their own segmentation scheme to classify cardiac cycle dynamic features, with an accuracy of 0.92. From this study, the benefit of the addition of temporal analysis is remarkable and provides a strong argument to be exploited further in future cine MRI studies. Cetin et al. (14) used SVM to classify a complete pool of radiomic features from manual segmentation, obtaining also an accuracy of 0.92. Additional research has been done later using the same dataset. Snaauw et al. (54) proposed a novel approach, using CNN bottleneck representations to discriminate between the five categories, obtaining a modest accuracy of 0.78. Another interesting approach was taken by Biffi et al. (55). Their VAE architecture was trained with two multi-center cohorts of 537 and 200 patients and tested on their own dataset and on the ACDC dataset, obtaining an accuracy of 1.0 and 0.9, respectively.

Later, Puyol-Antón et al. (56) combined MRI and echo data and per-segment motion analysis to diagnose DCM by means of LDA, achieving an accuracy of 0.94 (Sp = 0.96, Se = 0.93). Recently, Neisius et al. presented two complementary works approaching HCM and HHD diagnosis from two different perspectives, Neisius et al. (15, 57). In the first work, a complete strain analysis and a LR achieved an accuracy of 0.67 (Sp = 0.64, Se = 0.68). The second one applied an exhaustive texture analysis for T1 mapping. A selection of 6 radiomic texture features and a linear SVM model showed an improved accuracy of 0.86 (Sp = 0.91, Se = 0.77). A summary of cardiomyopathy studies can be found in **Table 7**.

TABLE 7 | Selected studies using image-based ML analysis for diagnosis of various cardiomyopathies.

Publication	Modality	Biomarker	ML technique	Diagnostic	Sample size	Performance
(49)	MRI	Conventional	BN	HCM/DCM/ARV/MYO	83	AUC = 0.79
(28)	MRI	Radiomics	RF/LR	HCM	62	AUC = 0.95
(52)	MRI	Conventional	RF	MI/HCM/DCM/ARV	100	ACC = 0.86
(14)	MRI	Radiomics	SVM	MI/HCM/DCM/ARV	100	ACC = 0.92
(53)	MRI	Conventional	RF	MI/HCM/DCM/ARV	100	ACC = 0.92
(4)	MRI	Conventional	RF	MI/HCM/DCM/ARV	100	ACC = 0.96
(55)	MRI	Deep Learning	VAE	HCM	737	ACC = 1.00
(10)	MRI	Conventional	LR	MI/HCM/DCM/ARV	100	ACC = 0.94
(57)	MRI	Conventional	LR	HHD/HCM	224	ACC = 0.67
(15)	MRI	Radiomics	SVM	HHD/HCM	224	ACC = 0.86
(54)	MRI	Deep Learning	CNN	MI/HCM/DCM/ARV	100	ACC = 0.78
(9)	MRI	Conventional	SVM/RF	MI/HCM	45	ACC = 0.94
(31)	MRI	Conventional	CL	CHD	60	ACC = 0.89
(50)	Echo	Conventional	SVM/RF/ANN	HCM/ATHCM	139	ACC = 0.91
(33)	Echo	Radiomics	ANN/GA	HCM/DCM	90	ACC = 0.95
(18)	Echo	Deep Learning	CNN	HCM/DCM	927	AUC = 0.84
(56)	Echo/MRI	Conventional	SVM	DCM	69	ACC = 0.94
(26)	Echo	Radiomics	SVM	DCM/ASD	439	ACC = 0.98
(37)	Echo	Deep Learning	CNN/GAN	HCM	772	ACC = 0.92
(36)	Echo	Deep Learning	CNN	HCM/CA/PH	14,035	AUC = 0.93

3.3. Coronary Artery Disease

Non-invasive imaging assessment for detection of CAD has a great potential impact on clinical practice. If ischemia can be discarded with a high probability, invasive coronary angiography (ICA) may be avoided. Advanced ML image analysis techniques can improve the diagnostic accuracy of myocardial ischemia and through this improve CAD management and reduce unnecessary downstream testing.

A very first approach dating from 1999 showed promising results. Considering ICA as reference standard, Kukar et al. (58) used scintigraphy, ECG and data on symptoms from 327 patients to detect CAD. Different ML models and feature selections were tested and in some cases the ML model outperformed clinicians in accuracy (0.92 vs. 0.91, respectively), but not in sensitivity. An exhaustive approach by Kurgan et al. (59) sets the base for a semi-automated diagnosis pipeline in perfusion SPECT. In their work, a pseudo-DT was crafted from intensity-based features, for 267 subjects, achieving an overall accuracy of 0.8. Another similar work was conducted in perfusion SPECT ($n = 115$) and Equilibrium Radionuclide Angiocardigraphy ($n = 58$) by Bagher-Ebadian et al. (8). Using ICA as ground truth for both studies, CAD was assessed using mean and variance intensity features extracted from stress and rest studies in anterior, left anterior oblique and left lateral projections, obtaining accuracies of 0.77 and 0.73 with an ANN. A similar methodology was covered in detail by Guner et al. (60). A cohort of 308 patients with clinical coronary CTA assessment was utilized to train an ensemble of ANNs for CAD discrimination. A combination of demographic information and frequency, phase and brightness features provided as input variables resulted in model accuracy of 0.74, outperforming some of the

non-expert clinicians. The results revealed that single-vessel CAD was more difficult to identify. Recently, complementary work by Shibutani et al. (61), including per-segment analysis, was performed on 21 patients who underwent perfusion SPECT. A total of 109 abnormal regions were examined and an ANN achieved better results than two independent observers for stress defect and ischemia detection, with respect to ICA as gold standard.

Alternatively, resting CT can be used for CAD diagnosis without additional contrast injection for stress imaging. Han et al. (62) used 3 quantitative features and the 17-segment model to obtain 51 input variables for training a gradient boosting algorithm, a ML technique that builds an ensemble of classifiers to improve the final accuracy. Invasive angiography and FFR were used as gold standard. This study based on a 252 patients' cohort from 5 countries and 17 centers, obtained an AUC of 0.75. Another state-of-the-art approach using cardiac CT, by Coenen et al. (63), showed that improved reclassification of non-significant stenosis is possible with ML-based image analysis. Three hundred and fifty-one patients, including 525 vessels with invasive FFR comparison were included in this study. A set of 28 anatomical features were computed from semi-automatic 3D CT reconstructions. On a per-vessel basis, diagnostic accuracy improved from 0.58 (CTA) to 0.78 (ML model). The per-patient accuracy improved from 0.71 to 0.85. A summary of CAD studies can be found in **Table 8**.

3.4. Atherosclerosis

Atherosclerosis is a strong and independent predictor of cardiovascular events. Plaque is often scored manually by experts, which leads to an increase in workload, is prone to

TABLE 8 | Selected studies using image-based ML analysis for diagnosis of coronary artery disease.

Publication	Modality	Biomarker	ML technique	Diagnostic	Sample size	Performance
(20)	SPECT	Conventional	LB	CAD	1,181	AUC = 0.94
(16)	SPECT	Deep Learning	CNN	CAD	1,160	AUC = 0.81
(34)	SPECT	Conventional	ANN	CAD	1,365	AUC = 0.75
(35)	SPECT	Conventional	ANN	CAD	106	AUC = 0.96
(60)	SPECT	Conventional	ANN	CAD	65	AUC = 0.74
(97)	SPECT	Conventional	DT/GA	CAD	267	ACC = 0.83
(24)	SPECT	Qualitative	SVM	CAD	267	ACC = 0.92
(8)	SPECT	Deep Learning	ANN/CL	CAD	173	AUC = 0.80
(61)	SPECT	Conventional	ANN	CAD	109	AUC = 0.88
(6)	PET	Conventional	N/A	CAD/MACE	1,234	AUC = 0.72
(63)	CT	Conventional	N/A	CAD	352	AUC = 0.84
(62)	CT	Conventional	GBRT	CAD	252	AUC = 0.75
(58)	echo/SCI	Qualitative	ANN	CAD	327	ACC = 0.80
(98)	echo	Radiomics	SVM	CAD	61	AUC = 0.88
(25)	echo	Qualitative	SVM	CAD	228	ACC = 0.99

false positives and to inter-observer variability regarding CAC detection. Hence, the ability to quickly and reliably quantify calcification using ML models provides additive value to clinical risk scoring tools and will enable superior prognostication of individuals. To overcome these issues and bring robustness to such procedures, intensive cardiac imaging feature extraction may be utilized.

Işgum et al. (30) designed an automated method for detection of aortic calcification, an indicator of established atherosclerotic disease, based on shape and intensity features. Forty abdominal scans contained a total of 249 CAC determined by a human observer. The method detected 209 CAC (Se = 0.84) at the expense of 1.0 false-positive object per scan on average, while the presence of contrast increased the number of incorrect classifications. This work was complemented by Işgum et al. (64), analysing cardiac CT with a more sophisticated feature set to obtain a final accuracy of 0.74 for CAC detection. Feature selection showed that no shape features were included in the classification stage, highlighting the discriminating power of texture analysis in CT.

Wolterink et al. (65) used cardiac CT scans thresholded at 130 Hounsfield units and a connected-component analysis to obtain candidate regions in the coronary arteries for 164 subjects with expert annotations. Their texture analysis was similar to Işgum et al. (64), and the resulting accuracy with DTs was 0.86 for risk stratification. This work also introduced a guided review where the most uncertain CAC were manually inspected again, increasing the overall accuracy up to 0.92. Later, a large radiomic pool of 4,440 features was extracted from a group of 60 subjects with Napkin Ring Sign (NRS) and non-NRS plaques with similar degree of manually segmented CAC by Kolossváry et al. (66). This research unveils the value of radiomics to find discriminative features: almost half of them reached an AUC of 0.8, short- and long-run low gray-level emphasis and surface ratio of high attenuation voxels had the highest AUC values (0.92 and 0.89, respectively). Finally, in a

recent work, Zreik et al. (67) used recurrent CNNs in multi-planar reformatted coronary CTA images previously annotated by an expert, achieving accuracies of 0.77 and 0.8 for plaque and stenosis characterization, respectively. A summary of ATH studies can be found in **Table 9**.

3.5. Valvular Heart Disease

Heart valve disease is an increasingly common pathology of the cardiovascular system and an increasing number of patients are expected to require heart valve replacement. Such diverse group of disorders can benefit from cardiac imaging ML integration through early diagnosis, treatment or surgery planning. For instance, Elalfi et al. (68) used imaging preprocessing techniques (Gaussian and Gabor filtering) and intensity and texture features to generate an ANN model with 120 echo images. These images were organized in 8 types of valvular diseases. The obtained accuracy was high at 0.93. This is encouraging particularly considering the diversity of outcomes.

A similar approach was addressed for mitral regurgitation (MR) severity estimation using echo videos. Moghaddasi et al. (69) took advantage of binary patterns as image descriptors which include details from different viewpoints of the heart. kNN and SVM models were trained with 102 patients divided in four groups: mild MR ($n = 34$), moderate MR ($n = 32$), severe MR ($n = 36$), and control ($n = 37$). SVM obtained the best accuracy, 0.99. Another interesting work mentioned in previous sections was conducted by Wojnarski et al. (32). A summary of HVD studies can be found in **Table 10**.

3.6. Heart Failure

Heart failure with preserved ejection fraction (HFpEF) is a heterogeneous group of disorders with variable treatment response and poor outcomes. There has been increasing interest in improved phenotyping of HFpEF to aid understanding of underlying disease mechanisms and also to guide treatments toward subtypes who may derive benefit. Given

TABLE 9 | Selected studies using image-based ML analysis for diagnosis of aortic and coronary atherosclerosis.

Publication	Modality	Biomarker	ML technique	Diagnostic	Sample size	Performance
(66)	CT	Radiomics	N/A	ATH	60	AUC = 0.91
(67)	CT	Deep Learning	CNN	ATH	163	ACC = 0.80
(65)	CT	Radiomics	DT	ATH	164	ACC = 0.86
(17)	CT	Deep Learning	CNN	ATH	250	ACC = 0.72
(64)	CT	Conventional	CL	ATH	615	ACC = 0.74
(30)	CT	Conventional	CL	ATH	249	ACC = 0.83

TABLE 10 | Selected studies using image-based ML analysis for diagnosis of valvular heart disease.

Publication	Modality	Biomarker	ML technique	Diagnostic	Sample size	Performance
(68)	echo	Radiomics	ANN	HVD	120	ACC = 0.93
(69)	echo	Radiomics	SVM/CL	HVD	102	ACC = 0.99
(32)	CT	Conventional	CL	HVD	656	N/A

TABLE 11 | Selected studies using image-based ML analysis for diagnosis of heart failure.

Publication	Modality	Biomarker	ML technique	Diagnostic	Sample size	Performance
(9)	MRI	Conventional	SVM/RF	MI/HCM/HF	45	ACC = 0.77
(21)	MRI	Radiomics	LR	HF	79	AUC = 0.85
(70)	echo	Conventional	CL	HHD/HFePF	100	ACC = 0.81
(71)	echo	Conventional	CL/SVM	HFePF	397	AUC = 0.76
(72)	echo	Conventional	CL	HF	1,106	N/A

the heterogeneous nature of HFpEF, ML techniques are a very suitable tool for diagnosis and image phenotype stratification. Some of the reviewed studies in previous sections were also related to the characterization of heart failure (9, 70). Additional work in this field was presented by Shah et al. (71), that prospectively studied 397 HFpEF patients and performed detailed clinical, laboratory, electrocardiographic and echocardiographic phenotyping of the study participants. Clustering techniques were applied to divide the cohort into 3 pheno-groups. Phenomapping was helpful for improved classification and categorization of HFpEF patients and risk stratification by means of SVM, obtaining an AUC of 0.76. ML applied to HF phenogrouping is also used for prognostic tasks by Cikes et al. (72). A summary of HF studies can be found in **Table 11**.

3.7. Abnormal Wall Motion

Most of the existing quantitative techniques for wall motion characterization involve laborious post-processing and image analysis. For this reason, ML approaches with a minimum user input and a correlation with the segmental cardiac function can improve clinical routine and triage.

For instance, Mantilla et al. (7) detected wall motion abnormalities in the left ventricle by means of spatiotemporal profiles obtained with pseudo delineations of 20 MRI patients. Wavelet and Fourier transforms were applied and the subsequent spaces were used to generate two models: SVM and dictionary

learning (DICTL). Dictionary Learning at mid-cavity level obtained the best accuracy, 0.96 ($Sp = Se = 0.96$). Afshin et al. (73) exploited intensity distributions per segment. In their work, a reference frame automatically propagated to each cardiac phase generated the 16 segments for the whole cardiac cycle. LDA reduced feature dimensionality and linear SVM obtained an accuracy of 0.86 in a cohort of 58 MRI subjects.

Kusunose et al. (19) used a total of 300 patients with a history of myocardial infarction and 100 age-matched control patients. Each case contained echo from short-axis views at end-diastolic, mid-systolic, and end-systolic phases. An ensemble of 10 CNN models were trained. AUC obtained by the ML ensemble was similar to that produced by the cardiologists and sonographer readers (0.99 vs. 0.98, respectively), and the same occurred for territory detection (0.97 vs. 0.95, respectively). A summary of AWM studies can be found in **Table 12**.

4. DISCUSSION AND FUTURE PERSPECTIVES

Reflected by the large amount of already published data reviewed above, AI in general and ML in particular have been shown to exhibit a huge potential to significantly influence diagnostic decision making in cardiology. In contrast to “traditional” statistical methods, the techniques from the field of AI are able to deal with large amounts of data (“big data”) and to

TABLE 12 | Selected studies using image-based ML analysis for diagnosis of wall motion abnormalities.

Publication	Modality	Biomarker	ML technique	Diagnostic	Sample size	Performance
(7)	MRI	Conventional	SVM/DICTL	AWM	20	ACC = 0.96
(73)	MRI	Conventional	SVM	AWM	58	ACC = 0.86
(19)	echo	Deep Learning	CNN	AWM	400	AUC = 0.99

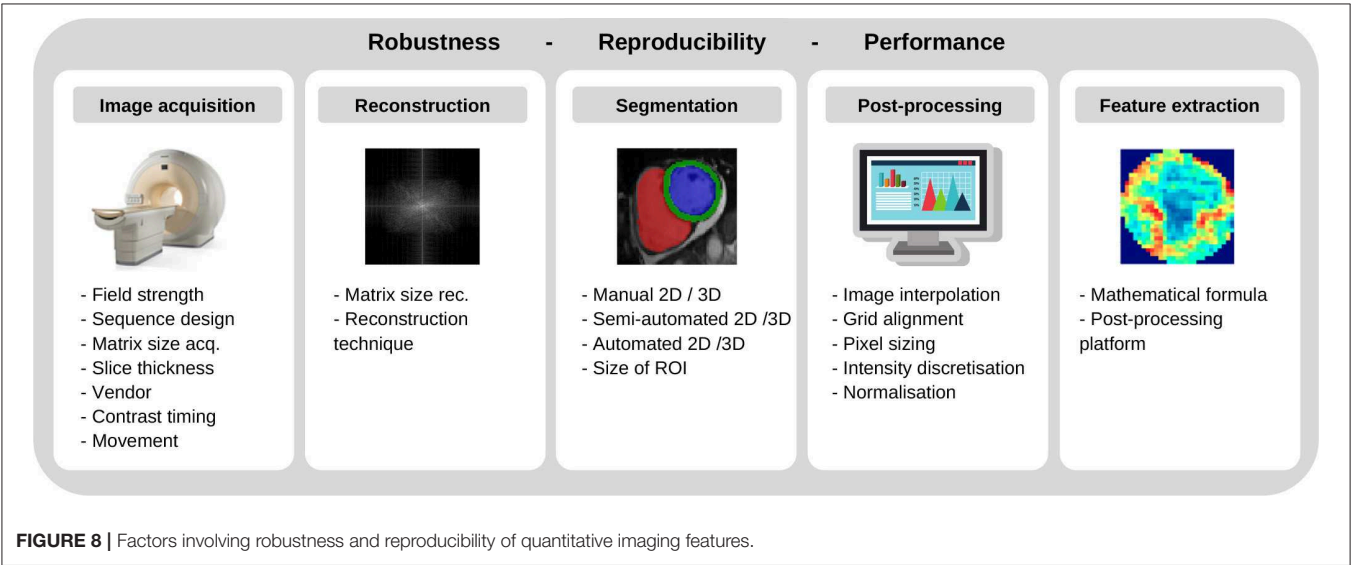


FIGURE 8 | Factors involving robustness and reproducibility of quantitative imaging features.

integrate information from all fields of clinical care, including e.g., clinical parameters (“clinomics”), genetic information (“genomics”), protein metabolism (“proteomics”), and imaging data (“radiomics”) within one large all-encompassing analysis framework. The steadily increasing computational power and the increasing availability of data through mobile applications and the digital transformation of the global healthcare systems further contribute to the advancement of the field. Consequently, future studies will continue the use of these techniques in order to allow translation into routine clinical practice and thus pave the way toward improved diagnostic decision making tailored to individual patient-specific needs (subsumed under the heading “precision medicine”).

Yet, in today’s clinical routine, diagnostic decisions are still drawn from stand-alone parameters [e.g., LV ejection fraction, (74)], despite many encouraging research studies from the field of AI. On a per-patient basis, the diagnostic and prognostic value of such independent functional parameters was found to be low, Park and Kim (75). Given the diversity of cardiovascular imaging modalities, their potential additive value for more accurate diagnostics and risk stratification remains unclear. Besides, continued reliance on subjective visual interpretation, has resulted in considerable observer-dependencies and lack of standardization. The application of AI and precision medicine to CVD, however, is currently still in its infancy, and faces huge challenges which have to be overcome by future research. To establish novel imaging biomarkers and AI techniques, the robustness and reproducibility of quantitative imaging features must be ensured, Zwanenburg et al. (76).

Up to now, trained models and algorithms have limited generalizability due to the multiplicity of potential influencing factors (including differing scanners, vendors, CT radiation doses, MRI field strengths, sequences, sequence parameters, spatial and temporal resolutions, reconstruction algorithms, reconstruction parameters, and so forth; **Figure 8**).

For CT and positron emission tomography (PET) imaging, a variety of studies have highlighted difficulties in producing reliably reproducible radiomic features when using different vendors, scanners, and acquisition or reconstruction settings (48, 77–84). While the “image biomarker standardization initiative” (IBSI) has established certain standards for radiomic studies, Zwanenburg et al. (76), the specific needs of cardiac imaging have not yet been met. For cardiac CT, Hinzpeter et al. and Mannil et al. have investigated the influence of slice thickness, Hinzpeter et al. (84), and iterative reconstruction algorithms, Mannil et al. (48), on the robustness and comparability of radiomics features – observing considerable feature variations for differing technical settings. In contrast to this evolving body of literature on CT imaging, little evidence exists concerning the robustness of radiomic features in MRI (75, 85–87). Given the qualitative nature of most MRI sequences and the absence of absolute signal intensities (in contrast to CT imaging for instance), the robustness of radiomic features seems to heavily depend on acquisition sequences as well as acquisition and reconstruction parameters. In a recent phantom study, Baeßler et al. sought to evaluate the influence of different acquisition sequences, spatial resolution, and postprocessing settings (88) revealing that the robustness of radiomic features was heavily

influenced by the acquisition sequence and image resolution as well as image processing settings. Future work not only needs to add to the understanding of such influencing factors but should also merge into extensive standardization efforts to ensure reliability of all imaging measures.

Several attempts to improve radiomic feature robustness through image normalization have been made. For more reliable quantification of emphysema, normalization was proposed for chest CT images reconstructed with different kernels, Gallardo-Estrella et al. (89). The proposed method decomposed each scan into multiple frequency bands, the energy of which was then normalized to the average energies observed in a set of scans reconstructed with a reference kernel. Building on these results, Jin et al. used a deep learning-based strategy for CT image normalization by means of a U-Net, Jin et al. (90). For harmonization of MRI images, similar deep learning algorithms were proposed for dynamic contrast enhanced (DCE) images in breast, Samala et al. (91), and brain MRI, Dewey et al. (92). Although yielding promising results, the applicability of such approaches in cardiovascular applications remains elusive, which is due to inherent particularities of cardiac imaging. Other than breast and brain, the human heart is steadily moving because of breathing and myocardial contraction. Second, the contrast bolus inside the ventricular lumen may influence the myocardial features. Aside from these specific characteristics, the impact of image normalization on extracted radiomic features has not been fully investigated yet. Besides lack of standardization of technical factors, the recent trend to train ML classifiers on relatively small datasets is a major issue of current methodology and hampers translation of the novel techniques into routine clinical practice. The small sample sizes in most cardiovascular imaging studies (usually $N < 100$ with $> 1,000$ variables in the models) lead to a considerable risk of overfitting. Overfitting leads to poor generalisability of the classification models when deployed to different datasets. Besides the current lack of imaging feature standardization and the problem of model-overfitting, other challenges should be acknowledged when it comes to translation of AI to daily patient care. While big data aims to integrate data from various sources, the current lack of interoperability of many systems used in clinical care poses huge obstacles for data pooling approaches. Several national and international attempts are currently under way to solve interoperability issues for medical care and to allow a seamless integration of different databases and informatic systems used in healthcare.

The ability to understand the rationale behind ML generated diagnostic grouping may be crucial in order to achieve widespread clinical use of this novel technology. However, especially with DL techniques, those are usually considered as being “black boxes,” which do not deliver any insights or explanations on how they reached their conclusions and upon which, e.g., imaging features, they based their decision. Although several attempts and ongoing research exist on delivering insights into an algorithm’s decision making (such as heatmaps), these attempts are not sufficiently elaborated so far to convince most cardiology practitioners to use a diagnostic black box in daily clinical patient management.

Thus, interpretability of DL models including the psychological aspects of digital transformation itself should represent one major aim of future research. Radiomics might represent a valid alternative for the meantime, since radiomic models—in cases where an appropriate and stepwise feature reduction is performed before training the ML algorithm—deliver more insights into the specific imaging features which were important for the model’s classification performance. In summary, solutions achieving better standardization or normalization resulting in better generalisability are an important condition to bring radiomics and AI into cardiac precision medicine with concomitant improved diagnostic approaches to CVDs. In addition, better interoperability of healthcare informatics systems should be achieved. Finally, the steadfast progression of AI approaches to clinical decision making represent an abrupt change from conventional medical reasoning, as such, it is essential to engage with the psychological impact of the ongoing digital transformation in order to facilitate the transition of medical practice in line with advancing technologies. The extensive and encouraging work reviewed in this article above pursues one common goal for the future of cardiovascular medicine: to pave the way toward better diagnosis and precision medicine in cardiology. The application of AI to cardiology holds the promise to revolutionize individual disease monitoring and treatment (93), thus overcoming the currently used “one size fits all” approach derived from large clinical studies.

AUTHOR CONTRIBUTIONS

CM-I performed the literature search and its categorization, developed statistics, wrote the diagnostics literature section, and coordinated the remaining sections. VC wrote the machine learning section and contributed to the remaining sections. CI wrote the data section and designed the figures. KL supervised the manuscript development, wrote the abstract and the introduction, and reviewed all sections. ZR-E, BB, and SP contributed to drafting and critical appraisal of the manuscript. All authors reviewed the manuscript.

FUNDING

This work was partly funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement No 825903 (euCanSHare project). SP acts as a paid consultant to Circle Cardiovascular Imaging Inc., Calgary, Canada and Servier. SP acknowledges support from the National Institute for Health Research (NIHR) Cardiovascular Biomedical Research Centre at Barts, from the SmartHeart EPSRC programme grant (EP/P001009/1) and the London Medical Imaging and AI Centre for Value-Based Healthcare. SP and KL acknowledge support from the CAP-AI programme, London’s first AI enabling programme focused on stimulating growth in the capital’s AI sector. ZR-E was supported by a British Heart Foundation Clinical Research Training Fellowship (FS/17/81/33318).

REFERENCES

- Wilkins E, Wilson L, Wickramasinghe K, Bhatnagar P, Leal J, Luengo-Fernandez R, et al. *European Cardiovascular Disease Statistics 2017*. Brussels: European Heart Network (2017).
- Ritchie H, Roser M. Our world in data. In: *Causes of Death* (2018). Retrieved from: <https://ourworldindata.org/causes-of-death>.
- Leiner T, Rueckert D, Suinesiaputra A, Baeßler B, Nezafat R, Išgum I, et al. Machine learning in cardiovascular magnetic resonance: basic concepts and applications. *J Cardiovasc Magn Resonan.* (2019) 21:61. doi: 10.1186/s12968-019-0575-y
- Khened M, Alex V, Krishnamurthi G. Densely connected fully convolutional network for short-axis cardiac cine MR image segmentation and heart diagnosis using random forest. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Quebec City, QC: Springer (2017). p. 140–51.
- Chen R, Lu A, Wang J, Ma X, Zhao L, Wu W, et al. Using machine learning to predict one-year cardiovascular events in patients with severe dilated cardiomyopathy. *Eur J Radiol.* (2019). 117:178–83. doi: 10.1016/j.ejrad.2019.06.004
- Juarez-Orozco LE, Knol RJ, Sanchez-Catasus CA, Martinez-Manzanera O, van der Zant FM, Knuuti J. Machine learning in the integration of simple variables for identifying patients with myocardial ischemia. *J Nucl Cardiol.* (2018). doi: 10.1007/s12350-018-1304-x. [Epub ahead of print].
- Mantilla J, Garreau M, Bellanger JJ, Paredes JL. Machine learning techniques for LV wall motion classification based on spatio-temporal profiles from cardiac cine MRI. In: *2013 12th International Conference on Machine Learning and Applications*. Vol. 1. IEEE (2013). p. 167–72.
- Bagher-Ebadian H, Soltanian-Zadeh H, Setayeshi S, Smith ST. Neural network and fuzzy clustering approach for automatic diagnosis of coronary artery disease in nuclear medicine. *IEEE Trans Nucl Sci.* (2004) 51:184–92. doi: 10.1109/TNS.2003.823047
- Moreno A, Rodriguez J, Martinez F. Regional multiscale motion representation for cardiac disease prediction. In: *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*. Bucaramanga, CO: IEEE (2019). p. 1–5.
- Zheng Q, Delingette H, Ayache N. Explainable cardiac pathology classification on cine MRI with motion characterization by semi-supervised learning of apparent flow. *Med Image Anal.* (2019) 56:80–95. doi: 10.1016/j.media.2019.06.001
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
- Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magn Reson Imaging.* (2012) 30:1234–48. doi: 10.1016/j.mri.2012.06.010
- Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* (2014) 5:4006. doi: 10.1038/ncomms5644
- Cetin I, Sanroma G, Petersen SE, Napel S, Camara O, Ballester MAG, et al. A radiomics approach to computer-aided diagnosis with cardiac cine-MRI. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Quebec City, QC: Springer (2017). p. 82–90.
- 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 Cardiovasc Imaging.* (2019) 12:1946–1954. doi: 10.1016/j.jcmg.2018.11.024
- Betancur J, Hu LH, Commandeur F, Sharir T, Einstein AJ, Fish MB, et al. Deep learning analysis of upright-supine high-efficiency spect myocardial perfusion imaging for prediction of obstructive coronary artery disease: a multicenter study. *J Nucl Med.* (2019) 60:664–70. doi: 10.2967/jnumed.118.213538
- Wolterink JM, Leiner T, de Vos BD, van Hamersvelt RW, Viergever MA, Išgum I. Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Med Image Anal.* (2016) 34:123–36. doi: 10.1016/j.media.2016.04.004
- Lu A, Dehghan E, Veni G, Moradi M, Syeda-Mahmood T. Detecting anomalies from echocardiography using multi-view regression of clinical measurements. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE (2018). p. 1504–8.
- Kusunose K, Abe T, Haga A, Fukuda D, Yamada H, Harada M, et al. A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images. *JACC Cardiovasc Imaging.* (2019). [Epub ahead of print].
- Arsanjani R, Xu Y, Dey D, Vahistha V, Shalev A, Nakanishi R, et al. Improved accuracy of myocardial perfusion SPECT for detection of coronary artery disease by machine learning in a large population. *J Nucl Cardiol.* (2013) 20:553–62. doi: 10.1007/s12350-013-9706-2
- Baeßler B, Luecke C, Lurz J, Klingel K, Das A, von Roeder M, et al. Cardiac MRI and texture analysis of myocardial T1 and T2 maps in myocarditis with acute versus chronic symptoms of heart failure. *Radiology.* (2019) 292:608–17. doi: 10.1148/radiol.2019190101
- Conforti D, Guido R. Kernel-based support vector machine classifiers for early detection of myocardial infarction. *Optimizat Methods Softw.* (2005) 20:401–13. doi: 10.1080/10556780512331318164
- Arsanjani R, Xu Y, Dey D, Fish M, Dorbala S, Hayes S, et al. Improved accuracy of myocardial perfusion SPECT for the detection of coronary artery disease using a support vector machine algorithm. *J Nucl Med.* (2013) 54:549–55. doi: 10.2967/jnumed.112.111542
- Ciecholewski M. Support vector machine approach to cardiac SPECT diagnosis. In: *International Workshop on Combinatorial Image Analysis*. Madrid: Springer (2011). p. 432–43.
- Berikol GB, Yildiz O, Özcan IT. Diagnosis of acute coronary syndrome with a support vector machine. *J Med Syst.* (2016) 40:84. doi: 10.1007/s10916-016-0432-6
- Borkar S, Annadate M. Supervised machine learning algorithm for detection of cardiac disorders. In: *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*. Pune: IEEE (2018). p. 1–4.
- Wong KC, Tee M, Chen M, Bluemke DA, Summers RM, Yao J. Regional infarction identification from cardiac CT images: a computer-aided biomechanical approach. *Int J Comput Assist Radiol Surg.* (2016) 11:1573–83. doi: 10.1007/s11548-016-1404-5
- Baeßler B, Mannil M, Maintz D, Alkadhi H, Manka R. Texture analysis and machine learning of non-contrast T1-weighted MR images in patients with hypertrophic cardiomyopathy—preliminary results. *Eur J Radiol.* (2018) 102:61–7. doi: 10.1016/j.ejrad.2018.03.013
- Tabassian M, Alessandrini M, Herbots L, Mirea O, Pagourelas ED, Jasaityte R, et al. Machine learning of the spatio-temporal characteristics of echocardiographic deformation curves for infarct classification. *Int J Cardiovasc Imaging.* (2017) 33:1159–67. doi: 10.1007/s10554-017-1108-0
- Išgum I, van Ginneken B, Olree M. Automatic detection of calcifications in the aorta from CT scans of the abdomen1: 3D computer-aided diagnosis. *Acad Radiol.* (2004) 11:247–57. doi: 10.1016/S1076-6332(03)00673-1
- Bruse JL, Zuluaga MA, Khushnood A, McLeod K, Ntsinjana HN, Hsia TY, et al. Detecting clinically meaningful shape clusters in medical image data: metrics analysis for hierarchical clustering applied to healthy and pathological aortic arches. *IEEE Trans Biomed Eng.* (2017) 64:2373–83. doi: 10.1109/TBME.2017.2655364
- Wojnarski CM, Roselli EE, Idrees JJ, Zhu Y, Carnes TA, Lowry AM, et al. Machine-learning phenotypic classification of bicuspid aortopathy. *J Thorac Cardiovasc Surg.* (2018) 155:461–9. doi: 10.1016/j.jtcvs.2017.08.123
- Tsai DY, Lee Y, Sekiya M, Ohkubo M. Medical image classification using genetic-algorithm based fuzzy-logic approach. *J Electr Imaging.* (2004) 13:780–9. doi: 10.1117/1.1786607
- Nakajima K, Kudo T, Nakata T, Kiso K, Kasai T, Taniguchi Y, et al. Diagnostic accuracy of an artificial neural network compared with statistical quantitation of myocardial perfusion images: a Japanese multicenter study. *Eur J Nucl Med Mol Imaging.* (2017) 44:2280–9. doi: 10.1007/s00259-017-3834-x

35. Nakajima K, Okuda K, Watanabe S, Matsuo S, Kinuya S, Toth K, et al. Artificial neural network retrained to detect myocardial ischemia using a Japanese multicenter database. *Ann Nucl Med.* (2018) 32:303–10. doi: 10.1007/s12149-018-1247-y
36. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, et al. Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. *Circulation.* (2018) 138:1623–35. doi: 10.1161/CIRCULATIONAHA.118.034338
37. Madani A, Ong JR, Tibrewal A, Mofrad MR. Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *npj Digit Med.* (2018) 1:59. doi: 10.1038/s41746-018-0065-x
38. Cetin I, Petersen SE, Napel S, Camara O, Ballester MÁG, Lekadir K. A radiomics approach to analyze cardiac alterations in hypertension. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. Venice: IEEE (2019). p. 640–3.
39. Nakada A, Shiozaki A, Hirano Y, Uehara H, Masuyama T. Using neural networks in color kinesis image process to automate diagnosis of cardiac disease. *Electr Commun Jpn (Part II: Electr).* (2006) 89:46–54. doi: 10.1002/ecjb.20245
40. Ungur K, Tenbrinck D, Jiang X, Stypmann J. Automatic classification of left ventricular wall segments in small animal ultrasound imaging. *Comput Methods Progr Biomed.* (2014) 117:2–12. doi: 10.1016/j.cmpb.2014.06.015
41. Agani N, Abu-Bakar S, Salleh SS. Application of texture analysis in echocardiography images for myocardial infarction tissue. *Jurnal Teknologi.* (2007) 46:61–76. doi: 10.11113/jt.v46.295
42. Sudarshan VK, Acharya UR, Ng E, San Tan R, Chou SM, Ghista DN. An integrated index for automated detection of infarcted myocardium from cross-sectional echocardiograms using texton-based features (Part 1). *Comput Biol Med.* (2016) 71:231–40. doi: 10.1016/j.compbiomed.2016.01.028
43. Vidya KS, Ng E, Acharya UR, Chou SM, San Tan R, Ghista DN. Computer-aided diagnosis of myocardial infarction using ultrasound images with DWT, GLCM and HOS methods: a comparative study. *Comput Biol Med.* (2015) 62:86–93. doi: 10.1016/j.compbiomed.2015.03.033
44. Baessler B, Mannil M, Oebel S, Maintz D, Alkadhi H, Manka R. Subacute and chronic left ventricular myocardial scar: accuracy of texture analysis on nonenhanced cine MR images. *Radiology.* (2017) 286:103–12. doi: 10.1148/radiol.2017170213
45. Larroza A, López-Lereu MP, Monmeneu JV, Gavara J, Chorro FJ, Bodí V, et al. Texture analysis of cardiac cine magnetic resonance imaging to detect nonviable segments in patients with chronic myocardial infarction. *Med Phys.* (2018) 45:1471–80. doi: 10.1002/mp.12783
46. Zhang N, Yang G, Gao Z, Xu C, Zhang Y, Shi R, et al. Deep learning for diagnosis of chronic myocardial infarction on nonenhanced cardiac Cine MRI. *Radiology.* (2019) 291:606–17. doi: 10.1148/radiol.2019182304
47. Mannil M, von Spiczak J, Manka R, Alkadhi H. Texture analysis and machine learning for detecting myocardial infarction in noncontrast low-dose computed tomography: unveiling the invisible. *Investigat Radiol.* (2018) 53:338–43. doi: 10.1097/RLI.0000000000000448
48. Mannil M, von Spiczak J, Muehlethaler UJ, Thanabalasingam A, Keller DI, Manka R, et al. Texture analysis of myocardial infarction in CT: comparison with visual analysis and impact of iterative reconstruction. *Eur J Radiol.* (2019) 113:245–50. doi: 10.1016/j.ejrad.2019.02.037
49. Gopalakrishnan V, Menon PG, Madan S. cMRI-BED: a novel informatics framework for cardiac MRI biomarker extraction and discovery applied to pediatric cardiomyopathy classification. *Biomed Eng Online.* (2015) 14:S7. doi: 10.1186/1475-925X-14-S2-S7
50. Narula S, Shameer K, Omar AMS, Dudley JT, Sengupta PP. Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography. *J Am Coll Cardiol.* (2016) 68:2287–95. doi: 10.1016/j.jacc.2016.08.062
51. Bernard O, Lalande A, Zotti C, Cervensky F, Yang X, Heng PA, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE Trans Med Imaging.* (2018) 37:2514–25. doi: 10.1109/TMI.2018.2837502
52. Wolterink JM, Leiner T, Viergever MA, Išgum I. Automatic segmentation and disease classification using cardiac cine MR images. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Quebec City, QC: Springer (2017). p. 101–10.
53. Isensee F, Jaeger PF, Full PM, Wolf I, Engelhardt S, Maier-Hein KH. Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Quebec City, QC: Springer (2017). p. 120–9.
54. Snaauw G, Gong D, Maicas G, van den Hengel A, Niessen WJ, Verjans J, et al. End-to-end diagnosis and segmentation learning from cardiac magnetic resonance imaging. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. Venice: IEEE (2019). p. 802–5.
55. Biffi C, Oktay O, Tarroni G, Bai W, De Marvao A, Doumou G, et al. Learning interpretable anatomical features through deep generative models: application to cardiac remodeling. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Granada: Springer (2018). p. 464–71.
56. Puylol-Antón E, Ruijsink B, Gerber B, Amzulescu MS, Langet H, De Craene M, et al. Regional multi-view learning for cardiac motion analysis: application to identification of dilated cardiomyopathy patients. *IEEE Trans Biomed Eng.* (2018) 66:956–66. doi: 10.1109/TBME.2018.2865669
57. Neisius U, Myerson L, Fahmy AS, Nakamori S, El-Rewaidy H, Joshi G, et al. Cardiovascular magnetic resonance feature tracking strain analysis for discrimination between hypertensive heart disease and hypertrophic cardiomyopathy. *PLoS ONE.* (2019) 14:e0221061. doi: 10.1371/journal.pone.0221061
58. Kukar M, Kononenko I, Grošelj C, Kralj K, Feticch J. Analysing and improving the diagnosis of ischaemic heart disease with machine learning. *Artif Intell Med.* (1999) 16:25–50. doi: 10.1016/S0933-3657(98)00063-3
59. Kurgan LA, Cios KJ, Tadeusiewicz R, Ogiela M, Goodenday LS. Knowledge discovery approach to automated cardiac SPECT diagnosis. *Artif Intell Med.* (2001) 23:149–69. doi: 10.1016/S0933-3657(01)00082-3
60. Guner LA, Karabacak NI, Akdemir OU, Karagoz PS, Kocaman SA, Cengel A, et al. An open-source framework of neural networks for diagnosis of coronary artery disease from myocardial perfusion SPECT. *J Nucl Cardiol.* (2010) 17:405–13. doi: 10.1007/s12350-010-9207-5
61. Shibutani T, Nakajima K, Wakabayashi H, Mori H, Matsuo S, Yoneyama H, et al. Accuracy of an artificial neural network for detecting a regional abnormality in myocardial perfusion SPECT. *Ann Nucl Med.* (2019) 33:86–92. doi: 10.1007/s12149-018-1306-4
62. Han D, Lee JH, Rizvi A, Gransar H, Baskaran L, Schulman-Marcus J, et al. Incremental role of resting myocardial computed tomography perfusion for predicting physiologically significant coronary artery disease: a machine learning approach. *J Nucl Cardiol.* (2018) 25:223–33. doi: 10.1007/s12350-017-0834-y
63. Coenen A, Kim YH, Kruk M, Tesche C, De Geer J, Kurata A, et al. Diagnostic accuracy of a machine-learning approach to coronary computed tomographic angiography-based fractional flow reserve: result from the MACHINE consortium. *Circulation.* (2018) 11:e007217. doi: 10.1161/CIRCIMAGING.117.007217
64. Išgum I, Rutten A, Prokop M, van Ginneken B. Detection of coronary calcifications from computed tomography scans for automated risk assessment of coronary artery disease. *Med Phys.* (2007) 34:1450–61. doi: 10.1118/1.2710548
65. Wolterink JM, Leiner T, Takx RA, Viergever MA, Išgum I. An automatic machine learning system for coronary calcium scoring in clinical non-contrast enhanced, ECG-triggered cardiac CT. In: *Medical Imaging 2014: Computer-Aided Diagnosis*, Vol. 9035. San Diego, CA: International Society for Optics and Photonics (2014). p. 90350E. doi: 10.1117/12.2042226
66. Kolossváry M, Karády J, Szilveszter B, Kitslaar P, Hoffmann U, Merkely B, et al. Radiomic features are superior to conventional quantitative computed tomographic metrics to identify coronary plaques with napkin-ring sign. *Circulation.* (2017) 136:e006843. doi: 10.1161/CIRCIMAGING.117.006843
67. Zreik M, van Hamersvelt RW, Wolterink JM, Leiner T, Viergever MA, Išgum I. A recurrent CNN for automatic detection and classification of coronary artery

- plaque and stenosis in coronary CT angiography. *IEEE Trans Med Imaging*. (2018) 38:1588–1598. doi: 10.1109/TMI.2018.2883807
68. Elalfi A, Eisa M, Ahmed H. Artificial neural networks in medical images for diagnosis heart valve diseases. *Int J Comput Sci Issues*. (2013) 10:83. Available online at: <https://www.ijcsi.org>
 69. Moghaddasi H, Nourian S, Nourian S. Automatic assessment of mitral regurgitation severity based on extensive textural features on 2D echocardiography videos. *Comput Biol Med*. (2016) 73:47–55. doi: 10.1016/j.combiomed.2016.03.026
 70. Tabassian M, Sunderji I, Erdei T, Sanchez-Martinez S, Degiovanni A, Marino P, et al. Diagnosis of heart failure with preserved ejection fraction: machine learning of spatiotemporal variations in left ventricular deformation. *J Am Soc Echocardiogr*. (2018) 31:1272–84. doi: 10.1016/j.echo.2018.07.013
 71. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation*. (2015) 131:269–79. doi: 10.1161/CIRCULATIONAHA.114.010637
 72. Cikes M, Sanchez-Martinez S, Claggett B, Duchateau N, Piella G, Butakoff C, et al. Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. *Eur J Heart Fail*. (2019) 21:74–85. doi: 10.1002/ejhf.1333
 73. Afshin M, Ayed IB, Punithakumar K, Law M, Islam A, Goela A, et al. Regional assessment of cardiac left ventricular myocardial function via MRI statistical features. *IEEE Trans Med Imaging*. (2013) 33:481–94. doi: 10.1109/TMI.2013.2287793
 74. Ibáñez B, James S, Agewall S, Antunes MJ, Bucciarelli-Ducci C, Bueno H, et al. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation. *Revista Espanola de Cardiologia (English ed)*. (2017) 70:1082. doi: 10.1016/j.rec.2017.11.010
 75. Park JE, Kim HS. Radiomics as a quantitative imaging biomarker: practical considerations and the current standpoint in neuro-oncologic studies. *Nucl Med Mol Imaging*. (2018) 52:99–108. doi: 10.1007/s13139-017-0512-7
 76. Zwanenburg A, Leger S, Vallières M, Lööck S. Image biomarker standardisation initiative. *arXiv:161207003*. (2016).
 77. Balagurunathan Y, Kumar V, Gu Y, Kim J, Wang H, Liu Y, et al. Test–retest reproducibility analysis of lung CT image features. *J Digit Imaging*. (2014) 27:805–23. doi: 10.1007/s10278-014-9716-x
 78. Lubner MG, Smith AD, Sandrasegaran K, Sahani DV, Pickhardt PJ. CT texture analysis: definitions, applications, biologic correlates, and challenges. *Radiographics*. (2017) 37:1483–503. doi: 10.1148/rg.2017170056
 79. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring CT scanner variability of radiomics features. *Investigat Radiol*. (2015) 50:757. doi: 10.1097/RLI.0000000000000180
 80. Nyflot MJ, Yang F, Byrd D, Bowen SR, Sandison GA, Kinahan PE. Quantitative radiomics: impact of stochastic effects on textural feature analysis implies the need for standards. *J Med Imaging*. (2015) 2:041002. doi: 10.1117/1.JMI.2.4.041002
 81. Shafiq-ul Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys*. (2017) 44:1050–62. doi: 10.1002/mp.12123
 82. Zhao B, Tan Y, Tsai WY, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep*. (2016) 6:23428. doi: 10.1038/srep23428
 83. Zhao B, Tan Y, Tsai WY, Schwartz LH, Lu L. Exploring variability in CT characterization of tumors: a preliminary phantom study. *Transl oncol*. (2014) 7:88. doi: 10.1593/tlo.13865
 84. Hinzpeter R, Wagner MW, Wurnig MC, Seifert B, Manka R, Alkadhi H. Texture analysis of acute myocardial infarction with CT: first experience study. *PLoS ONE*. (2017) 12:e0186876. doi: 10.1371/journal.pone.0186876
 85. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn Reson Imaging*. (2004) 22:81–91. doi: 10.1016/j.mri.2003.09.001
 86. Mayerhoefer ME, Szomolanyi P, Jirak D, Berg A, Materka A, Dirisamer A, et al. Effects of magnetic resonance image interpolation on the results of texture-based pattern classification: a phantom study. *Investigat Radiol*. (2009) 44:405–11. doi: 10.1097/RLI.0b013e3181a50a66
 87. Saha A, Harowicz MR, Mazurowski MA. Breast cancer MRI radiomics: an overview of algorithmic features and impact of inter-reader variability in annotating tumors. *Med Phys*. (2018) 45:3076–85. doi: 10.1002/mp.12925
 88. Baeßler B, Weiss K, dos Santos DP. Robustness and reproducibility of radiomics in magnetic resonance imaging: a phantom study. *Investigat Radiol*. (2019) 54:221–8. doi: 10.1097/RLI.0000000000000530
 89. Gallardo-Estrella L, Lynch DA, Prokop M, Stinson D, Zach J, Judy PF, et al. Normalizing computed tomography data reconstructed with different filter kernels: effect on emphysema quantification. *Eur Radiol*. (2016) 26:478–86. doi: 10.1007/s00330-015-3824-y
 90. Jin H, Heo C, Kim JH. Deep learning-enabled accurate normalization of reconstruction kernel effects on emphysema quantification in low-dose CT. *Phys Med Biol*. (2019) 64:135010. doi: 10.1088/1361-6560/ab28a1
 91. Samala RK, Chan HP, Hadjiiski L, Paramagul C, Helvie MA, Neal CH. Homogenization of breast MRI across imaging centers and feature analysis using unsupervised deep embedding. In: *Medical Imaging 2019: Computer-Aided Diagnosis*, Vol. 10950. San Diego, CA: International Society for Optics and Photonics (2019). p. 1095020. doi: 10.1117/12.2512209
 92. Dewey BE, Zhao C, Reinhold JC, Carass A, Fitzgerald KC, Sotirchos ES, et al. DeepHarmony: a deep learning approach to contrast harmonization across scanner changes. *Magn Reson Imaging*. (2019) 64:160–70. doi: 10.1016/j.mri.2019.05.041
 93. Leopold JA, Loscalzo J. Emerging role of precision medicine in cardiovascular disease. *Circulat Res*. (2018) 122:1302–15. doi: 10.1161/CIRCRESAHA.117.310782
 94. Rohé MM, Duchateau N, Serresant M, Pennec X. Combination of polyaffine transformations and supervised learning for the automatic diagnosis of LV infarct. In: *Statistical Atlases and Computational Models of the Heart*. Munich: Springer (2015). p. 190–8.
 95. Lekadir K, Albà X, Pereañez M, Frangi AF. Statistical shape modeling using partial least squares: application to the assessment of myocardial infarction. In: *Statistical Atlases and Computational Models of the Heart*. Munich: Springer (2015). p. 130–9.
 96. Sacha JP, Goodenday LS, Cios KJ. Bayesian learning for cardiac SPECT image interpretation. *Artif Intell Med*. (2002) 26:109–43. doi: 10.1016/S0933-3657(02)00055-6
 97. To C, Pham TD. Analysis of cardiac imaging data using decision tree based parallel genetic programming. In: *2009 Proceedings of 6th International Symposium on Image and Signal Processing and Analysis*. Salzburg: IEEE (2009). p. 317–20.
 98. Zhang L, Wahle A, Chen Z, Lopez JJ, Kovarnik T, Sonka M. Predicting locations of high-risk plaques in coronary arteries in patients receiving statin therapy. *IEEE Trans Med Imaging*. (2017) 37:151–61. doi: 10.1109/TMI.2017.2725443

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Martin-Isla, Campello, Izquierdo, Raisi-Estabragh, Baeßler, Petersen and Lekadir. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



From Compressed-Sensing to Artificial Intelligence-Based Cardiac MRI Reconstruction

Aurélien Bustin^{1†}, Niccolo Fuin^{1*†}, René M. Botnar^{1,2} and Claudia Prieto^{1,2}

¹ Department of Biomedical Engineering, School of Biomedical Engineering and Imaging Sciences, King's College London, London, United Kingdom, ² Escuela de Ingeniería, Pontificia Universidad Católica de Chile, Santiago, Chile

OPEN ACCESS

Edited by:

Steffen Erhard Petersen,
Queen Mary University of London,
United Kingdom

Reviewed by:

Reza Nezafat,
Harvard University, United States
Michael Jerosch-Herold,
Harvard Medical School,
United States

Daniel K. Sodickson,
New York University, United States

*Correspondence:

Niccolo Fuin
niccolo.fuin@kcl.ac.uk

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Cardiovascular Imaging,
a section of the journal
Frontiers in Cardiovascular Medicine

Received: 30 September 2019

Accepted: 31 January 2020

Published: 25 February 2020

Citation:

Bustin A, Fuin N, Botnar RM and
Prieto C (2020) From
Compressed-Sensing to Artificial
Intelligence-Based Cardiac MRI
Reconstruction.
Front. Cardiovasc. Med. 7:17.
doi: 10.3389/fcvm.2020.00017

Cardiac magnetic resonance (CMR) imaging is an important tool for the non-invasive assessment of cardiovascular disease. However, CMR suffers from long acquisition times due to the need of obtaining images with high temporal and spatial resolution, different contrasts, and/or whole-heart coverage. In addition, both cardiac and respiratory-induced motion of the heart during the acquisition need to be accounted for, further increasing the scan time. Several undersampling reconstruction techniques have been proposed during the last decades to speed up CMR acquisition. These techniques rely on acquiring less data than needed and estimating the non-acquired data exploiting some sort of prior information. Parallel imaging and compressed sensing undersampling reconstruction techniques have revolutionized the field, enabling 2- to 3-fold scan time accelerations to become standard in clinical practice. Recent scientific advances in CMR reconstruction hinge on the thriving field of artificial intelligence. Machine learning reconstruction approaches have been recently proposed to learn the non-linear optimization process employed in CMR reconstruction. Unlike analytical methods for which the reconstruction problem is explicitly defined into the optimization process, machine learning techniques make use of large data sets to learn the key reconstruction parameters and priors. In particular, deep learning techniques promise to use deep neural networks (DNN) to learn the reconstruction process from existing datasets in advance, providing a fast and efficient reconstruction that can be applied to all newly acquired data. However, before machine learning and DNN can realize their full potentials and enter widespread clinical routine for CMR image reconstruction, there are several technical hurdles that need to be addressed. In this article, we provide an overview of the recent developments in the area of artificial intelligence for CMR image reconstruction. The underlying assumptions of established techniques such as compressed sensing and low-rank reconstruction are briefly summarized, while a greater focus is given to recent advances in dictionary learning and deep learning based CMR reconstruction. In particular, approaches that exploit neural networks as implicit or explicit priors are discussed for 2D dynamic cardiac imaging and 3D whole-heart CMR imaging. Current limitations, challenges, and potential future directions of these techniques are also discussed.

Keywords: cardiac MRI, AI, reconstruction, dictionary learning, deep learning, undersampling

INTRODUCTION

Magnetic resonance imaging (MRI) is a valuable tool for the non-invasive assessment of cardiovascular disease. Cardiac MR (CMR) imaging has been established as a clinically important technique for the assessment of cardiac morphology, function, perfusion, viability, and more recently quantitative myocardial tissue characterization (1–3). CMR is currently used to diagnose congenital heart disease (CHD), ischemic heart disease, valvular heart disease, pericardial lesions, cardiac tumors and cardiomyopathies, among others (4, 5). However, CMR suffers from long acquisition times due to the need of obtaining images with high temporal and spatial resolution, different contrasts, and/or whole-heart coverage. In addition, both cardiac and respiratory-induced motion of the heart during the acquisition need to be accounted for, further increasing the scan time.

Several technical advances have been proposed during the last decades to improve CMR, including the development of efficient pulse sequences to speed up the scan and improve the contrast of the images, the development of motion compensation techniques to account for the respiratory and cardiac induced movement of the heart, the use of multiple radio-frequency receiver coils for parallel imaging (PI), and the development of undersampled reconstruction techniques to acquire less data than needed (in the Nyquist sense) and thus accelerate the acquisition. PI allows to decrease the scan time by reducing the number of phase increment steps (undersampling) and exploiting the sensitivity encoding of the multiple receiver coils to recover the non-acquired data. PI has been widely integrated into commercial MR systems and is routinely used in clinical practice. Undersampled reconstruction techniques such as compressed sensing (CS) have been also employed to accelerate CMR imaging. CS works under the assumption that the k-space data is randomly undersampled, the image has a sparse representation in some pre-defined basis or dictionary, and a non-linear reconstruction is performed to enforce the sparsity of the image and consistency with the acquired MR data. In practice, CS-based reconstruction techniques employ pseudo-random trajectories (usually with variable density) along with one or several (e.g., spatial and temporal dimensions) sparse transforms such as finite differences (e.g., total variation) or wavelets operators. Early 2017, the U.S. Food and Drug administration (FDA) cleared the CS technology to enable the fast acquisition of CMR images, thus officially opening the door to the broader clinical use of this technique (6–8).

Recent efforts have been made to further improve CS-based reconstruction quality by learning dictionary-based representations of the sparse domain from the acquired data itself (or jointly during reconstruction) instead of exploiting known analytical transform domains. However, CS-based reconstruction techniques usually suffer from long computational times and their performance depends on the choice of the sparsity representation and the tuning of the corresponding reconstruction parameters. More recently deep neural networks (DNN) have been proposed to overcome these challenges by learning optimal reconstruction parameters

and/or transforms from the data itself and enabling extremely fast computational times (after training), promising to further advance the field of CMR reconstruction.

In this review paper, we first briefly discuss the CS and dictionary learning models, which offer a framework for sparse signal recovery and low-dimensional signal models and serve as a background for the following section. Recent representative advances in deep learning (DL) for CMR reconstruction are next discussed, highlighting theoretical developments and cardiac applications.

TRANSFORM AND DATA-DRIVEN CMR RECONSTRUCTION

This section briefly introduces the key concepts that underlie MR image reconstruction as an inverse problem, that will serve as background material to the rest of the review. CS-based and dictionary learning models for CMR reconstruction are also discussed. We refer the reader to Ye (9) and Jaspán et al. (10) for further discussion on the application of CS to MR image reconstruction.

MR Reconstruction as an Inverse Problem

The general (discretized) principles of MR signal generation and image formation can be expressed as a system of linear equations (11):

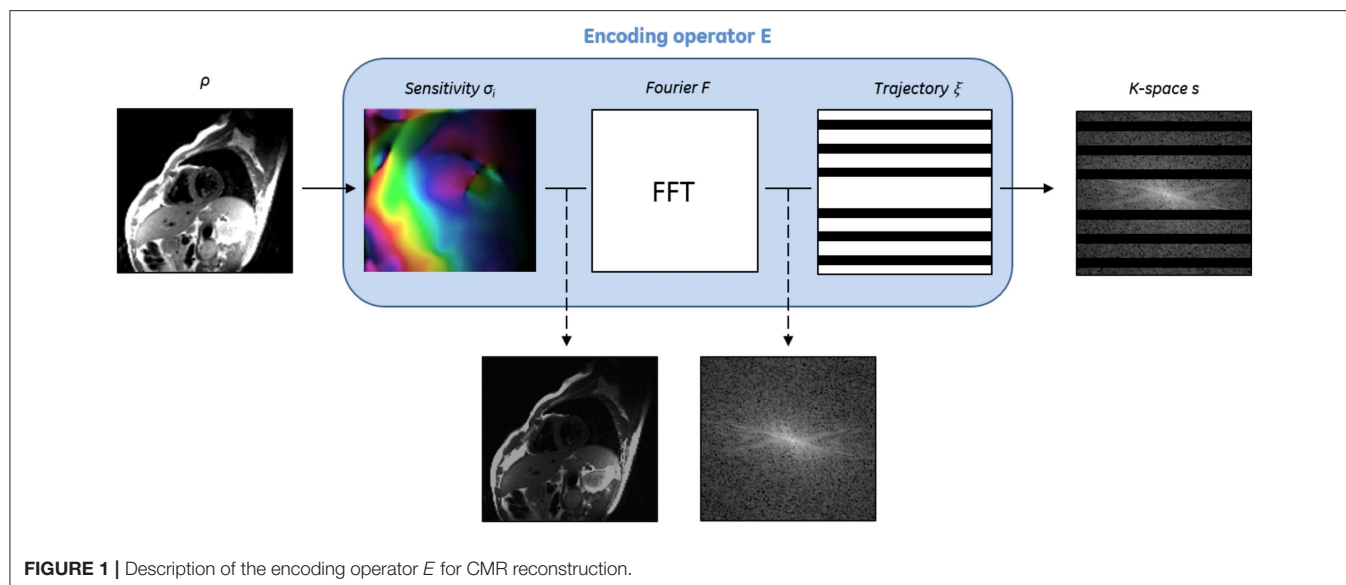
$$s = E\rho \quad (1)$$

Where the MR encoding operator E includes the coil sensitivity profiles, the Fourier transform and the sampling mask, ρ is the image to be recovered and s is the acquired k-space data (Figure 1). The image ρ is thus reconstructed by solving an inverse problem that aims to recover an estimate of ρ from the *known* encoding operator E and the acquired signal s . This inverse problem is ill-posed, i.e., not all the following well-posedness conditions are satisfied: (i) existence of the solution, (ii) uniqueness of the solution, and (iii) stability of the solution (i.e., small disturbances in s do not lead to large perturbations in ρ). The main factors that make MR reconstruction an ill-posed problem include the large scale of the optimization, the system imperfections (e.g., coils sensitivities, signal model simplifications), the limited amount of phase increment steps (undersampling) and the acquisition noise which corrupts the signal.

To overcome the ill-posed nature of the MR image reconstruction problem, this is typically reformulated as a regularized optimization:

$$\hat{\rho} = \operatorname{argmin}_{\rho} \|E\rho - s\|_2^2 + \lambda R(\rho) \quad (2)$$

where the image $\hat{\rho}$ is recovered by balancing between a regularization term $R(\rho)$, which is added as an additional constraint to stabilize the solution, and a data consistency $\|E\rho - s\|_2^2 < \epsilon$, where ϵ is the noise level. The weighting parameter λ controls the degree of regularization and needs to be chosen according to the noise level of the acquired data.



Especially, considering sparsity priors and statistical properties of the MR images to regularize the reconstruction problem have shown great promise. The application of these techniques to speed CMR imaging is the topic of the following subsections.

CS for CMR Imaging

CS MRI reconstruction assumes that the k-space data is pseudo-randomly undersampled, the image admits a sparse representation in some transform domain Φ , and a non-linear reconstruction is performed to enforce data consistency and sparsity of the MR image in the transform domain. A natural approach to enforce sparsity is by replacing the regularization term in Equation (2) by the l_0 (pseudonorm) of the sparse coefficients (12), which counts the number of non-zero entries. However, since the l_0 “norm” does not satisfy the convexity property of a norm and leads to an NP-hard combinatorial problem, approximate solutions are considered instead by replacing the l_0 term by the convex l_1 -norm (13):

$$\hat{\rho} = \underset{\rho}{\operatorname{argmin}} \|E\rho - s\|_2^2 + \lambda \|\Phi\rho\|_1 \quad (3)$$

The problem in Equation (3) is convex and can be solved with a variety of regularization and convex optimization techniques. In cardiac MRI, Φ can be chosen e.g., as the temporal Fourier transform, spatio-temporal total variation, or spatio-temporal wavelets (Figure 2). CS has been extensively used in numerous cardiac applications, such as cardiac cine imaging (14, 15), first-pass cardiac perfusion (16), 3D late gadolinium enhancement (LGE) imaging (17), 3D whole-heart coronary MR angiography (CMRA), and more recently for 4D and 5D free-running CMRA (18–21), among many others. We briefly review some of those techniques in the next paragraphs.

Cardiac Cine Imaging

Cardiac cine MRI with CS reconstruction has demonstrated accurate estimation of cardiac function in a single-breath-hold (22). The study enrolled 81 patients with different

cardiac conditions who were imaged using 2D cine acquisition, under three heart beats per slice, with high spatial ($1.7 \times 1.7 \text{ mm}^2$) and temporal resolution (41 ms). A non-linear iterative SENSE-type reconstruction was performed with spatio-temporal regularization using redundant Haar wavelets. The reconstruction was performed inline in ~ 3 min for a stack of eight continuous short-axis image. CS reconstruction led to slightly worse image quality compared to conventional PI cardiac cine. A similar acquisition/reconstruction framework was performed on 100 patients referred for CMR in Vermersch et al. (23). Free-breathing 2D motion-corrected cine CMR has been also studied in Usman et al. (14). Acquisition was performed on five healthy subjects using a golden radial pseudo-random sampling and non-rigid respiratory motion-corrected reconstruction with CS temporal regularization was performed offline (reconstruction time ~ 2 – 2.5 h).

A 3D cardiac cine acquisition with CS reconstruction has been proposed to image the left ventricle in a single breath-hold (15). Ten healthy subjects and three patients were imaged at $1.9 \times 1.9 \times 2.5 \text{ mm}^3$ spatial and 42–48 ms temporal resolution in ~ 19 s using a Cartesian spiral phyllotaxis sampling (24). Reconstruction times were ~ 4 min employing a soft-gated iterative SENSE reconstruction with spatial and temporal redundant Haar wavelet transforms. Free-breathing 3D cardiac cine has also been proposed to alleviate the requirement of breath-holding in Usman et al. (25). Whole-heart cardiac cine images were acquired in eight healthy subjects and three patients in ~ 4 – 5 min using an accelerated 3D free-running sequence with 2 mm^3 isotropic resolution and ~ 31 – 70 ms temporal resolution. A CS-SENSE reconstruction with total variation regularization and translational respiratory motion correction was performed offline in ~ 2.5 h.

3D Late Gadolinium Enhancement Imaging

CS has been employed to increase the spatial resolution and accelerate scan time of LGE imaging for myocardial scar and

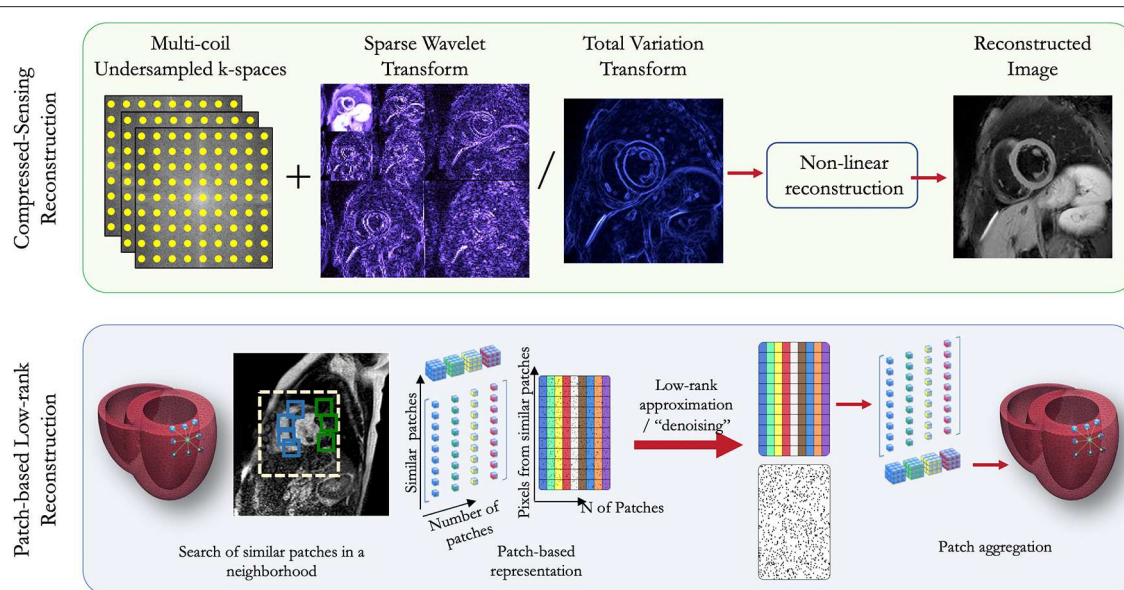


FIGURE 2 | Schematic diagram of compressed sensing and patch-based low-rank reconstructions for CMR.

fibrosis visualization. Kamesh Iyer et al. (17) proposed a CS technique for rapid 3D LGE imaging for visualization of ablation-induced scar in the left atrium wall in patients with a history of atrial fibrillation and ablation therapy. 3D LGE data was acquired fully sampled on 8 patients and retrospectively undersampled using a variable density sampling with a 3.5-fold acceleration at a resolution of $1.25 \times 1.25 \times 2.5 \text{ mm}^3$ (acquisition time of $\sim 10\text{--}15 \text{ min}$). CS reconstruction was performed offline after coil compression (four virtual channels reconstructed) using an efficient Split Bregman optimization (26) for fast reconstruction ($\sim 8 \text{ s}$ for 44 slices) with 3D total variation regularization. The Split Bregman method has shown to be an efficient solver for many regularized inverse problems with good convergence properties and fast minimization (26).

Basha et al. (27) proposed a patch-based CS technique (“LOST,” see next section) to acquire and reconstruct isotropic spatial resolution $1.4 \times 1.4 \times 1.4 \text{ mm}^3$ 3D LGE data in 270 patients referred for myocardial viability assessment, using a pseudo random k-space undersampling pattern (28) with up to 5-fold accelerated acquisition ($\sim 4 \text{ min}$ total acquisition time). LOST reconstruction was performed inline (via CPU cluster) in $\sim 1 \text{ h}$.

Whole-Heart CMRA

Forman et al. proposed a free-breathing (29) and multi-breath-hold (28) Cartesian spiral phyllotaxis (6.5-fold) acquisition combined with an inline multi-coil SENSE reconstruction and 3D total variation regularization to reconstruct high-resolution ($\sim 1 \text{ mm}^3$ isotropic) CMRA images in $\sim 52 \text{ s}$. Accelerated non-rigid motion-compensated isotropic (1.2 mm^3 , 3-fold acceleration) 3D CMRA was also performed in $\sim 5 \text{ min}$ using 3D total variation regularization (reconstruction time $\sim 44 \text{ min}$) and variable density Cartesian acquisition (30). Haar wavelets

combined with an efficient FISTA optimization were used for whole-heart navigator-gated CMRA imaging at 3T, employing a Cartesian spiral phyllotaxis sampling at 9-fold acceleration (effective scan time of $\sim 3 \text{ min } 45 \text{ s}$ at a resolution of $1.3 \times 1.3 \times 1.2 \text{ mm}^3$) (31). A similar optimization was employed at 1.5T to reconstruct CMRA images with an isotropic resolution of 0.8 mm^3 (32). CS techniques based on discrete wavelet transform were also implemented on GPU to bring whole-heart CMRA image reconstruction to $< 4 \text{ s}$ (33).

XD-GRASP (34) and its extensions have been proposed to enable free-breathing whole-heart motion-resolved 5D [($x-y-z$) spatial dimensions + respiratory and cardiac phases] CMRA in a single continuous acquisition by exploiting temporal total variation along the cardiac and respiratory dimensions (35–37). In Feng et al. (35), image acquisition was performed with a continuous 3D golden-angle pattern at isotropic 1.15 mm^3 resolution and $\sim 40\text{--}50 \text{ ms}$ temporal resolution (acquisition time $\sim 14 \text{ min}$). A conjugate gradient optimization was used to reach offline reconstruction times of $\sim 6 \text{ h } 48 \text{ min}$. Similar approaches were also proposed for time-resolved, cardiac-resolved, high-resolution flow imaging [XD flow (38)].

Drawbacks of CS for CMR

Although CS has shown noticeable success in CMR, as reflected by the many applications and recent integration into routine clinical scanners, there remains major drawbacks which may impede its full potential. Firstly, the non-linear nature of the optimization presents a barrier for fast reconstruction time, although notable improvement has been made on the maturation of the algorithms and the move toward GPU implementations to greatly reduced computational times. Another relevant weakness of CS-based reconstruction is the need for tuning regularization parameters that heavily depend on the type of

image, sampling trajectories, sparsifying transform, acceleration factor, etc. Finally, while choosing the appropriate transformation basis Φ can contribute to an efficient sparse representation, the robustness of the reconstruction will heavily depend on this specific operator.

Low-Rank-Based Approaches for CMR Imaging

Another model closely related to sparsity is the notion of low-rank matrices. Low-rank image reconstruction takes advantage of the fact that MR images have inherently a high degree of correlation (e.g., dynamically or locally on a patch scale) and thus can be represented by a union of low-dimensional subspaces. We provide below an overview of some reconstruction techniques incorporating low-rank models employed for CMR imaging.

Globally low-rank (GLR) reconstructions, exploiting low-rankness on the entire image series, have been exploited in many cardiac applications such as dynamic cine MRI (39–41), real-time CMR (42), cardiac perfusion (43), or simultaneous multislice CMR fingerprinting (44). GLR reconstruction techniques are particularly suited for image series that exhibit strong correlation over time. A Casorati matrix is usually formed from the undersampled image sequence, and the missing k - t samples are then estimated using low-rank matrix completion (41, 45, 46). Low-rank reconstruction has been combined with CS-based techniques to further improve image quality, particularly for high acceleration factors. Low-rank plus sparse ($L + S$) matrix decomposition, which separates the temporally correlated background (L) from the dynamic information (S), has been proposed for dynamic imaging (cardiac cine, cardiac perfusion, and time-resolved angiography) (43, 47). The recently proposed *multitasking* framework has extended global low-rank reconstruction to deal with multiple overlapping dynamics such as T1/T2 recovery and cardiac and respiratory motions, through tensor decomposition (48, 49).

Locally low-rank (LLR) regularization techniques have also been proposed for CMR reconstruction to further reduce spatial blurring often associated with the GLR techniques (50). In essence, LLR reconstruction techniques exploit low-rankness structure of an image series on local regions (i.e., patch), and have been efficiently used for dynamic CMR imaging (51, 52), high-resolution dynamic myocardial T1 mapping (53) and 5D flow (18).

More recently, patch-based image reconstructions exploiting local (i.e., within a patch) and non-local (i.e., between similar patches) similarities and low-rank matrix representations have been employed for CMR image reconstruction, leading to even sparser representations. In those techniques (a.k.a. LOST and PROST, **Figure 2**) the similarity of 2D/3D image patches have been exploited through block-matching and low-rank decomposition. These techniques have shown to reconstruct highly undersampled LGE (27, 54) and CMRA images with improved image quality compared to CS-based techniques (55, 56) (**Figure 3**). Accelerated free-breathing CMRA in concert with 3D-PROST reconstruction enables isotropic sub-millimeter (0.9 mm^3) whole-heart visualization of the coronary vasculature,

including small distal segments, in ~ 5 – 7 min acquisition time and ~ 3 min reconstruction time (**Figure 3**). Based on a similar idea, patch-based reconstruction has been used for the reconstruction of undersampled 2D cine MR images by extending the patch search to the cardiac temporal dimension (58). The technique has been also extended to multi-contrast CMR reconstruction through high-order tensor decomposition (59) and demonstrated for highly accelerated simultaneous 3D myocardial T1/T2 mapping and cine imaging (60), and 3D whole-heart myocardial T2 mapping (61).

Dictionary Learning-Based Approached for CMR Imaging

Dictionary learning based CS techniques (also referred as data-driven techniques) have been also proposed for CMR reconstruction. As opposed to conventional CS techniques, where sparse transforms or fixed dictionaries are known a priori, blind compressed sensing (BCS) techniques adaptively learn the sparse representation and dictionaries from the acquired undersampled data itself. These reconstruction techniques have the advantage to be highly adaptive to the image content at hand by learning dictionaries specific to the acquired data and without the need for training data. BCS has shown to outperform conventional CS approaches in several CMR applications such as cardiac cine MRI (62, 63) and contrast enhanced dynamic MRI (64).

Both dictionary learning and CS models can be leveraged to further increase acceleration factors. In Caballero et al. (62), a dictionary learning technique was combined with CS to speed up dynamic CMR imaging (~ 8 - to 16-fold acceleration). An optimal dictionary is learnt directly from undersampled data online, through processing of spatio-temporal 3D patches, and is used to fill the missing k -space lines. The algorithm was tested on 10 healthy subjects by retrospectively undersampling fully sampled dynamic CMR data. Enforcing temporal gradients with an additional constraint allows to reach higher undersampling factors and accelerate the convergence rate, while consistently showing improvement over non-dictionary-based CS techniques.

Those approaches, however, come at the cost of highly non-convex optimizations, which make theoretical analyses and convergence guarantees very hard, while being often associated with high computational burden and long reconstruction times.

DEEP LEARNING FOR CMR RECONSTRUCTION

Despite the high promise of CS approaches, robustness of the reconstruction will heavily depend on the choice of the sparsifying transform which may be incapable of capturing the complex structure of CMR images. This may lead to images that look overly smooth or unnatural when too high acceleration factors are considered. A further major drawback is the long computational time usually required with iterative reconstruction algorithms and the need for parameters tuning. An inaccurate choice of reconstruction parameters leads either to over-smoothing or to images with remaining undersampling

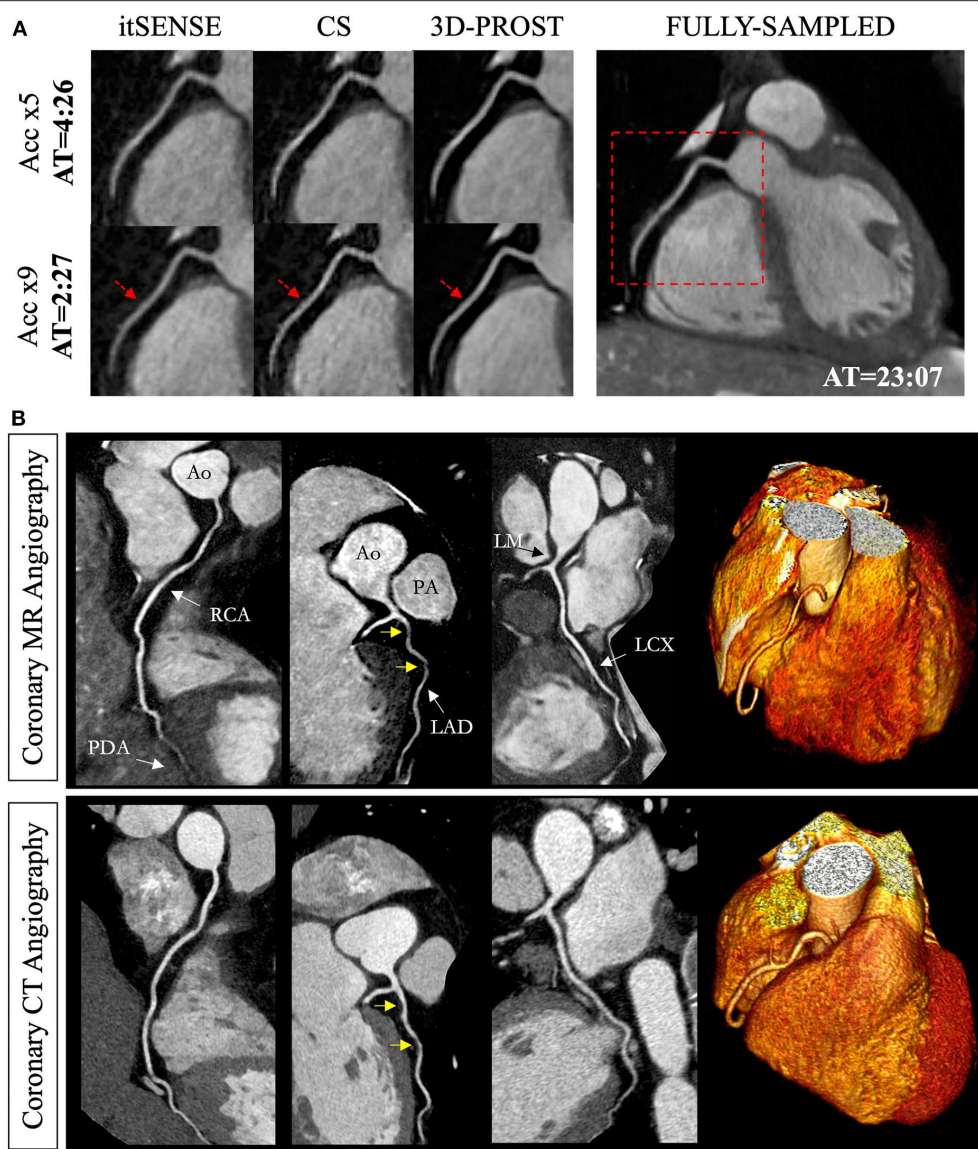


FIGURE 3 | Reconstruction comparisons for coronary MR angiography. **(A)** Example of reformatted images of the right coronary artery from three healthy subjects acquired at 1.2 mm^3 isotropic resolution with a fully sampled whole-heart coronary MR angiography sequence, and with 2 undersampled acquisitions (5- and 9-fold acceleration with variable density sampling), reconstructed using iterative SENSE (itSENSE), wavelet-based compressed sensing reconstruction (CS) and a 3D patch-based approach [3D-PROST (56)]. 3D-PROST provides higher image quality and sharpness (red arrows) than itSENSE and CS for both acceleration factors, achieving similar image quality to the fully sampled reference. Acquisition times (AT) are expressed as min:s. **(B)** Non-contrast whole-heart sub-millimeter isotropic CMRA images of 53-year-old female patient acquired in 10 min 7 s (5-fold undersampling) and reconstructed with 3D-PROST (56) and non-rigid motion correction (57). Visual comparison with contrast-enhanced Computed Tomography angiography (bottom row) shows good agreement and delineation of the coronary arteries with the free-breathing 3D patch-based motion corrected CMRA framework.

artifacts. Taking encouragement from early success in the use of DL in image classification and computer vision, several DL-based MRI reconstruction approaches have been recently proposed to learn models that better describe the reconstruction process and to shift the required optimization effort to an offline training stage, performed beforehand. In other words, rather than performing a reconstruction procedure to compute an appropriate transform between raw data and images for each new data set, DL reconstruction techniques propose to learn

the parameters of that reconstruction procedure in advance, so that it can be applied to all new undersampled data as a simple operation. When using an analytical approach to solve Equation (3) for MR image reconstruction, the applied regularization operator is explicitly described, and the optimization approach is carefully chosen. Generally, the more sophisticated the modeling adopted in reconstruction, the more demanding the optimization process. The aim in DL-based MRI reconstruction, is to replace this optimization with a convenient function

$f_\phi(\cdot)$ which is expressed as a DNN with parameters ϕ . Thus, a computationally efficient direct mapping from the acquired data s to the reconstructed image ρ can be obtained as a result of the neural network's training procedure. Training of a neural network implies changing its weights to optimize the network's output. This is performed by applying an optimization algorithm on a function measuring the difference of the outputs with respect to a target dataset, referred as loss function. Once these weights are learned, a network can be utilized to reconstruct new, unobserved data, and therefore learn to generalize. We will further discuss the training procedure in the section Training Procedure for DL-Based MRI Reconstruction. The main advantage of DL-based reconstruction techniques, with respect to conventional analytical reconstruction techniques, lies in the capability of a DNN to utilize the prior information learnt from the great number of routinely performed MRI exams, to help the reconstruction process. However, due to the problem's high dimensionality, a large dataset of raw k-space data s and target MRI images ρ need to be available to avoid over-fitting in the learning process. Collection of large MRI datasets can be challenging and proposed techniques for MRI reconstruction usually depend on the use of data-augmentation techniques, which is discussed in the section Data Availability for CMR Reconstruction. Given these preliminary remarks, a fundamental question may arise: Under which conditions would we expect DL approaches to outperform CS approaches in terms of reconstruction accuracy in CMR imaging (computational considerations aside)? In this section, we do not aim to provide a definitive answer to this question. Our objective is to provide the reader with a critical approach in reviewing the literature, to be used as guidance in solving their DL-based CMR reconstruction problems. DNN architectures and neural network training procedures will be described first for generic MRI reconstruction, followed by a review of the approaches that have been designed for cardiac applications.

Neural Networks Architectures for DL-Based MRI Reconstruction

Careful selection and design of the neural network architecture is fundamental to solve the MRI reconstruction problem at hand, since the architecture's design controls the set of available functions $f_\phi(\cdot)$ that are investigated during the learning process. A *Neural Network* is composed of an input layer, followed by hidden layers that transform the data in a new representation; and it ends with an output layer that generates the neural network's prediction. Each layer is composed of multiple neuron units. The output of the neurons in each layer is given by the weighted sum of the input neurons, followed by a non-linear function termed *Activation Function*. A series of fully connected layers and activation functions is referred to as fully-connected neural network. The major advantage of fully connected networks is that they are "structure agnostic," which means that no special assumptions need to be made about the network's input. In the following subsections we briefly discussed neural networks architectures that have been proposed to enable MR image reconstruction.

Convolutional Neural Networks

Convolutional neural networks (CNN) (65) differ from fully-connected neural networks by the application of convolutions to each layer. As multiple convolution kernels are applied, several feature maps are defining a novel image characterization. In CNNs, there are usually less parameters with respect to fully-connected neural networks, since the kernel's weights are fixed as they move across the input image. The reduction in number of parameters simplifies the network's optimization problem. CNNs have been shown to learn interesting features from medical images and to be particularly appropriate to capture their multiscale structure. The use of residual blocks (66) also plays a fundamental role in training DNNs. Instead of learning a complete mapping function between consecutive layers; by adding skip connections between two or more layers, it is possible to learn the residual from the input to the output of a residual block or to the output of the whole neural network. The use of skip connection has been shown to be particularly well-suited to learn image features, such as edges or noise-like artifacts (66).

Encoder-Decoder CNN

While for conventional CNNs feature map dimensions are fixed, for encoder-decoder CNNs the feature maps are gradually downsampled at each layer down to a convolution with a kernel of size 1×1 , and then upsampled to the output's size. The first half of the network, the encoder part, learns a representation in a smaller manifold of the input image, and is then given as input to the decoder part of the network to obtain an image with the most meaningful features. Since the encoder part of the network compresses the feature maps' spatial information, a loss of details in the output can be encountered using an encoding-decoding network (67). This issue can be overcome by inserting symmetric skip connections, therefore preserving the important details that are present in the input image. An encoder-decoder network with skip connections is commonly referred to as U-Net network (67).

Variational Neural Network

In the conventional CNN architectures described above, the input data is convolved with a set of filter kernels which are usually followed by a simple, non-learnable, activation function, e.g., rectified linear unit (ReLU). In a variational neural network (VNN), the regularization term R in Equation (3) is defined as a field of experts model (68):

$$R(\rho) = \sum_{k=1}^{FK} \langle \psi_k(\chi_k \rho), 1 \rangle \quad (4)$$

Where R is a linear operator that models convolutions of the image ρ with FK filter kernels $\chi_k \in R^{v \times v}$ of size v , and learnable non-linear activation function ψ_k . In the fields of experts model (68), the convolutional kernels and the parameters of the non-linear activation functions are learned from the data. In contrast to other techniques that make use of ReLU, the parametrizable activation functions ψ_k , used in Equation (4), are defined as a weighted combination of AF Gaussian radial basis functions. In a VNN architecture, the learning power is therefore shifted from

the sole learning of the filter kernels to the learning of both kernels and non-linear activation functions.

Training Procedure for DL-Based MRI Reconstruction

In the previous section, generic DNN architecture blocks have been described for solving MRI reconstruction problems. The choice of the architecture structure and of its constitutive elements determines a set of learnable functions, but it is during the training phase that the set of optimal functions for the given reconstruction task is determined. In general, the training procedure can be designed in a supervised or unsupervised fashion. Supervised methods are mostly used for MRI reconstruction, while unsupervised methods are an active topic of ongoing investigation. Therefore, for the rest of this section, we will focus on supervised approaches. In order to learn the network's parameters for the reconstruction procedure at hand, an optimization problem that minimizes a cost function needs to be defined. The training loss function can be defined as:

$$C(\phi) = \frac{1}{2B} \sum_{b=1}^B \|\rho_b^{\Upsilon}(\phi) - \rho_b^{\text{target}}\|_2^2 \quad (5)$$

Where ϕ are all the trainable parameters of the reconstruction network. Υ is the total number of layers in the network, corresponding to the network's gradient steps $v = 1, \dots, \Upsilon$. b is the current training output image. B is a randomly selected subset of the complete set of training data, referred as data batch. To solve the non-convex optimization problem in Equation (5), a variant of gradient descent, e.g., stochastic gradient descent or the ADAM optimizer are often used (69). The necessary computation of the gradient with respect to network parameters ϕ can be computed via backpropagation (70):

$$\frac{\delta C(\phi)}{\delta \phi^v} = \frac{\delta \rho^{v+1}}{\delta \phi^v} \cdot \frac{\delta \rho^{v+2}}{\delta \rho^{v+1}} \cdots \frac{\delta \rho^{\Upsilon}}{\delta \rho^{\Upsilon-1}} \cdot \frac{\delta C(\phi)}{\delta \rho^{\Upsilon}} \quad (6)$$

These optimization algorithms require the tuning of hyper-parameters, such as strength of regularization or learning rate decay. The choice of the loss function is also crucial for a successful outcome of the training procedure. Because the reconstruction problem is usually formulated as a regression problem, the mean squared error is conventionally utilized as a cost function. Other popular choices are the l_1 norm of the difference and the structural similarity index. Research on generative adversarial networks (71, 72) and learned content loss functions are currently in progress. Once the optimal parameters ϕ are learned, the reconstructed image ρ can then be estimated from the observed k-space data s by simply computing $\rho = f_{\phi}(s)$ using the trained network. This efficient functional relationship is a major advantage of neural networks over conventional CS techniques that may require complex inference procedures (73).

Data Availability for DL-Based CMR Reconstruction

The inference step between input and output of the reconstruction model is highly dependent on the set of

input k-space data and of reference images seen during training. This requires the availability of a large set of fully sampled multi-coil k-space data. Undersampled data can be obtained by retrospectively removing k-space data entries according to a sampling trajectory in the forward operator E . This data can be used as input for the reconstruction network during training. The lack of freely accessible databases of fully sampled multi-channel raw k-space data, is an open issue for DL-based CMR reconstruction. In addition, since the dataset used to train a certain model becomes an essential component that defines its performance, it is difficult to compare different approaches if the training data is not publicly available. Even if initiatives for release of annotated CMR images are growing (e.g., UK Biobank), very limited public or institutional k-space CMR raw data have been provided to the research community. Moreover, large data bases of annotated CMR images, such as UK Biobank, are limited to specific type of exams. The DL reconstruction techniques presented in the following section are therefore mostly applied to retrospectively simulated k-space data and are restricted to specific MRI sequences (e.g., cardiac cine MRI).

Neural Networks Architectures for DL-Based CMR Reconstruction

In this section, we review representative approaches proposed in the literature for MRI image reconstruction with a focus on CMR applications. The different approaches are summarized in Table 1.

Encoder-Decoder CNN for Image Dealiasing

U-net type of networks that perform an end-to-end mapping in image space have been successfully employed in many MRI post-processing applications (e.g., image segmentation) showing promising results. In the field of image recovery from undersampled k-space data, U-net architectures have been used by several groups to reduce noise-like image artifacts in post processing (see Figure 4A).

In Hauptmann et al. (74), a 3D residual U-net have been employed to reduce undersampling artifacts for 2D golden-angle radial cardiac cine MRI. This residual U-Net contains a contracting multi-scale decomposition path and a symmetric expanding path with skip connections at each scale (see Figure 5). The 3D-convolutions are trained on entire image sequences ($x - y - t$) to enforce temporal consistency between cardiac frames. This technique demonstrated robustness with respect to the flickering artifacts that would be present if 2D convolutions were separately applied to each frame. The proposed U-net architecture was trained from 13-fold retrospectively undersampled images using a simulated tiny golden angle radial trajectory. These images were obtained from Cartesian breath-hold (BH) bSSFP cine acquisitions of 250 patients with congenital heart disease (CHD). The trained 3D U-net was then applied to real-time 13-fold accelerated tiny golden angle 2D radial bSSFP data acquired under free-breathing in 10 previously unseen patients with CHD. The radial bSSFP data were recovered with the proposed 3D U-net and reconstructed with CS for image quality and computational time comparisons. Ventricular volume measurements for 10–15 contiguous slices,

TABLE 1 | Summary of methods that, to the best of our knowledge, have used a deep-learning-based approach for CMR reconstruction and which have been referred to in this article.

References	Application	Method/Network architecture	Training/Validation data
Hauptmann et al. (74)	Cine MRI	3D U-net applied in post-processing to reduce streaking artifacts	Single-coil retrospective/Single-coil prospective
Kofler et al. (75)	Cine MRI	2D U-net applied to the spatio-temporal domain in post-processing	Single-coil retrospective/Single-coil prospective
Schlemper et al. (76)	Cine MRI	End-to-end cascade of CNN regularization blocks and data-consistency blocks	Single-coil retrospective/Single-coil retrospective
Fuin et al. (77)	CMRA	End-to-end cascade of Multi-Scale VNN regularization blocks with data-consistency operators	Multi-coil retrospective/Multi-coil prospective
Biswas et al. (78)	Cine MRI	End-to-end cascade of CNN operators, an analytically defined STORM prior, and conjugate gradient data consistency steps	Multi-coil retrospective/Multi-coil retrospective
Qin et al. (79)	Cine MRI	End-to-end cascade of recurrent CNN regularization blocks and data-consistency blocks	Single-coil retrospective/Single-coil retrospective
Akçakaya (80)	Myocardial T1 mapping	CNN for k-space interpolation	Scan-specific Autocalibrating Signal data
Wang et al. (81)	Cine MRI	A first CNN for k-space interpolation followed by a concatenated CNN network architecture for image deblurring	Single-coil retrospective/Single-coil retrospective

obtained using both the CS reconstructed images and 3D U-net, were compared to a reference Cartesian fully sampled BH-bSSFP cardiac cine data. The overall reconstruction time with the residual 3D U-net implemented on graphics processing unit (GPU) was five times faster than conventional CS techniques implemented on CPU (74). Moreover, the overall image quality of the ventricular volume measurements from the 3D U-net recovered images were superior than the CS reconstructions (Figure 6). In this study, the validation data was acquired during free-breathing, while the training data was obtained during a breath-hold; the effects of cardiac and respiratory motions were therefore not taken into consideration.

The work presented in Hauptmann et al. (74) demonstrates that 3D CNNs can be employed to map entire undersampled 2D sequences to the corresponding fully-sampled 2D cardiac cine sequences. However, employing 3D convolutional layers requires a higher number of parameters and thus increases the amount of data needed to efficiently train a network and prevent overfitting. In Kofler et al. (75), the authors proposed a technique

to recover undersampled 2D golden-angle radial cine CMR by training a modified 2D U-net on the 2D spatio-temporal domain ($x - t$) extracted from the image sequences (Figure 7). This study suggests that the learning process can be improved by training the network on 2D $x - t$ images extracted from the spatio-temporal domain of the cardiac cine sequence. This technique obtained similar results with respect to the 3D U-Net (74) by training the network on a substantially smaller training data set and also proved to be robust with respect to rotations in image space.

The main limitation of the approaches presented in this section, as for all DL techniques applied in post-processing, is that the actual validation data consists of coil-combined magnitude images, instead of multi-coil complex k-space data. Therefore, these approaches do not learn a full reconstruction procedure that accounts for consistency with respect to the acquired k-space data (see Figure 4), but also do not take advantage of the full benefits of coil sensitivity encoding underlying parallel imaging.

Unrolled Convolutional Neural Networks

In this section, we describe how a DNN can be guided to learn operations that are similar to those performed in conventional iterative CS reconstruction, therefore bridging the gap with conventional iterative techniques. Incorporating domain expertise in a DNN framework can in fact facilitate the learning procedure of the model and result in better estimates of the MR images. For CS-based variable splitting techniques, the optimization problem in Equation (3) is usually solved using an alternating algorithm, iterating between a regularization stage and a data consistency stage. Instead of explicitly defining the regularization term, several DL techniques have been proposed to directly learn the regularization term by using CNNs. These techniques, such as Deep-ADMM net (83), VNN (84), or CascadeNet (76), represent a DL framework of an unrolled version of the iterative constrained reconstruction where the network parameters are trained in order to reconstruct the MR images directly from the undersampled k-space data as an input (see Figure 4B).

In particular, Schlemper et al. (76) proposed a framework for the reconstruction of 2D cardiac cine MR images from highly undersampled data using a cascade of CNNs, termed CascadeNet. Since a simple CNN is not efficient in learning the regularization operator iteratively; the authors proposed to concatenate a new CNN on the output of the previous CNN to create a DNN that iterates between CNN regularization operators and data consistency operators. The resulting network consists in convolutional layers, followed by ReLU, residual connections, and data consistency layers. The authors employed a hard-projection solution to enforce data consistency: for each stage of the unrolled model, if the k-space samples are initially unknown (non-acquired), then k-space values obtained from the FT of the previous layer's output are used. For the k-space entries that have been acquired, a linear combination between the estimated values from the previous layer and the original measurements is applied. Since the data consistency step has a simple expression, it is possible to treat it as a layer of a network and to specify the rules for forward and backward propagation for training. By defining the forward and back-backpropagation rules for the

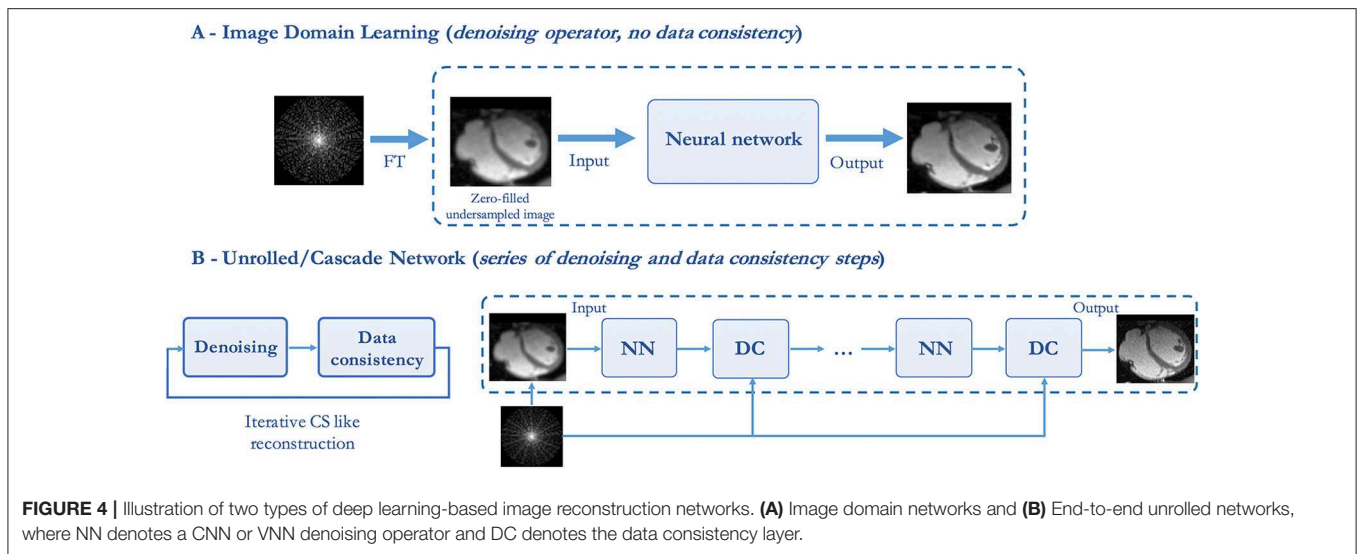
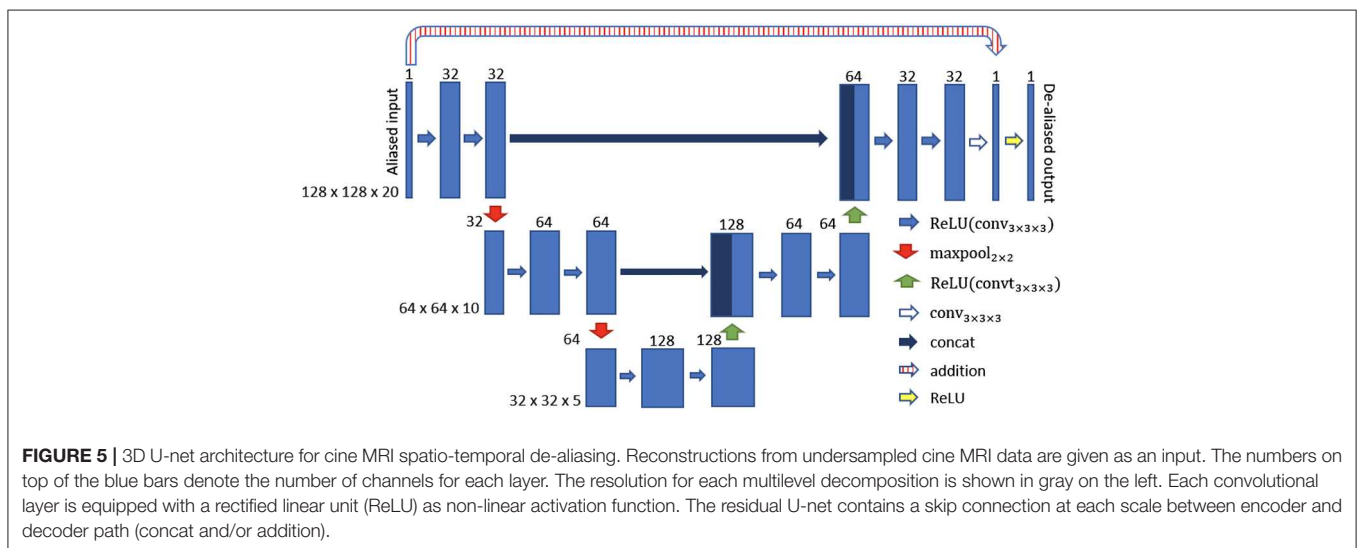


FIGURE 4 | Illustration of two types of deep learning-based image reconstruction networks. **(A)** Image domain networks and **(B)** End-to-end unrolled networks, where NN denotes a CNN or VNN denoising operator and DC denotes the data consistency layer.



data consistency layer, all stages of the network can be trained in an end-to-end fashion, therefore building one deep network. The authors also demonstrated that spatio-temporal correlations can be efficiently learned by CNNs, combining 3D ($x - y - t$) convolutions and data sharing approaches. Assuming that for adjacent cardiac frames the difference in data content is relatively small, the neighboring k -space frames along the temporal-axis share similar information. The missing k -space samples for each time frame can then be approximated using the samples from the adjacent cardiac frames. The authors therefore extended the proposed network architecture adding data “sharing layers that take an input image and generate multiple data-shared images” (76). The obtained images are then concatenated along the channel-axis of the network and fed into the proposed cascading network. For separate reconstruction of 2D cardiac single frames, this technique was compared to Dictionary Learning MRI (85), for retrospective undersampling factors of 3- and 9-fold.

For reconstruction of cardiac cine MRI, the technique was compared to state-of-the-art CS and low-rank approaches, such as dictionary learning with temporal gradient (62), k - t sparse and low-rank (kt-SLR) (46), and L+S matrix decomposition (43). The presented results demonstrated that the CascadeNet outperforms CS and low-rank approaches in terms of reconstruction error and perceptual quality, particularly for high undersampling rates (Figure 8). In addition, for 2D reconstruction, each image could be reconstructed in 23 ms, therefore enabling real-time applications, while for the reconstruction of cine MRI, an entire sequence was reconstructed within 10 s.

It is worth noting that in the experiments shown in Schlemper et al. (76), training and validation data were obtained by retrospectively undersampling single-coil data, thus further validations are required to understand the full potential of this technique for multi-coil prospective acquisitions. Other techniques have applied an unrolled end-to-end framework

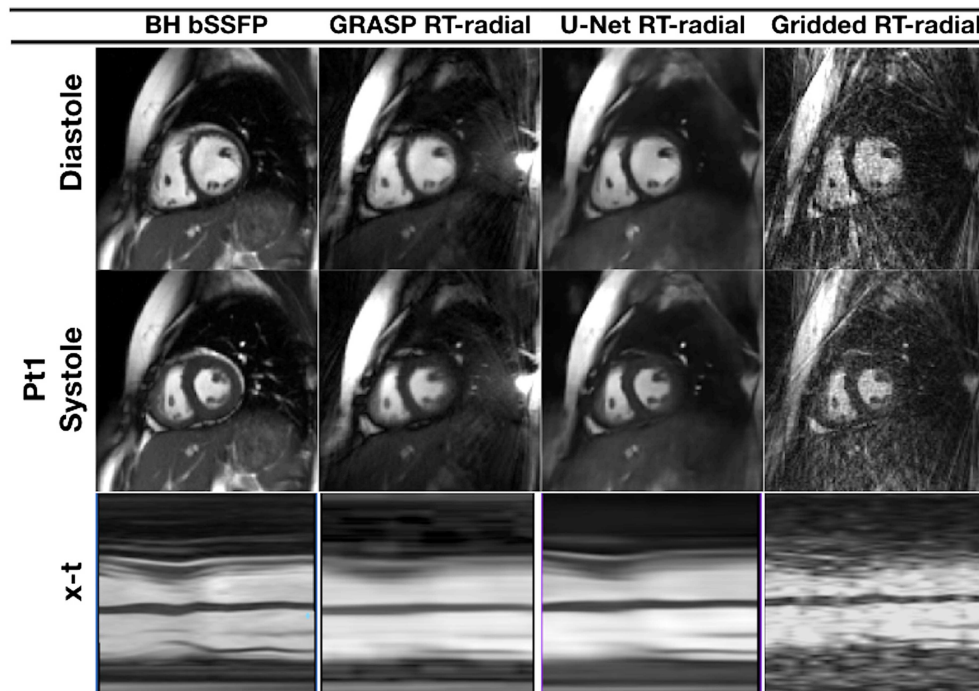


FIGURE 6 | Cine MRI images for one representative patient with congenital heart disease, acquired with prospective undersampling of 13-fold. Reconstructed images are presented in peak systole and peak diastole for a reference breath-held balanced steady-state free precession sequence (BH-bSSFP, first column), the real-time radial sequence reconstructed with GRASP (82) (second column) and the residual 3D U-net (third column), as proposed in Ronneberger et al. (67). Images reconstructed with GRASP and the proposed residual 3D U-Net show spatial and temporal blurring, that could be a result of undersampling and incomplete motion correction.

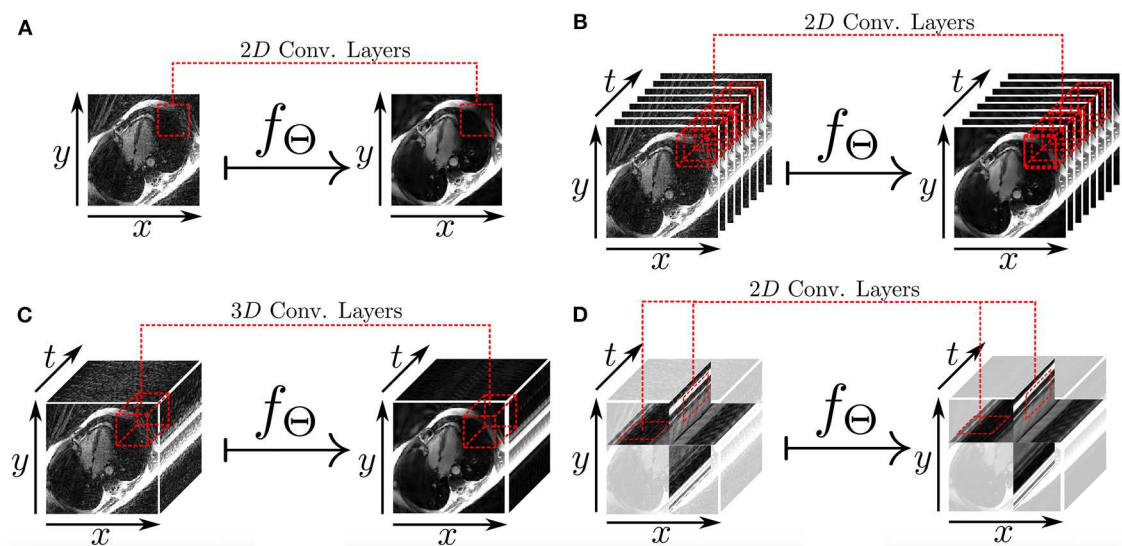


FIGURE 7 | Different 2D and 3D deep learning-based approaches for radial undersampling artifacts reduction (post-processing) presented in Kofler et al. (75). **(A)** 2D U-net for frame-to-frame mapping. **(B)** 2D U-net for sequence-to-sequence mapping with cardiac phases aligned along the channel dimension. **(C)** 3D U-net for sequence-to-sequence mapping with 3D convolutional kernels. **(D)** 2D U-net for recovery of two-dimensional spatio-temporal images.

in the more realistic scenario of multi-channel coil complex MR data. For example, Hammernik et al. proposed a trainable formulation for undersampled MRI reconstruction (84), which

embedded a PI and a CS reconstruction within a DL unrolled end-to-end framework. Undersampled k-space data and coil sensitivity maps are provided as input to this unrolled model

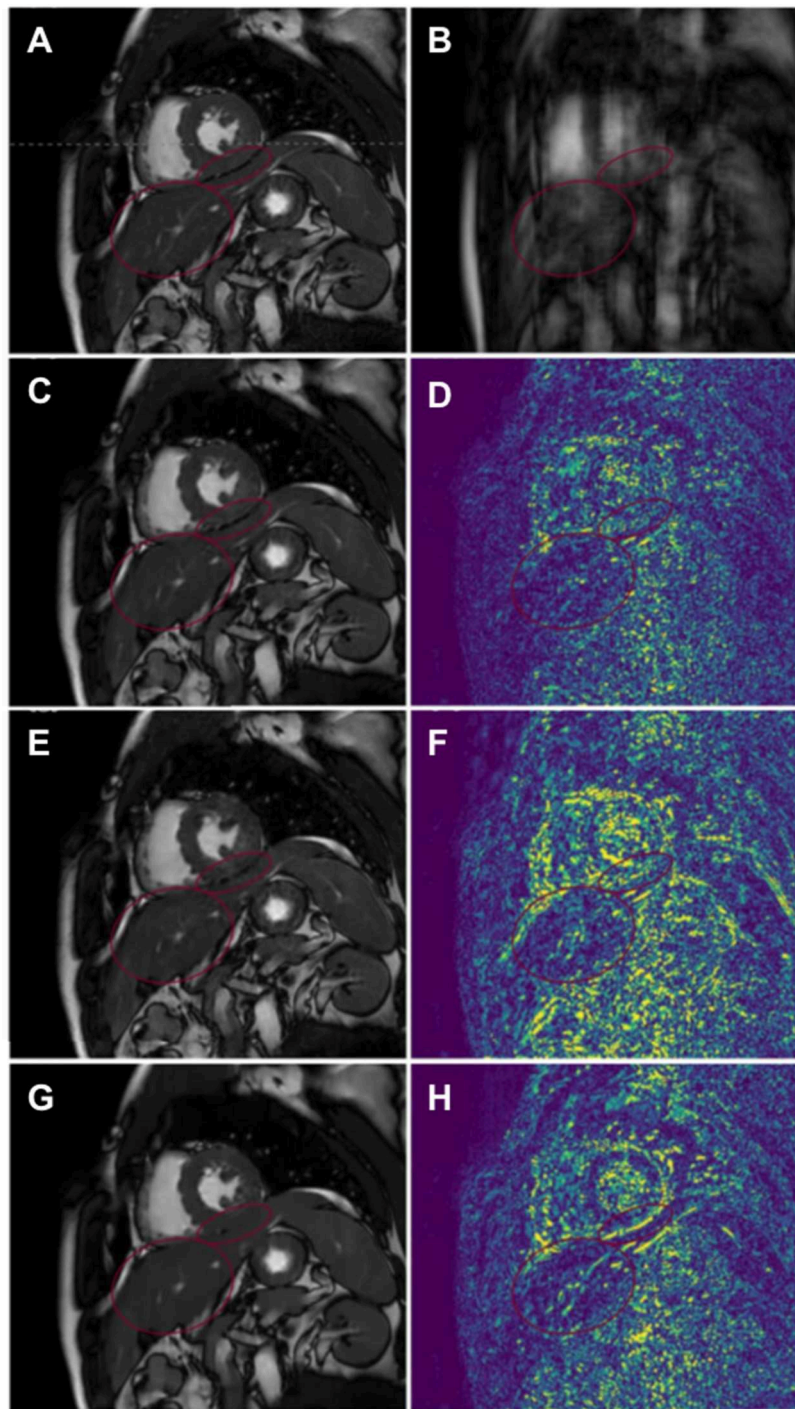


FIGURE 8 | Comparison of reconstructed 2D cardiac cine MR image sequences employing Dictionary Learning with Temporal Gradient (DLTG) (62) and CascadeNet (CNN-S) (76), from one representative healthy subject with retrospectively undersampling. **(A)** Ground truth fully-sampled cine MR image, **(B)** 9x retrospectively undersampled acquisition, **(C,D)** CascadeNet reconstruction with data sharing and its error map, **(E,F)** CascadeNet reconstruction without data sharing (CNN) and its error map, **(G,H)** DLTG reconstruction and its error map. Red ellipses highlight the anatomy that was reconstructed better by CNN than DLTG.

for DL reconstruction, and high-quality MR images are obtained as an output in an end-to-end fashion. The regularization term of this network was implemented as a VNN, and the data consistency term was implemented as the l_2 norm

with respect to the acquired k-space data, as in Equation (3). The use of a VNN was first introduced for multi-coil complex-valued MRI reconstruction of 2D static images of the knee.

Building on this work, Fuin et al. (77) extended the previously introduced VNN approach to enable fast reconstruction of undersampled motion-compensated free-breathing whole-heart 3D CMRA. A multi-scale VNN (MS-VNN) architecture was introduced in order to better capture the small caliber of the coronary arteries, as well as whole-heart structural features ($x - y - z$) in a 3D CMRA image. In order to increase the representation potential of the network, a wider network was implemented, using a multi-scale approach that can capture complementary and richer information at different resolutions. In addition, a training scheme suited for reconstruction of respiratory motion corrupted data was applied. The MS-VNN was trained on retrospectively undersampled (5- and 9-fold) translational motion corrected complex k-space data in an end-to-end fashion, in order to ensure that the effect of bulk, respiratory, and cardiac motion was identical in both output and target images during the training process. The MS-VNN reconstruction was then applied to newly acquired prospectively 5- and 9-fold undersampled data and compared to wavelet-based CS (12) reconstructions, as presented in **Figure 9**. MS-VNN outperformed the conventional CS in terms of quantitative right coronary artery sharpness and visible vessel length, with results comparable to the fully sampled scan. MS-VNN combined with 100% respiratory scan efficiency and variable density spiral-like Cartesian undersampling, allowed the acquisition of high-quality 1.2 mm³ isotropic CMRA images in a short and predictable scan time of ~2–4 min and their reconstruction in ~14 s.

Aggarwal et al. (86) introduced a similar network design, termed MoDL, where conventional CNNs are used for the implementation of the regularization term, but where all network stages share the same set of parameters. This unrolled technique with shared parameters, also applies a conjugate-gradient data consistency step instead of the simple gradient based approach utilized in Hammernik et al. (84). The use of a conjugate-gradient step within the network translates into improved results for a given number of iterations at the expense of a slightly longer run time. Another work from the same team combines DL MoDL reconstruction along with complementary analytical image regularization constraints to recover free-breathing cardiac cine MR images from highly undersampled multi-coil measurements (78). This framework alternates between a learned regularization of the image using CNN, an analytically defined Smoothness regularization on manifolds (SToRM) prior (87), and a conjugate gradient data consistency step. The method was tested on only two simulated datasets, but it promises to combine the benefits of CNNs with analytical image regularization priors, such as SToRM, which exploits subject-specific information including cardiac and respiratory patterns.

Unrolled Convolutional Recurrent Neural Networks

A recurrent neural network can be thought of as multiple copies of the same network stage, each passing a message to a successor stage. The stage of the recurrent network has a memory that stores the stage time states, and therefore it allows information to be reflected to the next time stage without overloading the system. Qin et al. (79) proposed a novel unrolled convolutional recurrent neural network architecture, termed CRNN-MRI,

which reconstructs cine CMR images from highly undersampled k-space data. The proposed CRNN-MRI architecture utilizes recurrent connections over each layer of an unrolled network with data consistency layers to reproduce the recurrence existing in the sequential steps of a reconstruction algorithm. Compared to independently learned CNN at each stage of an unrolled network (76), the iteration connections of the CRNN layers allow spatial information learned at a given iteration to be passed to the following iteration. Each stage of the network is therefore optimized depending on the resulting output but also depending on features from previous iterations that can memorize the learned feature and propagate them to the next stage. Secondly, at every stage of the network, the receptive field of a CNRR layer in the spatial domain increases, whereas for a conventional CNN it resets at each stage. Finally, since the network parameters are shared over iterations, the total number of parameters is greatly reduced in comparison to CNNs, potentially offering improved generalization properties. An additional limitation of CNNs is that they accept fixed-sized images as input and produce a fixed-sized image as output. Conversely, recurrent nets allow to operate over sequences of images: sequences in the input, the output, or in the most general case in both input and output. Exploiting this property of recurrent networks, the network architecture presented in Qin et al. (79) incorporates bidirectional recurrent convolutional layers that evolve over time to utilize the temporal correlations of the cardiac cine MRI. Consequently, the model architecture evolves in a recurrent manner over time and over steps/iterations. The CRNN-MRI network therefore comprises of bidirectional convolutional recurrent layers, residual connections and hard-projection data consistency layers [as in (76)]. The residual connections were added to address the potential problem of vanishing gradients during back-propagation. Training and validation data were produced by retrospective undersampling complex images obtained from single-coil data as in Schlemper et al. (76). The experimental results demonstrated that CRNN-MRI outperformed state-of-the-art CS-based dynamic MRI and low-rank reconstruction algorithms, such as k - t FOCUSS (88) and k - t SLR (46) for 9- and 16-fold retrospectively undersampled data. Additionally, CRNN-MRI demonstrated to outperform CascadeNet (76), that employs conventional CNNs in the regularization term.

DL Techniques for K-Space Based CMR Reconstruction

One of the most frequently used techniques for PI undersampled reconstruction in k-space is GRAPPA (89), which employs shift-invariant convolutions to recover/interpolate non-acquired k-space entries. The convolutional kernels, called autocalibrating signal (ACS), are estimated for each subject from either a fully sampled region at the k-space center or from a separate reference scan (autocalibrating signal or ACS). A CNNs based technique has been recently proposed to improve non-linear k-space interpolation for undersampled PI MRI reconstruction (80). Similar to existing approaches, such as non-linear GRAPPA (90), robust artificial-neural-networks for k-space interpolation (RAKI) (80) trains CNNs on ACS data with an l_2 norm loss; and uses these for interpolating missing k-space samples from

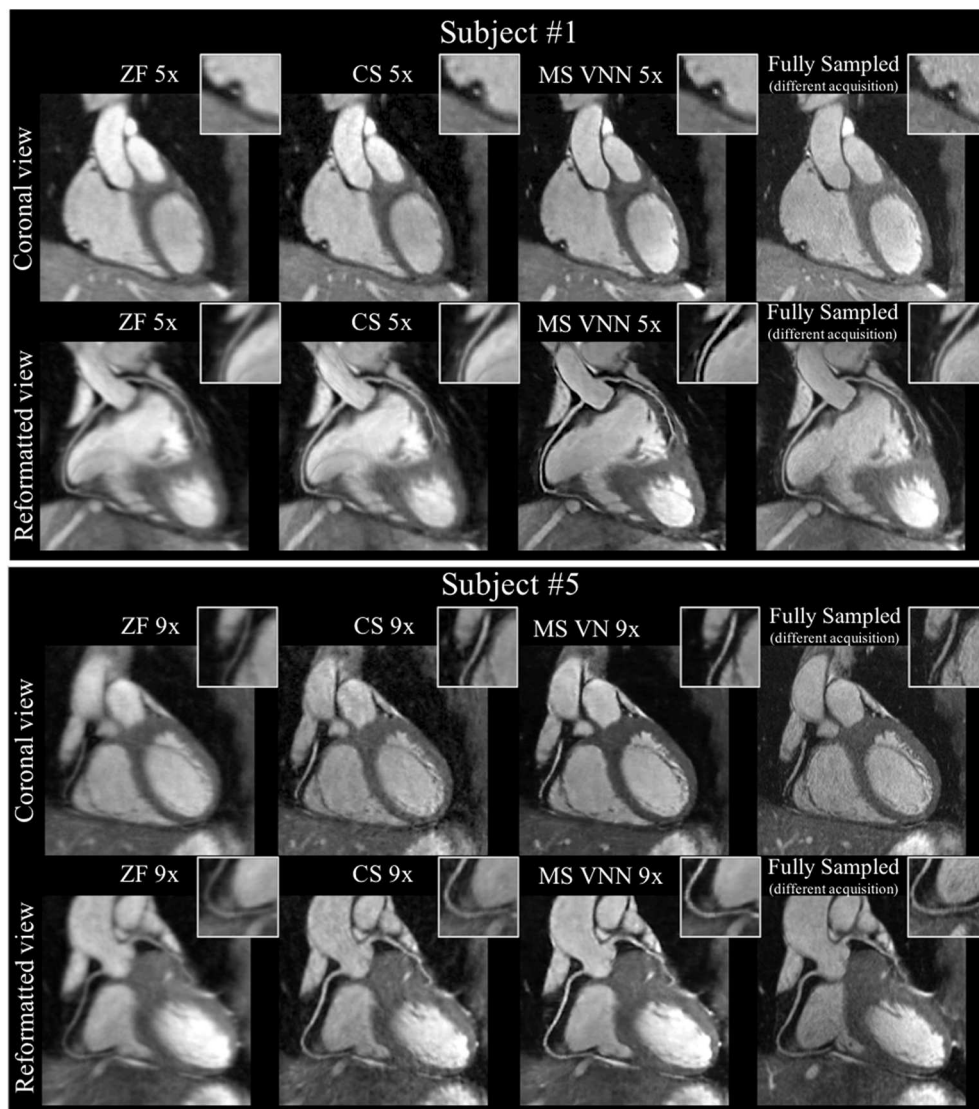


FIGURE 9 | Coronary MR angiography images in coronal view and reformatted along the left (LAD) and right (RCA) coronary arteries, for one representative healthy subject. Acquisitions were performed with isotropic resolution 1.2 mm^3 and 100% respiratory scan efficiency (no respiratory gating). Prospective undersampled acquisitions with acceleration factors 5x (first and second rows) and 9x (third and fourth rows) are shown. Images were reconstructed using zero-filling (ZF), a wavelet-based CS reconstruction (CS), and the Multi-Scale VNN (MS-VNN) reconstruction framework proposed in Fuin et al. (77). Corresponding (consecutively acquired) fully sampled acquisition are shown in the last column for comparison. Multi-scale VNN provides higher image quality than ZF and CS achieving similar image quality to the fully sampled scan. Reconstruction time was $\sim 14 \text{ s}$ with MS-VNN and $\sim 5 \text{ min}$ with wavelet-based CS.

acquired ones. The RAKI network architecture was applied for the reconstruction of myocardial 2D T1 mapping data. Eleven images with different T1 weights were acquired in a single breath-hold using a Cartesian fully sampled bSSFP sequence. Experimental results were then performed on 4- and 5-fold retrospectively undersampled data and RAKI showed improved noise resilience with respect to non-regularized GRAPPA reconstruction. As RAKI is a scan-specific technique and does not require a training data base, it could in theory be applied for the reconstruction of CMR data for which a fully sampled reference acquisition scan cannot be performed, as for example

in perfusion or real-time CMR. However, being scan-specific, this approach also comes with downsides, such as high computational burden, computationally expensive training of a neural network for each scan, and the requirement for additional calibration data.

Recently, a technique that combines DL for k-space interpolation and image dealiasing for retrospectively undersampled 2D cardiac cine MRI has been proposed (81). This approach consists of a first frequency domain network architecture for k-space data interpolation followed by a concatenated image domain network architecture for image dealiasing. Both networks consist of concatenated CNN and

ReLU layers, followed by a data consistency layer. The first and second networks are connected by a Fourier inversion and only one pass through the network is performed. Additionally, the authors propose a multi-supervised network training technique to constrain the frequency domain information and spatial domain information at different levels.

DISCUSSION

During the last decades, several undersampled MR reconstruction techniques have been developed to speed up CMR acquisition. These techniques rely on acquiring less data than needed (in the Nyquist sense) and estimating the non-acquired data exploiting some sort of prior information about the images. PI and CS undersampling reconstruction techniques have revolutionized the field, enabling high scan time accelerations to become standard in clinical practice. Despite of its maturity and recent FDA approval for clinical use, some major technical issues associated with CS reconstruction for CMR remain, including high complexity of the algorithms and long reconstruction times, image degradation at high accelerations, and the need for parameters tuning. Therefore, recent AI-based scientific advances have emerged as solutions to transfer the complexity of the CMR reconstruction from the inline side to the offline training side. Unlike analytical techniques for which the reconstruction problem is explicitly defined into the optimization process, DL-based techniques employ large data sets to learn the key reconstruction parameters and priors during an up-front training procedure, providing a fast and efficient reconstruction that can be applied to all newly-acquired cardiac data.

Strengths and Recent Advances in AI for CMR Reconstruction

The sudden resurgence and popularity of DL approaches for medical image reconstruction can be attributed to their ability to analyze high-dimensional datasets, the availability of computing power, algorithms, web-based storage information, and real-time reconstruction. Although the application of DL to CMR reconstruction is still at an early stage, promising cardiac applications (e.g., dynamic cine MRI or CMRA) have been proposed.

In particular, end-to-end unrolled neural networks models have shown great potential to obtain CMR images that are comparable, in terms of anatomical structure and features, to images obtained with conventional iterative techniques. For example, MS-VNN (77) has shown to obtain high quality static images for prospectively undersampled whole-heart 3D CMRA imaging. Cascade-Net (76) and CRNN-MRI (79), were specifically designed for dynamic imaging and have demonstrated to outperform conventional CS techniques for retrospectively undersampled 2D cardiac cine MRI. Fewer techniques exist for the use of DNN as a k-space estimation problem. This may be due to the non-uniform features of the k-space data (especially for non-Cartesian trajectories), which make it difficult to translate some of the DL techniques that have

been developed for image processing of natural images to CMR reconstruction. However, techniques such as RAKI (80) are scan-specific and do not require a training database; and thus, could in theory be applied to cases for which a reference fully-sampled acquisition cannot be performed.

Limitations and Pitfalls

Although DL-based reconstruction techniques for CMR are showing promising results, there are several remaining challenges that need to be addressed before enabling widespread clinical use.

Simulation and Lack of Clinical Validation

Most of the existing early DL-based techniques for CMR reconstruction are purely based on simulated data, using retrospective undersampling experiments on fully sampled datasets, and limited to single-coil MR acquisition model. Therefore, it remains to be seen how those techniques will work in a multi-coil setting with prospective undersampling, where additional factors can drastically disrupt the reconstruction and degrade the image quality (e.g., eddy current related effects due to gradient jump, blurring due to off resonant spins with spiral trajectories, more complex noise models, unknown coil sensitivity profiles, cardiac and respiratory motion) and intrinsically result in a reduction of the achievable acceleration factor. Furthermore, those different studies have been so far limited to healthy or small selected patient cohorts, which unfortunately limits their current clinical applicability and clinical impact in more complex scenarios. Further clinical validations are thus warranted to demonstrate the robustness of those techniques.

Generalization and Reconstruction Quality

A key strength of CMR is the ability to provide images with different contrast for a comprehensive assessment of the disease. Therefore, one open question regarding the applicability of DL-based reconstruction techniques, in practice, is generalization. The generalization potential and effectiveness of these reconstruction techniques should be further investigated in case of, for example, different imaging resolutions, pulse sequences, acquisition trajectories, magnetic fields strength, MR vendors or clinical sites. While it would be feasible to pre-train separate neural networks for different exams, the poor generalization performance of a DL model to different sequence settings, anatomy, physiology, or to unique pathologies, will limit its translation into clinical practice. On this account, there is still an open question that needs to be investigated: can we design a reconstruction network which accurately and precisely extract unique information from limited samples, while generalizing to different acquisition settings and pathologies?

Data Availability

Another major drawback of DL reconstruction approaches lies in the availability of a specific training data set. The approaches presented in the previous sections have been trained on small samples of hundreds of cases rather than millions, as it is often the case in DL for classification or computer vision. However, the training of reconstruction network still requires the availability

of organized and specific data sets that will allow the model to generalize toward new, unseen, test data. Moreover, most of the models presented are developed for few specific cardiac sequences, such as cardiac cine MRI, for which large image datasets are available to researchers (e.g., UK Biobank).

Quality of the Training Set

In addition to its size, the quality, and composition of the training set is of utmost importance. Several sequences in CMR, e.g., sub-millimeter CMRA or real-time CMR, cannot be acquired with fully-sampled data due to resolution and time constraints. This hinders the application of supervised training approaches for such datasets, justifying the necessity for future research in scan-specific strategies or unsupervised training. We anticipate that future research could focus on the development of neural networks architectures designed to learn features from different cardiac modalities or different MR acquisitions from other organs, in an unsupervised manner, and the incorporation of more conventional regularizations into the networks. The selection of the cost function also has an influence on the network training and optimization, and it is therefore the topic of currently ongoing research. Research on generative adversarial networks and learned content loss functions are also under progress.

Motion Compensated Reconstruction

Additionally, the considerable respiratory- and cardiac-induced motion of the heart during the MR acquisition can significantly impair image quality by showing blurring and/or ghosting like artifacts. Multiple accelerated motion corrected reconstruction frameworks have been developed to simultaneously accelerate scan time and correct for motion during reconstruction. In conventional iterative reconstruction approaches, it is more straightforward to account for motion correction in the reconstruction, as a non-rigid motion model can be directly included in the encoding operator E . Some preliminary simulation work in DL reconstruction have tackled the problem

of correcting motion-related artifacts in 2D cardiac cine images during reconstruction by adding an adversarial element to the network architecture (91). However, no DL reconstruction technique has yet explicitly modeled non-rigid motion directly in the reconstruction process. The efficient implementation of 3D non-rigid transformations in a DNN architecture could in fact prove to be challenging and research on the topic is currently in progress.

Workflow Integration

Finally, most of the DL techniques proposed for CMR reconstruction are implemented offline. Whilst this may be suitable for initial testing, the inline integration of those techniques will be key for their full adoption in clinical practice. Several frameworks, such as Gadgetron (92) or Yarra (<https://yarra.rocks>), have already been proposed for the easy integration of in-house reconstruction techniques into MR scanners; we expect them to play a key role for supporting DL-based reconstruction as well. Many clinical cardiac applications, such as real-time MR-guided cardiac interventions (93) will largely benefit from such inline real-time reconstruction.

AUTHOR CONTRIBUTIONS

AB, NF, RB, and CP devised and wrote the manuscript.

FUNDING

The authors acknowledge financial support from EPSRC EP/P001009/, EP/P032311/1, EPSRC EP/P007619, Wellcome EPSRC Centre for Medical Engineering (NS/A000049/1), and the Department of health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's and St. Thomas' NHS Foundation Trust. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

REFERENCES

- Kramer CM, Barkhausen J, Flamm SD, Kim RJ, Nagel E. Standardized cardiovascular magnetic resonance (CMR) protocols 2013 update. *J Cardiovasc Magn Reson.* (2013) 15:91. doi: 10.1186/1532-429X-15-91
- Messroghli DR, Moon JC, Ferreira VM, Grosse-Wortmann L, He T, Kellman P, et al. Clinical recommendations for cardiovascular magnetic resonance mapping of T1, T2, T2 and extracellular volume: a consensus statement by the Society for Cardiovascular Magnetic Resonance (SCMR) endorsed by the European Association for Cardiovascular Imaging. *J Cardiovasc Magn Reson.* (2017) 19:1–24. doi: 10.1186/s12968-017-0389-8
- Von Knobelsdorff-Brenkenhoff F, Schulz-Menger J. Role of cardiovascular magnetic resonance in the guidelines of the European Society of Cardiology. *J Cardiovasc Magn Reson.* (2016) 18:6. doi: 10.1186/s12968-016-0225-6
- Kaminaga T, Takeshita T, Kimura I. Role of magnetic resonance imaging for evaluation of tumors in the cardiac region. *Eur Radiol.* (2003) 13:L1–10. doi: 10.1007/s00330-002-1789-0
- Buechel ERV, Grosse-Wortmann L, Fratz S, Eichhorn J, Sarikouch S, Greil GF, et al. Indications for cardiovascular magnetic resonance in children with congenital and acquired heart disease: An expert consensus paper of the Imaging Working Group of the AEPC and the Cardiovascular Magnetic Resonance Section of the EACVI. *Eur Heart J Cardiovasc Imaging.* (2015) 16:281–97. doi: 10.1093/ehjci/jeu129
- FDA. 510k Premarket Notification of HyperSense (GE Medical Systems). (2017). Available online at: https://www.accessdata.fda.gov/cdrh_docs/pdf16/K162722.pdf (accessed February 11, 2020).
- FDA. 510k Premarket Notification of Compressed Sensing Cardiac Cine (Siemens). (2017). Available online at: https://www.accessdata.fda.gov/cdrh_docs/pdf16/K163312.pdf (accessed February 11, 2020).
- FDA. 510k Premarket Notification of Compressed SENSE (Philips). (2018). Available online at: https://www.accessdata.fda.gov/cdrh_docs/pdf17/K173079.pdf (accessed February 11, 2020).
- Ye JC. Compressed sensing MRI: a review from signal processing perspective. *BMC Biomed Eng.* (2019) 1:8. doi: 10.1186/s42490-019-0006-z
- Jaspan ON, Fleysher R, Lipton ML. Compressed sensing MRI: a review of the clinical literature. *Br J Radiol.* (2015) 88:20150487. doi: 10.1259/bjr.20150487
- Hansen MS, Kellman P. Image reconstruction: an overview for clinicians. *J Magn Reson Imaging.* (2015) 41:573–85. doi: 10.1002/jmri.24687
- Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Reson Med.* (2007) 58:1182–95. doi: 10.1002/mrm.21391

13. Candès EJ, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inf Theory*. (2006) 52:489–509. doi: 10.1109/TIT.2005.862083
14. Usman M, Atkinson D, Odille F, Kolbitsch C, Vaillant G, Schaeffter T, et al. Motion corrected compressed sensing for free-breathing dynamic cardiac MRI. *Magn Reson Med*. (2013) 70:504–16. doi: 10.1002/mrm.24463
15. Wetzl J, Schmidt M, Pontana F, Longère B, Lugauer F, Maier A, et al. Single-breath-hold 3-D CINE imaging of the left ventricle using Cartesian sampling. *Magn Reson Mater Phys Biol Med*. (2018) 31:19–31. doi: 10.1007/s10334-017-0624-1
16. Otazo R, Kim D, Axel L, Sodickson DK. Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI. *Magn Reson Med*. (2010) 64:767–76. doi: 10.1002/mrm.22463
17. Kamesh Iyer S, Tasdizen T, Burgon N, Kholmovski E, Marrouche N, Adluru G, et al. Compressed sensing for rapid late gadolinium enhanced imaging of the left atrium: a preliminary study. *Magn Reson Imaging*. (2016) 34:846–54. doi: 10.1016/j.mri.2016.03.002
18. Walheim J, Dillinger H, Kozerke S. Multipoint 5D flow cardiovascular magnetic resonance - accelerated cardiac- and respiratory-motion resolved mapping of mean and turbulent velocities. *J Cardiovasc Magn Reson*. (2019) 21:42. doi: 10.1186/s12968-019-0549-0
19. Ma LE, Markl M, Chow K, Huh H, Forman C, Vali A, et al. Aortic 4D flow MRI in 2 minutes using compressed sensing, respiratory controlled adaptive k-space reordering, and inline reconstruction. *Magn Reson Med*. (2019) 81:3675–90. doi: 10.1002/mrm.27684
20. Tariq U, Hsiao A, Alley M, Zhang T, Lustig M, Vasanawala SS. Venous and arterial flow quantification are equally accurate and precise with parallel imaging compressed sensing 4D phase contrast MRI. *J Magn Reson Imaging*. (2013) 37:1419–26. doi: 10.1002/jmri.23936
21. Dyvorne H, Knight-Greenfield A, Jajamovich G, Besa C, Cui Y, Stalder A, et al. Abdominal 4D flow MR imaging in a breath hold: combination of spiral sampling and dynamic compressed sensing for highly accelerated acquisition. *Radiology*. (2015) 275:245–54. doi: 10.1148/radiol.14140973
22. Kido T, Kido T, Nakamura M, Watanabe K, Schmidt M, Forman C, et al. Compressed sensing real-time cine cardiovascular magnetic resonance: accurate assessment of left ventricular function in a single-breath-hold. *J Cardiovasc Magn Reson*. (2016) 18:1–11. doi: 10.1186/s12968-016-0271-0
23. Vermersch M, Longère B, Coisne A, Schmidt M, Forman C, Monnet A, et al. Compressed sensing real-time cine imaging for assessment of ventricular function, volumes and mass in clinical practice. *Eur Radiol*. (2019) 30:609–19. doi: 10.1007/s00330-019-06341-2
24. Piccini D, Littmann A, Nelles-Vallespin S, Zenge MO. Spiral phyllotaxis: the natural way to construct a 3D radial trajectory in MRI. *Magn Reson Med*. (2011) 66:1049–56. doi: 10.1002/mrm.22898
25. Usman M, Ruijsink B, Nazir MS, Cruz G, Prieto C. Free breathing whole-heart 3D CINE MRI with self-gated Cartesian trajectory. *Magn Reson Imaging*. (2017) 38:129–37. doi: 10.1016/j.mri.2016.12.021
26. Goldstein T, Osher S. The split bregman method for L1-regularized problems. *SIAM J Imaging Sci*. (2009) 2:323. doi: 10.1137/080725891
27. Basha TA, Akçakaya M, Liew C, Tsao CW, Delling FN, Addae G, et al. Clinical performance of high-resolution late gadolinium enhancement imaging with compressed sensing. *J Magn Reson Imaging*. (2017) 46:1829–38. doi: 10.1002/jmri.25695
28. Forman C, Piccini D, Grimm R, Hutter J, Hornegger J, Zenge MO. High-resolution 3D whole-heart coronary MRA: a study on the combination of data acquisition in multiple breath-holds and 1D residual respiratory motion compensation. *Magn Reson Mater Phys Biol Med*. (2014) 27:435–43. doi: 10.1007/s10334-013-0428-x
29. Forman C, Piccini D, Grimm R, Hutter J, Hornegger J, Zenge MO. Reduction of respiratory motion artifacts for free-breathing whole-heart coronary MRA by weighted iterative reconstruction. *Magn Reson Med*. (2015) 73:1885–95. doi: 10.1002/mrm.25321
30. Correia T, Cruz G, Schneider T, Botnar RM, Prieto C. Technical note: accelerated nonrigid motion-compensated isotropic 3D coronary MR angiography. *Med Phys*. (2018) 45:214–22. doi: 10.1002/mp.12663
31. Nakamura M, Kido T, Kido T, Watanabe K, Schmidt M, Forman C, et al. Non-contrast compressed sensing whole-heart coronary magnetic resonance angiography at 3T: a comparison with conventional imaging. *Eur J Radiol*. (2018) 104:43–8. doi: 10.1016/j.ejrad.2018.04.025
32. Addy NO, Ingle RR, Wu HH, Hu BS, Nishimura DG. High-resolution variable-density 3D cones coronary MRA. *Magn Reson Med*. (2015) 74:614–21. doi: 10.1002/mrm.25803
33. Nam S, Akçakaya M, Basha T, Stehning C, Manning WJ, Tarokh V, et al. Compressed sensing reconstruction for whole-heart imaging with 3D radial trajectories: a graphics processing unit implementation. *Magn Reson Med*. (2013) 69:91–102. doi: 10.1002/mrm.24234
34. Feng L, Axel L, Chandarana H, Block KT, Sodickson DK, Otazo R. XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing. *Magn Reson Med*. (2016) 75:775–88. doi: 10.1002/mrm.25665
35. Feng L, Coppo S, Piccini D, Yerly J, Lim RP, Masci PG, et al. 5D whole-heart sparse MRI. *Magn Reson Med*. (2018) 79:826–38. doi: 10.1002/mrm.26745
36. Menchón-Lara RM, Royuela-del-Val J, Godino-Moya A, Cordero-Grande L, Simmross-Wattenberg F, Martin-Fernandez M, et al. An efficient multi-resolution reconstruction scheme with motion compensation for 5d free-breathing whole-heart MRI. *LNCS Lect Notes Comput Sci Ger*. (2017) 10555:136–45. doi: 10.1007/978-3-319-67564-0_14
37. Di Sopra L, Piccini D, Coppo S, Stuber M, Yerly J. An automated approach to fully self-gated free-running cardiac and respiratory motion-resolved 5D whole-heart MRI. *Magn Reson Med*. (2019) 82:2118–32. doi: 10.1002/mrm.27898
38. Cheng JY, Zhang T, Alley MT, Uecker M, Lustig M, Pauly JM, et al. Comprehensive multi-dimensional MRI for the simultaneous assessment of cardiopulmonary anatomy and physiology. *Sci Rep*. (2017) 7:5330. doi: 10.1038/s41598-017-04676-8
39. Pedersen H, Kozerke S, Ringgaard S, Nehrke K, Won YK. K-t PCA: temporally constrained k-t BLAST reconstruction using principal component analysis. *Magn Reson Med*. (2009) 62:706–16. doi: 10.1002/mrm.22052
40. Zhao B, Haldar JP, Christodoulou AG, Liang Z-P. Image reconstruction from highly undersampled (k,t)-space data with joint partial separability and sparsity constraints. *Med Imaging IEEE Trans*. (2012) 31:1809–20. doi: 10.1109/TMI.2012.2203921
41. Haldar JP, Liang Z. Low-rank approximations for dynamic imaging. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. Arlington, VA, (2011). p. 1052–5. doi: 10.1109/ISBI.2011.5872582
42. Zhao B, Haldar JP, Brinegar C, Liang ZP. Low rank matrix recovery for real-time cardiac MRI. In: *2010 7th IEEE International Symposium on Biomedical Imaging From Nano to Macro, ISBI 2010 - Proc*. San Francisco, CA, (2010). p. 996–9. doi: 10.1109/ISBI.2010.5490156
43. Otazo R, Candès E, Sodickson DK. Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. *Magn Reson Med*. (2015) 73:1125–36. doi: 10.1002/mrm.25240
44. Hamilton JI, Jiang Y, Ma D, Chen Y, Lo WC, Griswold M, et al. Simultaneous multislice cardiac magnetic resonance fingerprinting using low rank reconstruction. *NMR Biomed*. (2019) 32:1–16. doi: 10.1002/nbm.4041
45. Liang Z. Spatiotemporal imaging with partially separable function. In: *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 12-15 April, 2007*. Arlington, VA (2007). p. 988–91. doi: 10.1109/ISBI.2007.357020
46. Lingala SG, Hu Y, Dibella E, Jacob M. Accelerated dynamic MRI exploiting sparsity and low-rank structure: K-t SLR. *IEEE Trans Med Imaging*. (2011) 30:1042–54. doi: 10.1109/TMI.2010.2100850
47. Tremoulet B, Dikaos N, Atkinson D, Arridge SR. Dynamic MR image reconstruction-separation from undersampled (k,t)-Space via low-rank plus sparse prior. *IEEE Trans Med Imaging*. (2014) 33:1689–701. doi: 10.1109/TMI.2014.2321190
48. Christodoulou AG, Shaw JL, Nguyen C, Yang Q, Xie Y, Wang N, et al. Magnetic resonance multitasking for motion-resolved quantitative cardiovascular imaging. *Nat Biomed Eng*. (2018) 2:215–26. doi: 10.1038/s41551-018-0217-y
49. Shaw JL, Yang Q, Zhou Z, Deng Z, Nguyen C, Li D, et al. Free-breathing, non-ECG, continuous myocardial T1 mapping with cardiovascular magnetic resonance multitasking. *Magn Reson Med*. (2019) 81:2450–63. doi: 10.1002/mrm.27574

50. Trzasko J, Manduca A. Local versus global low-rank promotion in dynamic MRI series reconstruction. In: *Proceedings of the 19th Annual Meeting of ISMRM*. Montréal, QC (2011). p. 4371
51. Liu F, Li D, Jin X, Qiu W, Xia Q, Sun B. Dynamic cardiac MRI reconstruction using motion aligned locally low rank tensor (MALLRT). *Magn Reson Imaging*. (2019) 66:104–15. doi: 10.1016/j.mri.2019.07.002
52. Miao X, Lingala SG, Guo Y, Jao T, Usman M, Prieto C, et al. Accelerated cardiac cine MRI using locally low rank and finite difference constraints. *Magn Reson Imaging*. (2016) 34:707–14. doi: 10.1016/j.mri.2016.03.007
53. Yaman B, Weingärtner S, Kargas N, Sidiropoulos N, Akçakaya M. Locally low-rank tensor regularization for high-resolution quantitative dynamic MRI. In: *7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. Guadeloupe (2017). p. 1–5. doi: 10.1109/CAMSAP.2017.8313075
54. Akçakaya M, Rayatzadeh H, Basha T, Hong SN, Chan RH, Kissinger KV, et al. Accelerated late gadolinium enhancement cardiac MR imaging with isotropic spatial resolution using compressed sensing: Initial experience. *Radiology*. (2012) 264:691–9. doi: 10.1148/radiol.12112489
55. Akçakaya M, Basha TA, Goddu B, Goddu B, Goepfert LA, Kissinger KV, et al. Low-dimensional-structure self-learning and thresholding: regularization beyond compressed sensing for MRI reconstruction. *Magn Reson Med*. (2011) 66:756–67. doi: 10.1002/mrm.22841
56. Bustin A, Ginami G, Cruz G, Correia T, Ismail TF, Rashid I, et al. Five-minute whole-heart coronary MRA with sub-millimeter isotropic resolution, 100% respiratory scan efficiency, and 3D-PROST reconstruction. *Magn Reson Med*. (2019) 81:102–15. doi: 10.1002/mrm.27354
57. Bustin A, Correia Teresa, Rashid I, Cruz G, Neji R, Botnar R, et al. Highly accelerated 3D whole-heart isotropic sub-millimeter CMRA with non-rigid motion correction. In: *Proceedings of the 27th Annual Meeting of ISMRM, 11-16 May, 2019*. Montréal, QC (2019). p. 978.
58. Mohsin YQ, Lingala SG, DiBella E, Jacob M. Accelerated dynamic MRI using patch regularization for implicit motion compensation. *Magn Reson Med*. (2017) 77:1238–48. doi: 10.1002/mrm.26215
59. Bustin A, Cruz G, Jaubert O, Lopez K, Botnar RM, Prieto C. High-dimensionality undersampled patch-based reconstruction (HD-PROST) for accelerated multi-contrast magnetic resonance imaging. *Magn Reson Med*. (2019) 81:3705–19. doi: 10.1002/mrm.27694
60. Qi H, Bustin A, Cruz G, Jaubert O, Chen H, Botnar RM, et al. Free-running simultaneous myocardial T1/T2 mapping and cine imaging with 3D whole-heart coverage and isotropic spatial resolution. *Magn Reson Imaging*. (2019) 63:159–69. doi: 10.1016/j.mri.2019.08.008
61. Bustin A, Milotta G, Prieto C. Accelerated free-breathing whole-heart 3D T2 mapping with high isotropic resolution. *Magn Reson Med*. (2020) 83:988–1002. doi: 10.1002/mrm.27989
62. Caballero J, Price AN, Rueckert D, Hajnal JV. Dictionary learning and time sparsity for dynamic MR data reconstruction. *IEEE Trans Med Imaging*. (2014) 33:979–94. doi: 10.1109/TMI.2014.2301271
63. Wang Y, Ying L. Compressed sensing dynamic cardiac cine MRI using learned spatiotemporal dictionary. *IEEE Trans Biomed Eng*. (2014) 61:1109–20. doi: 10.1109/TBME.2013.2294939
64. Lingala SG, Jacob M. Blind compressive sensing dynamic MRI. *IEEE Trans Med Imaging*. (2013) 32:1132–45. doi: 10.1109/TMI.2013.2255133
65. Goodfellow I, Bengio Y, Courville A. Deep learning. *Nat Methods*. (2015) 13:35. doi: 10.1038/nmeth.3707
66. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV (2016). p. 770–8. doi: 10.1109/CVPR.2016.90
67. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *Med Image Comput Comput Interv*. (2015) 234–41. doi: 10.1007/978-3-319-24574-4_28
68. Roth S, Black MJ. Fields of experts: a framework for learning image priors. *Proc IEEE CVPR*. (2005) 2:860–7. doi: 10.1109/CVPR.2005.160
69. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv [Preprint]*. *arXiv 2014.1412.6980* (2014).
70. LeCun YA, Bottou L, Orr GB, Müller KR. Efficient BackProp. In: Montavon G, Orr GB, Müller KR, editors. *Neural Networks: Tricks of the Trade*. Lecture Notes in Computer Science. Berlin; Heidelberg: Springer. (2012). p. 7700. doi: 10.1007/978-3-642-35289-8_3
71. Yang G, Yu S, Dong H, Slabaugh G, Dragotti PL, Ye X, et al. DAGAN: Deep De-Aliasing Generative Adversarial Networks for fast compressed sensing MRI reconstruction. *IEEE Trans Med Imaging*. (2018) 37:1310–21. doi: 10.1109/TMI.2017.2785879
72. Quan TM, Nguyen-Duc T, Jeong WK. Compressed sensing MRI Reconstruction using a generative adversarial network with a cyclic loss. *IEEE Trans Med Imaging*. (2018) 37:1488–97. doi: 10.1109/TMI.2018.2820120
73. Lucas A, Iliadis M, Molina R, Katsaggelos AK. Using deep neural networks for inverse problems in imaging: beyond analytical methods. *IEEE Signal Process Mag*. (2018) 35:20–36. doi: 10.1109/MSP.2017.2760358
74. Hauptmann A, Arridge S, Lucka F, Muthurangu V, Steeden JA. Real-time cardiovascular MR with spatio-temporal artifact suppression using deep learning—proof of concept in congenital heart disease. *Magn Reson Med*. (2019) 81:1143–56. doi: 10.1002/mrm.27480
75. Kofler A, Dewey M, Schaeffter T, Wald C, Kolbitsch C. Spatio-temporal deep learning-based undersampling artefact reduction for 2D radial cine MRI with limited training data. *IEEE Trans Med Imaging*. (2019) 2019:2930318. doi: 10.1109/TMI.2019.2930318
76. Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging*. (2018) 37:491–503. doi: 10.1109/TMI.2017.2760978
77. Fuin N, Bustin A, Kuestner T, Botnar R, Prieto C. A variational neural network for accelerating free-breathing whole-heart coronary MR angiography. In: *Proceedings of the 27th Annual Meeting of ISMRM, 11-16 May, 2019*. Montréal, QC (2019). p. 478.
78. Biswas S, Aggarwal HK, Jacob M. Dynamic MRI using model-based deep learning and STORM priors: MoDL-STORM. *Magn Reson Med*. (2019) 82:485–94. doi: 10.1002/mrm.27706
79. Qin C, Schlemper J, Caballero J, Price AN, Hajnal JV, Rueckert D. Convolutional recurrent neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging*. (2019) 38:280–90. doi: 10.1109/TMI.2018.2863670
80. Akçakaya M, Moeller S, Weingärtner S, Ugurbil K. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: database-free deep learning for fast imaging. *Magn Reson Med*. (2019) 81:439–53. doi: 10.1002/mrm.27420
81. Wang S, Ke Z, Cheng H, Jia S, Ying L, Zheng H, et al. DIMENSION: dynamic MR imaging with both K-space and spatial prior knowledge obtained via multi-supervised network training. *NMR Biomed*. (2019) e4131. doi: 10.1002/nbm.4131. [Epub ahead of print].
82. Feng L, Grimm R, Block KT, Chandarana H, Kim S, Xu J, et al. Golden-angle radial sparse parallel MRI: combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI. *Magn Reson Med*. (2014) 72:707–17. doi: 10.1002/mrm.24980
83. Yang Y, Sun J, Li H, Xu Z. Deep ADMM-Net for compressive sensing MRI. In: *Proceedings of NIPS*. Barcelona (2016). p. 10–8.
84. Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*. (2018) 79:3055–71. doi: 10.1002/mrm.26977
85. Ravishanker S, Bresler Y. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans Med Imaging*. (2011) 30:1028–41. doi: 10.1109/TMI.2010.2090538
86. Aggarwal HK, Mani MP, Jacob M. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Trans Med Imaging*. (2019) 38:394–405. doi: 10.1109/TMI.2018.2865356

87. Poddar S, Jacob M. Dynamic MRI using SmooThness Regularization on Manifolds (SToRM). *IEEE Trans Med Imaging*. (2016) 35:1106–15. doi: 10.1109/TMI.2015.2509245
88. Jung H, Sung K, Nayak KS, Kim EY, Ye JC. k-t FOCUSS: a general compressed sensing framework for high resolution dynamic MRI. *Magn Reson Med*. (2009) 61:103–16. doi: 10.1002/mrm.21757
89. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med*. (2002) 47:1202–10. doi: 10.1002/mrm.10171
90. Chang Y, Liang D, Ying L. Nonlinear GRAPPA: a kernel approach to parallel MRI reconstruction. *Magn Reson Med*. (2012) 68:730–40. doi: 10.1002/mrm.23279
91. Oksuz I, Clough J, Bustin A, Cruz G, Prieto C, Botnar R, et al. Cardiac MR motion artefact correction from k-space using deep learning-based reconstruction. *Mach Learn Med Image Reconstr*. (2018) 11074:21–9. doi: 10.1007/978-3-030-00129-2_3
92. Hansen MS, Sørensen TS. Gadgetron: An open source framework for medical image reconstruction. *Magn Reson Med*. (2013) 69:1768–76. doi: 10.1002/mrm.24389
93. Pushparajah K, Chubb H, Razavi R. MR-guided cardiac interventions. *Top Magn Reson Imaging*. (2018) 27:115–28. doi: 10.1097/RMR.0000000000000156

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Bustin, Fuin, Botnar and Prieto. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Deep Learning for Cardiac Image Segmentation: A Review

Chen Chen^{1*}, Chen Qin¹, Huaqi Qiu¹, Giacomo Tarroni^{1,2}, Jinming Duan³, Wenjia Bai^{4,5} and Daniel Rueckert¹

¹ Biomedical Image Analysis Group, Department of Computing, Imperial College London, London, United Kingdom,

² CitAI Research Centre, Department of Computer Science, City University of London, London, United Kingdom, ³ School of Computer Science, University of Birmingham, Birmingham, United Kingdom, ⁴ Data Science Institute, Imperial College London, London, United Kingdom, ⁵ Department of Brain Sciences, Faculty of Medicine, Imperial College London, London, United Kingdom

OPEN ACCESS

Edited by:

Karim Lekadir,
University of Barcelona, Spain

Reviewed by:

Jichao Zhao,
The University of Auckland,
New Zealand

Marta Nuñez-García,
Institut de Rythmologie et
Modélisation Cardiaque
(IHU-Liryc), France

*Correspondence:

Chen Chen
chen.chen15@imperial.ac.uk

Specialty section:

This article was submitted to
Cardiovascular Imaging,
a section of the journal
Frontiers in Cardiovascular Medicine

Received: 30 October 2019

Accepted: 17 February 2020

Published: 05 March 2020

Citation:

Chen C, Qin C, Qiu H, Tarroni G,
Duan J, Bai W and Rueckert D (2020)
Deep Learning for Cardiac Image
Segmentation: A Review.
Front. Cardiovasc. Med. 7:25.
doi: 10.3389/fcvm.2020.00025

Deep learning has become the most widely used approach for cardiac image segmentation in recent years. In this paper, we provide a review of over 100 cardiac image segmentation papers using deep learning, which covers common imaging modalities including magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound and major anatomical structures of interest (ventricles, atria, and vessels). In addition, a summary of publicly available cardiac image datasets and code repositories are included to provide a base for encouraging reproducible research. Finally, we discuss the challenges and limitations with current deep learning-based approaches (scarcity of labels, model generalizability across different domains, interpretability) and suggest potential directions for future research.

Keywords: artificial intelligence, deep learning, neural networks, cardiac image segmentation, cardiac image analysis, MRI, CT, ultrasound

1. INTRODUCTION

Cardiovascular diseases (CVDs) are the leading cause of death globally according to World Health Organization (WHO). About 17.9 million people died from CVDs in 2016, from CVD, mainly from heart disease and stroke¹. The number is still increasing annually. In recent decades, major advances have been made in cardiovascular research and practice aiming to improve diagnosis and treatment of cardiac diseases as well as reducing the mortality of CVD. Modern medical imaging techniques, such as magnetic resonance imaging (MRI), computed tomography (CT) and ultrasound are now widely used, which enable non-invasive qualitative and quantitative assessment of cardiac anatomical structures and functions and provide support for diagnosis, disease monitoring, treatment planning, and prognosis.

Of particular interest, cardiac image segmentation is an important first step in numerous applications. It partitions the image into a number of semantically (i.e., anatomically) meaningful regions, based on which quantitative measures can be extracted, such as the myocardial mass, wall thickness, left ventricle (LV) and right ventricle (RV) volume as well as ejection fraction (EF) etc. Typically, the anatomical structures of interest for cardiac image segmentation include the LV, RV, left atrium (LA), right atrium (RA), and coronary arteries. An overview of typical tasks related to cardiac image segmentation is presented in **Figure 1**, where applications for the three most commonly used modalities, i.e., MRI, CT, and ultrasound, are shown.

¹ https://www.who.int/cardiovascular_diseases/about_cvd/en/

Before the rise of deep learning, traditional machine learning techniques, such as model-based methods (e.g., active shape and appearance models) and atlas-based methods had been shown to achieve good performance in cardiac image segmentation (1–4). However, they often require significant feature engineering or prior knowledge to achieve satisfactory accuracy. In contrast, deep learning (DL)-based algorithms are good at *automatically* discovering intricate features from data for object detection and segmentation. These features are directly learned from data using a general-purpose learning procedure and in end-to-end fashion. This makes DL-based algorithms easy to apply to other image analysis applications. Benefiting from advanced computer hardware [e.g., graphical processing units (GPUs) and tensor processing units (TPUs)] as well as increased available data for training, DL-based segmentation algorithms have gradually outperformed previous state-of-the-art traditional methods, gaining more popularity in research. This trend can be observed in **Figure 2A**, which shows how the number of DL-based papers for cardiac image segmentation has increased strongly in the last years. In particular, the number of the publications for MR image segmentation is significantly higher than the numbers of the other two domains, especially in 2017. One reason, which can be observed in **Figure 2B**, is that the publicly available data for MR segmentation has increased remarkably since 2016.

In this paper, we provide an overview of state-of-the-art deep learning techniques for cardiac image segmentation in the three most commonly used modalities (i.e., MRI, CT, ultrasound)

Abbreviations: Imaging-related terminology: CT, computed tomography; CTA, computed tomography angiography; LAX, long-axis; MPR, multi-planar reformatted; MR, magnetic resonance; MRI, magnetic resonance imaging; LGE, late gadolinium enhancement; RFCA, radio-frequency catheter ablation; SAX, short-axis; 2CH, 2-chamber; 3CH, 3-chamber; 4CH, 4-chamber.

Cardiac structures and indexes: AF, atrial fibrillation; AS, aortic stenosis; AO, aorta; CVD, cardiovascular diseases; CAC, coronary artery calcium; DCM, dilated cardiomyopathy; ED, end-diastole; ES, end-systole; EF, ejection fraction; HCM, hypertrophic cardiomyopathy; LA, left atrium; LV, left ventricle; LVEDV, left ventricular end-diastolic volume; LVESV, left ventricular end-systolic volume; MCP, mixed-calcified plaque; MI, myocardial infarction; Myo, left ventricular myocardium; NCP, non-calcified plaque; PA, pulmonary artery; PV, pulmonary vein; RA, right atrium; RV, right ventricle; RVEDV, right ventricular end-diastolic volume; RVESV, right ventricular end-systolic volume; RVEF, right ventricular ejection fraction; WHS, whole heart segmentation.

Machine learning terminology: AE, autoencoder; ASM, active shape model; BN, batch normalization; CONV, convolution; CNN, convolutional neural network; CRF, conditional random field; DBN, deep belief network; DL, deep learning; DNN, deep neural network; EM, expectation maximization; FCN, fully convolutional neural network; GAN, generative adversarial network; GRU, gated recurrent units; MSE, mean squared error; MSL, marginal space learning; MRF, markov random field; LSTM, Long-short term memory; ReLU, rectified linear unit; RNN, recurrent neural network; ROI, region-of-interest; SMC, sequential monte carlo; SRF, structured random forest; SVM, support vector machine.

Cardiac image segmentation datasets: ACDC, Automated Cardiac Diagnosis Challenge; CETUS, Challenge on Endocardial Three-dimensional Ultrasound Segmentation; MM-WHS, Multi-Modality Whole Heart Segmentation; LASC, Left Atrium Segmentation Challenge; LVSC, Left Ventricle Segmentation Challenge; RVSC, Right Ventricle Segmentation Challenge.

Others: EMBC, The International Engineering in Medicine and Biology Conference; GDPR, The General Data Protection Regulation; GPU, graphic processing unit; FDA, United States Food and Drug Administration; ISBI, The IEEE International Symposium on Biomedical Imaging; MICCAI, International Conference on Medical Image Computing and Computer-assisted Intervention; TPU, tensor processing unit; WHO, World Health Organization.

in clinical practice and discuss the advantages and remaining limitations of current deep learning-based segmentation methods that hinder widespread clinical deployment. To our knowledge, there have been several review papers that presented overviews about applications of DL-based methods for general medical image analysis (5–7), as well as some surveys dedicated to applications designed for cardiovascular image analysis (8, 9). However, none of them has provided a systematic overview focused on *cardiac segmentation applications*. This review paper aims at providing a comprehensive overview from the debut to the state-of-the-art of deep learning algorithms, focusing on a variety of cardiac image segmentation tasks (e.g., the LV, RV, and vessel segmentation) (section 3). Particularly, we aim to cover most influential DL-related works in this field published until 1st August 2019 and categorized these publications in terms of specific methodology. Besides, in addition to the basics of deep learning introduced in section 2, we also provide a summary of public datasets (see **Table 6**) as well as public code (see **Table 7**), aiming to present a good reading basis for newcomers to the topic and encourage future contributions. More importantly, we provide insightful discussions about the current research situations (section 3.4) as well as challenges and potential directions for future work (section 4).

1.1. Search Criterion

To identify related contributions, search engines like Scopus and PubMed were queried for papers containing (“convolutional” OR “deep learning”) and (“cardiac”) and (“image segmentation”) in title or abstract. Additionally, conference proceedings for MICCAI, ISBI, and EMBC were searched based on the titles of papers. Papers which do not primarily focus on segmentation problems were excluded. The last update to the included papers was on Aug 1, 2019.

2. FUNDAMENTALS OF DEEP LEARNING

Deep learning models are deep artificial neural networks. Each neural network consists of an input layer, an output layer, and multiple hidden layers. In the following section, we will review several deep learning networks and key techniques that have been commonly used in state-of-the-art segmentation algorithms. For a more detailed and thorough illustration of the mathematical background and fundamentals of deep learning we refer the interested reader to Goodfellow (43).

2.1. Neural Networks

In this section, we first introduce basic neural network architectures and then briefly introduce building blocks which are commonly used to boost the ability of the networks to learn features that are useful for image segmentation.

2.1.1. Convolutional Neural Networks (CNNs)

In this part, we will introduce convolutional neural network (CNN), which is the most common type of deep neural networks for image analysis. CNN have been successfully applied to advance the state-of-the-art on many image classification, object detection and segmentation tasks.

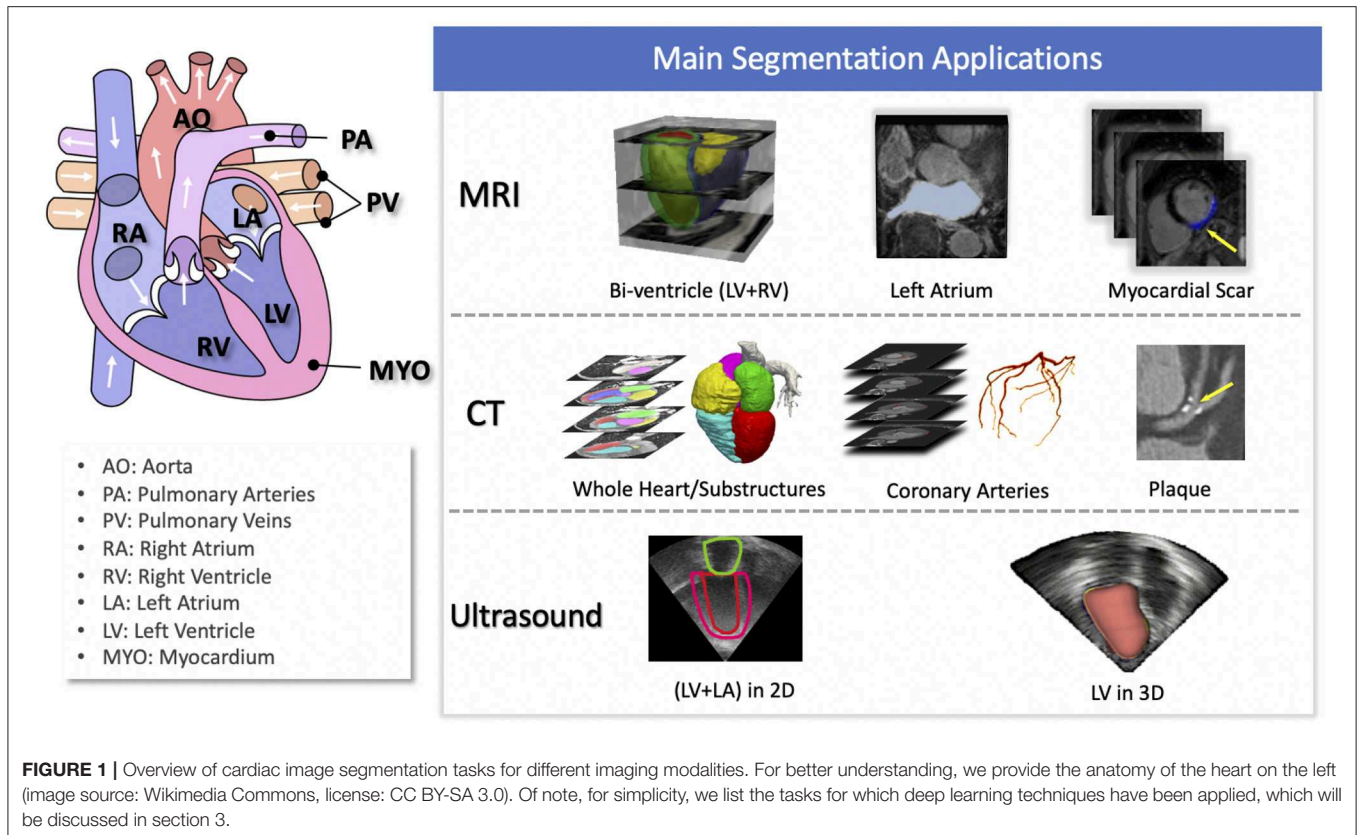


FIGURE 1 | Overview of cardiac image segmentation tasks for different imaging modalities. For better understanding, we provide the anatomy of the heart on the left (image source: Wikimedia Commons, license: CC BY-SA 3.0). Of note, for simplicity, we list the tasks for which deep learning techniques have been applied, which will be discussed in section 3.

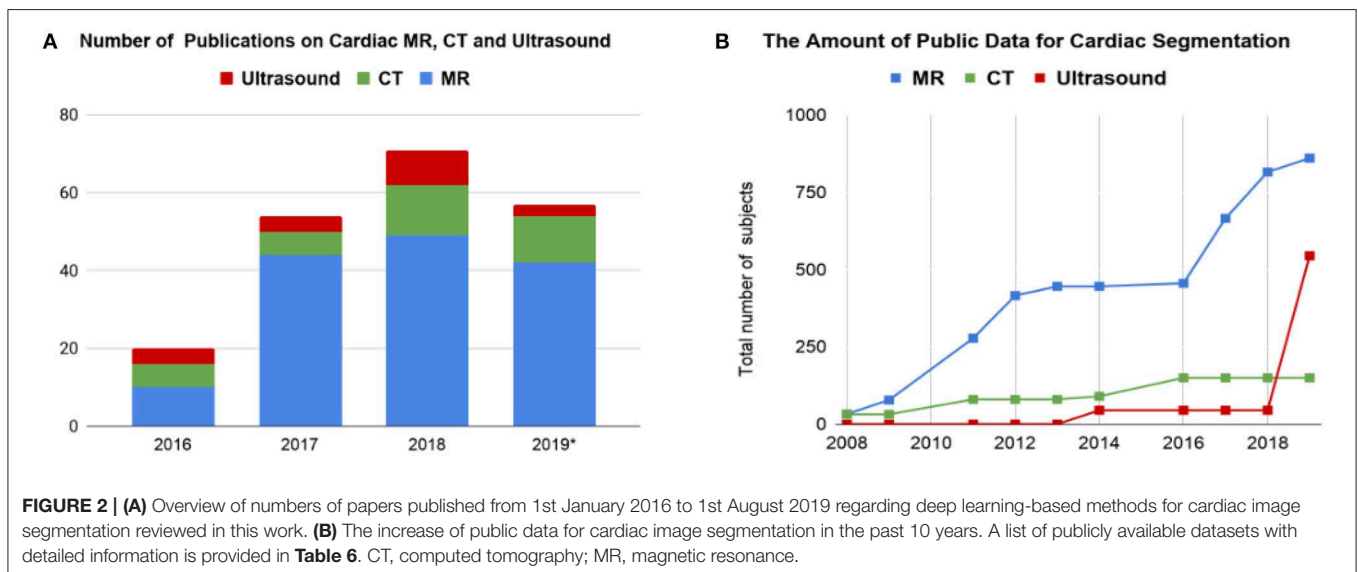


FIGURE 2 | (A) Overview of numbers of papers published from 1st January 2016 to 1st August 2019 regarding deep learning-based methods for cardiac image segmentation reviewed in this work. **(B)** The increase of public data for cardiac image segmentation in the past 10 years. A list of publicly available datasets with detailed information is provided in **Table 6**. CT, computed tomography; MR, magnetic resonance.

As shown in **Figure 3A**, a standard CNN consists of an input layer, an output layer and a stack of functional layers in between that transform an input into an output in a specific form (e.g., vectors). These functional layers often contains convolutional layers, pooling layers and/or fully-connected layers. In general, a convolutional layer $CONV_l$ contains k_l convolution kernels/filters, which is followed by a normalization

layer [e.g., batch normalization (44)] and a non-linear activation function [e.g., rectified linear unit (ReLU)] to extract k_l feature maps from the input. These feature maps are then downsampled by pooling layers, typically by a factor of 2, which remove redundant features to improve the statistical efficiency and model generalization. After that, fully connected layers are applied to reduce the dimension of features from its previous layer and

find the most task-relevant features for inference. The output of the network is a fix-sized vector where each element can be a probabilistic score for each category (for image classification), a real value for a regression task (e.g., the left ventricular volume estimation) or a set of values (e.g., the coordinates of a bounding box for object detection and localization).

A key component of CNN is the convolutional layer. Each convolutional layer has k_l convolution kernels to extract k_l feature maps and the size of each kernel n is chosen to be small in general, e.g., $n = 3$ for a 2D 3×3 kernel, to reduce the number of parameters². While the kernels are small, one can increase the receptive field (the area of the input image that potentially impacts the activation of a particular convolutional kernel/neuron) by increasing the number of convolutional layers. For example, a convolutional layer with large 7×7 kernels can be replaced by three layers with small 3×3 kernels (45). The number of weights is reduced by a factor of $7^2/(3 \times (3^2)) \approx 2$ while the receptive field remains the same (7×7). An online resource³ is referred here, which illustrates and visualizes the change of receptive field by varying the number of hidden layers and the size of kernels. In general, increasing the depth of convolution neural networks (the number of hidden layers) to enlarge the receptive field can lead to improved model performance, e.g., classification accuracy (45).

CNNs for image classification can also be employed for image segmentation applications without major adaptations to the network architecture (46), as shown in **Figure 3B**. However, this requires to divide each image into patches and then train a CNN to predict the class label of the center pixel for every patch. One major disadvantage of this patch-based approach is that, at inference time, the network has to be deployed for every patch individually despite the fact that there is a lot of redundancy due to multiple overlapping patches in the image. As a result of this inefficiency, the main application of CNNs with fully connected layers for cardiac segmentation is object localization, which aims to estimate the bounding box of the object of interest in an image. This bounding box is then used to crop the image, forming an image pre-processing step to reduce the computational cost for segmentation (47). For efficient, end-to-end pixel-wise segmentation, a variant of CNNs called fully convolutional neural network (FCN) is more commonly used, which will be discussed in the next section.

2.1.2. Fully Convolutional Neural Networks (FCNs)

The idea of FCN was first introduced by Long et al. (48) for image segmentation. FCNs are a special type of CNNs that

do not have any fully connected layers. In general, as shown in **Figure 4A**, FCNs are designed to have an encoder-decoder structure such that they can take input of arbitrary size and produce the output with the same size. Given an input image, the encoder first transforms the input into high-level feature representation whereas the decoder interprets the feature maps and recovers spatial details back to the image space for pixel-wise prediction through a series of upsampling and convolution operations. Here, upsampling can be achieved by applying transposed convolutions, e.g., 3×3 transposed convolutional kernels with a stride of 2 to up-scale feature maps by a factor of 2. These transposed convolutions can also be replaced by unpooling layers and upsampling layers. Compared to a patch-based CNN for segmentation, FCN is trained and applied to the entire images, removing the need for patch selection (50).

FCN with the simple encoder-decoder structure in **Figure 4A** may be limited to capture detailed context information in an image for precise segmentation as some features may be eliminated by the pooling layers in the encoder. Several variants of FCNs have been proposed to propagate features from the encoder to the decoder, in order to boost the segmentation accuracy. The most well-known and most popular variant of FCNs for biomedical image segmentation is the U-net (49). On the basis of the vanilla FCN (48), the U-net employs skip connections between the encoder and decoder to recover spatial context loss in the down-sampling path, yielding more precise segmentation (see **Figure 4B**). Several state-of-the-art cardiac image segmentation methods have adopted the U-net or its 3D variants, the 3D U-net (51) and the 3D V-net (52), as their backbone networks, achieving promising segmentation accuracy for a number of cardiac segmentation tasks (26, 53, 54).

2.1.3. Recurrent Neural Networks (RNNs)

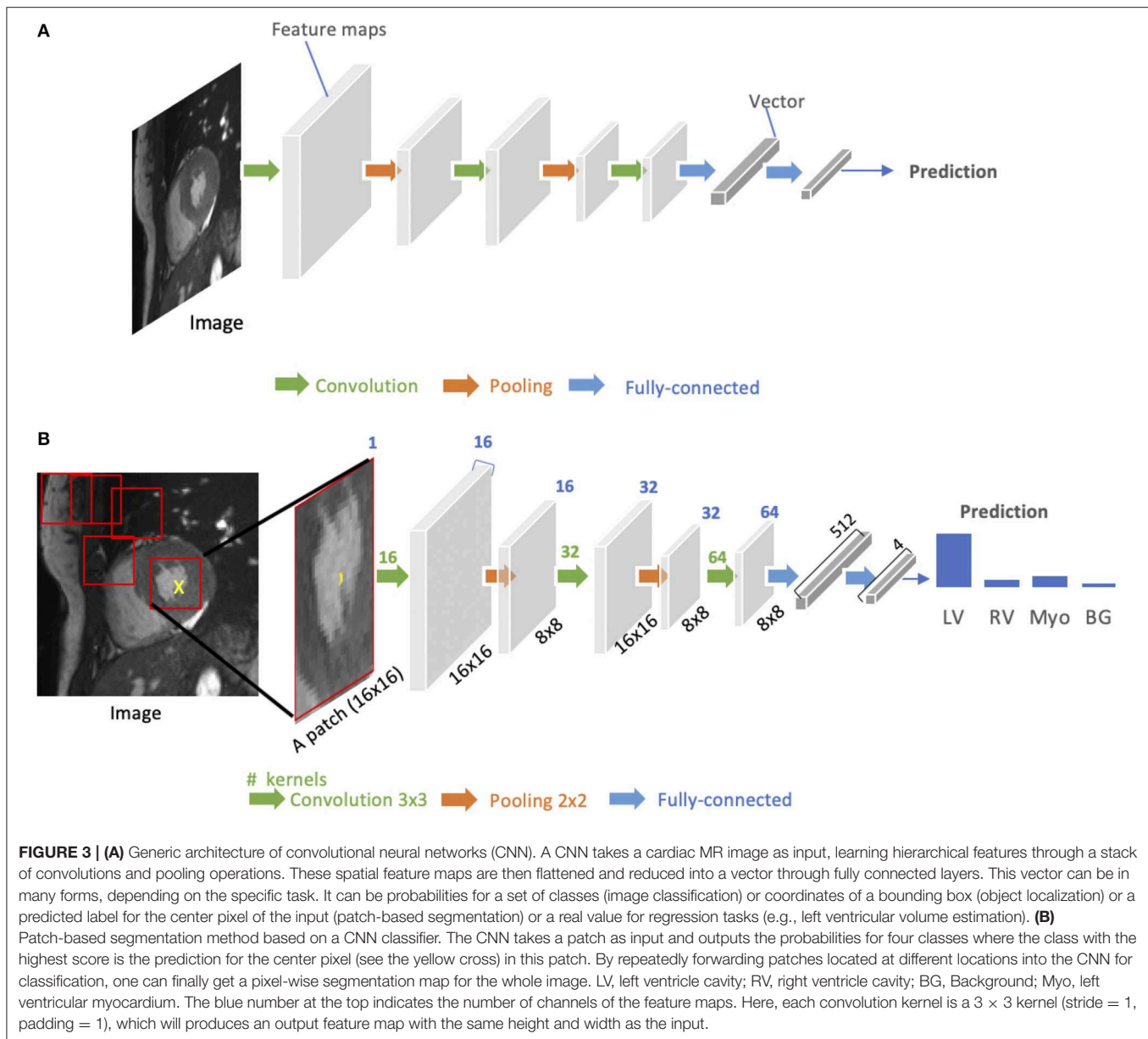
Recurrent neural networks (RNNs) are another type of artificial neural networks which are used for sequential data, such as cine MRI and ultrasound image sequences. An RNN can “remember” the past and use the knowledge learned from the past to make its present decision (see **Figures 5A,B**). For example, given a sequence of images, an RNN takes the first image as input, captures the information to make a prediction and then memorize this information which is then utilized to make a prediction for the next image. The two most widely used architectures in the family of RNNs are LSTM (56) and gated recurrent unit (GRU) (57), which are capable of modeling long-term memory. A use case for cardiac segmentation is to combine an RNN with a 2D FCN so that the combined network is capable of capturing information from adjacent slices to improve the inter-slice coherence of segmentation results (55).

2.1.4. Autoencoders (AE)

Autoencoders (AEs) are a type of neural networks that are designed to learn compact latent representations from data without supervision. A typical architecture of an autoencoder consists of two networks: an encoder network and a decoder network for the reconstruction of the input (see **Figure 6**). Since the learned representations contain generally useful information in the original data, many researchers have

²In a convolution layer l with k_l 2D $n \times n$ convolution kernels, each convolution kernel $\text{CONV}_l^{(i)}$, $i \in (1, k_l)$ has a weight matrix $\mathbf{w}_l^{(i)}$ and a bias term $b_l^{(i)}$ as parameters and can be formulated as: $\mathbf{y} = \mathbf{w}_l^{(i)} \circ \mathbf{x}_{in} + b_l^{(i)}$, where $\mathbf{w}_l^{(i)} \in \mathbb{R}^{n \times n \times l_{in}}$, $b_l^{(i)} \in \mathbb{R}$, $\mathbf{x}_{in} \in \mathbb{R}^{H \times W \times l_{in}}$, $\mathbf{y} \in \mathbb{R}^{H' \times W' \times k_l}$, l_{in} denotes the number of channels in the input \mathbf{x}_{in} and \circ denotes the convolution operation. Thus, the number of parameters in a convolutional layer is $k_l \times (n^2 \times l_{in} + 1)$. For a convolutional layer with $16 \times 3 \times 3$ filters where the input is a $28 \times 28 \times 1$ 2D gray image, the number of parameters in this layer is $16 \times (3^2 \times 1 + 1) = 160$. For more technical details about convolutional neural networks, an online tutorial is referred here: <http://cs231n.github.io/convolutional-networks>.

³<https://fomoro.com/research/article/receptive-field-calculator>



employed autoencoders to extract general semantic features or shape information from input images or labels and then use those features to guide the cardiac image segmentation (58, 62, 63).

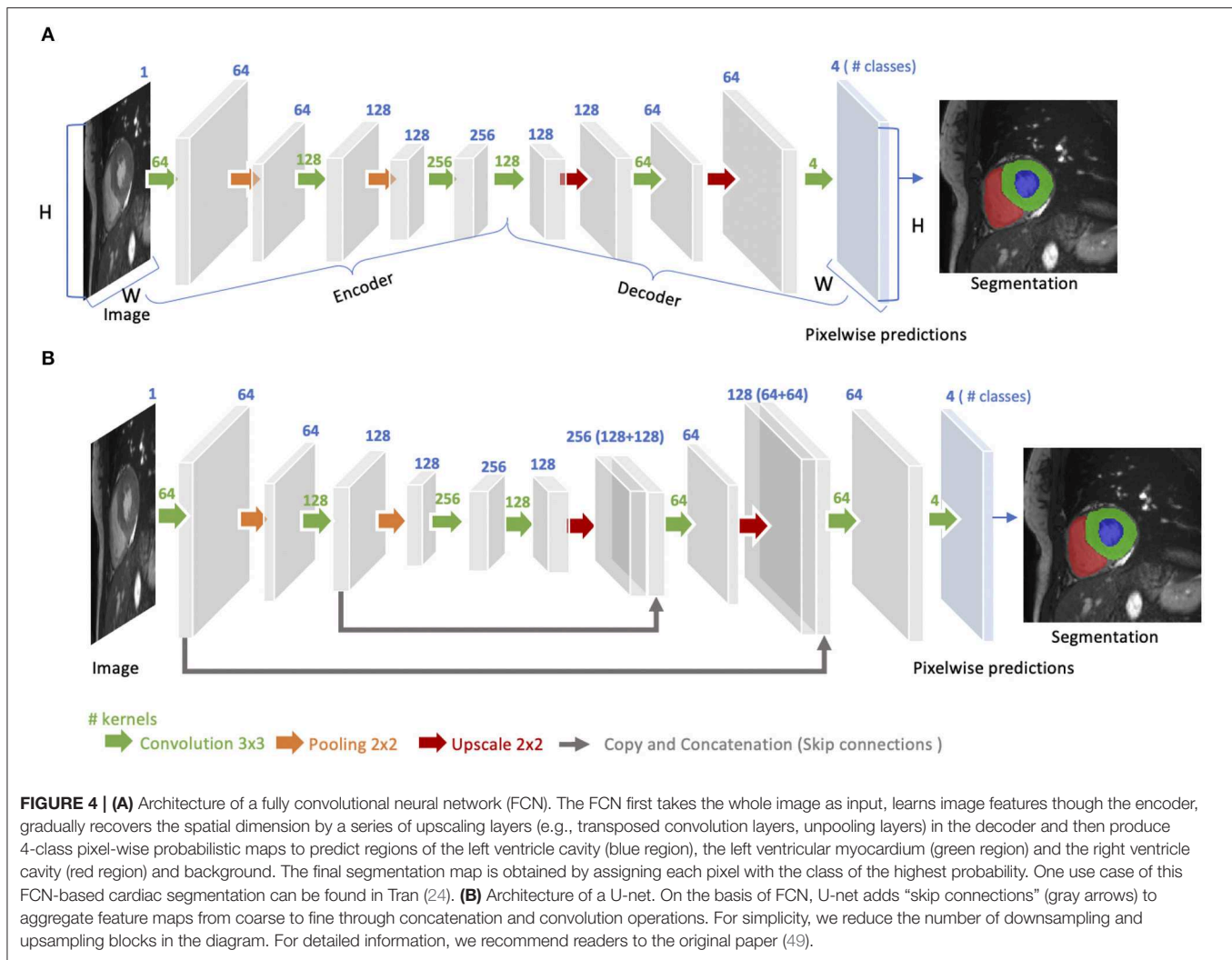
2.1.5. Generative Adversarial Networks (GAN)

The concept of Generative adversarial network (GAN) was proposed by Goodfellow et al. (64) for image synthesis from noise. GANs are a type of generative models that learn to model the data distribution of real data and thus are able to create new image examples. As shown in **Figure 7A**, a GAN consists of two networks: a generator network and a discriminator network. During training, the two networks are trained to compete against each other: the generator produces fake images aimed at fooling the discriminator, whereas the discriminator tries to identify real images from fake ones. This type of training is referred to as

“adversarial training,” since the two models are both set to win the competition. This training scheme can also be used for training a segmentation network. As shown in **Figure 7B**, the generator is replaced by a segmentation network and the discriminator is required to distinguish the generated segmentation maps from the ground truth ones (the target segmentation maps). In this way, the segmentation network is encouraged to produce more anatomically plausible segmentation maps (65, 66).

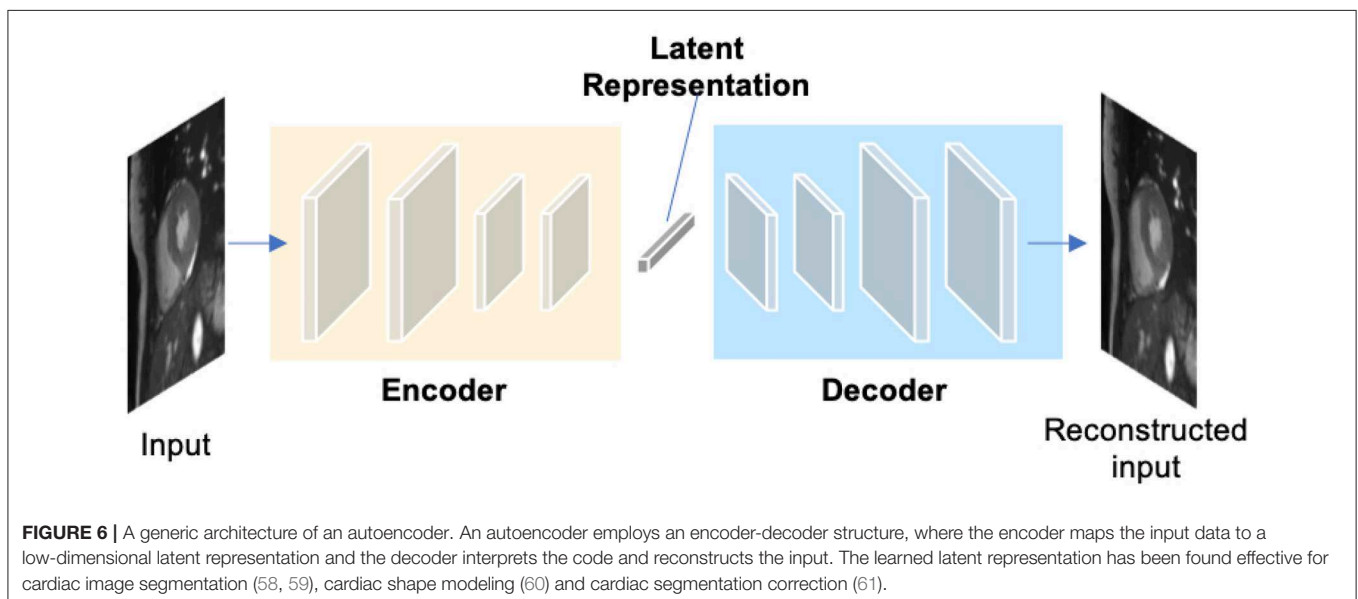
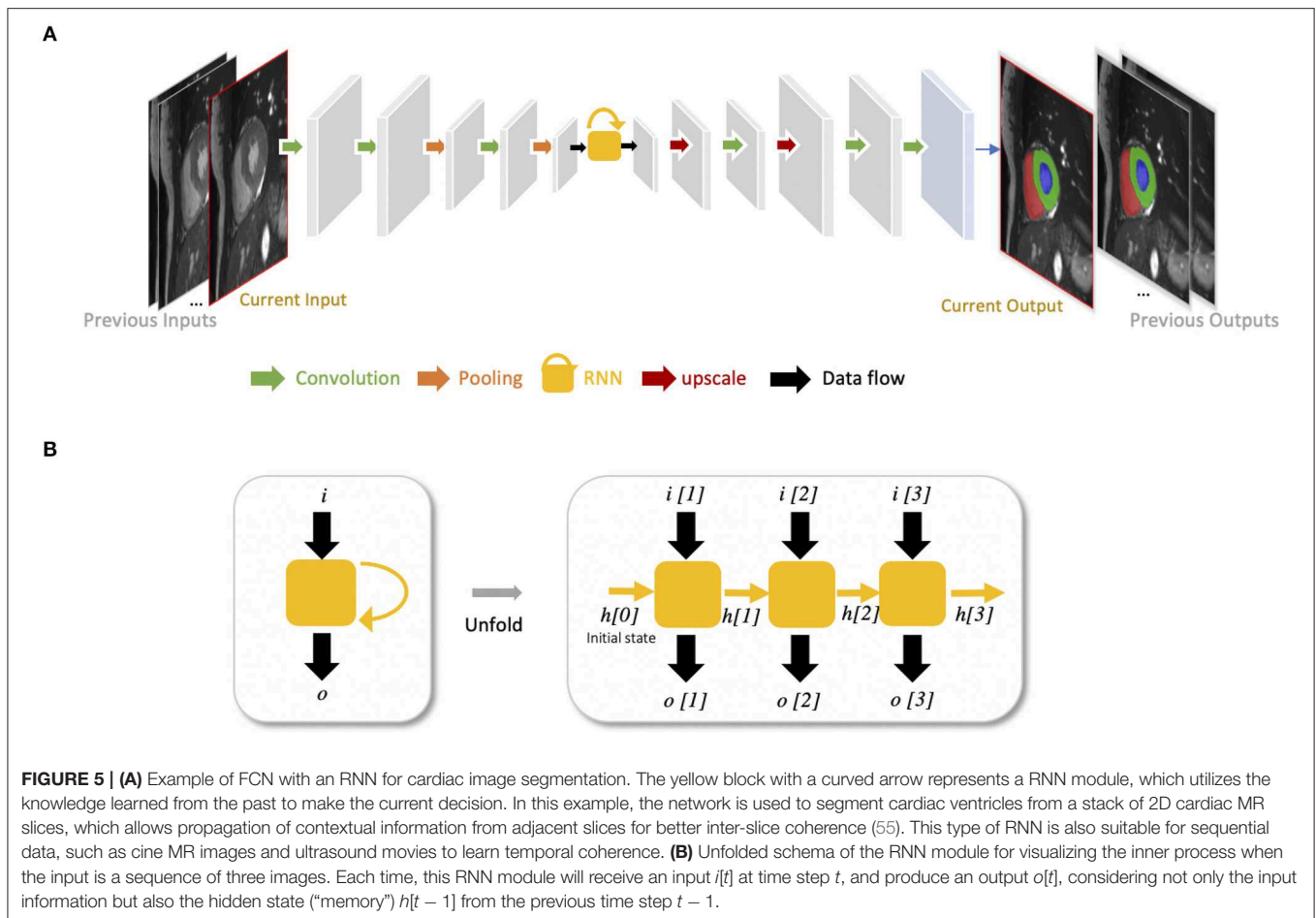
2.1.6. Advanced Building Blocks for Improved Segmentation

Medical image segmentation, as an important step for quantitative analysis and clinical research, requires high pixel-wise accuracy. Over the past years, many researchers have developed advanced building blocks to learn robust,



representative features for precise segmentation. These techniques have been widely applied to state-of-the-art neural networks (e.g., U-net) to improve cardiac image segmentation performance. Therefore, we identified several important techniques reported in the literature to this end and present them with corresponding references for further reading. These techniques are:

1. Advanced convolutional modules for multi-scale feature aggregation:
 - Inception modules (44, 67, 68), which concatenate multiple convolutional filter banks with different kernel sizes to extract multi-scale features in parallel (see **Figure 8A**);
 - Dilated convolutional kernels (72), which are modified convolution kernels with the same kernel size but different kernel strides to process input feature maps at larger scales;
 - Deep supervision (73), which utilizes the outputs from multiple intermediate hidden layers for multi-scale prediction;
2. Adaptive convolutional kernels designed to focus on important features:
 - Atrous spatial pyramid pooling (74), which applies spatial pyramid pooling (75) with various kernel strides to input feature maps for multi-scale feature fusion;
3. Interlayer connections designed to reuse features from previous layers:
 - Residual connections (71), which add outputs from a previous layer to the feature maps learned from the current layer (see **Figure 8C**);
 - Dense connections (78), which concatenate outputs from all preceding layers to the feature maps learned from the current layer.



2.2. Training Neural Networks

Before being able to perform inference, neural networks must be trained. Standard training process requires a dataset that contains paired images and labels $\{x, y\}$ for training and testing,

an optimizer (e.g., stochastic gradient descent, Adam) and a loss function to update the model parameters. This function accounts for the error of the network prediction in each iteration during training, providing signals for the optimizer to update the

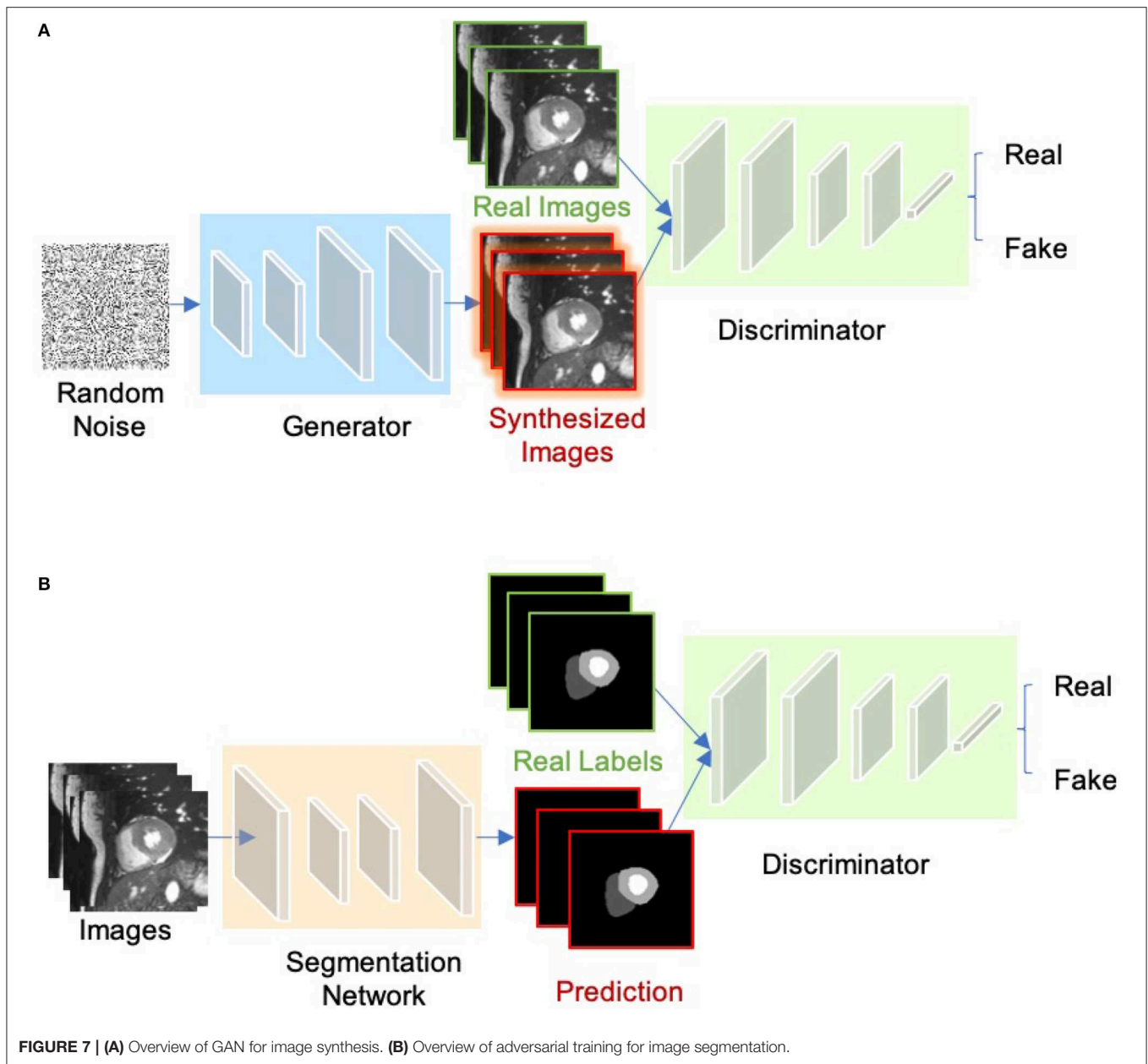


FIGURE 7 | (A) Overview of GAN for image synthesis. **(B)** Overview of adversarial training for image segmentation.

network parameters through backpropagation (43, 79). The goal of training is to find proper values of the network parameters to minimize the loss function.

2.2.1. Common Loss Functions

For regression tasks (e.g., heart localization, calcium scoring, landmark detection, image reconstruction), the simplest loss function is the mean squared error (MSE):

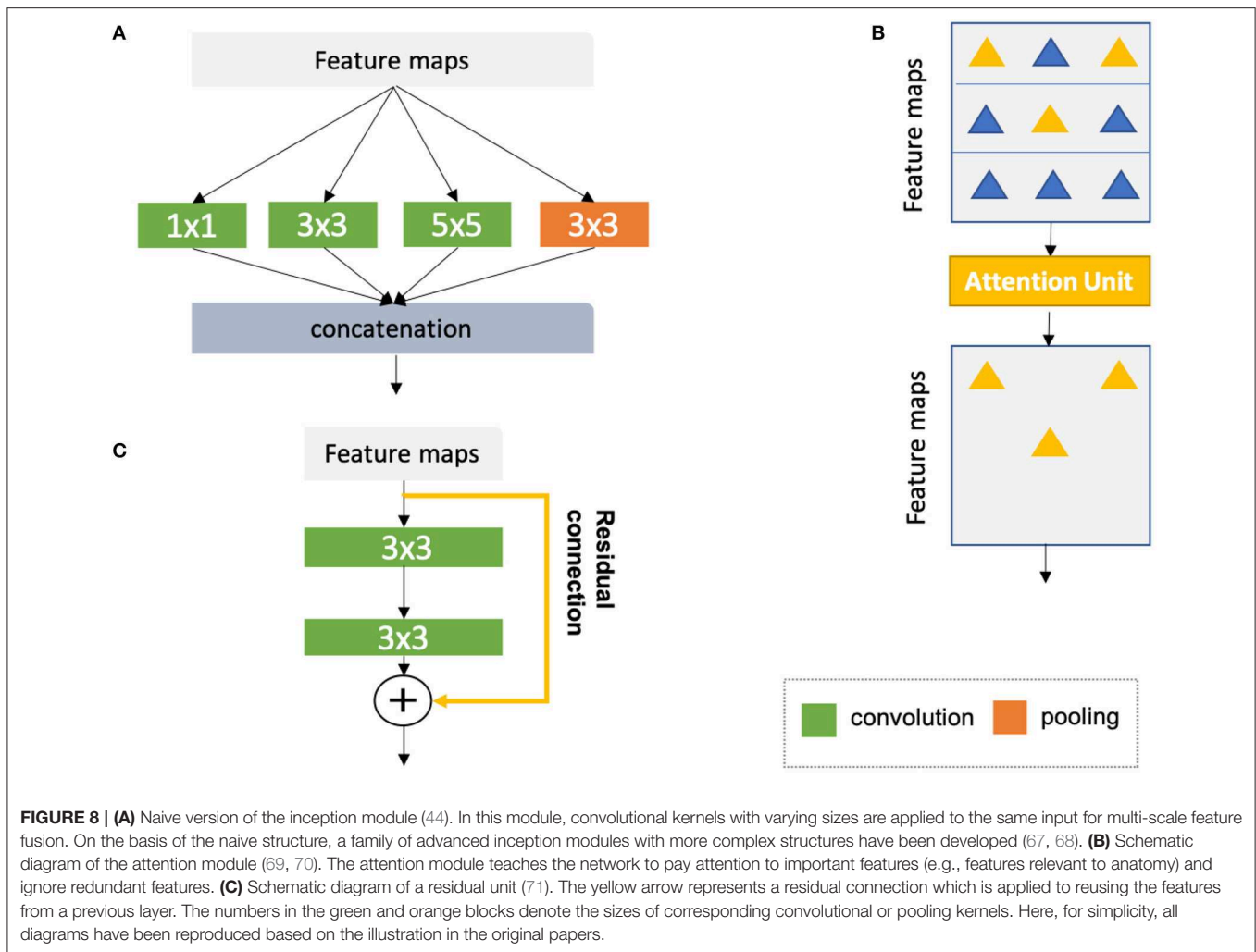
$$\mathcal{L}_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (1)$$

where y_i is the vector of target values and \hat{y}_i is the vector of the predicted values; n is the number of data samples at each iteration.

Cross-entropy is the most common loss for both image classification and segmentation tasks. In particular, the cross-entropy loss for segmentation summarizes pixel-wise probability errors between a predicted probabilistic output p_i^c and its corresponding target segmentation map y_i^c for each class c^4 :

$$\mathcal{L}_{\text{CE}} = -\frac{1}{n} \sum_{i=1}^n \sum_{c=1}^C y_i^c \log(p_i^c), \quad (2)$$

⁴At inference time, the predicted segmentation map for each image is obtained by assigning each pixel with the class of the highest probability: $\hat{y}_i = \text{argmax}_c p_i^c$.



where C is the number of all classes. Another loss function which is specifically designed for object segmentation is called soft-Dice loss function (52), which penalizes the mismatch between a predicted segmentation map and its target map at pixel-level:

$$\mathcal{L}_{Dice} = 1 - \frac{2 \sum_{i=1}^n \sum_{c=1}^C y_i^c p_i^c}{\sum_{i=1}^n \sum_{c=1}^C (y_i^c + p_i^c)}. \quad (3)$$

In addition, there are several variants of the cross-entropy or soft-Dice loss, such as the weighted cross-entropy loss (25, 80) and weighted soft-Dice loss (29, 81) that are used to address potential class imbalance problem in medical image segmentation tasks where the loss term is weighted to account for rare classes or small objects.

2.2.2. Reducing Over-Fitting

The biggest challenge of training deep networks for medical image analysis is over-fitting, due to the fact that there is often a limited number of training images in comparison with the number of learnable parameters in a deep network. A number of techniques have been developed to alleviate this problem. Some of the techniques are the following ones:

- **Weight regularization:** Weight regularization is a type of regularization techniques that add weight penalties to the loss function. Weight regularization encourages small or zero weights for less relevant or irrelevant inputs. Common methods to constrain the weights include L1 and L2 regularization, which penalize the sum of the absolute weights and the sum of the squared weights, respectively;
- **Dropout** (82): Dropout is a regularization method that randomly drops some units from the neural network during training, encouraging the network to learn a sparse representation;
- **Ensemble learning:** Ensemble learning is a type of machine learning algorithms that combine multiple trained models to obtain better predictive performance than individual models, which has been shown effective for medical image segmentation (83, 84);
- **Data augmentation:** Data augmentation is a training strategy that artificially generates more training samples to increase the diversity of the training data. This can be done via applying affine transformations (e.g., rotation, scaling), flipping or cropping to original labeled samples;

- **Transfer learning:** Transfer learning aims to transfer knowledge from one task to another related but different target task. This is often achieved by reusing the weights of a pre-trained model, to initialize the weights in a new model for the target task. Transfer learning can help to decrease the training time and achieve lower generalization error (85).

2.3. Evaluation Metrics

To quantitatively evaluate the performance of automated segmentation algorithms, three types of metrics are commonly used: (a) volume-based metrics (e.g., Dice metric, Jaccard similarity index); (b) surface distance-based metrics (e.g., mean contour distance, Hausdorff distance); (c) clinical performance metrics (e.g., ventricular volume and mass). For a detailed illustration of common used clinical indices in cardiac image analysis, we recommend the review paper by Peng et al. (2). In our paper, we mainly report the accuracy of methods in terms of the Dice metric for ease of comparison. The Dice score measures the ratio of overlap between two results (e.g., automatic segmentation vs. manual segmentation), ranging from 0 (mismatch) to 1 (perfect match). It is also important to note that the segmentation accuracy of different methods are not directly comparable in general, unless these methods are evaluated on the same dataset. This is because, even for the same segmentation task, different datasets can have different imaging modalities, different patient populations and different methods of image acquisition, which will affect the task complexities and result in different segmentation performances.

3. DEEP LEARNING FOR CARDIAC IMAGE SEGMENTATION

In this section, we provide a summary of deep learning-based applications for the three main imaging modalities: MRI, CT, and ultrasound regarding specific applications for targeted structures. In general, these deep learning-based methods provide an efficient and effective way to segmenting particular organs or tissues (e.g., the LV, coronary vessels, scars) in different modalities, facilitating follow-up quantitative analysis of cardiovascular structure and function. Among these works, a large portion of these methods are designed for ventricle segmentation, especially in MR and ultrasound domains. The objective of ventricle segmentation is to delineate the endocardium and epicardium of the LV and/or RV. These segmentation maps are important for deriving clinical indices, such as left ventricular end-diastolic volume (LVEDV), left ventricular end-systolic volume (LVESV), right ventricular end-diastolic volume (RVEDV), right ventricular end-systolic volume (RVESV), and EF. In addition, these segmentation maps are essential for 3D shape analysis (60, 86), 3D + time motion analysis (87), and survival prediction (88).

3.1. Cardiac MR Image Segmentation

Cardiac MRI is a non-invasive imaging technique that can visualize the structures within and around the heart. Compared to CT, it does not require ionizing radiation. Instead, it relies on the magnetic field in conjunction with radio-frequency waves

to excite hydrogen nuclei in the heart, and then generates an image by measuring their response. By utilizing different imaging sequences, cardiac MRI allows accurate quantification of both cardiac anatomy and function (e.g., cine imaging) and pathological tissues, such as scars (late gadolinium enhancement (LGE) imaging). Accordingly, cardiac MRI is currently regarded as the gold standard for quantitative cardiac analysis (89).

A group of representative deep learning based cardiac MR segmentation methods are shown in **Table 1**. From the table, one can see that a majority of works have focused on segmenting cardiac chambers (e.g., LV, RV, LA). In contrast, there are relatively fewer works on segmenting abnormal cardiac tissue regions, such as myocardial scars and atrial fibrosis from contrast-enhanced images. This is likely due to the limited relevant public datasets as well as the difficulty of the task. In addition, to the best of our knowledge, there are very few works that apply deep learning techniques to atrial wall segmentation, as also suggested by a recent survey paper (161). In the following sections, we will describe and discuss these methods regarding different applications in detail.

3.1.1. Ventricle Segmentation

3.1.1.1. Vanilla FCN-based segmentation

Tran (24) was among the first ones to apply a FCN (50) to segment the left ventricle, myocardium and right ventricle directly on short-axis cardiac magnetic resonance (MR) images. Their end-to-end approach based on FCN achieved competitive segmentation performance, significantly outperforming traditional methods in terms of both speed and accuracy. In the following years, a number of works based on FCNs have been proposed, aiming at achieving further improvements in segmentation performance. In this regard, one stream of work focuses on optimizing the network structure to enhance the feature learning capacity for segmentation (29, 80, 91, 162–165). For example, Khened et al. (29) developed a dense U-net with inception modules to combine multi-scale features for robust segmentation across images with large anatomical variability. Jang et al. (80), Yang et al. (81), Sander et al. (166), and Chen et al. (167) investigated different loss functions, such as weighted cross-entropy, weighted Dice loss, deep supervision loss and focal loss to improve the segmentation performance. Among these FCN-based methods, the majority of approaches use 2D networks rather than 3D networks for segmentation. This is mainly due to the typical low through-plane resolution and motion artifacts of most cardiac MR scans, which limits the applicability of 3D networks (25).

3.1.1.2. Introducing spatial or temporal context

One drawback of using 2D networks for cardiac segmentation is that these networks work slice by slice, and thus they do not leverage any inter-slice dependencies. As a result, 2D networks can fail to locate and segment the heart on challenging slices, such as apical and basal slices where the contours of the ventricles are not well-defined. To address this problem, a number of works have attempted to introduce additional contextual information to guide 2D FCN. This contextual information can include shape priors learned from labels or multi-view images (109, 110, 168).

TABLE 1 | A summary of representative deep learning methods on cardiac MRI segmentation.

Application	Selected works	Description	Type of images	Structure(s)
Ventricle segmentation	FCN-based			
	Tran (24)	2D FCN	SAX	Bi-ventricle
	Lieman-Sifry et al. (90)	A lightweight FCN (E-Net)	SAX	Bi-ventricle
	Isensee et al. (26)	2D U-net +3D U-net (ensemble)	SAX	Bi-ventricle
	Jang et al. (80)	2D M-Net with weighted cross entropy loss	SAX	Bi-ventricle
	Baumgartner et al. (25)	2D U-net with cross entropy	SAX	Bi-ventricle
	Bai et al. (31)	2D FCN trained and verified on a large dataset (~ 5000 subjects);	SAX, 2CH, 4CH	Four chambers
	Tao et al. (53)	2D U-net trained and verified on a multi-vendor, multi-scanner dataset	SAX	LV, Myo
	Khened et al. (29)	2D Dense U-net with inception module	SAX	Bi-ventricle
	Fahmy et al. (91)	2D FCN	SAX	LV, Myo
	Introducing spatial or temporal context			
	Poudel et al. (55)	2D FCN with RNN to model inter-slice coherency	SAX	Bi-ventricle
	Patravali et al. (92)	2D multi-channel FCN to aggregate inter-slice information	SAX	Bi-ventricle
	Wolterink et al. (93)	Dilated U-net to segment ED and ES simultaneously	SAX	Bi-ventricle
	Applying anatomical constraints			
	Oktay et al. (59)	FCN trained with additional anatomical shape-based regularization	SAX; Ultrasound	LV, Myo
	Multi-stage networks			
	Tan et al. (94)	Semi-automated method; CNN (localization) followed by another CNN to derive contour parameters	SAX	LV, Myo
	Zheng et al. (27)	FCN (localization) + FCN (segmentation); Propagate labels from adjacent slices	SAX	Bi-ventricle
	Vigneault et al. (95)	U-net (initial segmentation) + CNN (localization and transformation) + Cascaded U-net (segmentation)	SAX, 2CH, 4CH	Four chambers
	Hybrid segmentation methods			
	Avendi et al. (47, 96)	CNN (localization) + AE (shape initialization) + Deformable model	SAX	LV, Myo/RV
	Yang et al. (97)	CNN combined with Multi-atlas	SAX	LV, Myo
	Ngo et al. (98)	Level-set based segmentation with deep belief networks	SAX	LV, Myo
Atrial segmentation	Mortazi et al. (99)	Multi-view CNN with adaptive fusion strategy	3D scans	LA
	Xiong et al. (100)	Patch-based dual-stream 2D FCN	LGE MRI	LA
	Xia et al. (54)	Two-stage pipeline; 3D U-net (localization) + 3D U-net (segmentation)	LGE MRI	LA
Scar segmentation	Yang et al. (101)	Fully automated; Multi-atlas method for LA segmentation followed by an AE to find the atrial scars	LGE MRI	LA; atrial scars
	Chen et al. (102)	Fully automated; Multi-view two-task recursive attention model	LGE MRI	LA; atrial scars
	Zabihollahy et al. (103)	Semi-automated; 2D CNN for scar tissue classification	LGE MRI	Myocardial scars
	Moccia et al. (104)	Semi-automated; 2D FCN for scar segmentation	LGE MRI	Myocardial scars
	Xu et al. (105)	Fully automated; RNN for joint motion feature learning and scar segmentation	cine MRI	Myocardial scars
Aorta segmentation	Bai et al. (32)	RNN to learn temporal coherence; Propagate labels from labeled frames to unlabeled adjacent frames for semi-supervised learning;	cine MRI	Aorta
Whole heart segmentation	Yu et al. (30)	3D U-net with deep supervision	3D scans	Blood pool + Myocardium of the heart
	Li et al. (106)	3D FCN with deep supervision	3D scans	Blood pool + Myocardium of the heart
	Wolterink et al. (107)	dilated CNN with deep supervision	3D scans	Blood pool + Myocardium of the heart

By default, LV/RV and LA/RA segmentation refer to the left/right ventricle cavity segmentation and left/right atrium cavity segmentation, respectively. The same applies to **Tables 2–5**. SAX, short-axis view; 2CH, 2-chamber view; 4CH, 4-chamber view; ED, end-diastolic; ES, end-systolic; Myo, Left ventricular myocardium.

Others extract spatial information from adjacent slices to assist the segmentation, using recurrent units (RNNs) or multi-slice networks (2.5D networks) (27, 55, 92, 169). These networks can also be applied to leveraging information across different temporal frames in the cardiac cycle to improve spatial and temporal consistency of segmentation results (28, 93, 169–171).

3.1.1.3. Applying anatomical constraints

Another problem that may limit the segmentation performance of both 2D and 3D FCNs is that they are typically trained with pixel-wise loss functions only (e.g., cross-entropy or soft-Dice losses). These pixel-wise loss functions may not be sufficient to learn features that represent the underlying anatomical structures. Several approaches therefore focus on designing and applying anatomical constraints to train the network to improve its prediction accuracy and robustness. These constraints are represented as regularization terms which take into account the topology (172), contour and region information (173), or shape information (59, 63), encouraging the network to generate more anatomically plausible segmentations. In addition to regularizing networks at training time (61), proposed a variational AE to correct inaccurate segmentations, at the post-processing stage.

3.1.1.4. Multi-task learning

Multi-task learning has also been explored to regularize FCN-based cardiac ventricle segmentation during training by performing auxiliary tasks that are relevant to the main segmentation task, such as motion estimation (174), estimation of cardiac function (175), ventricle size classification (176), and image reconstruction (177–179). Training a network for multiple tasks simultaneously encourages the network to extract features which are useful across these tasks, resulting in improved learning efficiency and prediction accuracy.

3.1.1.5. Multi-stage networks

Recently, there is a growing interest in applying neural networks in a multi-stage pipeline which breaks down the segmentation problem into subtasks (27, 94, 95, 108, 180). For example, Zheng et al. (27) and Li et al. (108) proposed a region-of-interest (ROI) localization network followed by a segmentation network. Likewise, Vigneault et al. (95) proposed a network called Omega-Net which consists of a U-net for cardiac chamber localization, a learnable transformation module to normalize image orientation and a series of U-nets for fine-grained segmentation. By explicitly localizing the ROI and by rotating the input image into a canonical orientation, the proposed method better generalizes to images with varying sizes and orientations.

3.1.1.6. Hybrid segmentation methods

Another stream of work aims at combining neural networks with classical segmentation approaches, e.g., level-sets (98, 181), deformable models (47, 96, 182), atlas-based methods (97, 111), and graph-cut based methods (183). Here, neural networks are applied in the feature extraction and model initialization stages, reducing the dependency on manual interactions and improving the segmentation accuracy of the conventional segmentation methods deployed afterwards. For example, Avendi et al. (47) proposed one of the first DL-based methods for LV segmentation

in cardiac short-axis MR images. The authors first applied a CNN to automatically detect the LV and then used an AE to estimate the shape of the LV. The estimated shape was then used to initialize follow-up deformable models for shape refinement. As a result, the proposed integrated deformable model converges faster than conventional deformable models and the segmentation achieves higher accuracy. In their later work, the authors extended this approach to segment RV (96). While these hybrid methods demonstrated better segmentation accuracy than previous non-deep learning methods, most of them still require an iterative optimization for shape refinement. Furthermore, these methods are often designed for one particular anatomical structure. As noted in the recent benchmark study (17), most state-of-the-art segmentation algorithms for bi-ventricle segmentation are based on end-to-end FCNs, which allow the simultaneous segmentation of the LV and RV.

To better illustrate these developments for cardiac ventricle segmentation from cardiac MR images, we collate a list of bi-ventricle segmentation methods that have been trained and tested on the Automated Cardiac Diagnosis Challenge (ACDC) dataset, reported in **Table 2**. For ease of comparison, we only consider those methods which have been evaluated on the same online test set (50 subjects). As the ACDC challenge organizers keep the online evaluation platform open to the public, our comparison not only includes the methods from the original challenge participants [summarized in the benchmark study paper from Bernard et al. (17)] but also three segmentation algorithms that have been proposed after the challenge [i.e., (61, 108, 109)]. From this comparison, one can see that top algorithms are the ensemble method proposed by Isensee et al. (26) and the two-stage method proposed by Li et al. (108), both of which are based on FCNs. In particular, compared to the traditional level-set method (112), both methods achieved considerably higher accuracy even for the more challenging segmentation of the left ventricular myocardium (Myo), indicating the power of deep learning based approaches.

3.1.2. Atrial Segmentation

Atrial fibrillation (AF) is one of the most common cardiac electrical disorders, affecting around 1 million people in the UK⁵. Accordingly, atrial segmentation is of prime importance in the clinic, improving the assessment of the atrial anatomy in both pre-operative AF ablation planning and post-operative follow-up evaluations. In addition, the segmentation of atrium can be used as a basis for scar segmentation and atrial fibrosis quantification from LGE images. Traditional methods, such as region growing (184) and methods that employ strong priors [i.e., atlas-based label fusion (185) and non-rigid registration (186)] have been applied in the past for automated left atrium segmentation. However, the accuracy of these methods highly relies on good initialization and *ad-hoc* pre-processing methods, which limits the widespread adoption in the clinic.

Recently, Vigneault et al. (95) and Bai et al. (31) applied 2D FCNs to directly segment the LA and RA from standard 2D long-axis images, i.e., 2-chamber (2CH), 4-chamber (4CH)

⁵<https://www.nhs.uk/conditions/atrial-fibrillation/>

TABLE 2 | Segmentation accuracy of state-of-the-art segmentation methods verified on the cardiac bi-ventricular segmentation challenge (ACDC) dataset (17).

Methods	Description	LV	Myo	RV
Isensee et al. (26)	2D U-net + 3D U-net (ensemble)	0.950	0.911	0.923
Li et al. (108)	Two 2D FCNs for ROI detection and segmentation, respectively;	0.944	0.911	0.926
Zotti et al. (109)	2D GridNet-MD with registered shape prior	0.938	0.894	0.910
Khened et al. (29)	2D Dense U-net with inception module	0.941	0.894	0.907
Baumgartner et al. (25)	2D U-net with cross entropy loss	0.937	0.897	0.908
Zotti et al. (110)	2D GridNet with registered shape prior	0.931	0.890	0.912
Jang et al. (80)	2D M-Net with weighted cross entropy loss	0.940	0.885	0.907
Painchaud et al. (61)	FCN followed by an AE for shape correction	0.936	0.889	0.909
Wolterink et al. (93)	Multi-input 2D dilated FCN, segmenting paired ED and ES frames simultaneously	0.940	0.885	0.900
Patravali et al. (92)	2D U-net with a Dice loss	0.920	0.890	0.865
Rohé et al. (111)	Multi-atlas based method combined with 3D CNN for registration	0.929	0.868	0.881
Tziritis and Grinias (112)	Level-set + markov random field (MRF); <i>Non-deep learning method</i>	0.907	0.798	0.803
Yang et al. (81)	3D FCN with deep supervision	0.820	N/A	0.780

All the methods were evaluated on the same test set (50 subjects). Bold numbers are the highest overall Dice values for the corresponding structure. LV, left ventricle cavity; RV, right ventricle cavity; Myo, left ventricular myocardium; ED, end-diastolic; ES, end-systolic. Last update: 2019.8.1.

Note that for simplicity, we report the average Dice scores for each structure over ED and ES phases. More detailed comparison for different phases can be found on the public leaderboard in the post-testing part (<https://acdc.creatis.insa-lyon.fr>) as well as corresponding published works in this table.

views. Notably, their networks can also be trained to segment ventricles from 2D short-axis stacks without any modifications to the network architecture. Likewise, Xiong et al. (100), Preetha et al. (187), Bian et al. (188), and Chen et al. (34) applied 2D FCNs to segment the atrium from 3D LGE images in a slice-by-slice fashion, where they optimized the network structure for enhanced feature learning. 3D networks (54, 189–192) and multi-view FCN (99, 193) have also been explored to capture 3D global information from 3D LGE images for accurate atrium segmentation.

In particular, Xia et al. (54) proposed a fully automatic two-stage segmentation framework which contains a first 3D U-net to roughly locate the atrial center from down-sampled images followed by a second 3D U-net to accurately segment the atrium in the cropped portions of the original images at full resolution. Their multi-stage approach is both memory-efficient and accurate, ranking first in the left atrium segmentation challenge 2018 (LASC'18) with a mean Dice score of 0.93 evaluated on a test set of 54 cases.

3.1.3. Scar Segmentation

Scar characterization is usually performed using LGE MR imaging, a contrast-enhanced MR imaging technique. LGE MR imaging enables the identification of myocardial scars and atrial fibrosis, allowing improved management of myocardial infarction and atrial fibrillation (194). Prior to the advent of deep learning, scar segmentation was often performed using intensity thresholding-based or clustering methods which are sensitive to the local intensity changes (103). The main limitation of these methods is that they usually require the manual segmentation of the region of interest to reduce the search space and the computational costs (195). As a result, these semi-automated methods are not suitable for large-scale studies or clinical deployment.

Deep learning approaches have been combined with traditional segmentation methods for the purpose of scar segmentation: Yang et al. (101, 196) applied an atlas-based method to identify the left atrium and then applied deep neural networks to detect fibrotic tissue in that region. Relatively to end-to-end approaches, Chen et al. (102) applied deep neural networks to segment both the left atrium and the atrial scars. In particular, the authors employed a multi-view CNN with a recursive attention module to fuse features from complementary views for better segmentation accuracy. Their approach achieved a mean Dice score of 0.90 for the LA region and a mean Dice score of 0.78 for atrial scars.

In the work of Fahmy et al. (197), the authors applied a U-net based network to segment the myocardium and the scars at the same time from LGE images acquired from patients with hypertrophic cardiomyopathy (HCM), achieving a fast segmentation speed. However, the reported segmentation accuracy for the scar regions was relatively low (mean Dice: 0.58). Zabihollahy et al. (103) and Moccia et al. (104) instead adopted a semi-automated method which requires a manual segmentation of the myocardium followed by the application of a 2D network to differentiate scars from normal myocardium. They reported higher segmentation accuracy on their test sets (mean Dice >0.68). At the moment, fully-automated scar segmentation is still a challenging task since the infarcted regions in those contrast-enhanced images. Interestingly, Xu et al. (105) developed an RNN which leverages motion patterns to automatically delineate myocardial infarction area from cine MR image sequences without contrast agents. Their method achieved a high overall Dice score of 0.90 when compared to the manual annotations on LGE MR images, providing a novel approach for infarction assessment.

3.1.4. Aorta Segmentation

The segmentation of the aortic lumen from cine MR images is essential for accurate mechanical and hemodynamic characterization of the aorta. One common challenge for this task is the typical sparsity of the annotations in aortic cine image sequences, where only a few frames have been annotated. To address the problem, Bai et al. (32) applied a non-rigid image registration method (198) to propagate the labels from the annotated frames to the unlabeled neighboring ones in the cardiac cycle, effectively generating pseudo annotated frames that could be utilized for further training. This semi-supervised method achieved an average Dice metric of 0.96 for the ascending aorta and 0.95 for the descending aorta over a test set of 100 subjects. In addition, compared to a previous approach based on deformable models (199), their approach based on FCN and RNN can directly perform the segmentation task on a whole image sequence without requiring the explicit estimation of the ROI.

3.1.5. Whole Heart Segmentation

Apart from the above mentioned segmentation applications which target one particular structure, deep learning can also be applied to segmenting the main substructures of the heart in 3D MR images (30, 106, 107, 200). An early work from Yu et al. (30) adopted a 3D dense FCN to segment the myocardium and blood pool in the heart from 3D MR scans. Recently, more and more methods began to apply deep learning pipelines to segment more specific substructures [including four chambers, aorta, pulmonary vein (PV)] in both 3D CT and MR images. This has been facilitated by the availability of a public dataset for whole heart segmentation [Multi-Modality Whole Heart Segmentation (MM-WHS)] which consists of both CT and MRI images. We will discuss these segmentation methods in the next CT section in further detail (see section 3.2.1).

3.2. Cardiac CT Image Segmentation

CT is a non-invasive imaging technique that is performed routinely for disease diagnosis and treatment planning. In particular, cardiac CT scans are used for the assessment of cardiac anatomy and specifically the coronary arteries. There are two main imaging modalities: non-contrast CT imaging and contrast-enhanced coronary CT angiography (CTA). Typically, non-contrast CT imaging exploits density of tissues to generate an image, such that different densities using various attenuation values, such as soft tissues, calcium, fat, and air can be easily distinguished, and thus allows to estimate the amount of calcium present in the coronary arteries (201). In comparison, contrast-enhanced coronary CTA, which is acquired after the injection of a contrast agent, can provide excellent visualization of cardiac chambers, vessels and coronaries, and has been shown to be effective in detecting non-calcified coronary plaques. In the following sections, we will review some of the most commonly used deep learning-based cardiac CT segmentation methods. A summary of these approaches is presented in **Table 3**.

3.2.1. Cardiac Substructure Segmentation

Accurate delineation of cardiac substructures plays a crucial role in cardiac function analysis, providing important clinical variables, such as EF, myocardial mass, wall thickness etc. Typically, the cardiac substructures that are segmented include the LV, RV, LA, RA, Myo, aorta (AO), and pulmonary artery (PA).

3.2.1.1. Two-step segmentation

One group of deep learning methods relies on a two-step segmentation procedure, where a ROI is first extracted and then fed into a CNN for subsequent classification (113, 202). For instance, Zreik et al. (113) proposed a two-step LV segmentation process where a bounding box for the LV is first detected using the method described in de Vos et al. (203), followed by a voxel classification within the defined bounding box using a patch-based CNN. More recently, FCN, especially U-net (49), has become the method of choice for cardiac CT segmentation. Zhuang et al. (19) provides a comparison of a group of methods (36, 114, 115, 117, 118, 137) for whole heart segmentation (WHS) that have been evaluated on the MM-WHS challenge. Several of these methods (37, 114–116) combine a localization network, which produces a coarse detection of the heart, with 3D FCNs applied to the detected ROI for segmentation. This allows the segmentation network to focus on the anatomically relevant regions, and has shown to be effective for whole heart segmentation. A summary of the comparison between the segmentation accuracy of the methods evaluated on MM-WHS dataset is presented in **Table 4**. These methods generally achieve better segmentation accuracy on CT images compared to that of MR images, mainly because of the smaller variations in image intensity distribution across different CT scanners and better image quality (19). For a detailed discussion on these listed methods, please refer to Zhuang et al. (19).

3.2.1.2. Multi-view CNNs

Another line of research utilizes the volumetric information of the heart by training multi-planar CNNs (axial, sagittal, and coronal views) in a 2D fashion. Examples include Wang et al. (117) and Mortazi et al. (118) where three independent orthogonal CNNs were trained to segment different views. Specifically, Wang et al. (117) additionally incorporated shape context in the framework for the segmentation refinement, while Mortazi et al. (118) adopted an adaptive fusion strategy to combine multiple outputs utilizing complementary information from different planes.

3.2.1.3. Hybrid loss

Several methods employ a hybrid loss, where different loss functions (such as focal loss, Dice loss, and weighted categorical cross-entropy) are combined to address the class imbalance issue, e.g., the volume size imbalance among different ventricular structures, and to improve the segmentation performance (36, 119).

In addition, the work of Zreik et al. (120) has proposed a method for the automatic identification of patients with significant coronary artery stenoses through the segmentation

TABLE 3 | A summary of selected deep learning methods on cardiac CT segmentation.

Application	Selected works	Description	Imaging modality	Structure(s)
Cardiac substructure segmentation	Two-step segmentation			
	Zreik et al. (113)	Patch-based CNN	CTA	Myo
	Payer et al. (114)	A pipeline of two FCNs	MR/CT	WHS
	Tong et al. (115)	Deeply supervised 3D U-net	MR/CT	WHS
	Wang et al. (116)	Two-stage 3D U-net with dynamic ROI extraction	MR/CT	WHS
	Xu et al. (37)	Faster RCNN and U-net	CT	WHS
	Multi-view CNNs			
	Wang and Smedby (117)	Orthogonal 2D U-nets with shape context	MR/CT	WHS
	Mortazi et al. (118)	Multi-planar FCNs with an adaptive fusion strategy	MR/CT	WHS
	Hybrid loss			
	Yang et al. (36)	3D U-net with deep supervision	MR/CT	WHS
	Ye et al. (119)	3D deeply-supervised U-net with multi-depth fusion	CT	WHS
	Others			
Coronary artery segmentation	Zreik et al. (120)	Multi-scale FCN	CTA	Myo
	Joyce et al. (121)	Unsupervised segmentation with GANs	MR/CT	LV, Myo, RV
	End-to-end CNNs			
	Moeskops et al. (122)	Multi-task CNN	CTA	Vessel
	Merkow et al. (38)	3D U-net with deep multi-scale supervision	CTA	Vessel
	Lee et al. (123)	Template transformer network	CTA	Vessel
	CNN as pre-/post-processing			
	Gülsün et al. (124)	CNN as path pruning	CTA	coronary artery centerline
	Guo et al. (125)	Multi-task FCN with a minimal patch extractor	CTA	Coronary artery centerline
	Shen et al. (126)	3D FCN with level set	CTA	Vessel
	Others			
	Wolterink et al. (127)	CNN to estimate direction classification and radius regression	CTA	Coronary artery centerline
	Wolterink et al. (128)	Graph convolutional network	CTA	Vessel
Coronary artery calcium and plaque segmentation	Two-step segmentation			
	Wolterink et al. (129)	CNN pairs	CTA	CAC
	Lessmann et al. (130)	Multi-view CNNs	CT	CAC
	Lessmann et al. (131)	Two consecutive CNNs	CT	CAC
	Liu et al. (132)	3D vessel-focused ConvNets	CTA	CAC, NCP, MCP
	Direct segmentation			
	Santini et al. (133)	Patch-based CNN	CT	CAC
	Shadmi et al. (134)	U-net and FC DenseNet	CT	CAC
	Zhang et al. (135)	U-DenseNet	CT	CAC
	Ma and Zhang (136)	DenseRAU-net	CT	CAC

and analysis of the LV myocardium. In this work, a multi-scale FCN is first employed for myocardium segmentation, and then a convolutional autoencoder is used to characterize the LV myocardium, followed by a support vector machine (SVM) to classify patients based on the extracted features.

3.2.2. Coronary Artery Segmentation

Quantitative analysis of coronary arteries is an important step for the diagnosis of cardiovascular diseases, stenosis grading, blood flow simulation and surgical planning (204). Though this topic has been studied for years (4), only a small number of works

TABLE 4 | Segmentation accuracy of methods validated on MM-WHS dataset.

Methods	LV	RV	LA	RA	Myo	AO	PA	WHS
Payer et al. (114)	91.8/91.6	90.9 /86.8	92.9/85.5	88.8 /88.1	88.1/77.8	93.3/88.8	84.0/73.1	90.8 /86.3
Yang et al. (137)	92.3/75.0	85.7/75.0	93.0 /82.6	87.1/85.9	85.6/65.8	89.4/80.9	83.5/72.6	89.0/78.3
Mortazi et al. (118)	90.4/87.1	88.3/83.0	91.6/81.1	83.6/75.9	85.1/74.7	90.7/83.9	78.4/71.5	87.9/81.8
Tong et al. (115)	89.3/70.2	81.0/68.0	88.9/67.6	81.2/65.4	83.7/62.3	86.8/59.9	69.8/47.0	84.9/67.4
Wang et al. (116)	80.0/86.3	78.6/84.9	90.4/85.2	79.4/84.0	72.9/74.4	87.4/82.4	64.8/78.8	80.6/83.2
Ye et al. (119)	94.4 /–	89.5/–	91.6/–	87.8/–	88.9 /–	96.7 /–	86.2 /–	90.7/–
Xu et al. (37)	87.9/–	90.2/–	83.2/–	84.4/–	82.2/–	91.3/–	82.1/–	85.9/–

The training set contains 20 CT and 20 MRI whereas the test set contains 40 CT and 40 MRI. Reported numbers are Dice scores (CT/MRI) for different substructures on both CT and MRI scans. For more detailed comparisons, please refer to Zhuang et al. (19). The bold number in each column represents the highest score for the corresponding structure on CT images.

investigate the use of deep learning in this context. Methods relating to coronary artery segmentation can be mainly divided into two categories: centerline extraction and lumen (i.e., vessel wall) segmentation.

3.2.2.1. CNNs as a post-/pre-processing step

Coronary centerline extraction is a challenging task due to the presence of nearby cardiac structures and coronary veins as well as motion artifacts in cardiac CT. Several deep learning approaches employ CNNs as either a post-processing or pre-processing step for traditional methods. For instance, Gülsün et al. (124) formulated centerline extraction as finding the maximum flow paths in a steady state porous media flow, with a learning-based classifier estimating anisotropic vessel orientation tensors for flow computation. A CNN classifier was then employed to distinguish true coronary centerlines from leaks into non-coronary structures. Guo et al. (125) proposed a multi-task FCN centerline extraction method that can generate a single-pixel-wide centerline, where the FCN simultaneously predicted centerline distance maps and endpoint confidence maps from coronary arteries and ascending aorta segmentation masks, which were then used as input to the subsequent minimal path extractor to obtain the final centerline extraction results. In contrast, unlike the aforementioned methods that used CNNs either as a pre-processing or post-processing step, Wolterink et al. (127) proposed to address centerline extraction via a 3D dilated CNN, where the CNN was trained on patches to directly determine a posterior probability distribution over a discrete set of possible directions as well as to estimate the radius of an artery at the given point.

3.2.2.2. End-to-end CNNs

With respect to the lumen or vessel wall segmentation, most deep learning based approaches use an end-to-end CNN segmentation scheme to predict dense segmentation probability maps (38, 122, 126, 205). In particular, Moeskops et al. (122) proposed a multi-task segmentation framework where a single CNN can be trained to perform three different tasks including coronary artery segmentation in cardiac CTA and tissue segmentation in brain MR images. They showed that such a multi-task segmentation network in multiple modalities can achieve equivalent performance as a single task network. Merkow et al. (38) introduced deep multi-scale supervision into

a 3D U-net architecture, enabling efficient multi-scale feature learning and precise voxel-level predictions. Besides, shape priors can also be incorporated into the network (123, 206, 207). For instance, Lee et al. (123) explicitly enforced a roughly tubular shape prior for the vessel segments by introducing a template transformer network, through which a shape template can be deformed via network-based registration to produce an accurate segmentation of the input image, as well as to guarantee topological constraints. More recently, graph convolutional networks have also been investigated by Wolterink et al. (128) for coronary artery segmentation in CTA, where vertices on the coronary lumen surface mesh were considered as graph nodes and the locations of these tubular surface mesh vertices were directly optimized. They showed that such method significantly outperformed a baseline network that used only fully-connected layers on healthy subjects (mean Dice score: 0.75 vs. 0.67). Besides, the graph convolutional network used in their work is able to directly generate smooth surface meshes without post-processing steps.

3.2.3. Coronary Artery Calcium and Plaque Segmentation

Coronary artery calcium (CAC) is a direct risk factor for cardiovascular disease. Clinically, CAC is quantified using the Agatston score (208) which considers the lesion area and the weighted maximum density of the lesion (209). Precise detection and segmentation of CAC are thus important for the accurate prediction of the Agatston score and disease diagnosis.

3.2.3.1. Two-step segmentation

One group of deep learning approaches to segmentation and automatic calcium scoring proposed to use a two-step segmentation scheme. For example, Wolterink et al. (129) attempted to classify CAC in cardiac CTA using a pair of CNNs, where the first CNN coarsely identified voxels likely to be CAC within a ROI detected using De et al. (203) and then the second CNN further distinguished between CAC and CAC-like negatives more accurately. Similar to such a two-stage scheme, Lessmann et al. (130, 131) proposed to identify CAC in low-dose chest CT, in which a ROI of the heart or potential calcifications were first localized followed by a CAC classification process.

3.2.3.2. Direct segmentation

More recently, several approaches (133–136) have been proposed for the direct segmentation of CAC from non-contrast cardiac CT or chest CT: the majority of them employed combinations of U-net (49) and DenseNet (78) for precise quantification of CAC which showed that a sensitivity over 90% can be achieved (133). These aforementioned approaches all follow the same workflow where the CAC is first identified and then quantified. An alternative approach is to circumvent the intermediate segmentation and to perform direct quantification, such as in de Vos et al. (209) and Cano-Espinosa et al. (210), which have proven that this approach is effective and promising.

Finally, for non-calcified plaque (NCP) and mixed-calcified plaque (MCP) in coronary arteries, only a limited number of works have been reported that investigate deep learning methods for segmentation and quantification (132, 211). Yet, this is a very important task from a clinical point of view, since these plaques can potentially rupture and obstruct an artery, causing ischemic events and severe cardiac damage. In contrast to CAC segmentation, NCP and MCP segmentation are more challenging due to their similar appearances and intensities as adjacent tissues. Therefore, robust and accurate analysis often requires the generation of multi-planar reformatted (MPR) images that have been straightened along the centerline of the vessel. Recently, Liu et al. (132) proposed a vessel-focused 3D convolutional network with attention layers to segment three types of plaques on the extracted and reformatted coronary MPR volumes. Zreik et al. (211) presented an automatic method for detection and characterization of coronary artery plaques as well as determination of coronary artery stenosis significance, in which a multi-task convolutional RNN was used to perform both plaque and stenosis classification by analyzing the features extracted along the coronary artery in an MPR image.

3.3. Cardiac Ultrasound Image Segmentation

Cardiac ultrasound imaging, also known as echocardiography, is an indispensable clinical tool for the assessment of cardiovascular function. It is often used clinically as the first imaging examination owing to its portability, low cost and real-time capability. While a number of traditional methods, such as active contours, level-sets and active shape models have been employed to automate the segmentation of anatomical structures in ultrasound images (212), the achieved accuracy is limited by various problems of ultrasound imaging, such as low signal-to-noise ratio, varying speckle noise, low image contrast (especially between the myocardium and the blood pool), edge dropout and shadows cast by structures, such as dense muscle and ribs.

As in cardiac MR and CT, several DL-based methods have been recently proposed to improve the performance of cardiac ultrasound image segmentation in terms of both accuracy and speed. The majority of these DL-based approaches focus on LV segmentation, with only few addressing the problem of aortic valve and LA segmentation. A summary of the reviewed works can be found in **Table 5**.

3.3.1. 2D LV Segmentation

3.3.1.1. Deep learning combined with deformable models

The imaging quality of echocardiography makes voxel-wise tissue classification highly challenging. To address this challenge, deep learning has been combined with deformable model for LV segmentation in 2D images (138, 139, 141–145). Features extracted by trained deep neural networks were used instead of handcrafted features to improve accuracy and robustness.

Several works applied deep learning in a two-stage pipeline which first localizes the target ROI via rigid transformation of a bounding box, then segments the target structure within the ROI. This two-stage pipeline reduces the search region of the segmentation and increases robustness of the overall segmentation framework. Carneiro et al. (138, 139) first adopted this DL framework to segment the LV in apical long-axis echocardiograms. The method uses DBN (213) to predict the rigid transformation parameters for localization and the deformable model parameters for segmentation. The results demonstrated the robustness of DBN-based feature extraction to image appearance variations. Nascimento and Carneiro (140) further reduced the training and inference complexity of the DBN-based framework by using sparse manifold learning in the rigid detection step.

To further reduce the computational complexity, some works perform segmentation in one step without resorting to the two-stage approach. Nascimento and Carneiro (141, 142) applied sparse manifold learning in segmentation, showing a reduced training and search complexity compared to their previous version of the method, while maintaining the same level of segmentation accuracy. Veni et al. (143) applied a FCN to produce coarse segmentation masks, which is then further refined by a level-set based method.

3.3.1.2. Utilizing temporal coherence

Cardiac ultrasound data is often recorded as a temporal sequence of images. Several approaches aim to leverage the coherence between temporally close frames to improve the accuracy and robustness of the LV segmentation. Carneiro and Nascimento (144, 145) proposed a dynamic modeling method based on a sequential monte carlo (SMC) (or particle filtering) framework with a transition model, in which the segmentation of the current cardiac phase depends on previous phases. The results show that this approach performs better than the previous method (138) which does not take temporal information into account. In a more recent work, Jafari et al. (146) combined U-net, long-short term memory (LSTM) and inter-frame optical flow to utilize multiple frames for segmenting one target frame, demonstrating improvement in overall segmentation accuracy. The method was also shown to be more robust to image quality variations in a sequence than single-frame U-net.

3.3.1.3. Utilizing unlabeled data

Several works proposed to use non-DL based segmentation algorithms to help generating labels on unlabeled images, effectively increasing the amount of training data. To achieve this, Carneiro and Nascimento (147, 148) proposed on-line retraining strategies where segmentation network (DBN) is firstly initialized

TABLE 5 | A summary of reviewed deep learning methods for ultrasound image segmentation.

Application	Selected works	Method	Structure	Imaging modality
2D LV	Combined with deformable models			
	Carneiro et al. (138, 139)	DBN with two-step approach: localization and fine segmentation	LV	2D A2C, A4C
	Nascimento and Carneiro (140)	deep belief networks (DBN) and sparse manifold learning for the localization step	LV	2D A2C, A4C
	Nascimento and Carneiro (141, 142)	DBN and sparse manifold learning for one-step segmentation	LV	2D A2C, A4C
	Veni et al. (143)	FCN (U-net) followed by level-set based deformable model	LV	2D A4C
	Utilizing temporal coherence			
	Carneiro and Nascimento (144, 145)	DBN and particle filtering for dynamic modeling	LV	2D A2C, A4C
	Jafari et al. (146)	U-net and LSTM with additional optical flow input	LV	2D A4C
	Utilizing unlabeled data			
	Carneiro and Nascimento (147, 148)	DBN on-line retrain using external classifier as additional supervision	LV	2D A2C, A4C
	Smistad et al. (149)	U-Net trained using labels generated by a Kalman filter based method	LV and LA	2D A2C, A4C
	Yu et al. (150)	Dynamic CNN fine-tuning with mitral valve tracking to separate LV from LA	Fetal LV	2D
	Jafari et al. (151)	U-net with TL-net (152) based shape constraint on unannotated frames	LV	2D A4C
	Utilizing data from multiple domains			
	Chen et al. (153)	FCN trained using annotated data of multiple anatomical structures	Fetal head and LV	2D head, A2-5C
	Others			
	Smistad et al. (154)	Real time CNN view-classification and segmentation	LV	2D A2C, A4C
	Leclerc et al. (155)	U-net trained on a large heterogeneous dataset	LV, Myo	2D A4C
	Jafari et al. (156)	Real-time mobile software, lightweight U-Net, multitask and adversarial training	LV	2D A2C, A4C
3D LV	Dong et al. (157)	CNN for 2D coarse segmentation refined by 3D snake model	LV	3D (CETUS)
	Oktay et al. (59)	U-net with TL-net based shape constraint	LV	3D (CETUS)
	Dong et al. (158)	Atlas-based segmentation using DL registration and adversarial training	LV	3D
Others	Ghesu et al. (159)	Marginal space learning and adaptive sparse neural network	Aortic valves	3D
	Degel et al. (160)	V-net with TL-net based shape constraint and GAN-based domain adaptation	LA	3D
	Zhang et al. (42)	CNN for view-classification, segmentation and disease detection	Multi-chamber	2D PLAX, PSAX, A2-4C

A[X]C is short for Apical [X]-chamber view. PLAX/PSAX, parasternal long-axis/short-axis; CETUS, using the dataset from Challenge on Endocardial Three-dimensional Ultrasound Segmentation.

using a small set of labeled data and then applied to non-labeled data to propose annotations. The proposed annotations are then checked by external classifiers before being used to re-train the network. Smistad et al. (149) trained a U-net using images annotated by a Kalman filtering based method (214) and illustrated the potential of using this strategy for pre-training. Alternatively, some works proposed to exploit unlabeled data without using additional segmentation algorithm. Yu et al. (150) proposed to train a CNN on a partially labeled dataset of multiple sequences, then fine-tuned the network for each individual sequence using manual segmentation of the first frame as well as CNN-produced label of other frames. Jafari et al. (151) proposed

a semi-supervised framework which enables training on both the labeled and unlabeled images. The framework uses an additional generative network, which is trained to generate ultrasound images from segmentation masks, as additional supervision for the unlabeled frames in the sequences. The generative network forces the segmentation network to predict segmentation that can be used to successfully generate the input ultrasound image.

3.3.1.4. Utilizing data from multiple domains

Apart from exploiting unlabeled data in the same domain, leveraging manually annotated data from multiple domains (e.g., different 2D ultrasound views with various anatomical

structures) can also help to improve the segmentation in one particular domain. Chen et al. (153) proposed a novel FCN-based network to utilize multi-domain data to learn generic feature representations. Combined with an iterative refinement scheme, the method has shown superior performance in detection and segmentation over traditional database-guided method (215), FCN trained on single-domain and other multi-domain training strategies.

3.3.1.5. Others

The potential of CNN in segmentation has motivated the collection and labeling of large-scale datasets. Several methods have since shown that deep learning methods, most notably CNN-based methods, are capable of performing accurate segmentation directly without complex modeling and post-processing. Leclerc et al. (155) performed a study to investigate the effect of the size of annotated data for the segmentation of the LV in 2D ultrasound images using a simple U-net. The authors demonstrated that the U-net approach significantly benefits from larger amounts of training data. In addition to performance on accuracy, some work investigated the computational efficiency of DL-based methods. Smistad et al. (154) demonstrated the efficiency of CNN-based methods by successfully performing real-time view-classification and segmentation. Jafari et al. (156) developed a software pipeline capable of real-time automated LV segmentation, landmark detection and LV ejection fraction calculation on a mobile device taking input from point-of-care ultrasound (POCUS) devices. The software uses a lightweight U-net trained using multi-task learning and adversarial training, which achieves EF prediction error that is lower than inter- and intra- observer variability.

3.3.2. 3D LV Segmentation

Segmenting cardiac structures in 3D ultrasound is even more challenging than 2D. While having the potential to derive more accurate volume-related clinical indices, 3D echocardiograms suffer from lower temporal resolution and lower image quality compared to 2D echocardiograms. Moreover, 3D images dramatically increase the dimension of parameter space of neural networks, which poses computational challenges for deep learning methods.

One way to reduce the computational cost is to avoid direct processing of 3D data in deep learning networks. Dong et al. (157) proposed a two-stage method by first applying a 2D CNN to produce coarse segmentation maps on 2D slices from a 3D volume. The coarse 2D segmentation maps are used to initialize a 3D shape model which is then refined by 3D deformable model method (216). In addition, the authors used transfer learning to side-step the limited training data problem by pre-training network on a large natural image segmentation dataset and then fine-tuning to the LV segmentation task.

Anatomical shape priors have been utilized to increase the robustness of deep learning-based segmentation methods to challenging 3D ultrasound images. Oktay et al. (59) proposed an anatomically constrained network where a shape constraint-based loss is introduced to train a 3D segmentation network. The shape constraint is based on the shape prior

learned from segmentation maps using auto-encoders (152). Dong et al. (158) utilized shape prior more explicitly by combining a neural network with a conventional atlas-based segmentation framework. Adversarial training was also applied to encourage the method to produce more anatomically plausible segmentation maps, which contributes to its superior segmentation performance comparing to a standard voxel-wise classification 3D segmentation network (52).

3.3.3. Left Atrium Segmentation

Degel et al. (160) adopted the aforementioned anatomical constraints in 3D LA segmentation to tackle the domain shift problem caused by variation of imaging device, protocol and patient condition. In addition to the anatomically constraining network, the authors applied an adversarial training scheme (217) to improve the generalizability of the model to unseen domain.

3.3.4. Multi-Chamber Segmentation

Apart from LV segmentation, a few works (23, 42, 149) applied deep learning methods to perform multi-chamber (including LV and LA) segmentation. In particular, (42) demonstrated the applicability of CNNs on three tasks: view classification, multi-chamber segmentation and detection of cardiovascular diseases. Comprehensive validation on a large (non-public) clinical dataset showed that clinical metrics derived from automatic segmentation are comparable or superior than manual segmentation. To resemble real clinical situations and thus encourages the development and evaluation of robust and clinically effective segmentation methods, a large-scale dataset for 2D cardiac ultrasound has been recently made public (23). The dataset and evaluation platform were released following the preliminary data requirement investigation of deep learning methods (155). The dataset is composed of apical 4-chamber view images annotated for LV and LA segmentation, with uneven imaging quality from 500 patients with varying conditions. Notably, the initial benchmarking (23) on this dataset has shown that modern encoder-decoder CNNs resulted in lower error than inter-observer error between human cardiologists.

3.3.5. Aortic Valve Segmentation

Ghesu et al. (159) proposed a framework based on marginal space learning (MSL), Deep neural networks (DNNs) and active shape model (ASM) to segment the aortic valve in 3D cardiac ultrasound volumes. An adaptive sparsely-connected neural network with reduced number of parameters is used to predict a bounding box to locate the target structure, where the learning of the bounding box parameters is marginalized into sub-spaces to reduce computational complexity. This framework showed significant improvement over the previous non-DL MSL (218) method while achieving competitive run-time.

3.4. Discussion

So far, we have presented and discussed recent progress of deep learning-based segmentation methods in the three modalities (i.e., MR, CT, ultrasound) that are commonly used in the assessment of cardiovascular disease. To summarize, current state-of-the-art segmentation methods are mainly based on

CNNs that employ the FCN or U-net architecture. In addition, there are several commonalities in the FCN-based methods for cardiac segmentation which can be categorized into four groups: (1) enhancing network feature learning by employing advanced building blocks in networks (e.g., inception module, dilated convolutions), most of which have been mentioned earlier (section 2.1.6); (2) alleviating the problem of class imbalance with advanced loss functions (e.g., weighted loss functions); (3) improving the networks' generalization ability and robustness through a multi-stage pipeline, multi-task learning, or multi-view feature fusion; (4) forcing the network to generate more anatomically-plausible segmentation results by incorporating shape priors, applying adversarial loss or anatomical constraints to regularize the network during training. It is also worthwhile to highlight that for cardiac image sequence segmentation (e.g., cine MR images, 2D ultrasound sequences), leveraging spatial and temporal coherence from these sequences with advanced neural networks [e.g., RNN (32, 146), multi-slice FCN (27)] has been explored and shown to be beneficial for improving the segmentation accuracy and temporal consistency of the segmentation maps.

While the results reported in the literature show that neural networks have become more sophisticated and powerful, it is also clear that performance has improved with the increase of publicly available training subjects. A number of DL-based methods (especially in MRI) have been trained and tested on public challenge datasets, which not only provide large amounts of data to exploit the capabilities of deep learning in this domain, but also a platform for transparent evaluation and comparison. In addition, many of the participants in these challenges have shared their code with other researchers via open-source community websites (e.g., Github). Transparent and fair benchmarking and sharing of code are both essential for continued progress in this domain. We summarize the existing public datasets in **Table 6** and public code repositories in **Table 7** for reference.

An interesting conclusion supported by **Table 7** is that the target image type can affect the choice of network structures (i.e., 2D networks, 3D networks). For 3D imaging acquisitions, such as LGE-MRI and CT images, 3D networks are preferred whereas 2D networks are more popular approaches for segmenting cardiac cine short-axis or long-axis image stacks. One reason for using 2D networks for the segmentation of short-axis or long-axis images is their typically large slice thickness (usually around 7–8 mm) which can further exacerbated by inter-slice gaps. In addition, breath-hold related motion artifacts between different slices may negatively affect 3D networks. A study conducted by Baumgartner et al. (25) has shown that a 3D U-net performs worse than a 2D U-net when evaluated on the ACDC challenge dataset. By contrast, in the LASC'18 challenge mentioned in **Table 6**, which uses high-resolution 3D images, most participants applied 3D networks and the best performance was achieved by a cascaded network based on the 3D U-net (54).

It is well-known that training 3D networks is more difficult than training 2D networks. In general, 3D networks have significantly more parameters than 2D networks. Therefore, 3D networks are more difficult and computationally expensive to optimize as well as prone to over-fitting, especially if the training

data is limited. As a result, several researchers have tried to carefully design the structure of network to reduce the number of parameters for a particular application and have also applied advanced techniques (e.g., deep supervision) to alleviate the over-fitting problem (30, 54). For this reason, 2D-based networks (e.g., 2D U-net) are still the most popular segmentation approaches for all three modalities.

In addition to 2D and 3D networks, several authors have proposed “2D+” networks that have been shown to be effective in segmenting structures from cardiac volumetric data. These “2D+” networks are mainly based on 2D networks, but are adapted with increased capacity to utilize 3D context. These networks include multi-view networks which leverage multi-planar information (i.e., coronal, sagittal, axial views) (99, 117), multi-slice networks, and 2D FCNs combined with RNNs which incorporate context across multiple slices (33, 55, 92, 169). These “2D+” networks inherit the advantages of 2D networks while still being capable of leveraging through-plane spatial context for more robust segmentation with strong 3D consistency.

Finally, it is worth to note that there is no universally optimal segmentation method. Different applications have different complexities and different requirements, meaning that customized algorithms need to be optimized. For example, while anatomical shape constraints can be applied to cardiac anatomical structure segmentation (e.g., ventricle segmentation) to boost the segmentation performance, those constraints may not be suitable for the segmentation of pathologies or lesions (e.g., scar segmentation) which can have arbitrary shapes. Also, even if the target structure in two applications are the same, the complexity of the segmentation task can vary significantly from one to another, especially when their underlying imaging modalities and patient populations are different. For example, directly segmenting the left ventricle myocardium from contrast-enhanced MR images (e.g., LGE images) is often more difficult than from MR images without contrast agents, as the anatomical structures are more attenuated by the contrast agent. For cases with certain diseases (e.g., myocardial infarction), the border between the infarcted region and blood pool appears blurry and ambiguous to delineate. As a result, a segmentation network designed for non-contrast enhanced images may not be directly applied to contrast-enhanced images (100). A more sophisticated algorithm is generally required to assist the segmentation procedure. Potential solutions include applying dedicated image pre-processing, enhancing network capacity, adding shape constraints, and integrating specific knowledge about the application.

4. CHALLENGES AND FUTURE WORK

It is evident from the literature that deep learning methods have matched or surpassed the previous state of the art in various cardiac segmentation applications, mainly benefiting from the increased size of public datasets and the emergence of advanced network architectures as well as powerful hardware for computing. Given this rapid process, one may wonder if deep learning methods can be directly deployed to real-world

TABLE 6 | Summary of public datasets on cardiac segmentation for the three modalities.

Dataset Name/ References	Year	Main modalities	# of subjects	Target(s)	Main pathology
York (10)	2008	cine MRI	33	LV, Myo	Cardiomyopathy, aortic regurgitation, enlarged ventricles and ischemia
Sunnybrook (11)	2009	cine MRI	45	LV, Myo	Hypertrophy, heart failure w./w.o infarction
LVSC (12)	2011	cine MRI	200	LV, Myo	Coronary artery disease, myocardial infarction.
RVSC (1)	2012	cine MRI	48	RV	Myocarditis, ischemic cardiomyopathy, suspicion of arrhythmogenic, right ventricular dysplasia, dilated cardiomyopathy, hypertrophic cardiomyopathy, aortic stenosis
cDEMIRIS (13)	2012	LGE MRI	60	LA fibrosis and scar	Atrial fibrillation
LVIC (14)	2012	LGE MRI	30	Myocardial scars	Ischemic cardiomyopathy
LASC'13 (15)	2013	3D MRI	30	LA	N/A
HVSMR (16)	2016	3D MRI	4	Blood pool, myocardium of the heart	Congenital heart defects
ACDC (17)	2017	cine MRI	150	LV, Myo; RV	Myocardial infarction, dilated/hypertrophic cardiomyopathy, abnormal RV
LASC'18 (18)	2018	LGE MRI	154	LA	Atrial fibrillation
MM-WHS (19)	2017	CT/MRI	60/60	WHS	Myocardium infarction, atrial fibrillation, tricuspid regurgitation, aortic valve stenosis, Alagille syndrome, Williams syndrome, dilated cardiomyopathy, aortic coarctation, tetralogy of Fallot
CAT08 (20)	2008	CTA	32	Coronary artery centerline	Patients with presence of calcium scored as absent, modest or severe.
CLS12 (21)	2012	CTA	48	Coronary lumen and stenosis	Patients with different levels of coronary artery stenoses.
CETUS (22)	2014	3D Ultrasound	45	LV	Myocardial infarction, dilated cardiomyopathy
CAMUS (23)	2019	2D Ultrasound	500	LV, LA	Patients with EF < 45%

Most of the datasets listed above are from the MICCAI society.

applications to reduce the workload of clinicians. The current literature suggests that there is still a long way to go. In the following paragraphs, we summarize several major challenges in the field of cardiac segmentation and some recently proposed approaches that attempt to address them. These challenges and related works also provide potential research directions for future work in this field.

4.1. Scarcity of Labels

One of the biggest challenges for deep learning approaches is the scarcity of annotated data. In this review, we found that the majority of studies uses a fully supervised approach to train their networks, which requires a large number of annotated images. In fact, annotating cardiac images is time consuming and often requires significant amounts of expertise. These methods can be divided into five classes: data augmentation, transfer learning with fine-tuning, weakly and semi-supervised learning, self-supervised learning, and unsupervised learning.

- **Data augmentation.** Data augmentation aims to increase the size and the variety of training images by artificially generating new samples from existing labeled data. Traditionally, this can be achieved by applying a stack of geometric or photometric transformations to existing image-label pairs. These transformations can be affine transformations, adding random noise to the original data, or adjusting image contrast. However, designing an effective pipeline of data

augmentation often requires domain knowledge, which may not be easily extendable to different applications. And the diversity of augmented data may still be limited, failing to reflect the spectrum of real-world data distributions. Most recently, several researchers have begun to investigate the use of generative models [e.g., GANs, variational AE (219)], reinforcement learning (220), and adversarial example generation (221) to directly learn task-specific augmentation strategies from existing data. In particular, the generative model-based approach has been proven to be effective for one-shot brain segmentation (222) and few-shot cardiac MR image segmentation (223) and it is thus worth exploring for more applications in the future.

- **Transfer learning with fine-tuning.** Transfer learning aims at reusing a model pre-trained on one task as a starting point to train for a second task. The key of transfer learning is to learn features in the first task that are related to the second task such that the network can quickly converge even with limited data. Several researchers have successfully demonstrated the use of transfer learning to improve the model generalization performance for cardiac ventricle segmentation, where they first trained a model on a large dataset and then fine-tuned it on a small dataset (29, 31, 85, 91, 165).
- **Weakly and semi-supervised learning.** Weakly and semi-supervised learning methods aim at improving the learning accuracy by making use of both labeled and unlabeled or weakly-labeled data (e.g., annotations in forms of scribbles

TABLE 7 | Public code for DL-based cardiac image segmentation.

Modality	Application(s)	References	Basic network	Code repo (If not specified, the repository is located under github.com)
MR (SAX)	Bi-ventricular Segmentation	Tran (24)	2D FCN	vuptran/cardiac-segmentation
MR (SAX)	Bi-ventricular Segmentation	Baumgartner et al. (25)	2D/3D U-net	baumgach/acdc_segmenter
MR (SAX)	Bi-ventricular Segmentation; 1st rank in ACDC challenge	Isensee et al. (26)	2D + 3D U-net (ensemble)	MIC-DKFZ/ACDC2017
MR (SAX)	Bi-ventricular Segmentation	Zheng et al. (27)	Cascaded 2D U-net	julien-zheng/CardiacSegmentationPropagation
MR (SAX)	Bi-ventricular segmentation and Motion Estimation	Qin et al. (28)	2D FCN, RNN	cq615
MR (SAX)	Biventricular Segmentation	Khened et al. (29)	2D U-net	mahendrakhened
MR (3D scans)	WHS	Yu et al. (30)	3D CNN	yulequan/HeartSeg
MR (Multi-view)	Four-chamber Segmentation and Aorta Segmentation	Bai et al. (31, 32)	2D FCN, RNN	baiwenjia/ukbb_cardiac
MR	Cardiac segmentation and motion tracking	Duan et al. (33)	2.5D FCN + Atlas-based	j-duan/4Dsegment
LGE MRI	Left Atrial Segmentation	Chen et al. (34)	2D U-net	cherise215/atria_segmentation_2018
LGE MRI	Left Atrial Segmentation	Yu et al. (35)	3D V-net	yulequan/UA-MT
CT	WHS	Yang et al. (36)	3D U-net	xy0806/miccai17-mmwhs-hybrid
CT	WHS	Xu et al. (37)	Faster RCNN, 3D U-net	Wuziyi616/CFUN
CT, MRI	Coronary arteries	Merkow et al. (38)	3D U-net	jmerkow/I2I
CT, MRI	WHS	Dou et al. (39, 40)	2D CNN	carrenD/Medical-Cross-Modality-Domain-Adaptation
CT, MRI	WHS	Chen et al. (41)	2D CNN	cchen-cc/SIFA
Ultrasound	View classification and four-chamber segmentation	Zhang et al. (42)	2D U-net	bitbucket.org/raahuldeo/echocv

SAX, short-axis view; WHS, whole heart segmentation.

or bounding boxes). In this context, several works have been proposed for cardiac ventricle segmentation in MR images. One approach is to estimate full labels on unlabeled or weakly labeled images for further training. For example, Qin et al. (28) and Bai et al. (32) utilized motion information to propagate labels from labeled frames to unlabeled frames in a cardiac cycle whereas (224, 225) applied the expectation maximization (EM) algorithm to predict and refine the estimated labels recursively. Others have explored different approaches to regularize the network when training on unlabeled images, applying multi-task learning (177, 178), or global constraints (226).

- **Self-supervised learning.** Another approach is self-supervised learning which aims at utilizing labels that are *generated automatically* without human intervention. These labels, designed to encode some properties or semantics of the object, can provide strong supervisory signals to pre-train a network before fine-tuning for a given task. A very recent work from Bai et al. (227) has shown the effectiveness of self-supervised learning for cardiac MR image segmentation where the authors used auto-generated anatomical position labels to pre-train a segmentation network. Compared to a network trained from scratch, networks pre-trained on the self-supervised task performed better, especially when the training data was extremely limited.

- **Unsupervised learning.** Unsupervised learning aims at learning without paired labeled data. Compared to the former four classes, there is limited literature about unsupervised learning methods for cardiac image segmentation, perhaps because of the difficulty of the task. An early attempt has been made which applied adversarial training to train a network segmenting LV and RV from CT and MR images without requiring a training set of paired images and labels (121).

In general, transfer learning and self-supervised learning allow the network to be aware of general knowledge shared across different tasks to accelerate learning procedure and to encourage model generalization. On the other hand, data augmentation, weakly and semi-supervised learning allows the network to get more labeled training data in an efficient way. In practice, the two types of methods can be integrated together to improve the model performance. For example, transfer learning can be applied at the model initialization stage whereas data augmentation can be applied at the model fine-tuning stage.

4.2. Model Generalization Across Various Imaging Modalities, Scanners, and Pathologies

Another common limitation in DL-based methods is that they still lack generalization capabilities when presented with

previously unseen samples (e.g., data from a new scanner, abnormal, and pathological cases that have not been included in the training set). In other words, deep learning models tend to be biased by their respective training datasets. This limitation prevents models to be deployed in the real world and therefore diminishes their impact for improving clinical workflows.

To improve the model performance across MR images acquired from multiple vendors and multiple scanners (53), collected a large multi-vendor, multi-center, heterogeneous labeled training set from patients with cardiovascular diseases. However, this approach may not scale to the real world, as it implies the collection and labeling of a vastly large dataset covering all possible cases. Several researchers have recently started to investigate the use of unsupervised domain adaptation techniques that aim at optimizing the model performance on a target dataset without additional labeling costs. Several works have successfully applied adversarial training to cross-modality segmentation tasks, adapting a cardiac segmentation model learned from MR images to CT images and vice versa (39–41, 228, 229). These type of approaches can also be adopted for semi-supervised learning, where the target domain is a new set of unlabeled data of the same modality (230). Of note, these domain adaptation methods often require the access to unlabeled images in the target domain (e.g., a new scanner, a different hospital), which may not be easy to obtain due to the data privacy and ethics issues. How to collect and share data safely, fairly, and legally across different sites is still an open challenge.

On the other hand, some researchers have started to develop domain generalization algorithms, *without* requiring accessing images from new sites. One stream of works aims to improve the domain generalization ability by extracting domain-independent and robust features or disentangling learned features into domain-specific and domain-invariant components from various seen domains (e.g., multi-center data, multi-modality datasets) to improve the model performance on unseen domains (221, 228, 231). Other researchers have started to adopt data augmentation techniques to simulate various possible data distributions across different domains. For instance, Chen et al. (232) have proposed a data normalization and augmentation pipeline which enables a neural network for cardiac MR image segmentation trained from a single-scanner dataset to generalize well across multi-scanner and multi-site datasets. Zhang et al. (233) applied a similar data augmentation approach to improve the model generalization ability on unseen datasets. Their method has been verified on three tasks including left atrial segmentation from 3D MRI and left ventricle segmentation from 3D ultrasound images.

One bottleneck of augmenting training data for model generalization across different sites is that it is often required to increase the model capacity to compensate for the increased dataset size and variation (232). As a result, training becomes more expensive and challenging. To address this inefficiency problem, active learning (234) has been proposed, which selects the most representative images from a large-scale dataset, reducing labeling workload as well as computational costs. This technique is also related to incremental learning, which aims to improve the model performance by adding new classes incrementally while avoiding a dramatic decrease in overall performance (235). Given the increasing size of

the available medical imaging datasets and the practical challenges of collecting, labeling and storing large amounts of images from various sources, it is of great interest to combine domain generalization algorithms with active learning algorithms together to distill a large dataset into a small one but containing the most representative cases for effective and robust learning.

4.3. Lack of Model Interpretability

Unlike symbolic artificial intelligence systems, deep learning systems are difficult to interpret and not transparent. Once a network has been trained, it behaves like a “black box,” providing predictions which are not directly interpretable. This issue makes the model unpredictable, intractable for model verification, and ultimately untrustworthy. Recent studies have shown that deep learning-based vision recognition systems can be attacked by images modified with nearly imperceptible perturbations (236–238). These attacks can also happen in medical scenarios, e.g., a DL-based system may make a wrong diagnosis given an image with adversarial noise or even just small rotation, as demonstrated in a very recent paper (239). Although there is no denying that deep learning has become a very powerful tool for image analysis, building resilient algorithms robust to potential attacks remains an unsolved problem. One potential solution, instead of building the resilience into the model, is raising failure awareness of the deployed networks. This can be achieved by providing users with segmentation quality scores (240) or confidence maps, such as uncertainty maps (166) and attention maps (241). These scores or maps can be used as evidence to alert users when failure happens. For example, Sander et al. (166) built a network that is able to simultaneously predict the segmentation mask over cardiac structures and its associated spatial uncertainty map, where the latter one could be used to highlight potential incorrect regions. Such uncertainty information could alert human experts for further justification and refinement in a human-in-the-loop setting.

4.4. Future Work

4.4.1. Smart Imaging

We have shown that deep learning-based methods are able to segment images in real-time with good accuracy. However, these algorithms can still fail on those image acquisitions with low image quality or significant artifacts. Although there have been several algorithms developed to avoid this problem by either checking the image quality before follow-up studies (242, 243), or predicting the segmentation quality to detect failures (240, 244, 245), the development of algorithms that can give instant feedback to correct and optimize the image acquisition process is also important despite less explored. Improving the imaging quality can greatly improve the effectiveness of medical imaging as well as the accuracy of imaging-based diagnosis. For radiologists, however, finding the optimal imaging and reconstruction parameters to scan each patient can take a great amount of time. Therefore, a DL-based system that has the potential of efficiently and effectively improving the image quality with less noise is of great need. Some researchers have utilized learning-based methods (mostly are deep learning-based) for better image

resolution (62), view planning (246), motion correction (247, 248), artifacts reduction (249), shadow detection (250), and noise reduction (251) after image acquisition. However, combining these algorithms with segmentation algorithms and seamlessly integrating them into an efficient, patient-specific imaging system for high-quality image analysis and diagnosis is still an open challenge. An alternative approach is to directly predict cardiac segmentation maps from undersampled k-space data to accelerate the whole procedure, which bypasses the image reconstruction stage (58).

4.4.2. Data Harmonization

A number of works have reported the existence of missing labels and inconsistent labeling protocols among different cardiac image datasets (27, 232). Variations have been found in defining the end of basal slices as well as the endocardial wall of myocardium (some include papillary muscles as part of the endocardial contours whereas others do not). These inconsistencies can be a major obstacle for transferring, evaluating and deploying deep learning models trained from one domain (e.g., hospital) to another. Therefore, building a standard benchmark dataset like CheXpert (252) that (1) is *large* enough to have substantial data diversity that reflects the spectrum of real-world diversity; (2) has a *standard* labeling protocol approved by experts, is indeed a need. However, directly building such a dataset from scratch is time-consuming and expensive. A more promising way might be developing an automated tool to combine existing public datasets from multiple sources and then to harmonize them to a unified, high-quality dataset. This tool can not only open the door for crowd-sourcing but also enable the rapid deployment of those DL-based segmentation models.

4.4.3. Data Privacy

As deep learning is a data-driven approach, an unavoidable and rife concern is about the data privacy. Regulations, such as The General Data Protection Regulation (GDPR) now play an important role to protect users' privacy and have forced organizations to treat data ownership seriously. On the other hand, from a technical point of view, how to store, query, and process data such that there is no privacy concerns for building deep learning systems has now become an even more difficult but interesting challenge. Building a privacy-preserving algorithm requires to combine cryptography and deep learning together and to mix techniques from a wide range of subjects, such as data analysis, distributed computing, federated learning, differential privacy, in order to achieve models with strong security, fast run time, and great generalizability (253–256). In this respect, Papernot (257) published a report for guidance, which summarized a set of best practices for improving the privacy and security of machine learning systems. Yet, this field is still in its infancy.

5. CONCLUSION

In this review paper, we provided a comprehensive overview of these deep learning techniques used in three common imaging

modalities (MRI, CT, ultrasound), covering a wide range of existing deep learning approaches (mostly are CNN-based) that are designed for segmenting different cardiac anatomical structures (e.g., cardiac ventricle, atria, vessel). In particular, we presented and discussed recent progress of deep learning-based segmentation methods in the three modalities, outlined future potential and the remaining limitations of these deep learning-based cardiac segmentation methods that may hinder widespread clinical deployment. We hope that this review can provide an intuitive understanding of those deep learning-based techniques that have made a significant contribution to cardiac image segmentation and also increase the awareness of common challenges in this field that call for future contribution.

6. DATA AVAILABILITY STATEMENT

The datasets summarized in **Table 6** can be found in their corresponding websites listed below:

- York: <http://www.cse.yorku.ca/~mridataset/>
- Sunnybrook: <http://www.cardiacatlas.org/studies/sunnybrook-cardiac-data/>
- LVSC: <http://www.cardiacatlas.org/challenges/lv-segmentation-challenge/>
- RVSC: <http://www.litislab.fr/?projet=1rvsc>
- cDEMRS: https://www.doc.ic.ac.uk/~rkarim/la_lv_framework/fibrosis
- LVIC: https://www.doc.ic.ac.uk/~rkarim/la_lv_framework/lv_infarct
- LASC'13: www.cardiacatlas.org/challenges/left-atrium-segmentation-challenge/
- HVSMR: <http://segchd.csail.mit.edu/>
- ACDC: <https://acdc.creatis.insa-lyon.fr/>
- LASC'18: <http://atriaseg2018.cardiacatlas.org/data/>
- MM-WHS: <http://www.sdspeople.fudan.edu.cn/zhuangxiahai/0/mmwhs17/>
- CAT08: <http://coronary.bigr.nl/centerlines/>
- CLS12: <http://coronary.bigr.nl/stenoses>
- CETUS: <https://www.creatis.insa-lyon.fr/Challenge/CETUS>
- CAMUS: <https://www.creatis.insa-lyon.fr/Challenge/camus>.

AUTHOR CONTRIBUTIONS

CC, WB, and DR conceived and designed the work. CC, CQ, and HQ searched and read the MR, CT, Ultrasound literature, respectively, and drafted the manuscript together. WB, DR, GT, and JD provided the critical revision with insightful and constructive comments to improve the manuscript. All authors read and approved the manuscript.

FUNDING

This work was supported by the SmartHeart EPSRC Programme Grant (EP/P001009/1). HQ was supported by the EPSRC Programme Grant (EP/R005982/1).

ACKNOWLEDGMENTS

We would like to thank our colleagues: Karl Hahn, Qingjie Meng, James Batten, and Jonathan Passerat-Palmbach

who provided the insight and expertise that greatly assisted the work, and also constructive and thoughtful comments from Turkey Kart that greatly improved the manuscript.

REFERENCES

- Petitjean C, Zuluaga MA, Bai W, Dacher JN, Grosgeorge D, Caudron J, et al. Right ventricle segmentation from cardiac MRI: a collation study. *Med Image Anal.* (2015) 19:187–202. doi: 10.1016/j.media.2014.10.004
- Peng P, Lekadir K, Gooya A, Shao L, Petersen SE, Frangi AF. A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. *Magn Reson Mater Phys Biol Med.* (2016) 29:155–95. doi: 10.1007/s10334-015-0521-4
- Tavakoli V, Amini AA. A survey of shaped-based registration and segmentation techniques for cardiac images. *Comput Vis Image Understand.* (2013) 117:966–89. doi: 10.1016/j.cviu.2012.11.017
- Lesage D, Angelini ED, Bloch I, Funka-Lea G. A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes. *Med Image Anal.* (2009) 13:819–45. doi: 10.1016/j.media.2009.07.011
- Greenspan H, Van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans Med Imaging.* (2016) 35:1153–9. doi: 10.1109/TMI.2016.2553401
- Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng.* (2017) 19:221–48. doi: 10.1146/annurev-bioeng-071516-044442
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005
- Gandhi S, Mosleh W, Shen J, Chow CM. Automation, machine learning, and artificial intelligence in echocardiography: a brave new world. *Echocardiography.* (2018) 35:1402–18. doi: 10.1111/echo.14086
- Mazurowski MA, Buda M, Saha A, Bashir MR. Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI. *J Magn Reson Imaging.* (2019) 49:939–54. doi: 10.1002/jmri.26534
- Andreopoulos A, Tsotsos JK. Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI. *Med Image Anal.* (2008) 12:335–57. doi: 10.1016/j.media.2007.12.003. Data source: <http://www.cse.yorku.ca/~mridataset/>
- Radau P, Lu Y, Connelly K, Paul G, Dick AJ, Wright GA. Evaluation framework for algorithms segmenting short axis cardiac MRI. *MIDAS J.* (2009) 49. Available online at: <http://hdl.handle.net/10380/3070>. Data source: <http://www.cardiacatlas.org/studies/sunnybrook-cardiac-data/>
- Suinesiaputra A, Cowan BR, Al-Agamy AO, Elattar MA, Ayache N, Fahmy AS, et al. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images. *Med Image Anal.* (2014) 18:50–62. doi: 10.1016/j.media.2013.09.001. Data source: <http://www.cardiacatlas.org/challenges/lv-segmentation-challenge/>
- Karim R, Housden RJ, Balasubramaniam M, Chen Z, Perry D, Uddin A, et al. Evaluation of current algorithms for segmentation of scar tissue from late gadolinium enhancement cardiovascular magnetic resonance of the left atrium: an open-access grand challenge. *J Cardiovasc Magn Reson.* (2013) 15:105. doi: 10.1186/1532-429X-15-105. Data source: https://www.doc.ic.ac.uk/~rkarim/la_lv_framework/fibrosis
- Karim R, Bhagirath P, Claus P, James Housden R, Chen Z, Karimghaloo Z, et al. Evaluation of state-of-the-art segmentation algorithms for left ventricle infarct from late gadolinium enhancement MR images. *Med Image Anal.* (2016) 30:95–107. doi: 10.1016/j.media.2016.01.004. Data source: https://www.doc.ic.ac.uk/~rkarim/la_lv_framework/lv_infarct/
- Tobon-Gomez C, Geers AJ, Peters J, Pinto K, Karim R, et al. Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets. *IEEE Trans Med Imaging.* (2015) 34:1460–73. doi: 10.1109/TMI.2015.2398818. Data source: www.cardiacatlas.org/challenges/left-atrium-segmentation-challenge/
- Pace DF, Dalca AV, Geva T, Powell AJ, Moghari MH, Golland P. Interactive whole-heart segmentation in congenital heart disease. *Med Image Comput Assist Interv.* (2015) 9351:80–88. doi: 10.1007/978-3-319-24574-4_10. Data source: <http://segchd.csail.mit.edu/>
- Bernard O, Lalande A, Zotti C, Cervenansky F, Yang X, Heng PA, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE Trans Med Imaging.* (2018) 37:2514–25. doi: 10.1109/TMI.2018.2837502. Data source: <https://acdc.creatis.insa-lyon.fr/>
- Zhao J, Xiong Z. 2018 Left Atrial Segmentation Challenge Dataset (2018). available online at: <http://atriaseg2018.cardiacatlas.org/>. Data source: <http://atriaseg2018.cardiacatlas.org/>
- Zhuang X, Li L, Payer C, Stern D, Urschler M, Heinrich MP, et al. Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. *Med Image Anal.* (2019) 58:101537. doi: 10.1016/j.media.2019.101537. Data source: <http://www.sdspeople.fudan.edu.cn/zhuangxiahai/0/mmwhs17/>
- Schaap M, Metz CT, van Walsum T, van der Giessen AG, Weustink AC, Mollet NR, et al. Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *Med Image Anal.* (2009) 13:701–14. doi: 10.1016/j.media.2009.06.003. Data source: <http://coronary.bigr.nl/centerlines/>
- Kirişli HA, Schaap M, Metz CT, Dharampal AS, Meijboom WB, Papadopolou SL, et al. Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography. *Med Image Anal.* (2013) 17:859–76. doi: 10.1016/j.media.2013.05.007. Data source: <http://coronary.bigr.nl/stenoses>
- Bernard O, Bosch JG, Heyde B, Alessandrini M, Barbosa D, Camarasu-Pop S, et al. Standardized evaluation system for left ventricular segmentation algorithms in 3D echocardiography. *IEEE Trans Med Imaging.* (2016) 35:967–77. doi: 10.1109/TMI.2015.2503890. Data source: <https://www.creatis.insa-lyon.fr/Challenge/CETUS>
- Leclerc S, Smistad E, Pedrosa J, Ostvik A, Cervenansky F, Espinosa F, et al. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE Trans Med Imaging.* (2019) 38:2198–210. doi: 10.1109/TMI.2019.2900516
- Tran PV. A fully convolutional neural network for cardiac segmentation in short-axis MRI. *arXiv* (2016) abs/1604.00494. Available online at: <http://arxiv.org/abs/1604.00494> (accessed September 1, 2019).
- Baumgartner CF, Koch LM, Pollefeys M, Konukoglu E. An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. In: Pop M, Sermesant M, Jodoin P-M, Lalande A, Zhuang X, Yang G, Young AA, Bernard O, editors. *International Workshop on Statistical Atlases and Computational Models of the Heart*, Vol. 10663. Cham: Springer (2017). p. 1–8.
- Isensee F, Jaeger PF, Full PM, Wolf I, Engelhardt S, Maier-Hein KH. Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In: Pop M, Sermesant M, Jodoin P-M, Lalande A, Zhuang X, Yang G, Young AA, Bernard O, editors. *Proceedings of the 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Quebec City, QC: Springer International Publishing (2017). p. 120–9.
- Zheng Q, Delingette H, Duchateau N, Ayache N. 3-D consistent and robust segmentation of cardiac images by deep learning with spatial propagation. *IEEE Trans Med Imaging.* (2018) 37:2137–48. doi: 10.1109/TMI.2018.2820742

28. Qin C, Bai W, Schlemper J, Petersen SE, Piechnik SK, Neubauer S, et al. Joint learning of motion estimation and segmentation for cardiac MR image sequences. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Proceedings, Part I of the 21st International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI, 2018*. Granada: Springer International Publishing (2018). p. 472–80.
29. Khened M, Kollerathu VA, Krishnamurthi G. Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Med Image Anal.* (2019) 51:21–45. doi: 10.1016/j.media.2018.10.004
30. Yu L, Cheng JZ, Dou Q, Yang X, Chen H, Qin J, et al. Automatic 3D cardiovascular MR segmentation with densely-connected volumetric ConvNets. In: Descoteaux M, Maier-Hein L, Franz AM, Jannin P, Collins DL, Duchesne S, editors. *Proceedings, Part II of the 20th International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI, 2017*. Quebec City, QC: Springer International Publishing (2017). p. 287–95.
31. 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) 20:65. doi: 10.1186/s12968-018-0471-x
32. Bai W, Suzuki H, Qin C, Tarroni G, Oktay O, Matthews PM, et al. Recurrent neural networks for aortic image sequence segmentation with sparse annotations. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *21st International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI, 2018*. Granada: Springer International Publishing (2018). p. 586–94.
33. Duan J, Bello G, Schlemper J, Bai W, Dawes TJW, Biffi C, et al. Automatic 3D bi-ventricular segmentation of cardiac images by a shape-constrained multi-task deep learning approach. *IEEE Transactions on Medical Imaging.* (2019) PP:1. doi: 10.1109/TMI.2019.2894322
34. Chen C, Bai W, Rueckert D. Multi-task learning for left atrial segmentation on GE-MRI. In: Pop M, Sermesant M, Zhao J, Li S, McLeod K, Young AA, Rhode KS, Mansi T, editors. *Proceedings of the 9th International Workshop, STACOM 2018, Held in Conjunction With MICCAI 2018, Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*. Granada: Springer International Publishing (2018). p. 292–301.
35. Yu L, Wang S, Li X, Fu CW, Heng PA. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, Yap P-T, Khan A, editors. *Proceedings, Part II of the 22nd International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI, 2019*. Shenzhen: Springer International Publishing (2019). p. 605–13.
36. Yang X, Bian C, Yu L, Ni D, Heng PA. Hybrid loss guided convolutional networks for whole heart parsing. In: Pop M, Sermesant M, Jodoin P-M, Lalonde A, Zhuang X, Yang G, Young AA, Bernard O, editors. *Proceedings of the 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Quebec City, QC: Springer (2017). p. 215–23.
37. Xu Z, Wu Z, Feng J. CFUN: combining faster R-CNN and U-net network for efficient whole heart segmentation. *arxiv* (2018) abs/1812.04914. Available online at: <http://arxiv.org/abs/1812.04914> (accessed September 1, 2019).
38. Merkow J, Marsden A, Kriegman D, Tu Z. Dense volume-to-volume vascular boundary detection. In: Ourselin S, Joskowicz L, Sabuncu MR, Ünal GB, Wells W, editors. *Proceedings, Part III of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI, 2016*. Athens: Springer (2016). p. 371–9.
39. Dou Q, Ouyang C, Chen C, Chen H, Heng PA. Unsupervised cross-modality domain adaptation of ConvNets for biomedical image segmentations with adversarial loss. In: *International Joint Conferences on Artificial Intelligence* (2018). p. 691–7.
40. Dou Q, Ouyang C, Chen C, Chen H, Glocker B, Zhuang X, et al. PnP-AdaNet: plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation. *IEEE Access.* (2019) 7:99065–76. doi: 10.1109/ACCESS.2019.2929258
41. Chen C, Dou Q, Zhou J, Qin J, Heng PA. Synergistic image and feature adaptation: towards cross-modality domain adaptation for medical image segmentation. In: *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*. Honolulu, HI: AAAI Press (2019). p. 865–72.
42. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, et al. Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. *Circulation.* (2018) 138:1623–35. doi: 10.1161/CIRCULATIONAHA.118.034338
43. Goodfellow I. *Deep Learning. Adaptive Computation and Machine Learning*. Cambridge, MA; London: The MIT Press (2016).
44. Szegedy C, Liu W, Jia Y, Sermanet P, Reed SE, Anguelov D, et al. Going deeper with convolutions. In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015*. Boston, MA (2015). p. 1–9.
45. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *3rd International Conference on Learning Representations, ICLR 2015*. San Diego, CA (2015). p. 14. Available online at: <http://arxiv.org/abs/1409.1556>
46. Ciresan DC, Giusti A. Deep neural networks segment neuronal membranes in electron microscopy images. In: *Conference on Neural Information Processing Systems* (2012). p. 2852–60. Available online at: <http://papers.nips.cc/paper/4741-deep-neural-networks-segment-neuronal-membranes-in-electron-microscopy-images>
47. Avendi MR, Kheradvar A, Jafarkhani H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Med Image Anal.* (2016) 30:108–19. doi: 10.1016/j.media.2016.01.005
48. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Conference on Computer Vision and Pattern Recognition* (2015). p. 3431–40. doi: 10.1109/CVPR.2015.7298965
49. Ronneberger FP Olaf, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells III WM, Frangi AF, editors. *Proceedings, Part III of the 18th International Conference, Medical Image Computing and Computer-Assisted Intervention—MICCAI, 2015*. Munich: Springer International Publishing (2015). p. 234–41.
50. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* (2017) 39:640–51. doi: 10.1109/TPAMI.2016.2572683
51. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: *19th International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI, 2016*. Athens (2016). p. 424–32.
52. Milletari F, Navab N, Ahmadi S. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision, 3DV*. Stanford, CA: IEEE Computer Society (2016). p. 565–71.
53. Tao Q, Yan W, Wang Y, Paiman EHM, Shamonin DP, Garg P, et al. Deep learning-based method for fully automatic quantification of left ventricle function from cine MR images: a multivendor, multicenter study. *Radiology.* (2019) 290:180513. doi: 10.1148/radiol.2018180513
54. Xia Q, Yao Y, Hu Z, Hao A. Automatic 3D atrial segmentation from GE-MRIs using volumetric fully convolutional networks. In: Pop M, Sermesant M, Zhao J, Li S, McLeod K, Young AA, Rhode KS, Mansi T, editors. *Proceedings of the 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges—9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018*. Granada: Springer International Publishing (2018). p. 211–20.
55. Poudel RPK, Lamata P, Montana G. Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. In: *1st International Workshops on Reconstruction and Analysis of Moving Body Organs, RAMBO 2016 and 1st International Workshops on Whole-Heart and Great Vessel Segmentation from 3D Cardiovascular MRI in Congenital Heart Disease, HVSMR 2016* (2016). p. 83–94. Available online at: <http://segchd.csail.mit.edu/> (accessed September 1, 2019). Data source: <http://segchd.csail.mit.edu/>
56. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* (1997) 9:1735–80.

57. Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: *Conference on Empirical Methods in Natural Language Processing*. ACL (2014). p. 1724–34. Available online at: <https://www.aclweb.org/anthology/D14-1179/>
58. Schlemper J, Oktay O, Bai W, Castro DC, Duan J, Qin C, et al. Cardiac MR segmentation from undersampled k-space using deep latent representation learning. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *21st International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*. Granada: Springer International Publishing (2018). p. 259–67.
59. Oktay O, Ferrante E, Kamnitsas K, Heinrich M, Bai W, Caballero J, et al. Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. *IEEE Trans Med Imaging*. (2018) 37:384–95. doi: 10.1109/TMI.2017.2743464
60. Biffi C, Oktay O, Tarroni G, Bai W, De Marvao A, Doumou G, et al. Learning interpretable anatomical features through deep generative models: application to cardiac remodeling. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *21st International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*. Vol. 11071 LNCS. Granada: Springer International Publishing (2018). p. 464–71.
61. Painchaud N, Skandarani Y, Judge T, Bernard O, Lalande A, Jodoin PM. Cardiac MRI segmentation with strong anatomical guarantees. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, Yap P-T, Khan A, editors. *Proceedings, Part II of the 22nd International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI, 2019*. Shenzhen: Springer International Publishing (2019). p. 632–40.
62. Oktay O, Bai W, Lee M, Guerrero R, Kamnitsas K, Caballero J, et al. Multi-input cardiac image super-resolution using convolutional neural networks. In: Ourselin S, Joskowicz L, Sabuncu MR, Ünal GB, Wells W, editors. *19th International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI, 2016*. Vol. 9902 LNCS. Athens: Springer International Publishing (2016). p. 246–54.
63. Yue Q, Luo X, Ye Q, Xu L, Zhuang X. Cardiac segmentation from LGE MRI using deep neural network incorporating shape and spatial priors. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, Yap P-T, Khan A, editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI, 2019*. Cham: Springer International Publishing (2019). p. 559–67.
64. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. In: *Conference on Neural Information Processing Systems*. Curran Associates, Inc. (2014). p. 2672–80. Available online at: <http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>
65. Luc P, Couprie C, Chintala S, Verbeek J. Semantic segmentation using adversarial networks. In: *NIPS Workshop on Adversarial Training* (2016). p. 1–12.
66. Savioli N, Vieira MS, Lamata P, Montana G. A generative adversarial model for right ventricle segmentation. *arXiv* (2018) abs/1810.03969. Available online at: <http://arxiv.org/abs/1810.03969> (accessed September 1, 2019).
67. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*. Las Vegas, NV: IEEE Computer Society (2016). p. 2818–26.
68. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. San Francisco, CA (2017). p. 4278–84. Available online at: <http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14806>
69. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. *Conference on Neural Information Processing Systems*. Curran Associates, Inc. (2017). p. 5998–6008.
70. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention U-Net: learning where to look for the pancreas. In: *Medical Imaging With Deep Learning* (2018). p. 1804.03999.
71. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016*. IEEE Computer Society: Las Vegas, NV (2016). p. 770–8.
72. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. In: *International Conference on Learning Representations* (2016). p. 1–13.
73. Lee CY, Xie S, Gallagher P, Zhang Z, Tu Z. Deeply-supervised nets. In: Lebanon G, Vishwanathan SVN, editors. *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research*, Vol. 38. San Diego, CA: PMLR (2015). p. 562–70. Available online at: <http://proceedings.mlr.press/v38/lee15a.html>
74. Chen LC, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. *arXiv* (2017) abs/1706.05587. Available online at: <http://arxiv.org/abs/1706.05587> (accessed September 1, 2019).
75. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell*. (2015) 37:1904–16. doi: 10.1109/TPAMI.2015.2389824
76. Jetley S, Lord NA, Lee N, Torr PHS. Learn to pay attention. In: *6th International Conference on Learning Representations, ICLR 2018, Conference Track Proceedings*. Vancouver, BC (2018).
77. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2018*, Salt Lake City, UT: IEEE Computer Society (2018). p. 7132–41.
78. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Conference on Computer Vision and Pattern Recognition* (2017). p. 2261–9.
79. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. (1986) 323:533–6.
80. Jang Y, Hong Y, Ha S, Kim S, Chang HJ. Automatic segmentation of LV and RV in cardiac MRI. In: Pop M, Sermesant M, Jodoin P-M, Lalande A, Zhuang X, Yang G, Young AA, Bernard O, editors. *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer (2017). p. 161–9.
81. Yang X, Bian C, Yu L, Ni D, Heng PA. Class-balanced deep neural network for automatic ventricular structure segmentation. In: Pop M, Sermesant M, Jodoin P-M, Lalande A, Zhuang X, Yang G, Young AA, Bernard O, editors. *Proceedings of the 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Quebec City, QC: Springer International Publishing (2017). p. 152–60.
82. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. (2014) 15:1929–58. doi: 10.5555/2627435.2670313
83. Kamnitsas K, Bai W, Ferrante E, McDonagh S, Sinclair M. Ensembles of multiple models and architectures for robust brain tumour segmentation. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries—Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017*. Quebec City, QC (2017). p. 450–62.
84. Zheng H, Zhang Y, Yang L, Liang P, Zhao Z, Wang C, et al. A new ensemble learning framework for 3D biomedical image segmentation. In: *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*, Honolulu, HI: AAAI Press (2019). p. 5909–16.
85. Chen S, Ma K, Zheng Y. Med3D: transfer learning for 3D medical image analysis. *arXiv* (2019) abs/1904.00625. Available online at: <http://arxiv.org/abs/1904.00625> (accessed September 1, 2019).
86. Xue W, Brahm G, Pandey S, Leung S, Li S. Full left ventricle quantification via deep multitask relationships learning. *Med Image Anal*. (2018) 43:54–65. doi: 10.1016/j.media.2017.09.005
87. Zheng Q, Delingette H, Ayache N. Explainable cardiac pathology classification on cine MRI with motion characterization by semi-supervised learning of apparent flow. *Med Image Anal*. (2019) 56:80–95. doi: 10.1016/j.media.2019.06.001
88. Bello GA, Dawes TJW, Duan J, Biffi C, de Marvao A, Howard LS, et al. Deep learning cardiac motion analysis for human survival prediction. *Nat Mach Intell*. (2019) 1:95–104. doi: 10.1038/s42256-019-0019-2
89. Van Der Geest RJ, Reiber JH. Quantification in cardiac MRI. *J Magn Reson Imaging*. (1999) 10:602–8.
90. Lieman-Sifry J, Le M, Lau F, Sall S, Golden D. FastVentricle: cardiac segmentation with ENet. In: Pop M, Wright GA, editors. *Functional Imaging*

- and Modelling of the Heart. Vol. 10263 LNCS of Lecture Notes in Computer Science. Cham: Springer International Publishing (2017). p. 127–38.
91. Fahmy AS, El-Rewaify H, Nezafat M, Nakamori S, Nezafat R. Automated analysis of cardiovascular magnetic resonance myocardial native T1 mapping images using fully convolutional neural networks. *J Cardiovasc Magn Reson.* (2019) 21:1–12. doi: 10.1186/s12968-018-0516-1
 92. Patravali J, Jain S, Chilamkurthy S. 2D–3D fully convolutional neural networks for cardiac MR segmentation. In: Pop M, Sermesant M, Jodoin P-M, Lalonde A, Zhuang X, Yang G, Young AA, Bernard O, editors. *Proceedings of the 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges.* Quebec City, QC: Springer (2017). p. 130–9.
 93. Wolterink JM, Leiner T, Viergever MA, Išgum I. Automatic segmentation and disease classification using cardiac cine MR images. In: Pop M, Sermesant M, Jodoin P-M, Lalonde A, Zhuang X, Yang G, Young AA, Bernard O, editors. *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges—8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017.* Vol. 10663 LNCS of Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Quebec City, QC: Springer International Publishing (2017). p. 101–10.
 94. Tan LK, Liew YM, Lim E, McLaughlin RA. Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences. *Med Image Anal.* (2017) 39:78–86. doi: 10.1016/j.media.2017.04.002
 95. Vigneault DM, Xie W, Ho CY, Bluemke DA, Noble JA. Ω -Net (Omega-Net): fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks. *Med Image Anal.* (2018) 48:95–106. doi: 10.1016/j.media.2018.05.008
 96. Avendi MR, Kheradvar A, Jafarikhani H. Automatic segmentation of the right ventricle from cardiac MRI using a learning-based approach. *Magn Reson Med.* (2017) 78:2439–48. doi: 10.1002/mrm.26631
 97. Yang H, Sun J, Li H, Wang L, Xu Z. Deep fusion net for multi-atlas segmentation: application to cardiac MR images. In: Ourselin S, Joskowicz L, Sabuncu MR, Ünal GB, Wells W, editors. *19th International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI, 2016.* Athens: Springer International Publishing (2016). p. 521–8. doi: 10.1007/978-3-319-46723-8_60
 98. Ngo TA, Lu Z, Carneiro G. Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. *Med Image Anal.* (2017) 35:159–71. doi: 10.1016/j.media.2016.05.009
 99. Mortazi A, Karim R, Rhode K, Burt J, Bagci U. CardiacNET: segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN. In: Descoteaux M, Maier-Hein L, Franz AM, Jannin P, Collins DL, Duchesne S, editors. *20th International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI 2017.* Quebec City, QC: Springer (2017). p. 377–85.
 100. Xiong Z, Fedorov VV, Fu X, Cheng E, Macleod R, Zhao J. Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging using a dual fully convolutional neural network. *IEEE Trans Med Imaging.* (2019) 38:515–24. doi: 10.1109/TMI.2018.2866845
 101. Yang G, Zhuang X, Khan H, Haldar S, Nyktari E, Li L, et al. Fully automatic segmentation and objective assessment of atrial scars for long-standing persistent atrial fibrillation patients using late gadolinium-enhanced MRI. *Med Phys.* (2018) 45:1562–76. doi: 10.1002/mp.12832
 102. Chen J, Yang G, Gao Z, Ni H, Firmin D, others. Multiview two-task recursive attention model for left atrium and atrial scars segmentation. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *21st International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI 2018, Vol. 11071.* Granada: Springer (2018). p. 455–63.
 103. Zabihollahy F, White JA, Ukwatta E. Myocardial scar segmentation from magnetic resonance images using convolutional neural network. In: *Medical Imaging 2018: Computer-Aided Diagnosis.* Vol. 10575. International Society for Optics and Photonics (2018). p. 105752Z.
 104. Moccia S, Banali R, Martini C, Muscogiuri G, Pontone G, Pepi M, et al. Development and testing of a deep learning-based strategy for scar segmentation on CMR-LGE images. *Magn Reson Mater Phys Biol Med.* (2019) 32:187–95. doi: 10.1007/s10334-018-0718-4
 105. Xu C, Xu L, Gao Z, Zhao S, Zhang H, Zhang Y, et al. Direct delineation of myocardial infarction without contrast agents using a joint motion feature learning architecture. *Med Image Anal.* (2018) 50:82–94. doi: 10.1016/j.media.2018.09.001
 106. Li J, Zhang R, Shi L, Wang D. Automatic whole-heart segmentation in congenital heart disease using deeply-supervised 3D FCN. In: Zuluaga MA, Bhatia KK, Kainz B, Moghari MH, Pace DF, editors. *Proceedings of the First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Reconstruction, Segmentation, and Analysis of Medical Images.* Athens: Springer International Publishing (2017). p. 111–8.
 107. Wolterink JM, Leiner T, Viergever MA, Išgum I. Dilated convolutional neural networks for cardiovascular MR segmentation in congenital heart disease. In: Zuluaga MA, Bhatia KK, Kainz B, Moghari MH, Pace DF, editors. *Proceedings of the First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Reconstruction, Segmentation, and Analysis of Medical Images.* Athens: Springer International Publishing (2017). p. 95–102. doi: 10.1007/978-3-319-52280-7
 108. Li C, Tong Q, Liao X, Si W, Chen S, Wang Q, et al. APCP-NET: aggregated parallel cross-scale pyramid network for CMR segmentation. In: *16th IEEE International Symposium on Biomedical Imaging, ISBI, Venice* (2019). p. 784–8.
 109. Zotti C, Luo Z, Lalonde A, Jodoin PM. Convolutional neural Network with shape prior applied to cardiac MRI segmentation. *IEEE J Biomed Health Inform.* (2019) 23:1119–28. doi: 10.1109/JBHI.2018.2865450
 110. Zotti C, Luo Z, Lalonde A, Humbert O, Jodoin PM. GridNet with automatic shape prior registration for automatic MRI cardiac segmentation. In: Pop M, Sermesant M, Jodoin P-M, Lalonde A, Zhuang X, Yang G, Young AA, Bernard O, editors. *Proceedings of the 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges.* Vol. 10663. Quebec City, QC: Springer (2017). p. 73–81.
 111. Rohé MM, Sermesant M, Pennec X. Automatic multi-atlas segmentation of myocardium with SVF-Net. In: Pop M, Sermesant M, Jodoin P-M, Lalonde A, Zhuang X, Yang G, Young AA, Bernard O, editors. *Proceedings of the 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017 Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges.* Vol. 10663. Quebec City, QC: Springer International Publishing (2017). p. 170–7.
 112. Tziritis G, Grinias E. Fast fully-automatic localization of left ventricle and myocardium in MRI using MRF model optimization, substructures tracking and B-spline smoothing. In: Pop M, Sermesant M, Jodoin P-M, Lalonde A, Zhuang X, Yang G, Young AA, Bernard O, editors. *Proceedings of the 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges.* Quebec City, QC: Springer International Publishing (2017). p. 91–100.
 113. Zreik M, Leiner T, De Vos BD, Van Hamersvelt RW, Viergever MA, Išgum I. Automatic segmentation of the left ventricle in cardiac CT angiography using convolutional neural networks. In: *International Symposium on Biomedical Imaging* (2016). p. 40–3.
 114. Payer C, Štern D, Bischof H, Urschler M. Multi-label whole heart segmentation using CNNs and anatomical label configurations. In: Pop M, Sermesant M, Jodoin P-M, Lalonde A, Zhuang X, Yang G, Young AA, Bernard O, editors. *Proceedings of the 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges.* Quebec City, QC: Springer International Publishing (2018). p. 190–8.
 115. Tong Q, Ning M, Si W, Liao X, Qin J. 3D deeply-supervised U-net based whole heart segmentation. In: Pop M, Sermesant M, Jodoin P-M, Lalonde A, Zhuang X, Yang G, Young AA, Bernard O, editors. *Proceedings of the 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges.* Quebec City, QC: Springer (2017). p. 224–32.

116. Wang C, MacGillivray T, Macnaught G, Yang G, Newby D. A two-stage 3D Unet framework for multi-class segmentation on full resolution image. *arxiv* (2018) 1–10. Available online at: <http://arxiv.org/abs/1804.04341> (accessed September 1, 2019).
117. Wang C, Smedby Ö. Automatic whole heart segmentation using deep learning and shape context. In: Pop M, Sermesant M, Jodoin P-M, Lalande A, Zhuang X, Yang G, Young AA, Bernard O, editors. *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges—8th International Workshop, STACOM 2017, Held in Conjunction With MICCAI 2017*. Quebec City, QC: Springer (2017). p. 242–9.
118. Mortazi A, Burt J, Bagci U. Multi-planar deep segmentation networks for cardiac substructures from MRI and CT. In: Pop M, Sermesant M, Jodoin P-M, Lalande A, Zhuang X, Yang G, Young AA, Bernard O, editors. *Proceedings of the 8th International Workshop, STACOM 2017, Held in Conjunction With MICCAI 2017, Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Quebec City, QC: Springer International Publishing (2017). p. 199–206.
119. Ye C, Wang W, Zhang S, Wang K. Multi-depth fusion network for whole-heart CT image segmentation. *IEEE Access*. (2019) 7:23421–9. doi: 10.1109/ACCESS.2019.2899635
120. Zreik M, Lessmann N, van Hamersvelt RW, Wolterink JM, Voskuil M, Viergever MA, et al. Deep learning analysis of the myocardium in coronary CT angiography for identification of patients with functionally significant coronary artery stenosis. *Med Image Anal*. (2018) 44:72–85. doi: 10.1016/j.media.2017.11.008
121. Joyce T, Chartsias A, Tsafaris SA. Deep multi-class segmentation without ground-truth labels. In: *Medical Imaging With Deep Learning* (2018). p. 1–9.
122. Moeskops P, Wolterink JM, van der Velden BH, Gilhuijs KG, Leiner T, Viergever MA, et al. Deep learning for multi-task medical image segmentation in multiple modalities. In: Ourselin S, Joskowicz L, Sabuncu MR, Ünal GB, Wells W, editors. *19th International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI, 2016*. Athens: Springer (2016). p. 478–86.
123. Lee MCH, Petersen K, Pawlowski N, Glocker B, Schaap M. TETRIS: template transformer networks for image segmentation with shape priors. *IEEE Trans Med Imaging*. (2019) 38:2596–606. doi: 10.1109/TMI.2019.2905990
124. Gülsün MA, Funka-Lea G, Sharma P, Rapaka S, Zheng Y. Coronary centerline extraction via optimal flow paths and CNN path pruning. In: Ourselin S, Joskowicz L, Sabuncu MR, Ünal GB, Wells W, editors. *19th International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI, 2016*. Athens: Springer (2016). p. 317–25.
125. Guo Z, Bai J, Lu Y, Wang X, Cao K, Song Q, et al. DeepCenterline: a multi-task fully convolutional network for centerline extraction. In: Chung ACS, Gee JC, Yushkevich PA, Bao S, editors. *Proceedings of the 26th International Conference, Information Processing in Medical Imaging—IPMI, 2019*. Hong Kong: Springer (2019). p. 441–53.
126. Shen Y, Fang Z, Gao Y, Xiong N, Zhong C, Tang X. Coronary arteries segmentation based on 3D FCN with attention gate and level set function. *IEEE Access*. (2019) 7:42826–35. doi: 10.1109/ACCESS.2019.2908039
127. Wolterink JM, van Hamersvelt RW, Viergever MA, Leiner T, Išgum I. Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier. *Med Image Anal*. (2019) 51:46–60. doi: 10.1016/j.media.2018.10.005
128. Wolterink JM, Leiner T, Išgum I. Graph convolutional networks for coronary artery segmentation in cardiac CT angiography. *arxiv* (2019) abs/1908.05343. Available online at: <http://arxiv.org/abs/1908.05343> (accessed September 1, 2019).
129. Wolterink JM, Leiner T, de Vos BD, van Hamersvelt RW, Viergever MA, Išgum I. Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Med Image Anal*. (2016) 34:123–36. doi: 10.1016/j.media.2016.04.004
130. Lessmann N, Išgum I, Setio AA, de Vos BD, Ciompi F, de Jong PA, et al. Deep convolutional neural networks for automatic coronary calcium scoring in a screening study with low-dose chest CT. In: Tourassi GD, Armato SG III, editors. *Medical Imaging 2016: Computer-Aided Diagnosis*. Vol. 9785. San Diego, CA: International Society for Optics and Photonics; SPIE (2016). p. 978511.
131. Lessmann N, van Ginneken B, Zreik M, de Jong PA, de Vos BD, Viergever MA, et al. Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions. *IEEE Trans Med Imaging*. (2017) 37:615–25. doi: 10.1109/TMI.2017.2769839
132. Liu J, Jin C, Feng J, Du Y, Lu J, Zhou J. A vessel-focused 3D convolutional network for automatic segmentation and classification of coronary artery plaques in cardiac CTA. In: Pop M, Sermesant M, Zhao J, Li S, McLeod K, Young AA, Rhode KS, Mansi T, editors. *Proceedings of the 9th International Workshop, STACOM 2018, Held in Conjunction With MICCAI 2018, Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*. Granada: Springer (2018). p. 131–41.
133. Santini G, Della Latta D, Martini N, Valvano G, Gori A, Ripoli A, et al. An automatic deep learning approach for coronary artery calcium segmentation. In: *International Federation for Medical and Biological Engineering*. Vol. 65 (2017). p. 374–7.
134. Shadmi R, Mazo V, Bregman-Amitai O, Elnekave E. Fully-convolutional deep-learning based system for coronary calcium score prediction from non-contrast chest CT. In: *15th IEEE International Symposium on Biomedical Imaging, ISBI, 2018*. Washington, DC: IEEE (2018). p. 24–8.
135. Zhang W, Zhang J, Du X, Zhang Y, Li S. An end-to-end joint learning framework of artery-specific coronary calcium scoring in non-contrast cardiac CT. *Computing*. (2019) 101:667–78. doi: 10.1007/s00607-018-0678-6
136. Ma J, Zhang R. Automatic calcium scoring in cardiac and chest CT using DenseRAUnet. *arxiv* (2019) abs/1907.11392. Available online at: <http://arxiv.org/abs/1907.11392> (accessed September 1, 2019).
137. Yang X, Bian C, Yu L, Ni D, Heng PA. 3D convolutional networks for fully automatic fine-grained whole heart partition. In: Pop M, Sermesant M, Jodoin P-M, Lalande A, Zhuang X, Yang G, Young AA, Bernard O, editors. *Proceedings of the 8th International Workshop, STACOM 2017, Held in Conjunction With MICCAI 2017, Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Quebec City, QC: Springer (2017). p. 181–9.
138. Carneiro G, Nascimento J, Freitas A. Robust left ventricle segmentation from ultrasound data using deep neural networks and efficient search methods. In: *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2010). p. 1085–8.
139. Carneiro G, Nascimento JC, Freitas A. The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. *IEEE Trans Image Process*. (2012) 21:968–82. doi: 10.1109/TIP.2011.2169273
140. Nascimento JC, Carneiro G. Deep learning on sparse manifolds for faster object segmentation. *IEEE Trans Image Process*. (2017) 26:4978–90. doi: 10.1109/TIP.2017.2725582
141. Nascimento JC, Carneiro G. Non-rigid segmentation using sparse low dimensional manifolds and deep belief networks. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition—CVPR, 2014*. Columbus, OH: IEEE Computer Society (2014). p. 288–95. doi: 10.1109/CVPR.2014.44
142. Nascimento JC, Carneiro G. One shot segmentation: unifying rigid detection and non-rigid segmentation using elastic regularization. *IEEE Trans Pattern Anal Mach Intell*. (2019). doi: 10.1109/TPAMI.2019.2922959
143. Veni G, Moradi M, Bulu H, Narayan G, Syeda-Mahmood T. Echocardiography segmentation based on a shape-guided deformable model driven by a fully convolutional network prior. In: *15th IEEE International Symposium on Biomedical Imaging—ISBI, 2018*. Washington, DC: IEEE (2018). p. 898–902. doi: 10.1109/ISBI.2018.8363716
144. Carneiro G, Nascimento JC. Multiple dynamic models for tracking the left ventricle of the heart from ultrasound data using particle filters and deep learning architectures. In: *Conference on Computer Vision and Pattern Recognition*. IEEE (2010). p. 2815–22. doi: 10.1109/CVPR.2010.5540013
145. Carneiro G, Nascimento JC. Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data. *IEEE Trans Pattern Anal Mach Intell*. (2013) 35:2592–607. doi: 10.1109/TPAMI.2013.96
146. Jafari MH, Girgis H, Liao Z, Behnami D, Abdi A, Vaseli H, et al. A unified framework integrating recurrent fully-convolutional networks and optical flow for segmentation of the left ventricle in echocardiography data. In: *Deep*

- Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer International Publishing (2018). p. 29–37.
147. Carneiro G, Nascimento JC. Incremental on-line semi-supervised learning for segmenting the left ventricle of the heart from ultrasound data. In: *2011 International Conference on Computer Vision*. Barcelona: IEEE (2011). p. 1700–7. doi: 10.1109/ICCV.2011.6126433
 148. Carneiro G, Nascimento JC. The use of on-line co-training to reduce the training set size in pattern recognition methods: application to left ventricle segmentation in ultrasound. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. Providence, RI: IEEE (2012). p. 948–55.
 149. Smistad E, Ostvik A, Haugen BO, Lovstakken L. 2D left ventricle segmentation using deep learning. In: *2017 IEEE International Ultrasonics Symposium (IUS)*. IEEE (2017). p. 1–4.
 150. Yu L, Guo Y, Wang Y, Yu J, Chen P. Segmentation of fetal left ventricle in echocardiographic sequences based on dynamic convolutional neural networks. *IEEE Trans Biomed Eng.* (2017) 64:1886–95. doi: 10.1109/TBME.2016.2628401
 151. Jafari MH, Girgis H, Abdi AH, Liao Z, Pesteie M, Rohling R, et al. Semi-supervised learning for cardiac left ventricle segmentation using conditional deep generative models as prior. In: *16th IEEE International Symposium on Biomedical Imaging ISBI 2019*. Venice: IEEE (2019). p. 649–52.
 152. Girdhar R, Fouhey DF, Rodriguez M, Gupta A. Learning a predictable and generative vector representation for objects. In: Leibe B, Matas J, Sebe N, Welling M, editors. *Proceedings, Part VI of the 14th European Conference, Computer Vision—ECCV 2016*. Amsterdam: Springer International Publishing (2016). p. 484–99.
 153. Chen H, Zheng Y, Park JH, Heng PA, Zhou SK. Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images. In: Ourselin S, Leo Joskowicz, Mert R. Sabuncu, Ünal GB, William Wells, editors. *Proceedings, Part II of the 19th International Conference, Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016*. Athens: Springer International Publishing (2016). p. 487–95.
 154. Smistad E, Østvik A, Mjal Salte I, Leclerc S, Bernard O, Lovstakken L. Fully automatic real-time ejection fraction and MAPSE measurements in 2D echocardiography using deep neural networks. In: *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE (2018). p. 1–4.
 155. Leclerc S, Smistad E, Grenier T, Lartizien C, Ostvik A, Espinosa F, et al. Deep learning applied to multi-structure segmentation in 2D echocardiography: a preliminary investigation of the required database size. In: *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE (2018). p. 1–4.
 156. Jafari MH, Girgis H, Van Woudenberg N, Liao Z, Rohling R, Gin K, et al. Automatic biplane left ventricular ejection fraction estimation with mobile point-of-care ultrasound using multi-task learning and adversarial training. *Int J Comput Assist Radiol Surg.* (2019) 14:1027–37. doi: 10.1007/s11548-019-01954-w
 157. Dong S, Luo G, Wang K, Cao S, Li Q, Zhang H. A combined fully convolutional networks and deformable model for automatic left ventricle segmentation based on 3D echocardiography. *Biomed Res Int.* (2018) 2018:5682365. doi: 10.1155/2018/5682365
 158. Dong S, Luo G, Wang K, Cao S, Mercado A, Shmuelovich O, et al. VoxelAtlasGAN: 3D left ventricle segmentation on echocardiography with atlas guided generation and voxel-to-voxel discrimination. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *21st International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*. Granada: Springer International Publishing (2018). p. 622–9.
 159. Ghesu FC, Krubasik E, Georgescu B, Singh V, Yefeng Zheng, Hornegger J, et al. Marginal space deep learning: efficient architecture for volumetric image parsing. *IEEE Trans Med Imaging.* (2016) 35:1217–28. doi: 10.1109/TMI.2016.2538802
 160. Degel MA, Navab N, Albarqouni S. Domain and geometry agnostic CNNs for left atrium segmentation in 3D ultrasound. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *21st International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*. Springer International Publishing (2018). p. 630–7.
 161. Karim R, Blake LE, Inoue J, Tao Q, Jia S, James Housden R, et al. Algorithms for left atrial wall segmentation and thickness-evaluation on an open-source CT and MRI image database. *Med Image Anal.* (2018) 50:36–53. doi: 10.1016/j.media.2018.08.004
 162. Li J, Yu Z, Gu Z, Liu H, Li Y. Dilated-inception net: multi-scale feature aggregation for cardiac right ventricle segmentation. *IEEE Trans Biomed Eng.* (2019) 66:3499–508. doi: 10.1109/TBME.2019.2906667
 163. Zhou XY, Yang GZ. Normalization in training U-Net for 2D biomedical semantic segmentation. *IEEE Robot Autom Lett.* (2019) 4:1792–9. doi: 10.1109/LRA.2019.2896518
 164. Zhang J, Du J, Liu H, Hou X, Zhao Y, Ding M. LU-NET: an Improved U-Net for ventricular segmentation. *IEEE Access.* (2019) 7:92539–46. doi: 10.1109/ACCESS.2019.2925060
 165. Cong C, Zhang H. Invert-U-Net DNN segmentation model for MRI cardiac left ventricle segmentation. *J Eng.* (2018) 2018:1463–7. doi: 10.1049/joe.2018.8302
 166. Sander J, de Vos BD, Wolterink JM, Išgum I. Towards increased trustworthiness of deep learning segmentation methods on cardiac MRI. In: *Medical Imaging 2019: Image Processing*. Vol. 10949. International Society for Optics and Photonics (2019). p. 1094919.
 167. Chen M, Fang L, Liu H. FR-NET: focal Loss constrained deep residual networks for segmentation of cardiac MRI. In: *16th IEEE International Symposium on Biomedical Imaging, ISBI, 2019*. Venice: IEEE (2019). p. 764–7.
 168. Chen C, Biffi C, Tarroni G, Petersen S, Bai W, Rueckert D. Learning shape priors for robust cardiac MR segmentation from multi-view images. In: *Medical Image Computing and Computer Assisted Intervention* (2019). p. 523–31.
 169. Du X, Yin S, Tang R, Zhang Y, Li S. Cardiac-DeepIED: automatic pixel-level deep segmentation for cardiac bi-ventricle using improved end-to-end encoder-decoder network. *IEEE J Transl Eng Health Med.* (2019) 7:1–10. doi: 10.1109/JTEHM.2019.2900628
 170. Yan W, Wang Y, Li Z, van der Geest RJ, Tao Q. Left ventricle segmentation via optical-flow-net from short-axis cine MRI: preserving the temporal coherence of cardiac motion. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *21st International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI, 2018*. Vol. 11073 LNCS. Granada: Springer International Publishing (2018). p. 613–21.
 171. Savioli N, Vieira MS, Lamata P, Montana G. Automated segmentation on the entire cardiac cycle using a deep learning work-flow. In: *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)* (2018). p. 153–8.
 172. Clough JR, Oksuz I, Byrne N, Schnabel JA, King AP. Explicit topological priors for deep-learning based image segmentation using persistent homology. In: Chung ACS, Gee JC, Yushkevich PA, Bao S, editors. *Proceedings of the 26th International Conference, IPMI 2019, Information Processing in Medical Imaging*. Vol. 11492 LNCS. Hong Kong (2019). p. 16–28.
 173. Chen X, Williams BM, Vallabhaneni SR, Czanner G, Williams R, Zheng Y. Learning active contour models for medical image segmentation. In: *Conference on Computer Vision and Pattern Recognition* (2019). p. 11632–40.
 174. Qin C, Bai W, Schlemper J, Petersen SE, Piechnik SK, Neubauer S, et al. Joint motion estimation and segmentation from undersampled cardiac MR image. In: Knoll F, Maier AK, Rueckert D, editors. *Machine Learning for Medical Image Reconstruction - First International Workshop, MLMIR 2018, Held in Conjunction with MICCAI 2018*. Granada: Springer International Publishing (2018). p. 55–63.
 175. Dangi S, Yaniv Z, Linte CA. Left ventricle segmentation and quantification from cardiac cine MR images via multi-task learning. In: Pop M, Sermesant M, Zhao J, Li S, McLeod K, Young AA, Rhode KS, Mansi T, editors. *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges—9th International Workshop, STACOM 2018, Held in Conjunction With MICCAI 2018*. Granada: Springer (2018). p. 21–31.
 176. Zhang L, Karanikolas GV, Akçakaya M, Giannakis GB. Fully automatic segmentation of the right ventricle via multi-task deep neural networks. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2018*. Calgary, AB: IEEE (2018). p. 6677–81.

177. Chartsias A, Joyce T, Papanastasiou G, Semple S, Williams M, Newby D, et al. Factorised spatial representation learning: application in semi-supervised myocardial segmentation. In: *Proceedings, Part II of the 21st International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI, 2018*. Vol. 11071 LNCS. Granada: Springer International Publishing (2018). p. 490–8.
178. Chartsias A, Joyce T, Papanastasiou G, Semple S, Williams M, Newby DE, et al. Disentangled representation learning in cardiac image analysis. *Med Image Anal.* (2019) 58:101535. doi: 10.1016/j.media.2019.101535
179. Huang Q, Yang D, Yi J, Axel L, Metaxas D. FR-Net: joint reconstruction and segmentation in compressed sensing cardiac MRI. In: Coudière Y, Ozenne V, Vigmond EJ, Zemzemi N, editors. *10th International Conference on Functional Imaging and Modeling of the Heart—FIMH 2019*. Bordeaux: Springer International Publishing (2019). p. 352–60.
180. Liao F, Chen X, Hu X, Song S. Estimation of the volume of the left ventricle from MRI images using deep neural networks. *IEEE Trans Cybernet.* (2019) 49:495–504. doi: 10.1109/TCYB.2017.2778799
181. Duan J, Schlemper J, Bai W, Dawes TJW, Bello G, Doumou G, et al. Deep nested level sets: fully automated segmentation of cardiac MR images in patients with pulmonary hypertension. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *21st International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI, 2018*. Granada: Springer International Publishing (2018). p. 595–603.
182. Medley DO, Santiago C, Nascimento JC. Segmenting the left ventricle in cardiac MRI: from handcrafted to deep region based descriptors. In: *16th IEEE International Symposium on Biomedical Imaging ISBI, 2019*. Venice: IEEE (2019). p. 644–8.
183. Lu X, Chen X, Li W, Qiao Y. Graph cut segmentation of the right ventricle in cardiac MRI using multi-scale feature learning. In: Wang Y, Chang C-C, editors. *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy—ICCSPP 2019*. Kuala Lumpur: ACM (2019). p. 231–5.
184. Karim R, Mohiaddin R, Rueckert D. Left atrium segmentation for atrial fibrillation ablation. In: *Medical Imaging 2008: Visualization, Image-Guided Procedures, and Modeling*. San Diego, CA: SPIE (2008). p. 69182U.
185. Tao Q, Shahzad R, Ipek EG, Berendsen FF, Nazarian S, van der Geest RJ. Fully automatic segmentation of left atrium and pulmonary veins in late gadolinium-enhanced MRI: towards objective atrial scar assessment. *J Magn Reson Imaging.* (2016) 44:346–54. doi: 10.1002/jmri.25148
186. Zhuang X, Rhode KS, Razavi RS, Hawkes DJ, Ourselin S. A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. *IEEE Trans Medical Imaging.* (2010) 29:1612–25. doi: 10.1109/TMI.2010.2047112
187. Preetha CJ, Haridasan S, Abdi V, Engelhardt S. Segmentation of the left atrium from 3D gadolinium-enhanced MR images with convolutional neural networks. In: Pop M, Sermesant M, Zhao J, Li S, McLeod K, Young AA, Rhode KS, Mansi T, editors. *Proceedings of the 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*. Granada: Springer International Publishing (2018). p. 265–72.
188. Bian C, Yang X, Ma J, Zheng S, Liu YA, Nezafat R, et al. Pyramid network with online hard example mining for accurate left atrium segmentation. In: *Proceedings of the 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges 2018*. Granada: Springer International Publishing (2018). p. 237–45.
189. Savioli N, Montana G, Lamata P. V-FCNN: volumetric fully convolution neural network for automatic atrial segmentation. In: *Proceedings of the 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*. Granada: Springer International Publishing (2018). p. 273–81.
190. Jia S, Despinasse A, Wang Z, Delingette H, Pennec X, Jaïs P, et al. Automatically segmenting the left atrium from cardiac images using successive 3D U-Nets and a contour loss. In: Pop M, Sermesant M, Zhao J, Li S, McLeod K, Young AA, Rhode KS, Mansi T, editors. *Proceedings of the 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*. Granada: Springer International Publishing (2018). p. 221–9.
191. Vesal S, Ravikumar N, Maier A. Dilated convolutions in neural networks for left atrial segmentation in 3D gadolinium enhanced-MRI. In: Pop M, Sermesant M, Zhao J, Li S, McLeod K, Young AA, Rhode KS, Mansi T, editors. *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges—9th International Workshop, STACOM 2018, Held in Conjunction With MICCAI 2018*. Granada: Springer (2018). p. 319–28.
192. Li C, Tong Q, Liao X, Si W, Sun Y, Wang Q, et al. Attention based hierarchical aggregation network for 3D left atrial segmentation: 9th international workshop, STACOM 2018, held in conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers. In: Pop M, Sermesant M, Zhao J, Li S, McLeod K, Young A, et al., editors. *Proceedings of the 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*. Vol. 11395 of *Lecture Notes in Computer Science*. Cham; Granada: Springer International Publishing (2018). p. 255–64.
193. Yang G, Chen J, Gao Z, Zhang H, Ni H, Angelini E, et al. Multiview sequential learning and dilated residual learning for a fully automatic delineation of the left atrium and pulmonary veins from late gadolinium-enhanced cardiac MRI images. In: *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2018*. Honolulu, HI: IEEE (2018) 2018:1123–7. doi: 10.1109/EMBC.2018.8512550
194. Kim RJ, Fieno DS, Parrish TB, Harris K, Chen EL, Simonetti O, et al. Relationship of MRI delayed contrast enhancement to irreversible injury, infarct age, and contractile function. *Circulation.* (1999) 100:1992–2002.
195. Carminati MC, Boniotti C, Fusini L, Andreini D, Pontone G, Pepi M, et al. Comparison of image processing techniques for nonviable tissue quantification in late gadolinium enhancement cardiac magnetic resonance images. *J Thorac Imaging.* (2016) 31:168–76. doi: 10.1097/RTI.0000000000000206
196. Yang G, Zhuang X, Khan H, Haldar S, Nytkari E, Ye X, et al. Segmenting atrial fibrosis from late gadolinium-enhanced cardiac MRI by deep-learned features with stacked sparse auto-encoders. In: Hernández MCV, González-Castro V, editors. *Proceedings of the 21st Annual Conference on Medical Image Understanding and Analysis—MIUA*. Edinburgh: Springer International Publishing (2017). p. 195–206.
197. Fahmy AS, Rausch J, Neisius U, Chan RH, Maron MS, Appelbaum E, et al. Automated cardiac MR scar quantification in hypertrophic cardiomyopathy using deep convolutional neural networks. *JACC Cardiovasc Imaging.* (2018) 11:1917–8. doi: 10.1016/j.jcmg.2018.04.030
198. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging.* (1999) 18:712–21.
199. Herment A, Kachenoura N, Lefort M, Bensalah M, Dogui A, Frouin F, et al. Automated segmentation of the aorta from phase contrast MR images: validation against expert tracing in healthy volunteers and in patients with a dilated aorta. *J Magn Reson Imaging.* (2010) 31:881–8. doi: 10.1002/jmri.22124
200. Shi Z, Zeng G, Zhang L, Zhuang X, Li L, Yang G, et al. Bayesian VoxDRN: a probabilistic deep voxelwise dilated residual network for whole heart segmentation from 3D MR images. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *21st International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI, 2018*. Vol. 49. Granada: Springer International Publishing (2018). p. 569–77. Available online at: <http://hdl.handle.net/10380/3070> (accessed September 1, 2019).
201. Kang DW, Woo J, Slomka PJ, Dey D, Germano G, Kuo J. Heart chambers and whole heart segmentation techniques: review. *J Electron Imaging.* (2012) 21:010901. doi: 10.1117/1.JEI.21.1.010901
202. Dormer JD, Ma L, Halicek M, Reilly CM, Schreibmann E, Fei B. Heart chamber segmentation from CT using convolutional neural networks. In: Fei B, Webster RJ, editors. *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*. Houston, TX: SPIE (2018). p. 105782S.

203. de Vos BD, Wolterink JM, De Jong PA, Leiner T, Viergever MA, Išgum I. ConvNet-based localization of anatomical structures in 3-D medical images. *IEEE Trans Med Imaging*. (2017) 36:1470–81. doi: 10.1109/TMI.2017.2673121
204. Zhang DP. *Coronary Artery Segmentation and Motion Modelling*. London: Imperial College (2010).
205. Huang W, Huang L, Lin Z, Huang S, Chi Y, Zhou J, et al. Coronary artery segmentation by deep learning neural networks on computed tomographic coronary angiographic images. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC*, 2018. Honolulu, HI: IEEE (2018). p. 608–11.
206. Chen YC, Lin YC, Wang CP, Lee CY, Lee WJ, Wang TD, et al. Coronary artery segmentation in cardiac CT angiography using 3D multi-channel U-net. In: Jorge Cardoso M, Feragen A, Glocker B, Konukoglu E, Oguz I, Unal GB, Vercauteren T, editors. *International Conference on Medical Imaging with Deep Learning, MIDL, 2019*. London (2019). p. 1907.12246.
207. Duan Y, Feng J, Lu J, Zhou J. Context aware 3D fully convolutional networks for coronary artery segmentation. In: Pop M, Serresant M, Zhao J, Li S, McLeod K, Young AA, Rhode KS, Mansi T, editors. *Proceedings of the 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*. Granada: Springer (2018). p. 85–93.
208. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M, Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol*. (1990) 15:827–32.
209. de Vos BD, Wolterink JM, Leiner T, de Jong PA, Lessmann N, Išgum I. Direct automatic coronary calcium scoring in cardiac and chest CT. *IEEE Trans Med Imaging*. (2019) 38:2127–38. doi: 10.1109/TMI.2019.2899534
210. Cano-Espinosa C, González G, Washko GR, Cazorla M, Estépar RSJ. Automated Agatston score computation in non-ECG gated CT scans using deep learning. In: Angelini ED, Landman BA, editors. *Medical Imaging 2018: Image Processing*. Vol. 10574. Houston, TX: International Society for Optics and Photonics; SPIE (2018). p. 105742K.
211. Zreik M, van Hamersvelt RW, Wolterink JM, Leiner T, Viergever MA, Išgum I. A recurrent CNN for automatic detection and classification of coronary artery plaque and stenosis in coronary CT angiography. *IEEE Trans Med Imaging*. (2018) 38:1588–98. doi: 10.1109/TMI.2018.2883807
212. Noble JA, Boukerroui D. Ultrasound image segmentation: a survey. *IEEE Trans Med Imaging*. (2006) 25:987–1010. doi: 10.1109/TMI.2006.877092
213. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science*. (2006) 313:504–7. doi: 10.1126/science.1127647
214. Smistad E, Lindseth F. Real-time tracking of the left ventricle in 3D ultrasound using Kalman filter and mean value coordinates. In: *Medical Image Segmentation for Improved Surgical Navigation* (2014). p. 189.
215. Georgescu B, Zhou XS, Comaniciu D, Gupta A. Database-guided segmentation of anatomical structures with complex appearance. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 2. San Diego, CA: IEEE (2005). p. 429–36.
216. Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. *Int J Comput Vis*. (1988) 1:321–31. doi: 10.1007/BF00133570
217. Kamnitsas K, Baumgartner C, Ledig C, Newcombe V, Simpson J, Kane A, et al. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: Niethammer M, Styner M, Aylward SR, Zhu H, Oguz I, Yap P-T, Shen D, editors. *Proceedings of the 25th International Conference on Information Processing in Medical Imaging—IPMI, 2017*. Boone, NC: Springer International Publishing (2017). p. 597–609. doi: 10.1007/978-3-319-59050-9_47
218. Zheng Y, Barbu A, Georgescu B, Scheuering M, Comaniciu D. Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. *IEEE Trans Med Imaging*. (2008) 27:1668–81. doi: 10.1109/TMI.2008.2004421
219. Kingma DP, Welling M. Auto-encoding variational bayes. In: Bengio Y, LeCun Y, editors. *2nd International Conference on Learning Representations, ICLR 2014*. Banff, AB (2013). p. 1–14.
220. Cubuk ED, Zoph B, Mane D, Vasudevan V, Le QV. AutoAugment: learning augmentation policies from data. In: *Conference on Computer Vision and Pattern Recognition* (2019). p. 113–23.
221. Volpi R, Namkoong H, Sener O, Duchi JC, Murino V, Savarese S. Generalizing to unseen domains via adversarial data augmentation. In: Bengio S, Wallach HM, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, editors. *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018*. Montréal, QC (2018). p. 5339–49.
222. Zhao A, Balakrishnan G, Durand F, Guttag JV, Dalca AV. Data augmentation using learned transformations for one-shot medical image segmentation. In: *Conference on Computer Vision and Pattern Recognition* (2019). p. 8543–53.
223. Chaitanya K, Karani N, Baumgartner C, Donati O, Becker A, Konukoglu E. Semi-supervised and task-driven data augmentation. In: Chung ACS, Gee JC, Yushkevich PA, Bao S, editors. *Proceedings of the 26th International Conference—IPMI, 2019, Information Processing in Medical Imaging*. Hong Kong (2019). p. 29–41.
224. Bai W, Oktay O, Sinclair M, Suzuki H, Rajchl M, Tarroni G, et al. Semi-supervised learning for network-based cardiac MR image segmentation. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors. *Proceedings, Part II of the 20th International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI, 2017*. Vol. 10434. Quebec City, QC: Springer International Publishing (2017). p. 253–60.
225. Can YB, Chaitanya K, Mustafa B, Koch LM, Konukoglu E, Baumgartner CF. Learning to segment medical images with scribble-supervision alone. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer International Publishing (2018). p. 236–44.
226. Kervadec H, Dolz J, Tang M, Granger E, Boykov Y, Ben Ayed I. Constrained-CNN losses for weakly supervised segmentation. *Med Image Anal*. (2019) 54:88–99. doi: 10.1016/j.media.2019.02.009
227. Bai W, Chen C, Tarroni G, Duan J, Guitton F, Petersen SE, et al. Self-supervised learning for cardiac MR image segmentation by anatomical position prediction. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, Yap P-T, Khan A, editors. *Proceedings, Part II of the 22nd International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI, 2019*. Shenzhen: Springer International Publishing (2019). p. 541–9.
228. Dou Q, Liu Q, Heng PA, Glocker B. Unpaired multi-modal segmentation via knowledge distillation. *arxiv* (2020) abs/2001.03111. Available online at: <https://arxiv.org/abs/2001.03111>
229. Ouyang C, Kamnitsas K, Biffi C, Duan J, Rueckert D. Data efficient unsupervised domain adaptation for cross-modality image segmentation. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, Yap P-T, Khan A, editors. *22nd International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI 2019*. Shenzhen: Springer International Publishing (2019). p. 669–77.
230. Chen J, Zhang H, Zhang Y, Zhao S, Mohiaddin R, Wong T, et al. Discriminative consistent domain generation for semi-supervised learning. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, Yap P, Khan A, editors. *22nd International Conference on Medical Image Computing and Computer Assisted Intervention - MICCAI 2019*, Vol. 11765. Shenzhen: Springer (2019). p. 595–604.
231. Dou Q, Castro DC, Kamnitsas K, Glocker B. Domain generalization via model-agnostic learning of semantic features. In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019*. Vancouver, BC (2019). p. 6447–58. Available online at: <http://arxiv.org/abs/1910.13580>
232. Chen C, Bai W, Davies RH, Bhuvan AN, Manisty C, Moon JC, et al. Improving the generalizability of convolutional neural network-based segmentation on CMR images. *arxiv* (2019) abs/1907.01268. Available online at: <http://arxiv.org/abs/1907.01268> (accessed September 1, 2019).
233. Zhang L, Wang X, Yang D, Sanford T, Harmon S, Turkbey B, et al. When unseen domain generalization is unnecessary? Rethinking data augmentation. *arxiv* (2019) abs/1906.03347. Available online at: <http://arxiv.org/abs/1906.03347> (accessed September 1, 2019).
234. Mahapatra D, Bozorgtabar B, Thiran JP, Reyes M. Efficient active learning for image classification and segmentation using a sample selection and conditional generative adversarial network. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *21st*

- International Conference on Medical Image Computing and Computer Assisted Intervention - MICCAI 2018*. Granada: Springer International Publishing (2018). p. 580–8.
235. Castro FM, Marín-Jiménez MJ, Guil N, Schmid C, Alahari K. End-to-end incremental learning. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. *Computer Vision—ECCV 2018 - 15th European Conference*. Munich: Springer (2018). p. 241–57.
 236. Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, et al. Intriguing properties of neural networks. In: *2nd International Conference on Learning Representations, ICLR 2014, Conference Track Proceedings*. Banff, AB (2014). p. 1–10.
 237. Kurakin A, Goodfellow I, Bengio S. Adversarial examples in the physical world. In: *5th International Conference on Learning Representations, ICLR 2017, Workshop Track Proceedings*. Toulon (2017). p. 1–14.
 238. Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. In: *International Conference on Learning Representations* (2015). p. 43405. Available online at: <http://arxiv.org/abs/1412.6572>
 239. Finlayson SG, Bowers JD, Ito J, Zittrain JL, Beam AL, Kohane IS. Adversarial attacks on medical machine learning. *Science*. (2019) 363:1287–9. doi: 10.1126/science.aaw4399
 240. Robinson R, Valindria VV, Bai W, Oktay O, Kainz B, Suzuki H, et al. Automated quality control in image segmentation: application to the UK biobank cardiovascular magnetic resonance imaging study. *J Cardiovasc Magn Reson*. (2019) 21:18. doi: 10.1186/s12968-019-0523-x
 241. Heo J, Lee HB, Kim S, Lee J, Kim KJ, Yang E, et al. Uncertainty-aware attention for reliable interpretation and prediction. In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, editors. *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018*. Montréal, QC: Curran Associates Inc. (2018). p. 909–18.
 242. Ruijsink B, Puyol-Antón E, Oksuz I, Sinclair M, Bai W, Schnabel JA, et al. Fully automated, quality-controlled cardiac analysis from CMR: validation and large-scale application to characterize cardiac function. *J Am Coll Cardiol*. (2019). doi: 10.1016/j.jcmg.2019.05.030
 243. Tarroni G, Oktay O, Bai W, Schuh A, Suzuki H, Passerat-Palmbach J, et al. Learning-based quality control for cardiac MR images. *IEEE Trans Med Imaging*. (2019) 38:1127–38. doi: 10.1109/TMI.2018.2878509
 244. Peng B, Zhang L. Evaluation of image segmentation quality by adaptive ground truth composition. In: *European Conference on Computer Vision*. Berlin; Heidelberg: Springer Berlin Heidelberg (2012). p. 287–300.
 245. Zhou L, Deng W, Wu X. Robust image segmentation quality assessment without ground truth. *arXiv* (2019) abs/1903.08773. Available online at: <http://arxiv.org/abs/1903.08773> (accessed September 1, 2019).
 246. Alansary A, Folgoc LL, Vaillant G, Oktay O, Li Y, Bai W, et al. Automatic view planning with multi-scale deep reinforcement learning agents. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *21st International Conference on Medical Image Computing and Computer Assisted Intervention - MICCAI, 2018*. Springer (2018). p. 277–85.
 247. Dangi S, Linte CA, Yaniv Z. Cine cardiac MRI slice misalignment correction towards full 3D left ventricle segmentation. In: Fei B, Webster RJ. *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*. Houston, TX: SPIE (2018). p. 1057607.
 248. Tarroni G, Oktay O, Sinclair M, Bai W, Schuh A, Suzuki H, et al. A comprehensive approach for learning-based fully-automated inter-slice motion correction for short-axis cine cardiac MR image stacks. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *21st International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*. Granada (2018). p. 268–76.
 249. Oksuz I, Clough J, Bai W, Ruijsink B, Puyol-Antón E, Cruz G, et al. High-quality segmentation of low quality cardiac MR images using k-space artefact correction. In: Cardoso MJ, Feragen A, Glocker B, Konukoglu E, Oguz I, Unal G, et al., editors. *International Conference on Medical Imaging with Deep Learning, MIDL, 2019*. London: PMLR (2019). p. 380–9.
 250. Meng Q, Sinclair M, Zimmer V, Hou B, Rajchl M, Toussaint N, et al. Weakly supervised estimation of shadow confidence maps in fetal ultrasound imaging. *IEEE Trans Med Imaging*. (2019) 38:2755–67. doi: 10.1109/TMI.2019.2913311
 251. Wolterink JM, Leiner T, Viergever MA, Isgum I. Generative adversarial networks for noise reduction in low-dose CT. *IEEE Trans Med Imaging*. (2017) 36:2536–45. doi: 10.1109/TMI.2017.2708987
 252. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, et al. CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. In: *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*. Honolulu, HI: AAAI Press (2019). p. 590–7. doi: 10.1609/aaai.v33i01.3301590
 253. Dwork C, Roth A. The algorithmic foundations of differential privacy. *Foundat Trends Theor Comput Sci*. (2014) 9:211–407. doi: 10.1561/04000000042
 254. Abadi M, Chu A, Goodfellow IJ, McMahan HB, Mironov I, Talwar K, et al. Deep learning with differential privacy. In: *The 2016 ACM SIGSAC Conference on Computer and Communications Security Vienna* (2016). p. 308–18.
 255. Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, et al. Practical secure aggregation for privacy-preserving machine learning. In: *The 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017*. Dallas, TX (2017). p. 1175–91.
 256. Ryffel T, Trask A, Dahl M, Wagner B, Mancuso J, Rueckert D, et al. A generic framework for privacy preserving deep learning. In: *Privacy Preserving Machine Learning* (2018). p. 1–8.
 257. Papernot N. A Marauder's map of security and privacy in machine learning: an overview of current and future research directions for making machine learning secure and private. In: *The 11th ACM Workshop on Artificial Intelligence and Security, CCS 2018*. Toronto, ON (2018). p. 1.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Chen, Qin, Qiu, Tarroni, Duan, Bai and Rueckert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



AI in Cardiac Imaging: A UK-Based Perspective on Addressing the Ethical, Social, and Political Challenges

Matthew E. Fenech* and Olly Buston

Future Advocacy, London, United Kingdom

OPEN ACCESS

Edited by:

Tim Leiner,
University Medical Center
Utrecht, Netherlands

Reviewed by:

Fabien Hyafil,
Assistance Publique Hôpitaux De
Paris, France
Steffen Erhard Petersen,
Queen Mary University of London,
United Kingdom

*Correspondence:

Matthew E. Fenech
matthew.fenech@ada.com

Specialty section:

This article was submitted to
Cardiovascular Imaging,
a section of the journal
Frontiers in Cardiovascular Medicine

Received: 31 October 2019

Accepted: 20 March 2020

Published: 15 April 2020

Citation:

Fenech ME and Buston O (2020) AI in
Cardiac Imaging: A UK-Based
Perspective on Addressing the Ethical,
Social, and Political Challenges.
Front. Cardiovasc. Med. 7:54.
doi: 10.3389/fcvm.2020.00054

Imaging and cardiology are the healthcare domains which have seen the greatest number of FDA approvals for novel data-driven technologies, such as artificial intelligence, in recent years. The increasing use of such data-driven technologies in healthcare is presenting a series of important challenges to healthcare practitioners, policymakers, and patients. In this paper, we review ten ethical, social, and political challenges raised by these technologies. These range from relatively pragmatic concerns about data acquisition to potentially more abstract issues around how these technologies will impact the relationships between practitioners and their patients, and between healthcare providers themselves. We describe what is being done in the United Kingdom to identify the principles that should guide AI development for health applications, as well as more recent efforts to convert adherence to these principles into more practical policy. We also consider the approaches being taken by healthcare organizations and regulators in the European Union, the United States, and other countries. Finally, we discuss ways by which researchers and frontline clinicians, in cardiac imaging and more broadly, can ensure that these technologies are acceptable to their patients.

Keywords: artificial intelligence, ethics, policy, principles, regulation

INTRODUCTION

Technological change is certainly not a new phenomenon. 3.3 million-year old stone tools made by *Australopithecus*, one of the earliest hominid species, have been found in Kenya (1), indicating that the drive to use tools to make tasks easier, and hence to improve quality of life, has not changed over the millions of years of humanity's history. One thing that has certainly changed over this time period, however, is the increased speed at which technological progress occurs. Gordon Moore's famous prediction that the number of transistors per square inch on an integrated circuit would double every one to two years (2) has stood the test of time, and other metrics of technological advancement, such as data storage per unit cost, speed of DNA sequencing, and internet bandwidth, have also increased at exponential rates over the last few decades (3).

With new technologies come new potential socioeconomic impacts, and new reactions to these real and imagined impacts by governments and international bodies. Once again, the impulse to regulate novel technologies is long-standing—the history of everything from the railways, to the automobile, to mining, to *in vitro* fertilization provides fascinating case studies in how societies of the day reacted to unfamiliar technology. Nevertheless, it is arguable that artificial intelligence (AI)

is unlike other technologies in that never before has there been such a general-purpose technology that makes us question what it means to be human (4–6). Although there is no sign of anything approaching artificial general intelligence (AGI), the very fact that AI poses such deeply existential questions is just one of the challenges that it poses to regulators and policymakers, particularly in the hugely sensitive area of healthcare. In this paper, we discuss the various ethical, social, and political challenges the application of AI to health and care presents, and how reactions to these challenges are being used to develop principles for action. In some jurisdictions, these principles are being translated into policy and regulation, clearly setting out what should and should not be allowed. Moreover, we outline what researchers and clinicians can do to help ensure that the use of these technologies is acceptable to patients and practitioners alike.

ETHICAL, SOCIAL, AND POLITICAL CHALLENGES

Future Advocacy, an independent think tank focused on policy development around the responsible use of emerging technology, conducted a series of interviews with expert clinicians, technologists, and ethicists, as well as focus groups with patients, and identified ten sets of questions that are raised by the application of AI to the health setting (**Table 1**) (7). In the following sections, we briefly discuss each in turn, and reflect on any advances in thinking and practice that have occurred since the publication of our original report.

Relationships

Healthcare is built on a complex network of relationships between various stakeholders. The primacy of the relationship between patients and their healthcare professionals (HCPs) is clear from the value still placed on it, even in the context of medical practice that is increasingly characterized by the use

of technology (5, 7). It is however but one of the relationships in healthcare—others include those between the HCP and caregivers/relatives; between different HCPs; between top-level administrators and HCPs “on the ground”; and between patients and wider society (8). All of these relationships could be impacted by the introduction of an AI algorithm, and the inferences or predictions it provides, as a “third party” in what were previously two-way interactions. What will patients do, for example, when faced with the scenario of their doctor’s recommendation clashing with the suggestion for treatment provided by an AI tool? How will patients react to an error in their care that is traced back to a decision made, or supported, by AI? The specific issue of liability for error is discussed in section 2.5 below, and a Royal Society-commissioned study found that many members of the public were optimistic about the possibility for AI to *reduce* medical error (5), but there is a need for more research aimed at understanding how patients are likely to respond to such AI-derived errors.

Another way by which AI may impact relationships in healthcare is through its potential to fundamentally change the role of doctors and other HCPs. To paraphrase Mark Twain, reports of the death of the radiologist are greatly exaggerated (9). Nevertheless, as AI tools become better at performing certain circumscribed tasks in healthcare, such as image recognition, the repertoire of tasks that make up a HCP’s job will change. Some have expressed their hope that the “delegation” of such tasks to algorithms will free up more time for HCPs to spend with patients and their relatives (10), but previous experience of the introduction of different technologies into the clinical space suggests that they may well increase clinician workload in both primary and secondary care (11–13). Whether AI is different remains unknown; various medical bodies are grappling with this question (14, 15), and at the time of writing, Health Education England (the body in England responsible for postgraduate training and development of NHS England’s workforce) was holding a consultation on the topic of the “Future Doctor” (16).

Data

AI is increasingly being used to identify patterns in and extract value from the vast amounts of data being generated by individuals, governments, and companies. Healthcare is no exception—the volume, complexity and longevity of healthcare data are all rising fast, with some estimates predicting that the total amount of healthcare data will reach 2.3 billion gigabytes by next year (17). With larger volumes and greater complexity come new questions about the implications of such data use and storage. Firstly, there is the pragmatic concern of how informed consent—the bedrock of interactions between patients and healthcare systems since at least the nineteenth century (18)—is obtained from each and every contributor to a dataset, which may number in the millions. Similarly, as the technology is developing so rapidly, new insights are derived from existing datasets that could not have been predicted before data analysis, as evidenced by the Google/Verily Life Science deep learning algorithm that can determine gender from retinal photographs (19). How do we obtain informed consent for future, unimagined uses of data? The European Union (EU)

TABLE 1 | Ten major ethical, social, and political challenges of the use of artificial intelligence technologies in health and care.

1. What effect will AI have on human relationships in health and care?
2. How is the use, storage and sharing of medical data impacted by AI?
3. What are the implications of issues around algorithmic transparency/explainability on health?
4. Will these technologies help eradicate or exacerbate existing health inequalities?
5. What is the difference between an algorithmic decision and a human decision?
6. What do patients and members of the public want from AI and related technologies?
7. How should these technologies be regulated?
8. Just because these technologies could enable access to new information, should we always use them?
9. What makes algorithms, and the entities that create them, trustworthy?
10. What are the implications of collaboration between public and private sector organizations in the the development of these tools?

General Data Protection Regulation (GDPR) already makes it clear that there are multiple “lawful bases” for data processing, and informed consent is only one of them (20). Clearly, the field of health and care needs to determine whether alternatives to individualized informed consent (including broad consent, “opt-out” consent, and presumed consent) are acceptable in the context of AI research and development, whilst maintaining their patients’ and research subjects’ trust (21). GDPR also sets clear restrictions on how identifiable information about particular data subjects can or cannot be shared, and these are particularly relevant in the age of establishment and curation of “big data” healthcare datasets. Patients and research subjects are right to expect that their data, donated in good faith for use in research, does not end up being used to determine health insurance premiums, for example (22). Although the regulation exists, this is only as good as its enforcement, and concerns about the rigor with which GDPR is being enforced have been raised in other sectors (23). Ultimately, when it comes to such sensitive subjects, a reliance on regulation alone is not sufficient; this must be backed up by education of, and dialogue between, all stakeholders, focusing on their data rights and responsibilities in law.

A consideration that is perhaps particularly relevant to radiologists was highlighted in the Joint Statement on “Ethics of Artificial Intelligence in Radiology,” issued by the American College of Radiologists, European Society of Radiology, Radiological Society of North America, Society for Imaging Informatics in Medicine, European Society of Medical Imaging Informatics, Canadian Association of Radiologists, and American Association of Physicists in Medicine (24). Radiologists are in great demand to provide accurate and replicable labeling of radiological images, which are then used in supervised learning, for example in training convolutional neural networks. Those with expertise in cardiac imaging will be particularly sought after, given the especially time-consuming and resource-intensive nature of interpreting cardiac imaging modalities such as cardiac magnetic resonance (25), and any difficulty in recruiting such experts may well slow the development of these tools in this area of radiology. As any practicing clinician knows, labeling and classification of real-world clinical imaging is similar to all medical decision-making in that it involves many assumptions, heuristics, and potential biases (26–28). When processing data for use in AI training, radiologists need to be aware of these biases, to avoid introducing additional bias into imperfect datasets, as well as recognizing the various incentives and pressures that may influence their decision-making, including commercial pressures to provide these data (24, 29, 30). Radiology training programmes will need to be updated to make sure the radiologists of the future are best prepared to spot and mitigate these problems (31).

Perhaps of all the challenges discussed in this review, those surrounding data are the ones best addressed by existing regulation, with the Privacy Rule created under the Health Insurance Portability and Accountability Act (HIPAA) well-established in the United States, and GDPR incentivizing businesses and public bodies to give their European clients greater control of their data (32). Nevertheless, gaps remain.

HIPAA’s Privacy Rule, for example, does not cover non-health information from which health-related conclusions can be drawn, or user-generated health information (33)—such omissions cannot be tolerated for long in an age of linked datasets and wearable technologies constantly monitoring parameters such as heart rate. Moreover, although Article 22 on automated decision-making is clearly relevant, the words “artificial intelligence” do not appear in the text of the GDPR once, as the regulation is relatively agnostic about the downstream use of the data. This is in contrast to, for example, the guidance on the regulation of data-driven technology published by the German Government’s Data Ethics Commission, which explicitly draws links between data ethics and algorithmic ethics (34). European policy experts have reason to believe that this document will prove influential as the European Commission (EC) develops widely-expected regulation on AI in 2020 (35). The framework for such regulation was laid out in the EC’s White Paper on AI, published in February, which is now open to public consultation (36).

Transparency and Explainability

The “black box” problem is one of the major foci of AI ethics (37). Besides referring to the inherent opacity of complex machine learning algorithms such as neural networks, it is also the case that the increasing size of datasets used in developing AI for health makes explanations of the relationships between input data and outputs difficult—understanding how each of millions of variables contributes to the final output may be computationally intractable (38). Questions that may therefore follow include: How can patients give meaningful informed consent to, or clinicians advise the use of, algorithms the internal workings of which are unclear? (39) Should we be using black box algorithms in healthcare at all?

It is easy to forget that the human brain is itself a “black box,” given the ease with which we explain our own decisions via *post hoc* rationalization (40, 41). The field of medicine has therefore been accustomed to dealing with black box decision-making for millennia. Of course, part of the difference between an opaque human decision and an opaque algorithmic one is the ability to have a conversation with the former, such that the decision itself can be probed and aspects of the decision that are important to its subject better understood. This highlights an important concept that should be considered when grappling with the issue of explainability, which is the distinction between “model-centric explanations” (where the focus is on providing a complete account of how the model works), and “subject-centric explanations” (where “only” those aspects of model functioning that are relevant to the subject are considered) (42). Given that different subjects may require different types of explanation, there is a very strong argument for addressing the black box problem through thorough user/stakeholder research, and their meaningful involvement throughout the development process. Thus, rather than a blanket requirement of full explainability, smart regulatory frameworks may opt to give regard to the application of the AI tool, its intended target group, and its risk profile, with higher risk applications in more vulnerable groups necessitating deeper explanations. Nevertheless, we contend that

one area of transparency should remain strictly enforced, namely that developers and healthcare system administrators make it absolutely clear to service users when an algorithm is being used to support or to independently provide decision-making.

Health Inequalities

A systematic review found significant aversion amongst the UK public to health inequalities, particularly when such inequalities are presented in the context of socioeconomic differences (43). Thus, any suggestion that the use of AI in medicine may exacerbate existing health inequalities, for example by automating existing bias and unfairness at speed and scale, is likely to decrease trust in and acceptability of these tools. Sadly, there is evidence that this is already occurring. For example, an algorithm in widespread use in the US to determine the likely healthcare needs of a patient, and thus access to onward services, exhibits significant racial bias—in short, African-American patients needed to be significantly “sicker” than Caucasian patients to get the same score, and thus the same access to services, via this algorithm (44).

In the context of cardiac imaging, a specific source of inequality may result from the geographic distribution of these technologies. Much cardiac imaging, particularly using newer modalities such as cardiac magnetic resonance, is largely performed in higher-income countries, and even there, in centers of excellence or high-volume practices (45). This means that training datasets used in the development of AI models for the analysis of these images will suffer from a relative lack of images from patients in low- and middle-income countries. Even disadvantaged patients in high-income countries, who may not have access to the best, most expensive imaging, may be relatively underrepresented in such datasets. Such excluded groups may find that cardiac imaging AI tools developed with these unrepresentative datasets are either less accurate when applied to their cases, or are excluded altogether from the potential benefits of these technologies due to decisions around deployment and marketing by their manufacturers.

The question also remains as to whether the use of AI will create new health inequalities. For example, consumer-facing AI tools presuppose some degree of digital literacy, and their use is likely to pose a personal financial cost to an individual, given the expensive hardware that is frequently required, such as a smartphone or wearable technology. More work is needed to better understand which groups may be excluded from the benefits these technologies could bring, and to develop strategies to avoid such outcomes.

Errors and Liability

Just like the black box problem, the question of “who is responsible when things go wrong with AI” has received a lot of attention in ethics and policy circles (46, 47). The Canadian Association of Radiologists has approached this discussion by focusing on degree of autonomy as a critical determinant. Seeing as most current applications of AI strictly define its role as assistive, including in “intended use” statements that carry regulatory weight, it is reasonable to suggest that ultimate liability for erroneous decisions such as misdiagnosis would

rest with the responsible clinician. However, as the degree of autonomy increases, the degree of liability should shift toward the manufacturer, provided that the clinician can prove that they were using the AI tool exactly as intended. Another potential player is the healthcare system or institution that implemented the AI algorithm, especially if it is determined that, as with any other tool or technology, the organization has a duty to deploy it appropriately (21). However, there are concerns that difficulties in explaining algorithmic decisions (see section **Transparency and Explainability**) may translate into difficulties for patients who suffer harm in proving causation by an algorithm, regardless of the latter’s autonomy (39). Thus, a *res ipsa loquitur* (“the facts speak for themselves”) approach may come to be preferred, where it is the manufacturer that has a *prima facie* case to rebut, and which has successfully been used in cases of harm caused by machinery (48).

Ensuring the Public’s Needs Are Met

Patients and members of the public have a more nuanced understanding of tasks and roles in healthcare than they are frequently given credit for. In research we commissioned, for example, we found that 45% of respondents (in a sample selected to be representative of the UK adult population) agreed that AI should be used to “help diagnose disease,” but only 17% agreed that it should be used to “take on other tasks performed by doctors and nurses,” such as breaking bad news; 63% said it should not be used for this purpose (7). Similarly, attitudes to data sharing for AI research are complex and nuanced. For example, in a workshop study conducted by the Wellcome Trust with 246 patients and HCPs, 17% of participants indicated opposition to giving commercial companies access to their data for the purposes of research. However, when data sharing was tied to the possibility of benefits from this research, 61% of the same study participants indicated they would rather share their data with commercial companies than miss out on potential positive outcomes (49). Many such studies of attitudes to data sharing exist [and the Understanding Patient Data initiative provides an excellent compendium of these studies and their major findings (50)], but two themes emerge across all of these studies as critical factors in determining readiness to share data: firstly, the importance of trust in the *institution* carrying out the research or development, and secondly, the importance of communicating potential benefits clearly.

Regulation

In our 2018 report, we discussed the looming potential of a clash between existing healthcare regulators, and the new regulators, oversight bodies, and advisory committees being set up by governments and multinational organizations to focus on AI more generally, such as the Centre for Data Ethics and Innovation in the UK, and the EU’s High Level Expert Group. As it turns out, no such clash has transpired, as newer AI-focused bodies have thus far been content to leave the realm of health and care to the more established regulators. However, this does not mean that regulatory certainty has followed. The healthcare regulatory space is crowded, and the speed of technological development means that these regulators have

been undertaking an exercise of rapid capacity building, to be able to consider the potential impacts of these technologies. Furthermore, communication between these regulators needs to occur to ensure clear responsibility for all parts of the development process, and to avoid regulatory gaps. In the UK, the think tank Reform has released a series of resources that definitively map each step in developing a data-driven healthcare tool (from idea generation, through to securing data access, through to undertaking clinical research, to ascertaining regulatory compliance and post-market surveillance) to specific regulators, and lays out the requirements at each stage (51). The CEO of NHSX, UK Government unit with responsibility for setting national policy and developing best practice for NHS technology, digital and data, has acknowledged the need for better regulatory alignment (52). The very fact that such discussions are being had indicates the shift in thinking that is occurring in the health technology (healthtech) space, where rather than “software” and “apps,” more enlightened technologists are realizing that what they are creating are medical devices, with the risks and benefits inherent in any medical intervention. Having first been expressed by the Software as a Medical Device (SaMD) initiative kicked off by International Medical Device Regulators Forum (IMDRF), this culture shift has arguably reached its zenith in the EU’s Medical Device Regulation 2017/745, the post-market surveillance requirements of which will be fully applicable by May 2020. The launch of a European database on medical devices (EUDAMED) in May 2022, with a much wider scope than the existing one, will mean that data on post-market surveillance of various devices, including AI tools, will be publicly available to an unprecedented degree.

Another area where regulators may contribute is in the development of standardized benchmarks to allow replicable assessment of the performance of AI tools, both over time and in comparison to one another. This is precisely the aim of the AI for Health (AI4H) Focus Group, a joint initiative of the World Health Organization and the International Telecommunications Union (53).

Consequences of Novel Insights

We have already alluded to the fact that the novel methods of data analysis these AI tools could provide can lead to unexpected insights being obtained from datasets (see section Data). Taken one step further, we can envisage a situation where these tools could potentially present patients and members of the public with information that (a) would not have been previously available, and (b) has the capacity to radically alter how they think about themselves and their health. A close analogy is genomic testing, with the new insights and attendant deep ethical questions it has forced us to consider (54). Just as with genes, if algorithmic predictions come to be equated with “destiny,” then this could lead to a perception of futility and diminishment of hope. Negative consequences could include an individual fearing that they may not have access to certain interventions, and therefore not seeking them. Moreover, not everyone would like to discover that they are at high risk of one condition or another, especially if the treatment or cure options are limited. Decisions on these questions are likely to be nuanced and vary

greatly in different situations and between different patients, but they should always be taken in the context of meaningful conversations between patients and their healthcare providers, and with deep appreciation of a patient’s autonomy.

On a population level, algorithmic predictions of this nature can easily translate into algorithm profiling, create new categories and subgroups within existing populations. People may be assigned to these groups, and inferences and choices made about them, possibly without their knowledge (55). It is unclear where the balance should be struck between capitalizing on the new insights these algorithms could provide, and the threats to autonomy and individuality that categorization of societies and communities could lead to. An interesting suggestion has been to invoke the concept of solidarity as a means to reinforce the community-based nature of healthcare, and underlining the importance of the pursuit of a collective “good” (56).

Trusting Algorithms

As referred to earlier (see section Ensuring the Public’s Needs Are Met), trust in data-driven technologies and in their development may be intimately related to trust in the institutions responsible for this development. Further evidence for this is provided by a survey of 2000 people across Europe carried out by the Open Data Institute, where 94% of respondents said that whether or not they trust the organization asking for their data is important in considering whether or not to share data (57). It follows, therefore, to ask what it is that makes organizations trusted, and there is evidence to suggest that a major factor in determining this trust is the degree of perceived openness. Being open reduces the sense that a system or process has been captured by a particular organization or body that may not have the system’s users’ best interests at heart (58, 59). Moreover, openness allows the organization to demonstrate its competence in data handling, and to share its motivations for doing so; both these factors have also been found to be important determinants of readiness to engage by a systematic review (60). In order to address the requirements for openness and transparency in clinical trials involving AI algorithms, an international project is underway that aims to develop AI extensions to the existing CONSORT and SPIRIT checklists and guidance documents (61). On the other hand, given that a lot of development of AI for healthcare occurs in the private sector (see next section), legitimate concerns remain on the part of developers that regulators mandating excessive openness pose a threat to their intellectual property, and thus reduce the incentives for investment in developing these data-driven tools.

Collaborations Between Public and Private Sector Organizations

The development of AI tools for widespread clinical use is dominated by partnerships between health and research institutions such as hospitals and universities, and private sector organizations. In the UK, there is a perception that such partnerships are needed as the healthcare system, the National Health Service, controls access to data, whereas capital for investment in R&D and the human talent required to create these tools is increasingly being concentrated in technology

companies (62). There is the additional complicating factor of ensuring value not only for the patients whose data is used to develop these tools, but also for the taxpayer who funds the health service that acts as the data custodian, but who may never be in a position to directly benefit from the algorithms that are derived from such partnerships. There have already been some policy responses to such challenges. For example, following its launch in July 2019, one of NHSX's first acts was to confirm and take responsibility for enforcing a ban on exclusive data-sharing agreements between hospitals and commercial companies (63). This move has been seen as addressing concerns that exclusivity deals signed in the past by NHS hospital did not represent good value for money, and as signaling a shift toward more national decision-making on data use for technological applications.

DEVELOPING PRINCIPLES, AND TRANSLATING THEM INTO POLICY

In some countries, the response to questions such as those posed by our 2018 report has been to develop frameworks outlining principles for the ethical use of data and AI in healthcare. At the time of writing, for example, the Royal Australian and New Zealand College of Radiologists has an open consultation on its Draft Standards of Practice for Artificial Intelligence; this will close on 29th November 2019 (64). Perhaps one of the more mature frameworks is the UK Department of Health and Social Care (DHSC)'s "Code of Conduct for data-driven health and care technology," which was developed using a Delphi methodology and was first published in September 2018. It is already in its third iteration following a process of expert review and public consultation (65). This principles-based document has been broadly well-received, and constitutes a world-first that is likely to serve as a global standard.

Nevertheless, it has been clear for some time that principles are a necessary but not sufficient condition to ensure safe and ethical development of healthtech tools. Specifically, it was realized that developers, predominantly coming from a technological background and therefore not imbued in the cultural norms and expectations specific to healthcare, needed support with demonstrating adherence to the principles laid out in documents such as the Code of Conduct. Put another way, if the Code of Conduct laid out *what* developers should aspire to, what they wanted was guidance on *how* to do it. It is on this background that in October 2019, NHSX launched a series of resources specifically aimed at addressing this question (66). This combination of principles and policy has been termed a "principled proportionate governance" model. Future Advocacy contributed to the development of this policy document by focusing specifically on Principle 7 of the Code of Conduct, which is concerned with transparency, openness, and ensuring safe integration of algorithms in existing healthcare systems. In order to address these issues, we signpost a number of existing resources that developers can use to demonstrate adherence to this principle, and classify them according to whether they are general processes that apply across all aspects of principle 7,

or recommendations for specific processes that apply to certain subsections. For example, in order to conduct a meaningful, useful, and relevant stakeholder analysis, we encourage the use of value and consequence matrices in the context of the SUM principles developed by the Alan Turing Institute (67). Likewise, in order to encourage transparency around the means of collecting, storing, using and sharing data, we recommend the use of the Open Data Institute's "Data Ethics Canvas," a freely-available resource from a highly-respected institution (68). What is apparent is that rather than attempting to reinvent the wheel, HCPs and technologists collaborating on the creation of data-driven tools for healthcare need to develop greater familiarity with the work that is already ongoing in the wider technology ethics and policy community, as this cross-disciplinary approach is likely to suggest solutions to problems the field of healthcare is only beginning to grapple with.

SAFE AND ACCEPTABLE: THE FUTURE OF AI IN HEALTH AND CARE

Two specific themes that run through the Code of Conduct, and that have been referred to at multiple points in this review are those of stakeholder engagement, and openness. Both these concepts are important when thinking about how developers can increase the likelihood of their tools being acceptable to patients and HCPs. For example, research with patients and members of the public has indicated that they do not want the development of these tools to come at the expense of the relationships that characterize good care (5, 7). Undertaking a robust and inclusive process of stakeholder analysis will help highlight relationships of importance in healthcare, and will ensure that the participants in these relationships involved in the development process. Furthermore, if development is guided by a deep understanding of the needs of the prospective user from an early stage, the product that comes out of the development process is more likely to be adopted and deployed. Similarly, as has been discussed previously, openness is a determinant of trust, which is itself a determinant of likelihood to engage with the development of these tools. It is therefore our recommendation that the principles of stakeholder engagement and openness run through the development of AI tools for all applications in health and care.

Although the principles and policies discussed above inspired by a drive to increase the safety of AI technologies as applied to health, they are not in themselves sufficient to guarantee safety. A detailed treatment of the various processes and standards related to the safety of these products is beyond the scope of this review, but it is noted that the shift in thinking toward treating these tools as medical devices, as described earlier and as encapsulated in the Medical Device Regulation, should go some way toward protecting users and patients, by placing more stringent requirements in terms of external audit, of developing and maintaining robust quality management systems, and of being more responsive to user feedback and field surveillance (69). Other safety issues that remain relatively unaddressed by current regulation

and HCP training programmes include those of automation complacency, and of the use of dynamic, continuously learning systems (70).

CONCLUSION

In this review we have updated the ten challenges we originally identified in 2018 with current thinking and practice, reflecting the rapid changes in the field of AI as applied to health and care. There is still some way to go in addressing these questions. It is clear to us that, given the iterative nature of technological development, the development of pathways for continuous review of principles and policy frameworks should be prioritized by governments and healthcare authorities. Furthermore, given the complexity of these technologies, a truly multidisciplinary approach is required. It is only by involving all stakeholders with a sincere desire to ensure the successful development and deployment of these tools that their risks will be minimized, and their opportunities maximized.

REFERENCES

1. Harmand S, Lewis JE, Feibel CS, Lepre CJ, Prat S, Lenoble A, et al. 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. *Nature*. (2015) 521:310. doi: 10.1038/nature14464
2. Moore GE. Cramming more components onto integrated circuits. *Electronics*. (1965) 38:114–7.
3. Kurzweil R. *The Singularity Is Near: When Humans Transcend Biology*. New York, NY: Penguin Books. (2006).
4. Brynjolfsson E, Rock D, Syverson C. Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. In: Agrawal A, Gans J, Goldfarb A, editors. *The Economics of Artificial Intelligence: An Agenda*. Chicago, IL: The University of Chicago Press. (2019).
5. IpsosMORI. *Public Views of Machine Learning: Findings from Public Research and Engagement Conducted on Behalf of the Royal Society*, Report. London: Royal Society (2017).
6. McLean I. The origin and strange history of regulation in the UK: three case studies in search of a theory. In: *ESF/SCSS Exploratory Workshop: The Politics of Regulation*. Barcelona. (2002).
7. Fenech ME, Strukelj N, Buston O. Ethical, Social, and Political Challenges of Artificial Intelligence in Health. London: Future Advocacy. (2018).
8. Selinger HA. Six essential relationships in health care. *Fam Pract Manag*. (2013) 20:40.
9. Harvey H. *Why AI Will Not Replace Radiologists [Internet]*. Towards Data Science. (2018). Available online at: <https://towardsdatascience.com/why-ai-will-not-replace-radiologists-c7736f2c7d80>.
10. Topol E. *Preparing the Healthcare Workforce to Deliver the Digital Future*. London: Health Education England. (2019).
11. Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health records on time efficiency of physicians and nurses: a systematic review. *J Am Med Inform Assoc*. (2005) 12:505–16. doi: 10.1197/jamia.M1700
12. Banks J, Farr M, Salisbury C, Bernard E, Northstone K, Edwards H, et al. Use of an electronic consultation system in primary care: a qualitative interview study. *Br J Gen Pract*. (2018) 68:e1–e8. doi: 10.3399/bjgp17X693509
13. Campbell JL, Fletcher E, Britten N, Green C, Holt T, Lattimer V, et al. The clinical effectiveness and cost-effectiveness of telephone triage for managing same-day consultation requests in general practice: a cluster randomised controlled trial comparing general practitioner-led and nurse-led management systems with usual care (the ESTEEM trial). *Health Technol Assess*. (2015) 19:1–212. doi: 10.3310/hta19130
14. Pitruzzella B, Leahy P. *Future of Surgery*. London: Royal College of Surgeons. (2018).

AUTHOR CONTRIBUTIONS

MF conducted the original research by Future Advocacy cited in this review, and wrote this paper. OB also conducted the original research cited, reviewed this paper, and provided senior approval for its publication.

FUNDING

The original work conducted by Future Advocacy that is cited in this paper was funded by the Wellcome Trust and by NHSX.

ACKNOWLEDGMENTS

We are extremely grateful to our colleague Nika Strukelj for her contribution to the original research cited in this paper. We are also grateful to the numerous people, including patients and clinicians, who were interviewed or who took part in roundtables and focus groups as part of the research that informed this review.

15. Hungin P. *The Changing Face of Medicine and the Role of Doctors in the Future*. London, UK: British Medical Association. (2017).
16. Future Doctor. *What do the NHS, Patients and the Public Require From 21st-Century Doctors?* Health Education England (2019). Available online at: <https://www.hee.nhs.uk/our-work/future-doctor>.
17. Dinov ID. Volume and value of big healthcare data. *J Med Stat Inform*. (2016) 4:3. doi: 10.7243/2053-7662-4-3
18. Cocanour CS. Informed consent-It's more than a signature on a piece of paper. *Am J Surg*. (2017) 214:993–7. doi: 10.1016/j.amjsurg.2017.09.015
19. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. (2018) 2:158–64. doi: 10.1038/s41551-018-0195-0
20. ICO. *Lawful Basis for Processing*. (2019). Available online at: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/>.
21. Jaremko JL, Azar M, Bromwich R, Lum A, Alicia Cheong LH, Gibert M, et al. Canadian Association of Radiologists White Paper on Ethical and Legal Issues Related to Artificial Intelligence in Radiology. *Can Assoc Radiol J*. (2019) 70:107–18. doi: 10.1016/j.carj.2019.03.001
22. Hopkins H, Kinsella S, van Mil A. *Foundations of Fairness: Views on Uses of NHS Patients' Data and NHS Operational Data*. London: Understanding Patient Data. (2020).
23. Cath-Speth C, Kaltheuner F. Risking everything: where the EU's white paper on AI falls short. *New Statesman*. (2020).
24. Geis JR, Brady AP, Wu CC, Spencer J, Ranshaert E, Jaremko JL, et al. Ethics of artificial intelligence in radiology: summary of the joint european and north american multisociety statement. *Can Assoc Radiol J*. (2019) 293:436–440. doi: 10.1148/radiol.2019191586
25. Petersen SE, Abdulkareem M, Leiner T. Artificial intelligence will transform cardiac imaging-opportunities and challenges. *Front Cardiovasc Med*. (2019) 6:133. doi: 10.3389/fcvm.2019.00133
26. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med*. (2003) 78:775–80. doi: 10.1097/00001888-200308000-00003
27. Groopman J. *How Doctors Think*. Boston, MA: Houghton Mifflin Co. (2007).
28. Brady AP. Error and discrepancy in radiology: inevitable or avoidable? *Insights Imaging*. (2017) 8:171–82. doi: 10.1007/s13244-016-0534-1
29. Jiang H, Nachum O. Identifying and correcting label bias in machine learning. (2019). *arXiv:1901.04966*.
30. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med*. (2018) 169:866–72. doi: 10.7326/M18-1990

31. Tang A, Tam R, Cadrin-Chenevert A, Guest W, Chong J, Barfett J, et al. Canadian association of radiologists white paper on artificial intelligence in radiology. *Can Assoc Radiol J.* (2018) 69:120–35. doi: 10.1016/j.carj.2018.02.002
32. Greengard S. Weighing the Impact of GDPR. *Communications of the ACM.* (2018) 61:16–8. doi: 10.1145/3276744
33. Price WN, 2nd, Cohen IG. Privacy in the age of medical big data. *Nat Med.* (2019) 25:37–43. doi: 10.1038/s41591-018-0272-7
34. Datenethikkommission. *Opinion of the Data Ethics Commission.* (2019). Available online at: https://datenethikkommission.de/wp-content/uploads/191023_DEK_Kurzfassung_en_bf.pdf.
35. Delcker J. *A German Blueprint for Europe's AI Rules.* Politico. (2019). Available online at: <https://www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-hello-world-a-german-blueprint-for-ai-rules-california-dreaming-2/>.
36. EuropeanCommission. *White Paper: On Artificial Intelligence - A European Approach to Excellence and Trust.* Brussels: European Commission (2020).
37. Castelvacci D. Can we open the black box of AI? *Nature.* (2016) 538:20–3. doi: 10.1038/538020a
38. Watson DS, Krutzinna J, Bruce IN, Griffiths CE, McInnes IB, Barnes MR, et al. Clinical applications of machine learning algorithms: beyond the black box. *BMJ.* (2019) 364:l886. doi: 10.1136/bmj.l886
39. Schönberger D. Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. *Int J Law Inform Tech.* (2019) 27:171–203. doi: 10.1093/ijlit/ez004
40. Brehm JW. Postdecision changes in the desirability of alternatives. *J Abnorm Psychol.* (1956) 52:384–9. doi: 10.1037/h0041006
41. Harmon-Jones E, Harmon-Jones C. Testing the action-based model of cognitive dissonance: the effect of action orientation on postdecisional attitudes. *Pers Soc Psychol Bull.* (2002) 28:711–23. doi: 10.1177/0146167202289001
42. Edwards L, Veale M. Slave to the algorithm? Why a right to explanation is probably not the remedy you are looking for Duke Law. *Tech Rev.* (2017) 16:18–84. doi: 10.31228/osf.io/97upg
43. McNamara S, Holmes J, Stevely AK, Tsuchiya A. How averse are the UK general public to inequalities in health between socioeconomic groups? A systematic review. *Eur J Health Econ.* (2019) 21:275–85. doi: 10.1007/s10198-019-01126-2
44. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* (2019) 366:447–53. doi: 10.1126/science.aax2342
45. Kramer CM. Potential for rapid and cost-effective cardiac magnetic resonance in the developing (and developed) world. *J Am Heart Assoc.* (2018) 7:e010435. doi: 10.1161/JAHA.118.010435
46. Neri E, de Souza N, Brady A, Bayarri AA, Becker CD, Coppola F, et al. What the radiologist should know about artificial intelligence - an ESR white paper. *Insights Imag.* (2019) 10:44. doi: 10.1186/s13244-019-0738-2
47. Allain JS. From Jeopardy! to jaundice: the medical liability implications of Dr. Watson and other artificial intelligence systems. *La L Rev.* (2012) 73:1049. Available online at: <https://digitalcommons.law.lsu.edu/lalrev/vol73/iss4/7>
48. Laurie G, Harmon S, Dove E. *Mason and McCall Smith's Law and Medical Ethics.* Eleventh edition. Oxford: Oxford University Press (2019). doi: 10.1093/he/9780198826217.001.0001
49. IpsosMORI. *The One-Way Mirror: Public Attitudes to Commercial Access to Health Data.* Wellcome Trust. (2016). Available online at: <https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf>.
50. UnderstandingPatientData. *How Do People Feel About the Use of Data?* (2019). Available online at: <https://understandingpatientdata.org.uk/how-do-people-feel-about-use-data>.
51. Reform. *Data-Driven Healthcare: Regulation and Regulators.* (2019). Available online at: <https://reform.uk/research/data-driven-healthcare-regulation-regulators>.
52. Gould M. *Regulating AI in Health and Care.* (2020). Available online at: digital.nhs.uk/blog/transformation-blog/2020/regulating-ai-in-health-and-care.
53. Wiegand T, Krishnamurthy R, Kuglitsch M, Lee N, Pujari S, Salathe M, et al. WHO and ITU establish benchmarking process for artificial intelligence in health. *Lancet.* (2019) 394:9–11. doi: 10.1016/S0140-6736(19)30762-7
54. Samuel GN, Farsides B. Public trust and 'ethics review' as a commodity: the case of Genomics England Limited and the UK's 100,000 genomes project. *Med Health Care Philos.* (2018) 21:159–68. doi: 10.1007/s11019-017-9810-1
55. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi F. The ethics of algorithms: mapping the debate. *Big Data Soc.* (2016) 3:2. doi: 10.1177/2053951716679679
56. Prainsack B, Buyx A. *Solidarity: Reflections on an Emerging Concept in Bioethics.* Nuffield Council on Bioethics. (2011). Available online at: http://nuffieldbioethics.org/wp-content/uploads/2014/07/Solidarity_report_FINAL.pdf
57. OpenDataInstitute. *Who Do We Trust With Personal Data?* (2018). Available online at: <https://theodi.org/article/who-do-we-trust-with-personal-data-odi-commissioned-survey-reveals-most-and-least-trusted-sectors-across-europe/>.
58. Rousseau DM, Sitkin SB, Burt RS, Camerer C. Not so different after all: a cross-discipline view of trust. *Acad Manag Rev.* (1998) 23:393–404. doi: 10.5465/amr.1998.926617
59. Mayer RC, Davis JH, Schoorman DF. An Integrative model of organizational trust. *Acad Manag Rev.* (1995) 20:709–34. doi: 10.5465/amr.1995.9508080335
60. Stockdale J, Cassell J, Ford E. 'Giving something back': A systematic review and ethical enquiry of public opinions on the use of patient data for research in the United Kingdom and the Republic of Ireland. *Wellcome Open Res.* (2018) 3:6. doi: 10.12688/wellcomeopenres.13531.1
61. Liu X, Faes L, Calvert MJ, Denniston AK, Group CS-AE. Extension of the CONSORT and SPIRIT statements. *Lancet.* (2019) 394:1225. doi: 10.1016/S0140-6736(19)31819-7
62. Boland H. *Britain Faces an AI Brain Drain As Tech Giants Raid Top Universities.* The Telegraph. (2018). Available online at: <https://www.telegraph.co.uk/technology/2018/09/02/britain-faces-artificial-intelligence-brain-drain/>.
63. Heather B. NHSX to oversee industry access to patient data. *Health Service J.* (2019).
64. RANZCR. *Consultation: RANZCR Standards of Practice for AI.* (2019). Available online at: <https://www.ranzcr.com/our-work/advocacy/position-statements-and-submissions/consultation-ranzcr-standards-standards-of-practice-for-ai>.
65. DHSC. *Code of Conduct for Data-Driven Health and Care Technology.* (2018). Available online at: <https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology>.
66. Joshi I, Morley J. *Artificial Intelligence: How to Get It Right. Putting Policy Into Practice for Safe Data-Driven Innovation in Health and Care.* London: NHSX. (2019). Available online at: https://www.nhs.uk/assets/NHSX_AI_report.pdf.
67. Leslie D. *Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector.* The Alan Turing Institute. (2019). doi: 10.5281/zenodo.3240529
68. OpenDataInstitute. *Data Ethics Canvas User Guide.* (2019). Available online at: https://docs.google.com/document/d/1MkvoAP86CwimbBD0dxySVCO0zeVQut_bu1A6kHV73M/edit.
69. Byrne RA. Medical device regulation in Europe - what is changing and how can I become more involved? *EuroIntervention.* (2019) 15:647–9. doi: 10.4244/EIJV15I8A118
70. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf.* (2019) 28:231–7. doi: 10.1136/bmjqs-2018-008370

Conflict of Interest: As of 1st September 2019, MF has been employed by Ada Health GmbH, which develops AI tools for use in health applications.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Fenech and Buston. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Mini Review: Deep Learning for Atrial Segmentation From Late Gadolinium-Enhanced MRIs

Kevin Jamart¹, Zhaohan Xiong¹, Gonzalo D. Maso Talou¹, Martin K. Stiles² and Jichao Zhao^{1*}

¹ Auckland Bioengineering Institute, The University of Auckland, Auckland, New Zealand, ² Waikato Clinical School, Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand

OPEN ACCESS

Edited by:

Steffen Erhard Petersen,
Queen Mary University of London,
United Kingdom

Reviewed by:

Wenjia Bai,
Imperial College London,
United Kingdom
Joao Bicho Augusto,
Barts Heart Centre, United Kingdom
Redha Boubertakh,
Singapore Bioimaging Consortium
(A*STAR), Singapore

*Correspondence:

Jichao Zhao
j.zhao@auckland.ac.nz

Specialty section:

This article was submitted to
Cardiovascular Imaging,
a section of the journal
Frontiers in Cardiovascular Medicine

Received: 08 January 2020

Accepted: 21 April 2020

Published: 27 May 2020

Citation:

Jamart K, Xiong Z, Maso Talou GD,
Stiles MK and Zhao J (2020) Mini
Review: Deep Learning for Atrial
Segmentation From Late
Gadolinium-Enhanced MRIs.
Front. Cardiovasc. Med. 7:86.
doi: 10.3389/fcvm.2020.00086

Segmentation and 3D reconstruction of the human atria is of crucial importance for precise diagnosis and treatment of atrial fibrillation, the most common cardiac arrhythmia. However, the current manual segmentation of the atria from medical images is a time-consuming, labor-intensive, and error-prone process. The recent emergence of artificial intelligence, particularly deep learning, provides an alternative solution to the traditional methods that fail to accurately segment atrial structures from clinical images. This has been illustrated during the recent 2018 Atrial Segmentation Challenge for which most of the challengers developed deep learning approaches for atrial segmentation, reaching high accuracy (>90% Dice score). However, as significant discrepancies exist between the approaches developed, many important questions remain unanswered, such as which deep learning architectures and methods to ensure reliability while achieving the best performance. In this paper, we conduct an in-depth review of the current state-of-the-art of deep learning approaches for atrial segmentation from late gadolinium-enhanced MRIs, and provide critical insights for overcoming the main hindrances faced in this task.

Keywords: atrial fibrillation, left atrium, machine learning, image segmentation, convolutional neural network, LGE-MRI

INTRODUCTION

The ability to perform body imaging has been described as one of the most important revolutions in medicine of the past 1,000 years for its contribution to medical prevention, diagnosis, and prognosis (1). Since then, medical imaging has never ceased to improve, allowing cardiologists, and researchers to assess heart size using chest x-rays (2), to evaluate heart mechanical work with echocardiography imaging (3–5) and to accurately determine the heart's dimensions using cardiac magnetic resonance imaging (MRI) (6). Due to its good image quality, excellent soft-tissue contrast, and absence of ionizing radiation, MRI has become the gold standard modality to precisely identify patients' cardiac structures and etiology, guiding diagnosis and therapy decisions (7).

Improvements of MRI techniques, particularly with the aid of contrast agents such as gadolinium, led to the development of late gadolinium-enhanced MRI (LGE-MRI), allowing for the detection of scar tissue located within the myocardium. This technique has been extensively employed for clinical studies at Utah University (8–10) to analyze and understand the role of fibrosis and underlying structures that sustain atrial fibrillation (AF), the most common cardiac arrhythmia predicted to become a new epidemic in the coming decades (11, 12). They notably

demonstrated the correlation between an increased amount of fibrosis present in the left atrial (LA) wall and a poor outcome of AF ablation (10). Over time, LGE-MRIs have become a widely accepted technique of choice allowing the detection and quantification of scar tissues located in the atrial wall.

The currently widely used clinical practice, including those conducted at Utah University, to analyze atrial structures and determine and quantify fibrosis distribution is by performing manual segmentation of the LA chamber from LGE-MRIs. However, the LA cavity represents a small volume ($73 \pm 14.9 \text{ cm}^3$), constrained by a thin atrial wall (2–3 mm) and comprised of complex anatomy (13–15). Moreover, the anatomical structures surrounding the atria display similar intensities that can mislead some segmentation algorithms (16) (**Figure 1**). As a consequence, manual segmentation of the atrium is a time-consuming, labor-intensive, and error-prone process (8, 17, 18).

Before the advent of deep learning, researchers tried to develop and improve automated approaches to alleviate the burden of manual segmentation (19, 20). Earlier algorithms proposed would require important manual tunings such as thresholding methods or region growing approaches (21, 22). Other methods were later developed to provide a higher degree of automation using classifiers or clustering approaches such as *k*-nearest-neighbor (23) or *k*-means clustering (24), respectively. More recent methods, using statistical classifiers like support vector machine (25), active shape model (26), or multi-atlases (27) approaches, gained increasing interest for medical image analysis and cardiac segmentation. Though many of these approaches showed promising results, none presented enough consistency to be implemented widely in clinical practice.

In recent years, the development of more powerful computational hardware and the growth of clinical databases enabled deep learning, a subset of artificial intelligence (AI) (28–32) capable of automatic feature extraction and learning, to achieve tremendous advances notably in image classification and segmentation (33, 34). When applied to clinical images, deep learning even surpassed human-level accuracies for the detection of cancer on cervical images (35). Certain architectures employed for deep learning have also been proven to be very effective when applied to cardiac imaging. For example, Avendi et al. (36, 37) used a three-stage approach combining convolutional neural network (CNN), stacked encoder, and deformable models to segment the left ventricle (and later the right ventricle) on a small MRI dataset of 45 patients. On the other hand, Bai et al. (38) used a large MRI dataset provided by the UK Biobank database to develop their CNN for ventricular chamber assessment (volume, mass, ejection fraction) and segmentation, obtaining accuracy scores competing with human-level precision.

This increasing interest around deep learning can also be seen in the number of participants using deep learning approaches for the various challenges designed to promote the development of more robust methods for cardiac image segmentation (39–41). Atrial segmentation is becoming a matter of greater importance and can highly benefit from the development of deep learning. As an example, during the 2018 Atrial Segmentation Challenge, 15 of the 17 published approaches used deep learning to segment the

LA cavity from LGE-MRI images, yielding high accuracy results and outperforming conventional segmentation approaches (42). The number is in sharp contrast with the previous atrial segmentation challenge held in 2013, during which only one approach used a learning algorithm (16). Thus, this growing interest for deep learning in research challenges illustrates the shift occurring in atrial segmentation and more broadly in clinical imaging development, moving more and more toward deep learning-based approaches that will revolutionize clinical practice in the coming years.

In this paper, we aim to provide an analysis of the current deep learning technique used for atrial segmentation on LGE-MRIs. Firstly, we will describe some of the fundamental concepts employed in deep learning for medical image segmentation. Subsequently, we will detail the various deep learning approaches addressing the main obstacles faced performing automated atrial segmentation. Finally, we will conclude our review with an outline of future developments for atrial segmentation using deep learning and more broadly the future of AI in clinical practice.

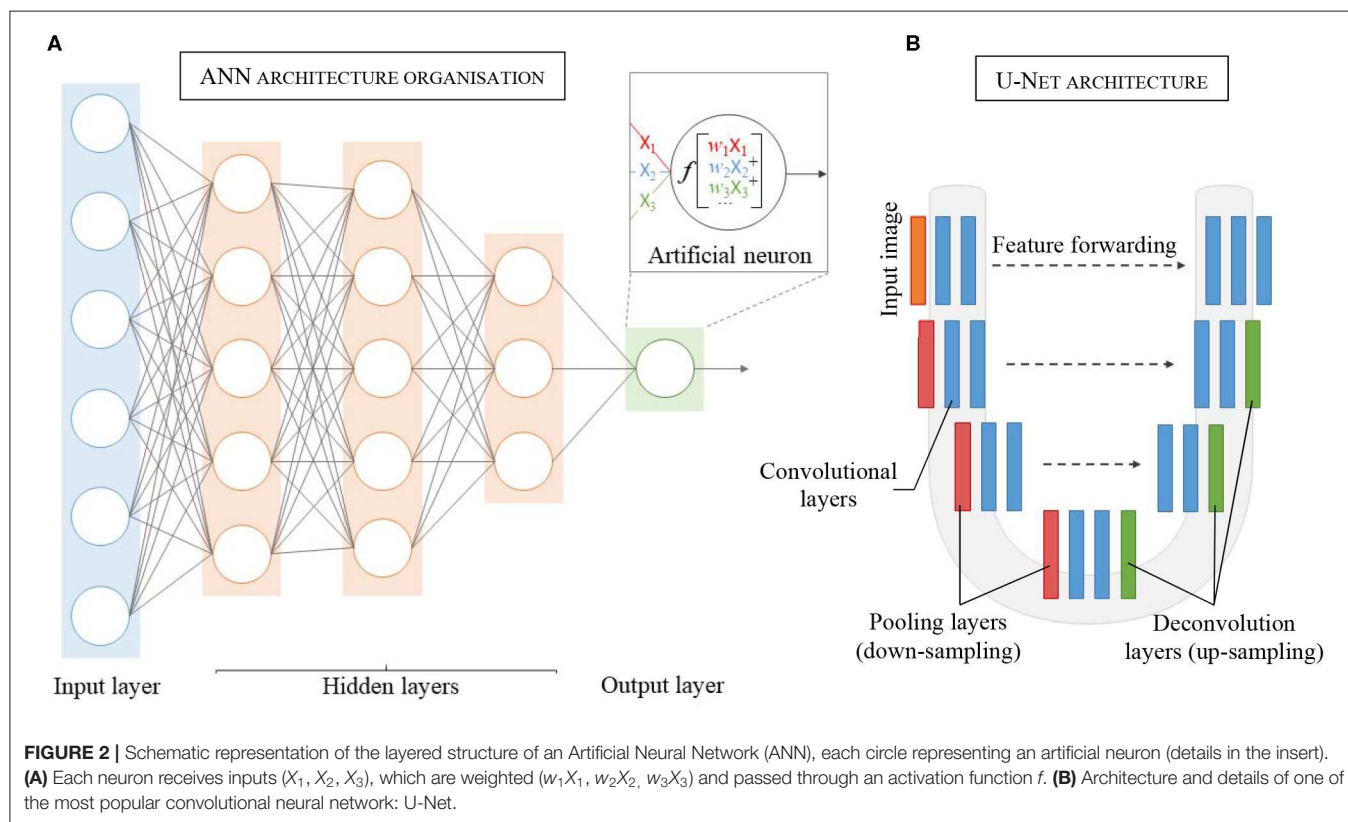
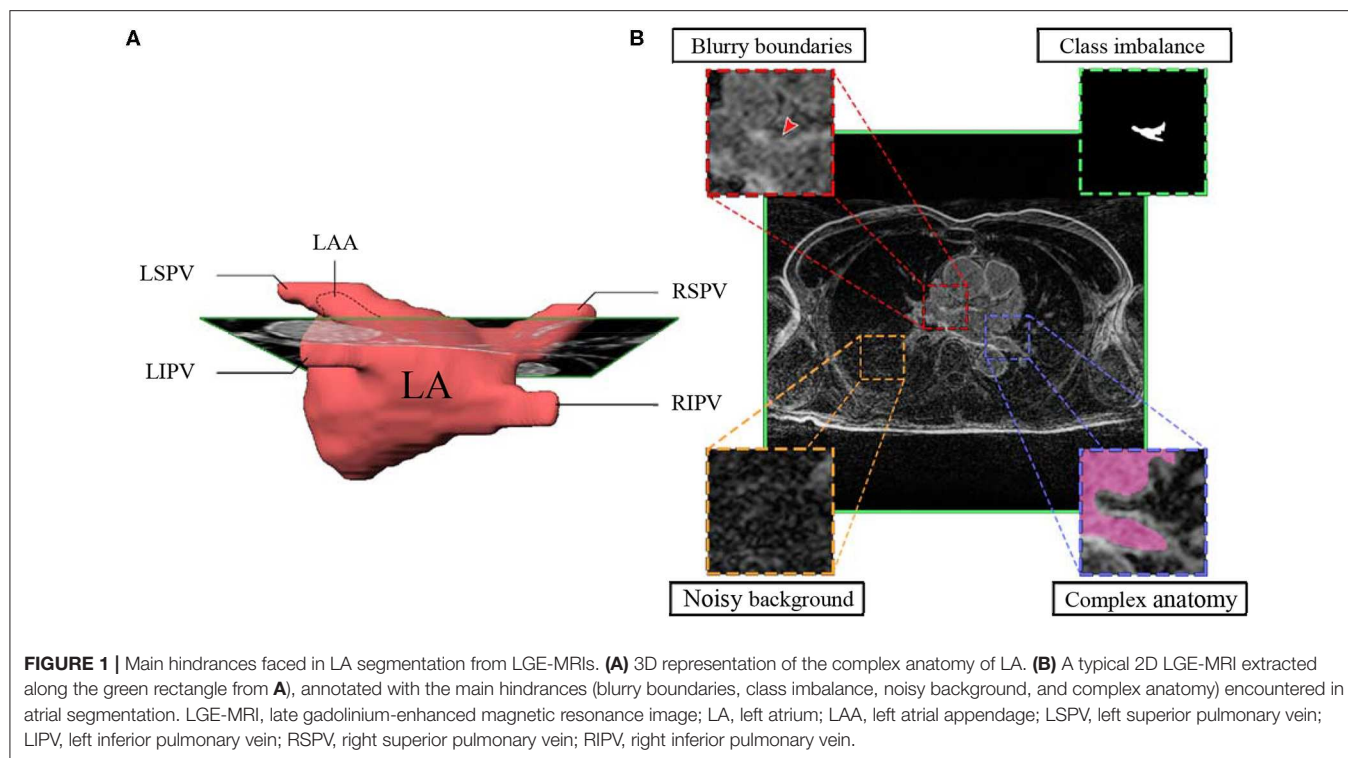
CORE CONCEPTS OF DEEP LEARNING

Since Alan Turing published his article “Computing Machinery and Intelligence” asking “Can machines think?” researchers have thrived to comprehend, develop, and achieve AI (43, 44) although today, after over 60 years, general AI is still not within reach. Nevertheless, in recent years, the growth of computer processing power and technologies has allowed researchers to develop algorithms capable of learning proficiently through deep learning using artificial neural networks (ANNs). As ANNs represent the most popular structure to perform deep learning, this section will describe the core concepts of ANNs and their various practical use in medical imaging.

Artificial Neural Networks

Inspired by the biological neural networks found in the human brain (45), an ANN represents a collection of connected and tunable computational units, called artificial neurons, organized in a layered structure comprising a network (**Figure 2A**). Each neuron is a processing unit that can take multiple inputs. Each input is multiplied by an adjustable parameter called weight. All weighted inputs are summed together and passed through a non-linear function to yield a single output (30). Neural networks can address complex, highly non-linear problems due to the layered and connected structure of ANNs. In particular, the introduction of more advanced feature learning tools such as convolutional layers, the improvement of large datasets and better activation functions, e.g., ReLU, greatly helped the development of deep learning for segmentation tasks.

The key attribute of an ANN lies in its ability to learn the unique traits of a dataset by adjusting its weights accordingly during a training process. Typically, the weights are randomly initialized at the start of training. The training process can then be described in three consecutive phases: (1) forward propagation, (2) error calculation, and (3) back-propagation. In the forward propagation stage, the input data (e.g., LGE-MRI image) is fed to the network and flows through the different layers that extract



the characteristic traits of the data, to ultimately yield a prediction (e.g., desired segmented image). The prediction is then compared to a reference data (e.g., manually segmented image by experts),

called labeled data, and error is calculated using a dedicated function (called loss function). Finally, the weights are modified to minimize the estimated error, improving prediction accuracy.

These three phases are repeated several times until the error converges to a significant minimized value.

Different Tasks, Different Networks

Medical imaging encompasses a wide field of applications, and different tasks can represent different aspects of a diagnosis. Examples include the detection of an abnormal ECG signal, its classification as AF (46), or even atrial segmentation for planning for AF ablation (47). Therefore, each task requires a specific ANN architecture to properly model the desired operator, as the inputs and output can be drastically different depending on the nature of the task to be performed.

The number, types, and connections of layers in an ANN defines the network architecture. The CNN model is one of the most widely employed architectures in image analysis. CNN is a specific ANN architecture in which its hidden layers comprise one or more convolutional layers. The convolutional layers act as feature extractors from the input image, applying different convolution kernels to the initial image to generate feature maps containing meaningful information. Moreover, in convolutional layers, each artificial neuron receives their inputs from multiple neighboring neurons from the previous layer, sharing their weights and keeping the most spatially relevant information. This feature also allows a reduction in the number of parameters to adjust and therefore lowers the computational processing cost. Generally inserted in between sets of successive convolutional layers are pooling layers that are used to reduce the dimensionality of each generated feature map while retaining the relevant information. This down-sampling of the feature maps, typically by a factor of two, allows reduction of the computational cost while enlarging the field of view for the later convolutional layers.

For CNNs dedicated to image classification or detection, the architectures usually incorporate a fully connected layer as an end layer to summarize all information contained in the feature maps into a unique final prediction (output). Furthermore, CNNs can also be adapted for segmentation tasks by discarding the final fully connected layer and incorporating up-convolution layers in the network (35). These networks are called fully convolutional networks (FCNs). Up-convolution layers allow up-sampling of the feature maps to produce, *in fine*, output with the same size as the original input size (48). Thus, FCNs using up-convolution layers can perform pixel-wise prediction and therefore image segmentation.

First proposed by Long et al. (33) for semantic segmentation, the FCN architecture has been adapted and further extended for medical imaging notably with U-Net, a U-shape architecture (Figure 2B) developed for segmentation of histological images (48). By using skip-connections between down-sampled feature maps and up-sampled feature maps, the U-Net architecture allows features forwarding between the encoding part and the decoding part of the network, preventing singularities and achieving higher accuracy (49–51). After winning the ISBI cell tracking challenge in 2015, U-Net became the principal FCN architecture for medical imaging segmentation. Other studies further developed the U-shape architecture to use 3D images as

input to render the spatial resolution of anatomical structures more accurately (52, 53).

ATRIAL SEGMENTATION USING DEEP LEARNING

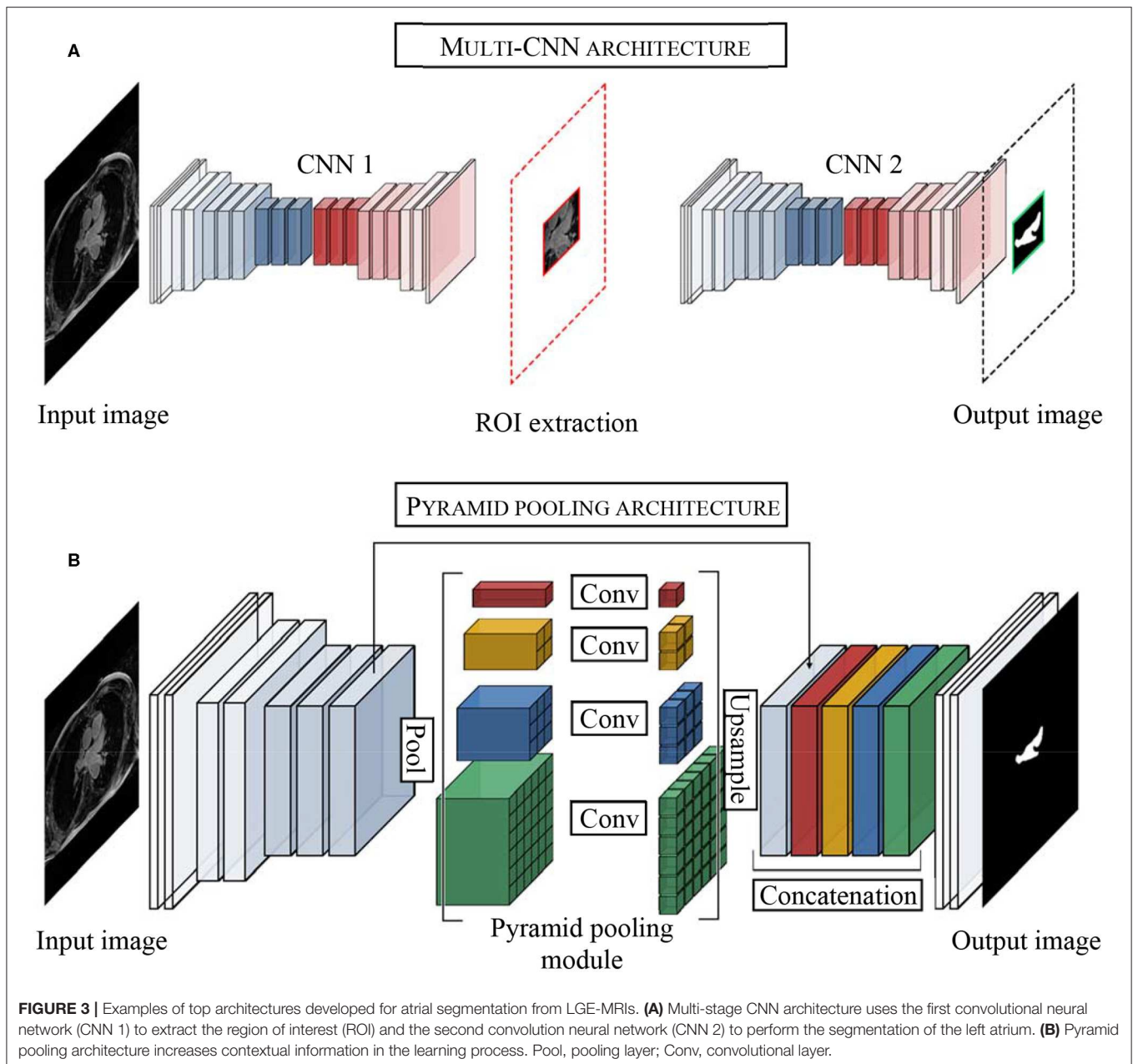
In this section, we provide a summary of the main difficulties encountered in atrial segmentation and the state-of-the-art deep learning approaches developed from LGE-MRIs to address them. To this regard, many of the methods reviewed were proposed for the MICCAI 2018 Atrial Segmentation Challenge which represented a cornerstone for the development of deep learning approaches for atrial segmentation from LGE-MRIs. Firstly, we will analyze the main methods employed to address class imbalance issues, a recurrent problem in segmentation of small structures such as the LA. Secondly, we will review the approaches developed to exploit image context providing more information for semantic segmentation of the LA using multi-scale strategies. Next, we will analyze the impact of loss function selection regarding either volumetric segmentation or surface segmentation. Finally, we will discuss the influence of the input dimensionality (2D/3D) for atrial segmentation when dataset size represents a significant shortcoming.

Multi-Stage CNN and Class Imbalance

One of the difficulties of atrial segmentation is that the atrial cavity represents only a small fraction of the image volume ($\sim 0.7\%$) and therefore creates a severe class imbalance between the over-represented background and the under-represented atrial structures, impairing the learning process. To address this issue, Vesal et al. (54) proposed to crop the input images from the center of the image, using fixed coordinates, to substantially remove the predominant background surrounding the LA. As a result, the learning process was entirely focused on a smaller region of interest (ROI), allowing better representation of the LA features. Based on a similar principle, other researchers (55–57), pushed this idea a step further by using a multi-CNN approach for atrial segmentation (Figure 3A). In their approaches, two consecutive networks were employed instead. The first CNN was specially trained to localize the LA on each input, allowing to subsequently crop out the unwanted background around the LA, as a prior step to segmentation. Then, the second network was dedicated to the segmentation task itself, focusing entirely on a small patch of each image.

Despite following a similar idea, it is important to distinguish these two methods. As the LA can show different positions on LGE-MRIs, using fixed coordinates from the center of the image to crop may result in unwanted cropping of relevant LA pixels. On the other hand, by dynamically centering the ROI on the LA for each input, multi-CNN approaches ensured the conservation of the atrial structures, cropping exclusively superfluous background pixels, and consequently optimizing background isotropy for the learning process.

To quantify the impact of each cropping approach, our recent study has investigated the importance of cropping the input patch



to the CNN either from the center of the image (image-centered) or from the center of the LA (center of mass/centroid of the atrium) using different patch sizes (ranging from 240×240 to 576×576) (58). When using center cropping of the image, we did not observe any significant influence of the patch size on the Dice score (92.03 vs. 91.95% Dice score for 240×240 and 512×512 image size, respectively). On the other hand, cropping the images from the centroid of the LA using dynamic cropping, we noticed a significant increase in the accuracy when using small patches (240×240) compared to large patches (576×576) (Dice score 92.86 vs. 92.26%, $p < 0.01$). The utilization of LA centroid-centered patches allows the CNN to process a more condensed region of the large LGE-MRI scan as the exact location of the LA

is known, reducing the class imbalance of each patch processed by the network.

Multi-Scale Approaches and Context Learning

Another problem that decreases segmentation performance and limits the extraction of relevant cues during the training phase is the inconsistency in the sizes of the LA anatomical structures such as the pulmonary veins or the left atrial appendage seen in LGE-MRIs from different patients.

He et al. (59) initially developed a pyramid pooling module, a multi-scale pooling, intended to prevent object misclassification by using image context information. By incorporating multi-scale

pooling, the CNN could associate contextual features, delivering more accurate classification. Based on this idea, Zhao et al. (60) proposed PSPNet, a neural network with pyramid pooling which incorporates object and image context to the learning process. These two approaches were developed using large miscellaneous datasets such as ImageNet (61), PASCAL VOC 2012 (62), or ADE20K dataset (63), and the pyramid pooling exploiting the context variability of the dataset, allowed to alleviate object misclassification or segmentation errors.

Inspired by He et al. (59) and the PSPNet developed by Zhao et al. (60), Bian et al. (64) proposed a multi-scale 2D CNN using spatial pyramid pooling to extract different scale features of the training dataset (**Figure 3B**). Thus, by means of different pooling kernel sizes and their combination, they proposed a CNN able to learn different cue size and improve network robustness against high shape variability usually encountered in clinical datasets.

However, the dataset employed for this approach (154 3D LGE-MRIs of the chest cavity) does not provide as much contextual variability as the large image database aforementioned, but rather displays the same object (the LA) in the same anatomical context (the thoracic cavity), providing only a few contextual variations to train on. Thus, arguably using pyramid pooling module for LA segmentation in the chest cavity might only show limited benefits from context learning.

Pyramid pooling also grants the ability to generate a fixed-length vector on a fully connected layer for classification tasks. This was illustrated by Chen et al. (65) using the pyramid pool module to extract more information from the dataset and classify the images between pre-/post-surgery, as they used a deeper U-Net to segment the LA simultaneously.

Based on the similar idea of incorporating multi-scale cues during the learning process, Vesal et al. (54) employed dilated convolution layers (also called atrous convolution layers) at the deepest level of their network. These convolution layers use dilatation rates to enlarge their receptive fields, allowing the network to learn different scale features (66). However, at each convolution the receptive field of each neuron is increased, therefore if not used wisely, receptive fields can become larger than the input image, resulting in a waste of memory while not improving the learning process.

These approaches ensure the incorporation of shallow features (spatial cues) and deep features (semantic cues) during the learning process. Therefore, combining effective class imbalance management with contextual cues could potentially improve even more the current methods. However, cropping to the smallest ROI possible using a first CNN of a two-stage approach, like Xia et al. drastically reduces the image context shown to the network. Therefore, the pyramid pooling module might not be able to provide contextual cues from the LA surrounding structures to improve the learning process. Moreover, during the cropping process, the input image size is significantly reduced. Thus, the use of dilated convolution for segmentation in the second network of this strategy becomes almost obsolete as the receptive fields would quickly grow larger than the input image during the learning process. Thus, fusing these strategies, although interesting, needs to be considered wisely.

Loss Function

The current main evaluation metrics employed in segmentation task using deep learning is the Dice score, for which a higher accuracy reflects almost exclusively a volume of pixel accurately annotated rather than well-defined anatomy. Hence, most of the deep learning approaches for segmentation employ pixel-wise segmentation relying either on cross-entropy loss function or dice loss function. However, these loss functions weigh more volume over contours, which can impair the learning of accurate boundaries in favor of a correct volume.

To improve boundary accuracy, several teams have developed contour-oriented loss functions. For example, Jia et al. (67) proposed a contour loss function (based on the pixel Euclidean distance) that decreases when the contour gets nearer to the reference contours of the label images during training, providing spatial distance information to the learning process. In their approach, they associated the dice function loss to obtain pixel-wise information, and their contour loss function for spatial information, achieving good shape consistency. In another strategy, Yang et al. (57) also defined a composite loss function, combining the overlap loss function (to reduce intersection between foreground and background) and a novel loss function called “focal positive loss” to guide the learning of voxel specific threshold and emphasize the foreground, improving, *in fine*, classification sensitivity. By recognizing ambiguous boundary location and enforcing positive prediction, this novel loss function improved the learning process and consequently the final atrial segmentation. However, these approaches did not obtain a better score than other approaches using more conventional loss function (e.g., dice loss, cross-entropy loss).

Therefore, it would be interesting to investigate the impact of a combined loss function allowing the network to learn from the volume (cross-entropy loss function or dice loss function) and from the contours of the LA. As segmentation tasks not only rely on minimizing volume error but also relies on boundaries accuracy (particularly for small structures). it is crucial to consider these two major aspects to ensure the reliability of the approach employed.

Spatial Context (2D vs. 3D)

Even if clinical datasets are becoming bigger and better with the creation of centralized databases, for example, the UK Biobank (with more than 90000 3D MRI scans) (68), most of the current clinical databases available remain of humble size, making it difficult for a CNN to provide robust generalized solutions for segmentation. As an example, the current largest LGE-MRI dataset with only 154 3D LGE-MRIs (which represent nearly 9,000 2D images for training) appears relatively small when compared to the hundreds of thousands of images used for the major classification challenges for which the proposed approaches reach outstanding accuracy (59, 69, 70).

Thus, in this race of performance, it is important to consider how to make the best of the dataset employed. To this regard, the choice of the image dimensions employed (2D and 3D) approaches must be considered wisely. As 2D approaches need considerably fewer trainable parameters to yield good results, they are less gluttonous regarding memory consumption, and

therefore require less time during the training process. Moreover, 2D approaches allow the processing of bigger batches of images compared to 3D approaches, as they require less memory to be processed. Therefore, 2D methods, using bigger batch size, help reduce gradient fluctuation and lead to faster convergence during the learning process. Additionally, 2D approaches can exploit more efficiently small datasets, reducing the risk of overfitting as the neural networks are fed with more images for the learning.

On the other hand, 3D approaches provide better spatial representation, fully exploiting data dimensionality as well as inter-slice continuity during training. This allows the network to learn major spatial features to render a more accurate 3D anatomy and yield, *in fine*, higher accuracy. Moreover, with the ever improvement of GPU technology, the current memory limitations will become of less importance in the near future; therefore, 3D approaches will become easier to use. Furthermore, as datasets are growing better and bigger, 3D approaches will be able to rely on more data and become more and more prominent in clinical imaging deep learning.

Nevertheless, relying on 2D images, Puybureau et al. (71) tried to improve the spatial representation of their dataset using a method called “pseudo-3D.” Their method employed the generation of color images from the 2D grayscale images, each slice being color expanded into the R, G, B space using slice $n-1$, slice n and slice $n+1$, to generate a three-channel image. This approach allows an improved spatial representation and alleviates low contrast intensity between atrial tissues and background and enrich the dataset. However, even if this approach does not provide the expected spatial representation, it can be a method of choice if resources are limited.

Following the multi-view approach developed by Mortazi et al. (72), Chen et al. investigated the possibility to combine 2D images and 3D representation (73). In their study, Chen et al. extracted the 2D images for each anatomical view (axial, coronal, and sagittal) from 100 3D LGE-MRIs. Then, they combined a first encoder-decoder network using long short term memory convolutional layers to preserve inter-slice correlation using the axial view, and a second network to learn complementary information from the sagittal and coronal views. Finally, the outputs for each view of the network were fused to yield LA and PV segmentation simultaneously. Using their approach, they obtained 90.83% Dice score accuracy for PV and atrial segmentation. Employing the same method, Yang et al. studied the influence of dilated convolution to counter image resolution variability encountered using a multi-view approach (74). Using 100 3D LGE-MRIs, they achieved 89.7% Dice score accuracy underlining the necessity to investigate systematic parameters tuning to obtain optimal performances on a task-specific basis.

In the present context, it is important to consider the trade-off using either a 2D approach requiring less memory and profiting more from the dataset (8,800 images rather than 154 3D LGE-MRIs) a 3D approach allowing more accurate spatial representation at the cost of longer and more difficult training. However, at the current stage, it is difficult to assess which method yields systematically better results. For example, during the 2018 Atrial Segmentation Challenge, the performances of 2D and 3D approaches remained very close (Table 1). Another

possibility is to use a multi-view approach combining 2D images from different views to improve the spatial representation. These methods require training each view separately before combining the different output for the final prediction. While interesting, these methods still need improvement to reach the current state-of-the-art for atrial segmentation. Therefore, further improvements need to be sought regarding the size of the dataset, the number of approaches compared and the metrics employed to be able to draw a better conclusion.

Evaluation Metrics

Another crucial point is to use metrics that provide a reliable evaluation of the final output using deep learning. One of the main scores employed is called Dice score and gauges the pixel-wise similarity between the predicted segmentation and the reference data. Dice score provides a good representation of the specificity and the sensitivity of the model. However, Dice score metric has some limitations as it only evaluates a percentage of pixel accurately annotated neglecting contours and shapes of organs that can be a critical part of diagnosis in clinical practice. Other metrics providing distance measurements, such as mean surface distance and Hausdorff maximum distance, are usually employed to provide an alternative evaluation. Mean surface distance estimates the average error (in mm) between the outer surfaces of the reference data and the predicted segmentation. Given the size and structure of LA, mean surface distance is a meaningful tool to reliably assess the anatomical boundaries of the predicted segmentation compared to the reference data. Hausdorff maximum distance (in mm) represents the maximum error between the surface of the predicted segmentation and the surface of the reference data. Therefore, Hausdorff distance indicates solely the distance at the worst part of the segmentation, providing only partial information of the correctness of the predicted segmentation. By combining mean surface distance and Hausdorff distance, it is possible to evaluate the fidelity of the boundaries of the segmented structures reliably. Finally, a more clinical aspect of the predictions can be examined to express the reliability of the approach by calculating volume error or anteroposterior atrial diameter error when comparing the segmented prediction with the reference image.

Atrial Wall and Scar Segmentation

While the deep learning methods for atrial cavity segmentation on LGE-MRIs are effective, the more clinically relevant tasks, such as LA wall and fibrosis (scar) segmentation, remain challenging. For LA wall segmentation, several approaches have been developed using traditional strategies such as multi-atlas segmentation or graph-cuts method (83, 84). However, currently no deep learning approaches have been proposed for direct LA wall segmentation from LGE-MRI. Yang et al. (85) proposed a hybrid approach combining multi-atlases and an unsupervised sparse auto-encoders for LA scar segmentation. A multi-atlas algorithm was used to segment the LA blood pool from the LGE-MRIs. Then, this initial LA cavity segmentation was dilated uniformly by 3 mm to include the LA wall. Next, they used a sparse auto-encoder to delineate and segment the fibrosis from the atrial wall. They achieved $90 \pm 0.12\%$ Dice score

TABLE 1 | Summary of deep learning approaches developed for the 2018 atrial segmentation challenge.

First Author	Summary	DC	Architecture	Pros/Cons
Xia et al. (56)	2 stage network (LA localization, LA segmentation), Dice loss	93.2	2x 3D U-Net	Good class imbalance management, highest performance/computationally expensive
Bian et al. (64)	LA segmentation using, ResNet101, atrous convolutional layers and pyramid pooling, online hard negative example mining (objective function)	92.6	2D Pyramid Network	Multi-scale representation/competitive training can reinforce overfitting
Vesal et al. (54)	LA segmentation using manual cropping, dilated convolution at the deepest level of U-Net, combination of Dice loss and cross-entropy loss function	92.6	3D U-Net	Class imbalance management, new loss function/Risk of loss of information using center cropping
Li et al. (55)	2 stage network: 3D U-Net for detection, Hierarchical Aggregation network (HAANet) for LA segmentation, Dice loss	92.3	3D U-Net + HAANet	Class imbalance management/Slow and small benefits from Hierarchical mechanism (0.4%)
Puybareau et al. (71)	Assembly of three 2D gray-scale images to create RGB 2D color image, transfer learning (VGG), multinomial loss function for LA segmentation	92.3	VGG-Net	Fast to train, pseudo-spatial representation/pseudo spatial representation not multi-view or 3D
Yang et al. (57)	2 stage approach: LA detection (Faster-RCNN), LA segmentation (U-Net). Deep supervision, transfer learning. Composite loss function: Overlap loss and Focal Positive loss	92.3	Faster-RCNN/3D U-Net	Good ROI detection, composite loss function /Recursive training with risk of overfitting
Chen et al. (73)	LA segmentation and classification (pre/post-ablation) of images, using cross-entropy and sigmoid loss function, respectively	92.1	2D U-Net	Fast to train (2D), interesting data augmentation
Jia et al. (67)	2 stage network (LA localization, LA segmentation), contour loss	90.7	3D U-Net	Contour loss/computationally expensive
Liu et al. (75)	Manual center cropping, evaluation of 2 different networks U-Net and FCN for LA segmentation, Dice loss	90.3	2D U-Net and FCN	Quick (2D)/Native Unet/FCN
Borra et al. (76)	Otsu's algorithm for cropping, LA and pulmonary veins joined segmentation, Dice loss	89.8	3D U-Net	Otsu's for cropping/computationally expensive
de Vente et al. (77)	U-net for LA segmentation, Dice loss	89.7	2D U-Net	Fast (2D)/Native Unet
Preetha et al. (78)	Deep supervision (79) and U-Net for LA segmentation	88.8	2D U-Net	Deep supervision, Fast (2D)/Native Unet
Qiao et al. (80)	Multi-atlas selection and registration for LA segmentation	86.2	Multi-atlas	Groupwise registration/Slow prediction(multi-atlas)
Núñez-García et al. (81)	Multi-atlas whole heart labeling and shape-based atlas selection	85.9	Multi-atlas	Registration using global-atlases, shape based clustering/Difficulties do manage high variability in small dataset
Savioli et al. (82)	LA segmentation using V-Net and combination of mean squared error and Dice loss	85.1	3D V-Net	Composite loss function/computationally expensive

DC, Dice Score; LA, Left Atrium.

for blood pool segmentation and $78 \pm 0.08\%$ Dice score for fibrosis segmentation. In their subsequent study (86), by fine-tuning the sparse auto-encoder parameters, the accuracy was improved to $82 \pm 0.05\%$ Dice score for fibrosis segmentation. While showing promising results, with these methods being only developed and tested on 20 3D LGE MRIs, they remain untested on larger datasets to assess their reliability against a broader range of anatomical variabilities regarding LA structures and fibrosis. Chen et al. (73) developed a CNN with an attention mechanism (87) to highlight salient features (in this case, the enhanced pixels of the scar tissues on LGE MRIs) and to force the model to focus on the scars locations. With this approach, Chen et al. obtained 77.64% Dice score for atrial scar segmentation using 100 3D LGE MRIs. This lower score (compared to that obtained from LA cavity segmentation) is potentially due to the scarcity of the LA scar pixels, which are small patches of inhomogeneous enhanced pixels within the atrial wall, impairing the extraction of meaningful features for fibrosis identification during the learning process of the CNN.

While these methods require atrial wall segmentation to be performed before fibrosis detection, Li et al. proposed a hybrid approach using a graph-cuts framework combined with a multi-scale CNN approach for direct scar identification (88). In their approach, the LA and PV were initially delineated using a multi-atlas segmentation method. Then fibrosis was segmented and quantified using a graph-cut network in which two neural networks were dedicated to predicting edge weights. The first network was dedicated to predicting the probabilities of a node belonging to scar or normal tissue, while the second network was devoted to evaluate the connection between two nodes, yielding, in fine, the fibrosis segmentation. By embedding the CNN networks in the graph-cut framework, Li et al. obtained a mean Dice score of 70.2% for scar tissue segmentation, showing the possibility of effectively assessing LA fibrosis without the need for prior wall segmentation. Thus, even if the two networks employed did not directly perform the fibrosis segmentation task, the CNNs contributed to the optimisation process refining the graph-cut approach used in this study. However, these

methods tended to find fibrotic tissue out of the atrial wall boundaries regions, resulting in a drastic decrease in the final scores. Hence, the current models remain insufficient to provide anatomically accurate assessments allowing reliable fibrosis quantification due to the low Dice scores obtained. Thus, these approaches still require improvements to reach reliability and clinical applicability.

DISCUSSION AND CONCLUSION

In this paper, we provided an in-depth analysis of the main automatic approaches using deep learning for atrial cavity segmentation from LGE-MRIs. Most of the proposed deep learning approaches for atrial segmentation used FCNs, most notably the very popular U-Net architecture. While U-Net is widely used for medical image segmentation in many disciplines (38, 89, 90), the discrepancy in the accuracy obtained between different studies still presents inherent issues involved in the generalized implementation of such architectures. By presenting a normalized survey of U-Net for the task of atrial segmentation, we showed the importance of proper class imbalance management, appropriate features extraction process, and meaningful loss function selection to yield precise and accurate atrial segmentation.

The current leading approach for LA segmentation from LGE-MRIs dataset involved a two-stage 3D CNN method which reached a remarkable Dice accuracy of 93.2%, currently the best-benchmarked performance using 100 3D LGE MRIs (42). In this approach, the first network reduces class imbalance effectively while optimizing background isotropy using dynamic cropping, providing the second network with a targeted region for more localized segmentation. Additionally, they employed extensive data augmentation to enhance the generalization capability of their approach. Finally, they employed a 3D approach reinforcing the features' spatial representation, allowing them to obtain the current highest score for LA segmentation using machine learning.

Small training datasets represent one of the main limitations of clinical datasets as annotation and data gathering remains difficult. For example, the current largest LGE-MRIs dataset only contains 154 cases and therefore cannot effectively represent human anatomical variability. In fact, in order to improve performance, most of the developed approaches rely heavily on data augmentation such as affine transformations, cropping and scaling to virtually enlarge the dataset, also taking the risk of introducing more artifacts in the dataset. Moreover, the annotation process of anatomical structures is a complex and tedious process, which can be seen in the inter/intra-observer variability reported in several studies (38, 91). For example, atrial structures such as the mitral valve are difficult to segment due to the lack of clear anatomical border between LA and left ventricle. Moreover, the PVs are a very thin structure and represent a challenge for experts to distinguish from other structures on poorly contrasted images, and current protocols for defining the degree of extension of the PVs from the LA wall still remains subjective. Thus, this labeling uncertainty leads to some label variability in the dataset used, impairing the training process and potentially misleading the deep learning algorithm for the

prediction process. However, despite all these difficulties the study shows the success of deep learning approaches reaching a high Dice score accuracy (>90% Dice score), showing the importance of careful parameter selection and architecture design for achieving the best performance (38).

In this study, we showed the potential of applying deep learning to perform automatic segmentation of the LA directly from clinical imaging data. The current accuracy of the various approaches presented is promising for future clinical implementation by providing highly accurate anatomical maps of the LA. Additionally, multiple teams already proposed auspicious solutions for fibrosis assessment using deep learning, providing particularly valuable information for AF ablation strategies that could highly benefit initial patient stratification, diagnosis, prognosis, and potential guidance for an optimized ablation strategy. Moreover, the ability to generate high fidelity segmentations such as the LA opens the way for further applications of deep learning to segment other anatomical structures. For instance, high accuracy left atrial appendage segmentation would provide crucial information for atrial thrombosis risk assessment (92). Thus, practitioners would be able to provide adapted treatment strategies on time, potentially reducing the number of stroke accidents caused by migrating atrial thrombus. Additionally, LA segmentation approaches could also be applied to the RA, providing a better understanding of the role of fibrotic extents spread through the RA myocardium notably in sinoatrial diseases (93).

Finally, it is important to underline the limitation of the current metrics employed. As most of the segmentation tasks rely on pixel-wise classification, Dice score proposes an efficient way to determine the correctness of the overlapping prediction. However, Dice score can be defined as a volumetric metric as it weighs more generously toward an accurate volume over precise anatomical delimitations. In clinical practice, Dice score and volume accuracy are important for assessing LA dilatation, but becomes irrelevant when assessing boundaries of fine structures such as LA. Therefore, other metrics such as mean surface distance representing the distance between the labeled surface and the predicted surface should be considered to produce better anatomical accuracy evaluation. The Hausdorff distance, representing the maximum distance between two surfaces, can also be used to evaluate the maximum error between prediction and label, potentially guiding algorithms to minimize their maximum error. Moreover, other limitations such as variations in image quality and resolution or the introduction of image artifacts intrinsic to scanner manufacturer have to be taken to account for future clinical deployment. At the current stage, no study has investigated the influence of LGE-MRI image quality on the Dice score but empirically, the best image quality tends to yield higher accuracy scores. However, in clinical practice image quality can vary tremendously as cardiac motion, body fat, and chest breathing motion, amongst others, can generate artifacts to various degrees on the final images. Therefore, to provide good generalization capacity, deep learning models have to be able to extract meaningful features regardless of the quality of the image. Similarly to the image quality issue, to obtain good generalization capacity, a network should be trained with many images from many different scanners. Thus, large multi-center datasets need

to be built to ensure satisfying scanner variability and image quality variability representation for the learning process. Finally, it is crucial to promote deep models with efficient inherent generalization capabilities, as different image resolutions can represent a major difficulty for deep learning models using large scale datasets. However, promising results were demonstrated using pyramid pooling architecture ensuring extraction of multi-scale features. Thus, at the current stage efforts remain to be made to develop a deep learning model satisfying these criteria for further clinical deployment.

With the development of computational hardware and the general effort to enrich medical image databases, the effectiveness of deep learning will only improve with time. Arguably, the current trend would lead to improve all fields of clinical practices as AI technologies become more widely developed and implemented. Furthermore, the current flourishing of the deep learning approaches in all areas of medical practice has already breached out research. Despite initial professional reluctance, AI technologies will become of major importance in the near future.

REFERENCES

- Angell M, Kassirer J, Relman AS. Looking back on the millennium in medicine [editorial]. *N Engl J Med*. (2000) 342:42–9. doi: 10.1056/NEJM200001063420108
- Joo HS, Wong J, Naik VN, Savodelli GL. The value of screening preoperative chest X-rays: a systematic review. *Can J Anaesthesia*. (2005) 52:568–74. doi: 10.1007/BF03015764
- Sahn DJ, DeMaria A, Kisslo J, Weyman A. Recommendations regarding quantitation in M-mode echocardiography: results of a survey of echocardiographic measurements. *Circulation*. (1978) 58:1072–83. doi: 10.1161/01.CIR.58.6.1072
- Nagueh SF, Appleton CP, Gillebert TC, Marino PN, Oh JK, Smiseth OA, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography. *J Am Soc Echocardiogr*. (2009) 10:165–93. doi: 10.1016/j.echo.2008.11.023
- Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of cardiovascular imaging. *J Am Soc Echocardiogr*. (2015) 16:233–71. doi: 10.1016/j.echo.2014.10.003
- La AG, Claessen G, de Bruaene Van A, Pattyn N, Van JC, Gewillig M, et al. Cardiac MRI: a new gold standard for ventricular volume quantification during high-intensity exercise. *Circul Cardiovasc Imaging*. (2013) 6:329–38. doi: 10.1161/CIRCIMAGING.112.980037
- Karamitsos TD, Francis JM, Myerson S, Selvanayagam JB, Neubauer S. The role of cardiovascular magnetic resonance imaging in heart failure. *J Am College Cardiol*. (2009) 54:1407–24. doi: 10.1016/j.jacc.2009.04.094
- Oakes RS, Badger TJ, Kholmovski EG, Akoum N, Burgon NS, Fish EN, et al. Detection and quantification of left atrial structural remodeling using delayed enhancement MRI in patients with atrial fibrillation. *Circulation*. (2009) 119:1758. doi: 10.1161/CIRCULATIONAHA.108.811877
- Akoum N, Fernandez G, Wilson B, McGann C, Kholmovski E, Marrouche N. Association of atrial fibrosis quantified using LGE-MRI with atrial appendage thrombus and spontaneous contrast on transesophageal echocardiography in patients with atrial fibrillation. *J Cardiovascul Electrophysiol*. (2013) 24:1104–9. doi: 10.1111/jce.12199
- McGann C, Akoum N, Patel A, Kholmovski E, Revelo P, Damal K, et al. Atrial fibrillation ablation outcome is predicted by left atrial

AUTHOR CONTRIBUTIONS

KJ and JZ conceived and designed the work. KJ searched and read the literature and drafted the manuscript. ZX, GM, MS, and JZ provided guidelines, critical revision, and insightful comments to improve the manuscript. All authors read and approved the manuscript.

FUNDING

This work was supported by the Health Research Council of New Zealand.

ACKNOWLEDGMENTS

We would like to thank our colleagues: Dr. Nawshin Dastagir, Joseph Ashby, and Christopher Walker for their precious comments and insights that greatly helped to improve the manuscript. We would also like to thank Vincent Guichot who provided great assistance for the creation of the figures.

- remodeling on MRI. *Circul Arrhythmia Electrophysiol*. (2014) 7:23–30. doi: 10.1161/CIRCEP.113.000689
- Chugh SS, Blackshear JL, Shen W-K, Hammill SC, Gersh BJ. Epidemiology and natural history of atrial fibrillation: clinical implications. *J Am College Cardiol*. (2001) 37:371–8. doi: 10.1016/S0735-1097(00)01107-4
- Anter E, Jessup M, Callans DJ. Atrial fibrillation and heart failure: treatment considerations for a dual epidemic. *Circulation*. (2009) 119:2516–25. doi: 10.1161/CIRCULATIONAHA.108.821306
- Maceira AM, Cosin-Sales J, Roughton M, Prasad SK, Pennell DJ. Reference left atrial dimensions and volumes by steady state free precession cardiovascular magnetic resonance. *J Cardiovascul Magnet Resonance*. (2010) 12:65. doi: 10.1186/1532-429X-12-65
- Zhao J, Hansen BJ, Wang Y, Csepe TA, Sul LV, Tang A, et al. Three-dimensional integrated functional, structural, and computational mapping to define the structural ‘fingerprints’ of heart-specific atrial fibrillation drivers in human heart *ex vivo*. *J Am Heart Assoc*. (2017) 6:e005922. doi: 10.1161/JAHA.117.005922
- Wang Y, Xiong Z, Nalar A, Hansen BJ, Kharche S, Seemann G, et al. A robust computational framework for estimating 3D Bi-Atrial chamber wall thickness. *Comput Biol Med*. (2019) 114:103444. doi: 10.1016/j.combiomed.2019.103444
- Tobon-Gomez C, Geers AJ, Peters J, Weese J, Pinto K, Karim R, et al. Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets. *Proc IEEE Med Imaging*. (2015) 34:1460–73. doi: 10.1109/TMI.2015.2398818
- Petitjean C, Dacher J-NJM. A review of segmentation methods in short axis cardiac MR images. *Med Image Analysis*. (2011) 15:169–84. doi: 10.1016/j.media.2010.12.004
- Caudron J, Fares J, Lefebvre V, Vivier P-H, Petitjean C, Dacher JN. Cardiac MRI assessment of right ventricular function in acquired heart disease: factors of variability. *Acad Radiol*. (2012) 19:991–1002. doi: 10.1016/j.acra.2012.03.022
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005
- Litjens G, Ciompi F, Wolterink JM, de Vos BD, Leiner T, Teuwen J, et al. State-of-the-art deep learning in cardiovascular image analysis. *JACC Cardiovascul Imaging*. (2019) 12:1549–65. doi: 10.1016/j.jcmg.2019.06.009
- Metz C, Schaap M, Weustink A, Mollet N, van Walsum T, Niessen WJM, et al. Coronary centerline extraction from CT coronary angiography images

- using a minimum cost path approach. *Med Phys.* (2009) 36:5568–79. doi: 10.1118/1.3254077
22. Feng C, Zhang S, Zhao D, Li C. Simultaneous extraction of endocardial and epicardial contours of the left ventricle by distance regularized level sets. *Med Phys.* (2016) 43:2741–55. doi: 10.1118/1.4947126
 23. Bezdek JC, Hall L, Clarke LP. Review of MR image segmentation techniques using pattern recognition. *Med Phys.* (1993) 20:1033–48. doi: 10.1118/1.597000
 24. Kaus MR, Von Berg J, Weese J, Niessen W, Pekar V. Automated segmentation of the left ventricle in cardiac MRI. *Med Image Anal.* (2004) 8:245–54. doi: 10.1016/j.media.2004.06.015
 25. Lao Z, Shen D, Liu D, Jawad AF, Melhem ER, Launer LJ, et al. Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine. *Acad Radiol.* (2008) 15:300–13. doi: 10.1016/j.acra.2007.10.012
 26. Van Assen HC, Danilouchkine MG, Behloul F, Lamb HJ, van der Geest RJ, Reiber JH, et al. Cardiac LV segmentation using a 3D active shape model driven by fuzzy inference. *Int Conf Med Image Comput Comput Assisted Interv.* (2003) 2878:533–40. doi: 10.1007/978-3-540-39899-8_66
 27. Isgm I, Staring M, Rutten A, Prokop M, Viergever MA, Van Ginneken B. Multi-atlas-based segmentation with local decision fusion-application to cardiac and aortic segmentation in CT scans. *Proc IEEE Med Imaging.* (2009) 28:1000–10. doi: 10.1109/TMI.2008.2011480
 28. Deng L, Yu D. Deep learning: methods and applications. *Found Trends Signal Proces.* (2014) 7:197–387. doi: 10.1561/20000000039
 29. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436–44. doi: 10.1038/nature14539
 30. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge, MA: MIT Press (2016).
 31. Garcia-Garcia A, Orts-Escobedo S, Oprea S, Villena-Martinez V, Garcia-Rodriguez J. A review on deep learning techniques applied to semantic segmentation. *arXiv.* (2017) 17, 41–65. doi: 10.1016/j.asoc.2018.05.018
 32. Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis. *IEEE Access.* (2017) 6:9375–89. doi: 10.1109/ACCESS.2017.2788044
 33. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA (2016). p. 3431–40. doi: 10.1109/CVPR.2015.7298965
 34. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV (2016). p. 770–8. doi: 10.1109/CVPR.2016.90
 35. Hu L, Bell D, Antani S, Xue Z, Yu K, Horning MP, et al. An observational study of deep learning and automated evaluation of cervical images for cancer screening. *J National Cancer Institue.* (2019) 74:343–4. doi: 10.1097/OGX.0000000000000687
 36. Avendi M, Kheradvar A, Jafarkhani H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Med Image Anal.* (2016) 30:108–19. doi: 10.1016/j.media.2016.01.005
 37. Avendi MR, Kheradvar A, Jafarkhani H. Automatic segmentation of the right ventricle from cardiac MRI using a learning-based approach. *Magnetic Resonance Med.* (2017) 78:2439–48. doi: 10.1002/mrm.26631
 38. 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 Cardiovascul Magnetic Resonance.* (2018) 20:65. doi: 10.1186/s12968-018-0471-x
 39. Suinesiaputra A, Cowan BR, Finn JP, Fonseca CG, Kadish AH, Lee DC, et al. Left ventricular segmentation challenge from cardiac MRI: a collation study. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. (2018). p. 88–97. doi: 10.1007/978-3-642-28326-0_9
 40. Ho SY, Cabrera JA, Sanchez-Quintana D. Left atrial anatomy revisited. *Circul Arrhythmia Electrophysiol.* (2012) 5:220–8. doi: 10.1161/CIRCEP.111.962720
 41. Zuluaga MA, Bhatia K, Kainz B, Moghari MH, Pace DF. Reconstruction, segmentation, and analysis of medical images: first international workshops. In: *RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016*. Athens, Greece (2017).
 42. Pop M, Sermesant M, Zhao J, Li S, McLeod K, Young AA, et al. Statistical atlases and computational models of the heart: atrial segmentation and LV quantification challenges: 9th international workshop. In: *STACOM 2018, Held in Conjunction With MICCAI 2018*, Granada, Spain, (2019).
 43. Turing AM. I.— Computing machinery and intelligence. *Mind LIX.* (1950) LIX:433–60. doi: 10.1093/mind/LIX.236.433
 44. Saygin AP, Cicekli I, Akman V. Turing test: 50 years later. *Minds Mach.* (2000) 10:463–518. doi: 10.1023/A:1011288000451
 45. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Mathemat Biol.* (1943) 5:115–33. doi: 10.1007/BF02478259
 46. Xiong Z, Nash MP, Cheng E, Fedorov VV, Stiles MK, Zhao J. ECG signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network. *Physiol Measurement.* (2018) 39:094006. doi: 10.1088/1361-6579/aad9ed
 47. Xiong Z, Fedorov VV, Fu X, Cheng E, Macleod R, Zhao J. Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging using a dual fully convolutional neural network. *IEEE on Medical Imaging.* (2018) 38:515–24. doi: 10.1109/TMI.2018.2866845
 48. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. (2015). p. 234–41. doi: 10.1007/978-3-319-24574-4_28
 49. Drozdal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The importance of skip connections in biomedical image segmentation. *Deep Learn Data Label Med Appl.* (2016) 10008:179–87. doi: 10.1007/978-3-319-46976-8_19
 50. Mao X, Shen C, Yang YB. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *Adv Neural Inform Proc Syst.* (2016) arXiv:1603.09056:2802–10.
 51. Orhan AE, Pitkow X. Skip Connections Eliminate Singularities. In: *Sixth International Conference on Learning Representations*. Toulon (2017).
 52. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. *Int Conf Med Image Comput Comput Assisted Interv.* (2016) 9901:424–32. doi: 10.1007/978-3-319-46723-8_49
 53. Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*: IEEE (2016). p. 565–71. doi: 10.1109/3DV.2016.79
 54. Vesal S, Ravikumar N, Maier A. Dilated convolutions in neural networks for left atrial segmentation in 3d gadolinium enhanced-MRI. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. (2018). p. 319–28. doi: 10.1007/978-3-030-12029-0_35
 55. Li C, Tong Q, Liao X, Si W, Sun Y, Wang Q, et al. Attention based hierarchical aggregation network for 3D left atrial segmentation. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer (2018). p. 255–64. doi: 10.1007/978-3-030-12029-0_28
 56. Xia Q, Yao Y, Hu Z, Hao A. Automatic 3D Atrial Segmentation from GE-MRIs Using Volumetric Fully Convolutional Networks. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. (2018). p. 211–20. doi: 10.1007/978-3-030-12029-0_23
 57. Yang X, Wang N, Wang Y, Wang X, Nezafat R, Ni D, et al. Combating uncertainty with novel losses for automatic left atrium segmentation. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer (2018). p. 246–54. doi: 10.1007/978-3-030-12029-0_27
 58. Jamart K, Xiong Z, Talou GM, Stiles MK, Zhao J. Two-stage 2D CNN for automatic atrial segmentation from LGE-MRIs. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer (2019). p. 81–9. doi: 10.1007/978-3-030-39074-7_9
 59. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In: *Proceedings of the IEEE on Pattern Analysis and Machine Intelligence*. (2014) 37:1904–16. doi: 10.1007/978-3-319-10578-9_23
 60. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2017). p. 2881–90. doi: 10.1109/CVPR.2017.660
 61. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vision.* (2015) 115:211–52. doi: 10.1007/s11263-015-0816-y
 62. Everingham M, Van Gool L, Williams C, Winn J, Zisserman A. *The Pascal Visual Object Classes Challenge 2012 (voc2012) Results*. (2012).

- Available online at: URL <http://www.pascal-network.org/challenges/VOC/voc2011/workshop/index.html>.
63. Zhou B, Zhao H, Puig X, Fidler S, Barriuso A, Torralba A. Scene parsing through ade20k dataset. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2017). p. 633–41. doi: 10.1109/CVPR.2017.544
 64. Bian C, Yang X, Ma J, Zheng S, Liu YA, Nezafat R, et al. Pyramid network with online hard example mining for accurate left atrium segmentation. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. (2018). p. 237–45. doi: 10.1007/978-3-030-12029-0_26
 65. Chen C, Bai W, Rueckert D. Multi-task learning for left atrial segmentation on GE-MRI. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. (2018). p. 292–301. doi: 10.1007/978-3-030-12029-0_32
 66. Yu F, Koltun V, Funkhouser T. Dilated residual networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2017). p. 472–80. doi: 10.1109/CVPR.2017.75
 67. Jia S, Despinasse A, Wang Z, Delingette H, Pennec X, Jais P, et al. Automatically segmenting the left atrium from cardiac images using successive 3D U-nets and a contour loss. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. (2018). p. 221–9. doi: 10.1007/978-3-030-12029-0_24
 68. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. (2015) 12:e1001779. doi: 10.1371/journal.pmed.1001779
 69. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft coco: common objects in context. In: *European Conference on Computer Vision*. (2015). p. 740–55.
 70. Wu B, Chen W, Fan Y, Zhang Y, Hou J, Huang J, et al. Tencent ML-images: a large-scale multi-label image database for visual representation learning. In: *IEEE Access*. (2019). p. 172683–93. doi: 10.1109/ACCESS.2019.2956775
 71. Puybareau É, Zhao Z, Khouli Y, Carlinet E, Xu Y, Lacotte J, et al. Left atrial segmentation in a few seconds using fully convolutional network and transfer learning. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. (2018). p. 339–47. doi: 10.1007/978-3-030-12029-0_37
 72. Mortazi A, Karim R, Rhode K, Burt J, Bagci U. CardiacNET: segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. (2017). p. 377–85. doi: 10.1007/978-3-319-66185-8_43
 73. Chen J, Yang G, Gao Z, Ni H, Angelini E, Mohiaddin R, et al. Multiview two-task recursive attention model for left atrium and atrial scars segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer (2018). p. 455–63. doi: 10.1007/978-3-030-00934-2_51
 74. Yang G, Chen J, Gao Z, Zhang H, Ni H, Angelini E, et al. Multiview sequential learning and dilated residual learning for a fully automatic delineation of the left atrium and pulmonary veins from late gadolinium-enhanced cardiac MRI images. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. (2018). p. 1123–7. doi: 10.1109/EMBC.2018.8512550
 75. Liu Y, Dai Y, Yan C, Wang K. Deep learning based method for left atrial segmentation in GE-MRI. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer (2018). p. 311–8. doi: 10.1007/978-3-030-12029-0_34
 76. Borra D, Masci A, Esposito L, Andalò A, Fabbri C, Corsi C. A semantic-wise convolutional neural network approach for 3-D left atrium segmentation from late gadolinium enhanced magnetic resonance imaging. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer (2018). p. 329–38. doi: 10.1007/978-3-030-12029-0_36
 77. de Vente C, Veta M, Razeghi O, Niederer S, Pluim J, Rhode K, et al. Convolutional neural networks for segmentation of the left atrium from gadolinium-enhancement MRI images. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer (2018). p. 348–56. doi: 10.1007/978-3-030-12029-0_38
 78. Preetha CJ, Haridasan S, Abdi V, Engelhardt S. Segmentation of the left atrium from 3D gadolinium-enhanced MR images with convolutional neural networks. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer (2018). p. 265–72. doi: 10.1007/978-3-030-12029-0_29
 79. Yu F, Wang D, Shelhamer E, Darrell T. Deep layer aggregation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2018). p. 2403–12. doi: 10.1109/CVPR.2018.00255
 80. Qiao M, Wang Y, van der Geest RJ, Tao Q. Fully automated left atrium cavity segmentation from 3D GE-MRI by multi-atlas selection and registration. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer (2018). p. 230–6. doi: 10.1007/978-3-030-12029-0_25
 81. Nuñez-García M, Zhuang X, Sanroma G, Li L, Xu L, Butakoff C, et al. Left atrial segmentation combining multi-atlas whole heart labeling and shape-based atlas selection. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer (2018). p. 302–10. doi: 10.1007/978-3-030-12029-0_33
 82. Savioli N, Montana G, Lamata P. V-FCNN: volumetric fully convolution neural network for automatic atrial segmentation. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*. Cham: Springer (2018). p. 273–81. doi: 10.1007/978-3-030-12029-0_30
 83. Veni G, Fu Z, Awate SP, Whitaker RT. Bayesian segmentation of atrium wall using globally-optimal graph cuts on 3D meshes. In: *International Conference on Information Processing in Medical Imaging*. (2013). p. 656–67. doi: 10.1007/978-3-642-38868-2_55
 84. Tao Q, Ipek EG, Shahzad R, Berendsen FF, Nazarian S, van der Geest RJ. Fully automatic segmentation of left atrium and pulmonary veins in late gadolinium-enhanced MRI: towards objective atrial scar assessment. *J Magnet Resonance Imaging*. (2016) 44:346–54. doi: 10.1002/jmri.25148
 85. Yang G, Zhuang X, Khan H, Haldar S, Nyktari E, Ye X, et al. A fully automatic deep learning method for atrial scarring segmentation from late gadolinium-enhanced MRI images. In: *2017 IEEE 14th International Symposium on Biomedical Imaging*. (2017). p. 844–8. doi: 10.1109/ISBI.2017.7950649
 86. Yang G, Zhuang X, Khan H, Haldar S, Nyktari E, Ye X, et al. Segmenting atrial fibrosis from late gadolinium-enhanced cardiac MRI by deep-learned features with stacked sparse auto-encoders. In: M. Valdés Hernández and V. González-Castro, *Medical Image Understanding and Analysis*. Cham: Springer International Publishing (2017). p. 195–206. doi: 10.1007/978-3-319-60964-5_17
 87. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. *arXiv [preprint]*. arXiv:1409.0473 (2014)
 88. Li L, Wu F, Yang G, Xu L, Wong T, Mohiaddin R, et al. Atrial scar quantification via multi-scale CNN in the graph-cuts framework. *Med Image Anal.* (2020) 60:101595. doi: 10.1016/j.media.2019.101595
 89. Dong H, Yang G, Liu F, Mo Y, Guo Y. Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. In: *Annual Conference on Medical Image Understanding and Analysis*. (2017). p. 506–17. doi: 10.1007/978-3-319-60964-5_44
 90. Li X, Chen H, Qi X, Dou Q, Fu C-W, Heng PA. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *Proc IEEE Med Imaging*. (2018) 37:2663–74. doi: 10.1109/TMI.2018.2845918
 91. Bernard O, Lalonde A, Zotti C, Cervenansky F, Yang X, Heng P-A, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *Proc IEEE Med Imag*. (2018) 37:2514–25. doi: 10.1109/TMI.2018.2837502
 92. Jin C, Feng J, Wang L, Yu H, Liu J, Lu J, et al. Left atrial appendage segmentation using fully convolutional neural networks and modified three-dimensional conditional random fields. *IEEE J Biomed Health Inform*. (2018) 22:1906–16. doi: 10.1109/JBHI.2018.2794552
 93. Csepe TA, Kalyanasundaram A, Hansen BJ, Zhao J, Fedorov VV. Fibrosis: a structural modulator of sinoatrial node physiology and dysfunction. *Front Physiol*. (2015) 6:37. doi: 10.3389/fphys.2015.00037

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Jamart, Xiong, Maso Talou, Stiles and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Artificial Intelligence in Cardiac Imaging With Statistical Atlases of Cardiac Anatomy

Kathleen Gilbert¹, Charlène Mauger^{1,2}, Alistair A. Young^{2,3*} and Avan Suinesiaputra^{2,4,5}

¹ Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand, ² Department of Anatomy and Medical Imaging, University of Auckland, Auckland, New Zealand, ³ Department of Biomedical Engineering, King's College London, London, United Kingdom, ⁴ Centre for Computational Imaging and Simulation Technologies in Biomedicine, School of Computing, University of Leeds, Leeds, United Kingdom, ⁵ School of Medicine, Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds, Leeds, United Kingdom

OPEN ACCESS

Edited by:

Matteo Cameli,
University of Siena, Italy

Reviewed by:

Bennett Allan Landman,
Vanderbilt University, United States
Nicolas Duchateau,
Université Claude Bernard
Lyon 1, France

*Correspondence:

Alistair A. Young
alistair.young@kcl.ac.uk

Specialty section:

This article was submitted to
Cardiovascular Imaging,
a section of the journal
Frontiers in Cardiovascular Medicine

Received: 15 November 2019

Accepted: 14 May 2020

Published: 30 June 2020

Citation:

Gilbert K, Mauger C, Young AA and
Suinesiaputra A (2020) Artificial
Intelligence in Cardiac Imaging With
Statistical Atlases of Cardiac Anatomy.
Front. Cardiovasc. Med. 7:102.
doi: 10.3389/fcvm.2020.00102

In many cardiovascular pathologies, the shape and motion of the heart provide important clues to understanding the mechanisms of the disease and how it progresses over time. With the advent of large-scale cardiac data, statistical modeling of cardiac anatomy has become a powerful tool to provide automated, precise quantification of the status of patient-specific heart geometry with respect to reference populations. Powered by supervised or unsupervised machine learning algorithms, statistical cardiac shape analysis can be used to automatically identify and quantify the severity of heart diseases, to provide morphometric indices that are optimally associated with clinical factors, and to evaluate the likelihood of adverse outcomes. Recently, statistical cardiac atlases have been integrated with deep neural networks to enable anatomical consistency of cardiac segmentation, registration, and automated quality control. These combinations have already shown significant improvements in performance and avoid gross anatomical errors that could make the results unusable. This current trend is expected to grow in the near future. Here, we aim to provide a mini review highlighting recent advances in statistical atlas of cardiac function in the context of artificial intelligence in cardiac imaging.

Keywords: cardiac anatomy, machine learning, left ventricle, MRI, statistical shape

INTRODUCTION

The main function of the heart is to pump blood to the lungs and body. In order to maintain the equilibrium state of normal blood circulation, the heart continuously adapts its structure, shape, and function in response to physiological challenges and long-term environmental factors. From the onset of injury or disease, the heart starts a cascade of structural and morphological adaptations, known as *cardiac remodeling*. Common cardiac remodeling includes left ventricular dilatation, increasing ventricular mass, hypertrophy, aortic dilation, and systolic/diastolic functional alterations. When this condition is prolonged, cardiac function may deteriorate until symptoms become clinically evident and may eventually lead to heart failure (1). Here, we define cardiac remodeling to encompass a wide spectrum of physiological processes from adaptive remodeling in athlete's hearts (2) and normal aging process (3) to adverse remodeling in hypertensive heart disorder (4) and ischemia (5). It is therefore critical in the management of patients with heart

disease to identify and quantify the different types of cardiac remodeling and associations with environmental and clinical factors and to predict the likelihood of adverse outcomes in the future.

The associations between traditional risk factors of cardiovascular disease (including smoking, raised blood pressure, raised serum cholesterol, and diabetes mellitus) and developing cardiac disease were discovered from large epidemiological studies such as the Framingham Heart Study (6). To better understand the mechanism of subclinical disease, before symptoms are clinically evident, modern imaging examinations were later included, such as in the Multi-Ethnic Study of Atherosclerosis (MESA) (7) and the UK Biobank study (8). These large-scale studies have enabled a massive increase of imaging data available for the investigation of variations in cardiac geometry and function by using statistical shape analysis, as well as providing training data for machine learning algorithms.

Modern cardiac imaging modalities include echocardiography, computed tomography (CT), and magnetic resonance imaging (MRI). Each modality has its own advantages and disadvantages, but MRI has unique attributes over the other modalities that have enabled large-scale imaging studies in the general population, including the study of 6,000 preclinical subjects in the MESA and 100,000 asymptomatic subjects in the UK Biobank. MR images are acquired without ionizing radiation, and tomographic analysis can be performed without any geometrical assumption. In a single examination session, cardiac MRI can provide anatomical and functional images of the heart and great vessels in multiple views with high contrast-to-noise ratio, as well as high spatiotemporal resolution blood flow, microstructural tissue characterization, myocardial strain, blood perfusion, and scar images.

In this mini review, we focus on the rapid developments of machine learning combined with cardiac atlases. Although examples were taken mainly from cardiac MRI studies, these methods are generally extensible to other modalities. We first show how statistical shape analysis has enabled better understanding of cardiac shape remodeling within and between pathological groups. We then discuss current developments in machine learning to utilize the robustness of cardiac anatomy derived from statistical atlases to improve image analysis, including motion atlases to highlight the utility of dynamic data analysis vs. static analysis. **Table 1** compares representative papers in each category. We conclude with a discussion of future perspectives of cardiac atlases in the context of artificial intelligence (AI) in cardiac imaging.

STATISTICAL CARDIAC ATLASES

Statistical atlases consist of maps of cardiac shape and function, which can be used to quantify the variation in the population and quantify the differences between cohorts. They can also be used to quantify shape scores in individual patients relative to standard population groups. For example, the Cardiac Atlas Project¹ (24)

¹<http://www.cardiacatlas.org>

provides repositories of thousands of cardiac MRI studies (25) and benchmark data for the development of automated analysis algorithms, including segmentation of images (26) and shape analysis (27).

Two common atlas construction pipelines are shown in **Figure 1**, where both approaches lead to a comparable statistical analysis (28). In the first approach (9), images are analyzed to obtain the locations of cardiac landmarks (valve positions and the margins of the interventricular septum) and ventricular contours. The points are then mapped into 3D, and slice shifts due to breath-hold mis-registration are corrected. A 3D shape model template is then customized to the location of the landmarks and contours by minimizing the point-to-surface distances between the landmarks/contours and the model surfaces. Homologous points are then sampled from the surfaces and used to construct a point distribution model. This surface template fitting approach has also been translated to echocardiographic images where temporal resolution is much higher, as demonstrated in (20).

The second approach uses 3D images to establish a mean image template before generating cardiac mesh data. In (11), a high-resolution 3D MR template image and myocardial mesh are used. Each short axis image stack is then corrected for breath-hold mis-registration and registered to the template image using non-rigid image registration methods. For each case, a registration map is stored to give a mapping from subject space to template space at each voxel. The template mesh is then propagated to each subject using the inverse registration map. A point distribution model can then be calculated from the resulting homologous points. A similar approach was demonstrated in (12) by using CT images, with the advantage of high resolution and no breath-hold mis-registration in CT data.

Both these approaches benefit from recent advances in machine learning methods. Firstly, deep learning segmentation networks for cardiac images have been developed to enable fast generation of contours and landmarks (17, 29); and secondly, deep learning has enabled fast computation of registration maps, which can be trained without extensive manual image annotation using image similarity as the loss function (23, 30, 31).

ATLAS MEASURES OF CARDIAC REMODELING

Let $s \in R^{3P}$ be a shape vector with P homologous points in 3D. To extract shape parameters from a cohort or pathology group, a linear generative model is commonly applied, that is,

$$s \approx \bar{s} + \Phi^T b \quad (1)$$

where $\bar{s} \in R^{3P}$ is the mean shape estimated from the cohort, $\Phi \in R^{M \times 3P}$ is the linear decomposition matrix (defining modes of shape variation), and $b \in R^M$ is the shape parameter vector. If N is the number of patient shapes in the cohort and $M < N$, then Equation (1) is called a dimension reduction technique. Because each 3D point in this point distribution model encapsulates approximately the same anatomical location in the heart, the relative locations of neighboring positions are highly correlated,

TABLE 1 | Summary of cardiac atlas construction and deep learning methods with cardiac shape priors.

Methods	Model	Strength	Weakness	Training cohorts	Availability
ATLAS CONSTRUCTIONS					
Medrano-Gracia et al. (9)	LV	Mathematically defined cardiac shape model	Requires contours	1,991 MESA	CAP ^a
Mauger et al. (10)	LV, RV	Diffeomorphic	Requires segmentation	4,329 UK Biobank	CAP ^a
Bai et al. (11)	LV, RV	Volumetric model	Long breath-hold	1,093 healthy	ICL ^b
Hoogendoorn et al. (12)	LV, RV, LA, RA	High spatial resolution from CT	Small cohort; no healthy reference for CT	138 CAD	CISTIB ^c
DEEP LEARNING IMAGE ANALYSIS WITH SHAPE PRIORS					
Oktaý et al. (13)	LV	Latent space regularization	Layers reduced for 3D	1,200 healthy	VitaLabAI ^d
Zotti et al. (14)	LV, RV	Simple adjustment of the U-Net	Single prior map	150 ACDC	
Chen et al. (15)	LV	Latent spaces for standard orientations	Each prior requires separate encoder; memory intensive	734 healthy	
Duan et al. (16)	LV, RV	2.5D	Computationally expensive; two stages of network (not end-to-end learning)	1,831 healthy; 649 pulmonary hypertension	Github ^e
DEEP LEARNING SHAPE ANALYSIS					
Attar et al. (17)	LV, RV	Direct prediction of the shape scores	Linear PCA shape model from non-linear deep learning network	3,500 healthy	Github ^f
Clough et al. (18)	LV, RV	Interpretable	Latent reconstruction blurred	10,038 healthy; 778 CAD	
Painchaud et al. (19)	LV, RV	Augmented latent space	Requires three networks to train	150 ACDC	
DYNAMIC ANALYSIS					
Puyol-Antón et al. (20)	LV	Multimodalities	Separate pipelines; small cohort	50 healthy	Github ^f
Bello et al. (21)	LV, RV	Survival loss on latent space	Displacement only	302 pulmonary hypertension	
Peressutti et al. (22)	LV	Motion and clinical features	Small cohort	34 dyssynchrony	Github ^g
Qin et al. (23)	LV	Joint motion and segmentation	2D + time	220 healthy	

LV, left ventricle; RV, right ventricle; LA, left atrium; RA, right atrium; CAD, coronary artery disease; MESA, multi-ethnic study of atherosclerosis; ACDC, automated cardiac diagnosis challenge.

^aCAP, <http://cardiacatlas.org>.

^bICL, <http://wp.doc.ic.ac.uk/wbai/data/>.

^cCISTIB, <http://www.cistib.org/full-heart-pca-model-all-phases/en/full-heart-pca-model-all-phases>.

^dVitaLabAI, https://bitbucket.org/vitalab/vitalabai_public/src/master/VITALabAI/model/.

^eGithub, <https://github.com/j-duan/4Dsegment>.

^fGithub, <https://github.com/UK-Digital-Heart-Project/4Dsurvival>.

^gGithub, <https://github.com/cq615/Joint-Learning-of-Motion-Estimation-and-Segmentation-for-Cardiac-MR-Image-Sequences>.

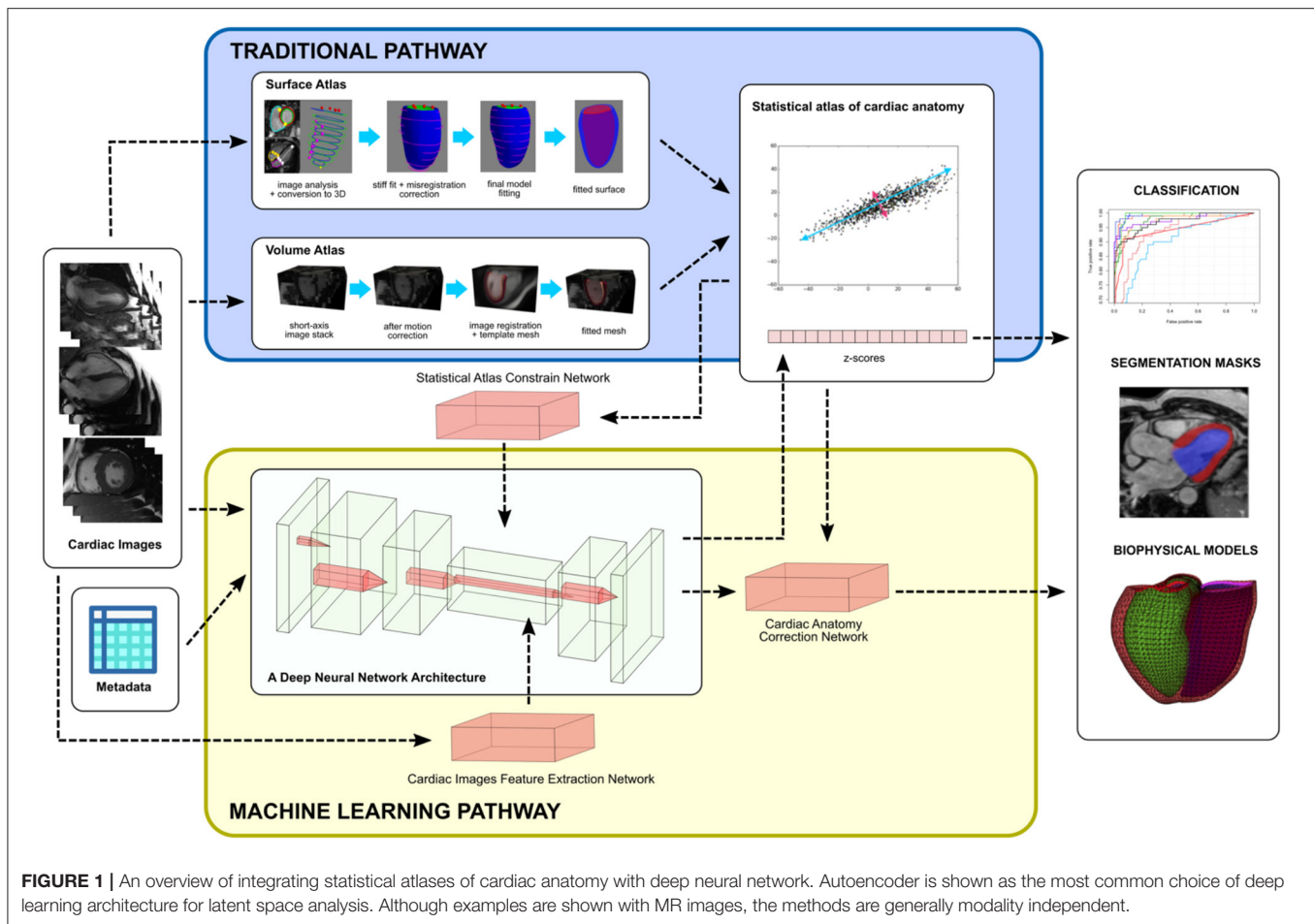
enabling the dimension reduction method to distill a small number of shape parameters.

The most common dimension reduction method is principal component analysis (PCA), whereby shape modes are ordered by the amount of variance explained. Most of the shape variations can then be explained in terms of the first few principal modes of variation. In the MESA baseline imaging study, the PCA mode explaining the most shape variation was associated with the size of the heart, even after correction for patient height (9). This is a common finding because the first mode often relates to the amplitude of the studied descriptors. The second mode was associated with sphericity. Clinically, these first two PCA modes are known to be associated with adverse outcomes in both symptomatic disease and asymptomatic cohorts (32–35).

PCA regression enables evaluation of the relationships between the PCA scores and clinical factors such as diabetes (9, 28). However, PCA is an unsupervised dimension reduction

method, and component modes do not in general map to recognizable shape characteristics (9, 28). Supervised dimension reduction methods such as information maximizing component analysis have shown promise for quantifying the differences between a patient group and a control group, or two patient groups (36). Another approach is to combine dimensionality reduction with direct correlation with clinically defined remodeling indices such as ventricular volumes, wall thickness, and sphericity, by using the partial least squares method. Zhang et al. (37) applied this method in conjunction with a sequential orthogonalization algorithm to construct orthogonal shape scores, which are optimally matched with known clinical indices of remodeling. More general ways of characterizing the shape probability distribution have been investigated (38).

Gilbert et al. (28) found that both volume and surface cardiac atlases showed similar morphometric characteristics and similar relationships between risk factors and left ventricular



shape. Thus, shape scores derived from atlases are robust to differences in construction methodology and quantify real anatomical relationships with cardiovascular risk factors. Morphometric scores were found to be more sensitive to cardiovascular risk factors than traditional measures of mass and volume. Mauger et al. (10) used a biventricular shape model to study right and left ventricular interactions in the UK Biobank study. A subdivision surface biventricular shape model was automatically customized to manually draw contours using a diffeomorphic least squares optimization algorithm. A control group sub-cohort consisting of 630 participants with no cardiovascular risk factors and normal cardiac parameters was used as a reference group to quantify shape differences due to traditional risk factors. Morphometric scores were computed using linear regression to quantify shape variations associated with prediction variables including sex, age, height, high cholesterol, high blood pressure, obesity, and smoking as well as diabetes, previous myocardial infarction, and angina. This regression approach enabled quantification of the effects of each prediction variable while controlling for the effects of the others.

In congenital heart disease, atlas-based analysis of shape variations can provide quantitative measures of deterioration

before detection of symptoms. Sheehan et al. (39) developed a method for patient customization using a linear combination of database templates. This knowledge-based reconstruction method has shown accurate and rapid analysis of right ventricular shapes and volumes in patients with tetralogy of Fallot (39), dextro-transposition of the great arteries (40), and other types of congenital heart disease (41). A more dilated and spherical right ventricle was found in patients with transposition of the great arteries after atrial switch, with regional reduction in function at the base (42, 43). These methods assume that the patient heart geometry is accurately represented by a linear combination of cases in the database. An alternative approach is to jointly estimate the shape and the underlying statistical shape model so that the statistical model can be automatically updated while analyzing new cases (44). Shape model templates have been constructed to describe common congenital pathologies, such as congenitally corrected transposition of the great arteries, enabling a wide range of pathologies to be accurately characterized (45). In single-ventricle pathologies, with tricuspid atresia and Fontan repair, shape mode scores were able to quantify differences in shape and function, with more spherical ED shapes being associated with reduced longitudinal shortening (46). Atlas analysis in

association with biomechanical analysis may be able to identify mechanisms underlying changes in function with developing disease (47).

DEEP LEARNING NETWORKS WITH CARDIAC SHAPE PRIORS

Deep learning is currently the state-of-the-art method for medical image feature extraction and supervised analysis. Its superior performance has surpassed any other traditional machine learning algorithms in many applications, including cardiac imaging (48, 49). This success is mainly attributed to the automatic generation of optimal features, rather than relying on handcrafted features. This means that without significantly modifying the architecture, deep learning allows transfer of techniques, thereby shifting the data domain from one application, for example, natural image analysis, to another, for example, cardiac imaging. In addition, transfer learning directly reuses a pretrained network and fine-tunes to a new application domain. Examples include transfer learning of retinal image segmentation into cardiac vessels (50) or predicting cardiovascular risk from retinal fundus images (51). This flexibility and reusability of deep neural network architectures have led to rapid development. However, there are some limitations. Deep learning is prone to overfitting and usually cannot infer the anatomical correctness of the prediction results. The network's parameters are also sensitive to the data or cohort used during training (implicit bias). Statistical atlases or shape priors can therefore be integrated with deep learning to overcome these limitations. Thus, anatomical correctness can be imposed by enabling the network to learn the biological constraints as well as the measurement correlations.

Machine learning methods can add new quantitative analysis techniques to examine the relationships between shape features and clinical status, in addition to the traditional methods of linear or logistic regression. These are now being applied to statistical shape atlases to characterize differences in patient groups and predict outcomes. In the STACOM 2015 shape analysis challenge (27), various machine learning algorithms were compared on a benchmark dataset, and 11 groups participated to determine cardiac shapes of patients with myocardial infarction from healthy subjects. Five groups used the z -scores (standardized b vector in Equation 1) in different ways to classify myocardial infarction shapes. The training accuracies ranged between 0.93 and 0.98, whereas the test accuracies were 0.83–0.98. Shape atlases have been useful in identifying genetic mutations affecting left ventricular (LV) mass (52). Shape features associated with disease can be interpreted through visualizations using deep generative networks (53).

Incorporating cardiac anatomy in deep learning was demonstrated by Oktay et al. (13) with an anatomically constrained neural network. Two separate autoencoder networks were appended after the final predicted segmentation mask and the ground truth mask layers, which extracted features from mask images separately. A global shape similarity loss function calculated from the output of autoencoder networks

was introduced as a way to constrain the optimization to follow the same shapes as the ground truth. Their results showed improved super resolution and segmentation accuracies in the long-axis view² by correcting mis-registration between image slices. Another shape-based loss function was also proposed by Yang et al. (54) to segment the right ventricle.

Alternatively, shape priors can be introduced directly inside a network (14–16). Zotti et al. (14) inserted a cardiac shape probability map before the final layer of a U-Net architecture to ensure that the output segmentation masks were valid. Chen et al. (15) also modified a U-Net architecture with cardiac shape priors, but they modified the bottom layer (feature extraction layer) by inserting short-axis and long-axis feature vectors trained independently from short-axis and long-axis cardiac MRI, respectively. Duan et al. (16) embedded a more specialized shape refinement subnetwork into the main segmentation and super resolution network. The subnetwork consisted of shape affine alignment, atlas selection, and non-rigid free form deformation registration operations. The network was able to generate smooth high-resolution 3D cardiac mesh data from low-resolution cardiac MRI.

DEEP LEARNING FOR STATISTICAL CARDIAC ATLASES

The ability of deep learning to learn non-linear relationships between different data domains and the high focus on segmentation have enabled several studies to directly link cardiac imaging and statistical shape analysis. In Equation (1), patient-specific shape parameters with population reference of Φ are represented by $b \in R^M$ vectors. A statistically plausible new shape of s can be generated by setting values of b within $\pm 2\sqrt{\sigma}$, where σ is the eigenvalues from the PCA. Shape generation can also be performed by sampling from a probability distribution function learned from an atlas (38).

Attar et al. (17) proposed a neural network model that learns how to directly predict shape parameters b given a combination of cardiac MRI and patient characteristics metadata [age, weight, height, body mass index (BMI), body surface area (BSA), heart rate, systolic blood pressure (SBP), diastolic blood pressure (DBP), sex, smoking status, and alcohol consumption]. Hence, the network was trained to predict statistically plausible b vector from images and metadata parameters to generate a 3D cardiac shape by using (Equation 1). Also, Clough et al. (18) used a variational autoencoder to generate interpretable representations of patients with low ejection fraction. This aids the interpretability of machine learning algorithm, which is vital to their acceptance in the clinical community.

A different approach to embed statistical shape parameterization into deep neural network was proposed by Painchaud et al. (19). A separate adversarial variational autoencoder was trained to generate a latent space of cardiac

²In standard cardiac imaging acquisition, short-axis views show an image of the left and right ventricular chambers, and long-axis views typically show either two chambers (left ventricle and left atrium) or all four chambers in a single image. Short-axis views are generally perpendicular to the long-axis views.

anatomy from mask images and was then connected to another anatomical variational autoencoder to correct errors after segmentation. Hence, this network (19) indirectly learned patient-specific parameters in the latent space without actually modeling how the latent space should be parameterized as in (17). The disentanglement of latent spaces is an active area of research and shows promise in factorizing anatomical representations from modality characteristics (55).

DYNAMIC ATLASES

Many of the features associated with cardiac pathology are manifest as changes in motion rather than changes in static shape. As the heart is responsible to deliver sufficient blood into the circulation system, the onset of cardiac diseases forces the heart to adapt its motion. Changes in cardiac shape deformation, myocardial strain, and strain rate are examples of important dynamic remodeling indices when building a cardiac motion or dynamic atlas. However, building a dynamic atlas is sometimes limited by the temporal resolution of the acquired imaging data, although combining two modalities, such as MRI and echocardiography (20), can increase the temporal resolution of the atlas considerably.

There are a significant number of cardiac applications that can get the benefit of machine learning from cardiac motion. In pulmonary hypertension, a motion atlas is combined with the latent space of autoencoder network to predict the survival rate (21). A machine learning system that combines motion atlas with non-motion data (ECG and clinical reports) has been demonstrated in the selection of patients with dyssynchrony for cardiac resynchronization therapy (22). The study of dynamic atlases will be a fruitful area of future research. Deep learning methods for combined shape and motion analysis are now being developed (23), which can be used to extend previous methods for motion atlas (11).

DISCUSSION

A statistical atlas of cardiac anatomy is a powerful tool to analyze a patient-specific remodeling compared with the reference population. An abnormal cardiac shape can be quantified against

a population reference, regional wall motion differences can be compared across pathological groups, and a hypothetical cardiac shape can further be predicted from a longitudinal study. Apart from that, a statistical atlas can be used as a reference by machine learning algorithms to constrain their analysis within valid anatomic boundaries.

In summary, we have reviewed three ways to integrate a statistical atlas into a machine learning framework. The first approach is to directly use individual shape atlas parameters, for example, the z-scores, as the training data. This approach needs homologous points generated from a shape modeling technique derived from images and a registration method to align points to remove variations in the global position and orientation. The effectiveness of this approach was demonstrated in the STACOM 2015 challenge. The second approach is to use statistical atlases as shape priors either as a way to measure shape similarity in a loss function or to add shape features to be learned inside the network. The third approach is to predict statistical shape parameters or a location in a shape-based feature space directly from images. This is a promising field for deep learning, because it can generate relationships between two completely different data domains.

In the future, statistical atlases show promise for augmenting deep learning methods, and vice versa. An atlas can add robustness to the prediction results because additional information on a reference population is included during the learning process. Atlases will also increase the interpretability of the AI process, which is critical for the acceptance of AI in health care.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was provided by the National Institutes of Health (USA) 1R01HL121754. AY acknowledges HRC Grant 17/234. CM acknowledges New Zealand Heart Foundation Grant 1695.

REFERENCES

- Gjesdal O, Bluemke DA, Lima JA. Cardiac remodeling at the population level—risk factors, screening, and outcomes. *Nat Rev Cardiol.* (2011) 8:673–85. doi: 10.1038/nrcardio.2011.154
- Maron BJ, Pelliccia A. The heart of trained athletes: cardiac remodeling and the risks of sports, including sudden death. *Circulation.* (2006) 114:1633–44. doi: 10.1161/CIRCULATIONAHA.106.613562
- Cohn JN, Ferrari R, Sharpe N. Cardiac remodeling—concepts and clinical implications: a consensus paper from an international forum on cardiac remodeling. *J Am Coll Cardiol.* (2000) 35:569–82. doi: 10.1016/S0735-1097(99)00630-0
- Nwabuo CC, Vasan RS. Pathophysiology of hypertensive heart disease: beyond left ventricular hypertrophy. *Curr Hypertens Rep.* (2020) 22:11. doi: 10.1007/s11906-020-1017-9
- Sutton MG, Sharpe N. Left ventricular remodeling after myocardial infarction: pathophysiology and therapy. *Circulation.* (2000) 101:2981–8. doi: 10.1161/01.CIR.101.25.2981
- Kannel WB. Factors of risk in the development of coronary heart disease—six-year follow-up experience: the framingham study. *Ann Intern Med.* (1961) 55:33. doi: 10.7326/0003-4819-55-1-33
- Bild DE. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol.* (2002) 156:871–81. doi: 10.1093/aje/kwf113
- Petersen SE, Matthews PM, Bamberg F, Bluemke DA, Francis JM, Friedrich MG, et al. Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK Biobank - rationale, challenges and approaches. *J Cardiovasc Magn Reson.* (2013) 15:46. doi: 10.1186/1532-429X-15-46
- Medrano-Gracia P, Cowan BR, Ambale-Venkatesh B, Bluemke DA, Eng J, Finn JP, et al. Left ventricular shape variation in asymptomatic populations:

- the multi-ethnic study of atherosclerosis. *J Cardiovasc Magn Reson.* (2014) 16:56. doi: 10.1186/s12968-014-0056-2
10. Mauger C, Gilbert K, Lee AM, Sanghvi MM, Aung N, Fung K, et al. Right ventricular shape and function: cardiovascular magnetic resonance reference morphology and biventricular risk factor morphometrics in UK Biobank. *J Cardiovasc Magn Reson.* (2019) 21:41. doi: 10.1186/s12968-019-0551-6
 11. Bai W, Shi W, de Marvao A, Dawes TJW, O'Regan DP, Cook SA, et al. A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of healthy subjects and an analysis of shape and motion. *Med Image Anal.* (2015) 26:133–45. doi: 10.1016/j.media.2015.08.009
 12. Hoogendoorn C, Duchateau N, Sanchez-Quintana D, Whitmarsh T, Sukno FM, De Craene M, et al. A high-resolution atlas and statistical model of the human heart from multislice CT. *IEEE Trans Med Imaging.* (2013) 32:28–44. doi: 10.1109/TMI.2012.2230015
 13. Oktay O, Ferrante E, Kamnitsas K, Heinrich M, Bai W, Caballero J, et al. Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. *IEEE Trans Med Imaging.* (2018) 37:384–95. doi: 10.1109/TMI.2017.2743464
 14. Zotti C, Luo Z, Lalande A, Jodoin P-M. Convolutional neural network with shape prior applied to cardiac MRI segmentation. *IEEE J Biomed Health Inform.* (2019) 23:1119–28. doi: 10.1109/JBHI.2018.2865450
 15. Chen C, Biffi C, Tarroni G, Petersen S, Bai W, Rueckert D. Learning shape priors for robust cardiac MR segmentation from multi-view images. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al. editors. *Medical Image Computing Computer Assisted Intervention – MICCAI 2019*. Cham: Springer International Publishing (2019). p. 523–31. doi: 10.1007/978-3-030-32245-8_58
 16. Duan J, Bello G, Schlemper J, Bai W, Dawes TJW, Biffi C, et al. Automatic 3D Bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach. *IEEE Trans Med Imaging.* (2019) 38:2151–64. doi: 10.1109/TMI.2019.2894322
 17. Attar R, Pereañez M, Bowles C, Piechnik SK, Neubauer S, Petersen SE, et al. 3D Cardiac shape prediction with deep neural networks: simultaneous use of images patient metadata. In: Shen DT, Liu TM, Peters LH, Staib C, Essert S, Zhou PT, Yap A, editors. *Medical Image Computing Computer Assisted Intervention – MICCAI 2019*. Cham: Springer International Publishing (2019). p. 586–94. doi: 10.1007/978-3-030-32245-8_65
 18. Clough JR, Oksuz I, Puyol-Antón E, Ruijsink B, King AP, Schnabel JA. Global local interpretability for cardiac MRI classification. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al. editors. *Medical Image Computing Computer Assisted Intervention – MICCAI 2019*. Cham: Springer International Publishing (2019). p. 656–64. doi: 10.1007/978-3-030-32251-9_72
 19. Painchaud N, Skandarani Y, Judge T, Bernard O, Lalande A, Jodoin P-M. Cardiac MRI segmentation with strong anatomical guarantees. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al. editors. *Medical Image Computing Computer Assisted Intervention – MICCAI 2019*. Cham: Springer International Publishing (2019). p. 632–40. doi: 10.1007/978-3-030-32245-8_70
 20. Puyol-Antón E, Sinclair M, Gerber B, Amzulescu MS, Langet H, Craene MD, et al. A multimodal spatiotemporal cardiac motion atlas from MR and ultrasound data. *Med Image Anal.* (2017) 40:96–110. doi: 10.1016/j.media.2017.06.002
 21. Bello GA, Dawes TJW, Duan J, Biffi C, de Marvao A, Howard LS, et al. Deep-learning cardiac motion analysis for human survival prediction. *Nat Mach Intell.* (2019) 1:95–104. doi: 10.1038/s42256-019-0019-2
 22. Peressutti D, Sinclair M, Bai W, Jackson T, Ruijsink J, Nordsletten D, et al. A framework for combining a motion atlas with non-motion information to learn clinically useful biomarkers: application to cardiac resynchronisation therapy response prediction. *Med Image Anal.* (2017) 35:669–84. doi: 10.1016/j.media.2016.10.002
 23. Qin C, Bai W, Schlemper J, Petersen SE, Piechnik SK, Neubauer S, et al. Joint learning of motion estimation segmentation for cardiac MR image sequences. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Medical Image Computing Computer Assisted Intervention – MICCAI 2018*. Cham: Springer International Publishing (2018). p. 472–80. doi: 10.1007/978-3-030-00934-2_53
 24. Fonseca CG, Backhaus M, Bluemke DA, Britten RD, Chung JD, Cowan BR, et al. The cardiac atlas project—an imaging database for computational modeling and statistical atlases of the heart. *Bioinformatics.* (2011) 27:2288–95. doi: 10.1093/bioinformatics/btr360
 25. Suinesiaputra A, Medrano-Gracia P, Cowan BR, Young AA. Big heart data: advancing health informatics through data sharing in cardiovascular imaging. *IEEE J Biomed Health Inform.* (2015) 19:1283–90. doi: 10.1109/JBHI.2014.2370952
 26. Suinesiaputra A, Cowan BR, Al-Agamy AO, Elattar MA, Ayache N, Fahmy AS, et al. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac MR images. *Med Image Anal.* (2014) 18:50–62. doi: 10.1016/j.media.2013.09.001
 27. Suinesiaputra A, Dhooge J, Duchateau N, Ehrhardt J, Frangi AF, Gooya A, et al. Statistical shape modeling of the left ventricle: myocardial infarct classification challenge. *IEEE J Biomed Health Inform.* (2018) 22:503–15. doi: 10.1109/JBHI.2017.2652449
 28. Gilbert K, Bai W, Mauger C, Medrano-Gracia P, Suinesiaputra A, Lee AM, et al. Independent left ventricular morphometric atlases show consistent relationships with cardiovascular risk factors: a UK biobank study. *Sci Rep.* (2019) 9:1130. doi: 10.1038/s41598-018-37916-6
 29. 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) 20:65. doi: 10.1186/s12968-018-0471-x
 30. Haskins G, Kruger U, Yan P. Deep learning in medical image registration: a survey. *Mach Vis Appl.* (2020) 31:8. doi: 10.1007/s00138-020-01060-x
 31. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I. A deep learning framework for unsupervised affine and deformable image registration. *Med Image Anal.* (2019) 52:128–43. doi: 10.1016/j.media.2018.11.010
 32. White HD, Norris RM, Brown MA, Brandt PW, Whitlock RM, Wild CJ. Left ventricular end-systolic volume as the major determinant of survival after recovery from myocardial infarction. *Circulation.* (1987) 76:44–51. doi: 10.1161/01.CIR.76.1.44
 33. Wong SP, French JK, Lydon A-M, Manda SOM, Gao W, Ashton NG, et al. Relation of left ventricular sphericity to 10-year survival after acute myocardial infarction. *Am J Cardiol.* (2004) 94:1270–75. doi: 10.1016/j.amjcard.2004.07.110
 34. Bluemke DA, Kronmal RA, Lima JAC, Liu K, Olson J, Burke GL, et al. The relationship of left ventricular mass and geometry to incident cardiovascular events. *J Am Coll Cardiol.* (2008) 52:2148–55. doi: 10.1016/j.jacc.2008.09.014
 35. Ambale-Venkatesh B, Yoneyama K, Sharma RK, Ohyama Y, Wu CO, Burke GL, et al. Left ventricular shape predicts different types of cardiovascular events in the general population. *Heart.* (2017) 103:499–507. doi: 10.1136/heartjnl-2016-310052
 36. Zhang X, Ambale-Venkatesh B, Bluemke DA, Cowan BR, Finn JP, Kadish AH, et al. Information maximizing component analysis of left ventricular remodeling due to myocardial infarction. *J Transl Med.* (2015) 13:343. doi: 10.1186/s12967-015-0709-4
 37. Zhang X, Medrano-Gracia P, Ambale-Venkatesh B, Bluemke DA, Cowan BR, Finn JP, et al. Orthogonal decomposition of left ventricular remodeling in myocardial infarction. *GigaScience.* (2017) 6:1–15. doi: 10.1093/gigascience/gix005
 38. Gooya A, Lekadir K, Castro-Mateos I, Pozo JM, Frangi AF. Mixture of probabilistic principal component analyzers for shapes from point sets. *IEEE Trans Pattern Anal Mach Intell.* (2018) 40:891–904. doi: 10.1109/TPAMI.2017.2700276
 39. Sheehan FH, Kilner PJ, Sahn DJ, Vick GW, Stout KK, Ge S, et al. Accuracy of knowledge-based reconstruction for measurement of right ventricular volume and function in patients with tetralogy of fallot. *Am J Cardiol.* (2010) 105:993–9. doi: 10.1016/j.amjcard.2009.11.032
 40. Morcos M, Kilner PJ, Sahn DJ, Litt HI, Valsangiacomo-Buechel ER, Sheehan FH. Comparison of systemic right ventricular function in transposition of the great arteries after atrial switch and congenitally corrected transposition of the great arteries. *Int J Cardiovasc Imaging.* (2017) 33:1993–2001. doi: 10.1007/s10554-017-1201-4
 41. Trzebiatowska-Krzynska A, Driessen M, Sieswerda GT, Wallby L, Swahn E, Meijboom F. Knowledge-based 3D reconstruction of the right ventricle:

- comparison with cardiac magnetic resonance in adults with congenital heart disease. *Echo Res Pract.* (2015) 2:109–16. doi: 10.1530/ERP-15-0029
42. Nyns ECA, Dragulescu A, Yoo S-J, Grosse-Wortmann L. Evaluation of knowledge-based reconstruction for magnetic resonance volumetry of the right ventricle in tetralogy of fallot. *Pediatr Radiol.* (2014) 44:1532–40. doi: 10.1007/s00247-014-3042-9
 43. Nyns ECA, Dragulescu A, Yoo S-J, Grosse-Wortmann L. Evaluation of knowledge-based reconstruction for magnetic resonance volumetry of the right ventricle after arterial switch operation for dextro-transposition of the great arteries. *Int J Cardiovasc Imaging.* (2016) 32:1415–23. doi: 10.1007/s10554-016-0921-1
 44. Stebbing RV, Namburete ALL, Upton R, Leeson P, Noble JA. Data-driven shape parameterization for segmentation of the right ventricle from 3D+t echocardiography. *Med Image Anal.* (2015) 21:29–39. doi: 10.1016/j.media.2014.12.002
 45. Gilbert K, Forsch N, Hegde S, Mauger C, Omens JH, Perry JC, et al. Atlas-based computational analysis of heart shape and function in congenital heart disease. *J Cardiovasc Transl Res.* (2018) 11:123–32. doi: 10.1007/s12265-017-9778-5
 46. Farrar G, Suinesiaputra A, Gilbert K, Perry JC, Hegde S, Marsden A, et al. Atlas-based ventricular shape analysis for understanding congenital heart disease. *Prog Pediatr Cardiol.* (2016) 43:61–9. doi: 10.1016/j.ppedcard.2016.07.010
 47. Salehyar S, Forsch N, Gilbert K, Young AA, Perry CJ, Hegde S, et al. A novel atlas-based strategy for understanding cardiac dysfunction in patients with congenital heart disease. *Mol Cell Biomech.* (2019) 16:179–83. doi: 10.32604/mcb.2019.07384
 48. Bizopoulos P, Koutsouris D. Deep learning in cardiology. *IEEE Rev Biomed Eng.* (2019) 12:168–93. doi: 10.1109/RBME.2018.2885714
 49. Leiner T, Rueckert D, Suinesiaputra A, Baeßler B, Nezafat R, Išgum I, et al. Machine learning in cardiovascular magnetic resonance: basic concepts and applications. *J Cardiovasc Magn Reson.* (2019) 21:61. doi: 10.1186/s12968-019-0575-y
 50. Yu F, Zhao J, Gong Y, Wang Z, Li Y, Yang F, et al. Annotation-free cardiac vessel segmentation via knowledge transfer from retinal images. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al. editors. *Medical Image Computing Computer Assisted Intervention – MICCAI 2019*. Cham: Springer International Publishing (2019). p. 714–22. doi: 10.1007/978-3-030-32245-8_79
 51. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng.* (2018) 2:158–64. doi: 10.1038/s41551-018-0195-0
 52. Biffi C, de Marvao A, Attard MI, Dawes TJW, Whiffin N, Bai W, et al. Three-dimensional cardiovascular imaging-genetics: a mass univariate framework. *Bioinformatics.* (2018) 34:97–103. doi: 10.1093/bioinformatics/btx552
 53. Biffi C, Doumou G, Duan J, Prasad SK, Cook SA, O Regan DP, et al. Explainable anatomical shape analysis through deep hierarchical generative models. *IEEE Trans Med Imaging.* (2020) 39:2088–99. doi: 10.1109/TMI.2020.2964499
 54. Yang H, Liu Z, Yang X. Right ventricle segmentation in short-axis MRI using a shape constrained dense connected U-Net. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al. editors. *Medical Image Computing Computer Assisted Intervention – MICCAI 2019*. Cham: Springer International Publishing (2019). p. 532–40. doi: 10.1007/978-3-030-32245-8_59
 55. Chartsias A, Joyce T, Papanastasiou G, Semple S, Williams M, Newby DE, et al. Disentangled representation learning in cardiac image analysis. *Med Image Anal.* (2019) 58:101535. doi: 10.1016/j.media.2019.101535

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Gilbert, Mauger, Young and Suinesiaputra. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership